SERIE RESEARCHMEMORANDA

EXPLANATORY DISCRETE SPATIAL DATA AND CHOICE ANALYSIS : A STATE-OF-THE-ART REVIEW

Manfred M. Fischer
Peter Nijkamp

Researchmemorandum 1985-6 Jan. 1985

VRIJE UNIVERSITEIT
EKONOMISCHE FAKULTEIT
AMSTERDAM

EXPLANATORY DISCRETE SPATIAL DATA AND CHOICE ANALYSIS : A STATE-OF-THE-ART REVIEW

Manfred M. Fischer
Peter Nijkamp

Paper presented at the Working
Group on Quantitative Methods
in Geography,
Osnabrück, May 1984.

1. INTRODUCTION

The increased use of quantitative methods in humanities has been stimulated by advances in such disciplines as applied mathematics, econometrics, and statistics which have basically provided the tools for formal and quantitative research methodologies. The 'quantitative revolution' was also favoured by the explosive growth of computer hardware and software which provided the conditions for a fast and flexible treatment of large data sets representing a wide variety of facets of complex problems under study. These developments have also had profound effects on the research methodology in geography and economics. Illustrations and surveys of useful methods and models in the spatial sciences (notably regional science, regional economics and human geography) can be found in Bahrenberg et al. (1984), Nijkamp et al. (1984) and Issaev et al. (1982). In the past decade, much attention has been given to the specific consequences of the spatial distribution and coherence of activities for the research methodology in the spatial sciences. Examples are found inter alia in spatially hierarchical models, spatial interaction models, spatial auto and cross-correlation analysis, multiregional conflict management models and spatial choice models.

In contrast to natural sciences, the measurement of many variables in social science studies are discrete (non-metric or categorical) rather than metric. This is a consequence of the fact that the measurement procedures such as interviews or questionnaires have only a rather limited degree of precision. The term 'discrete' (non-metric or categorical) is used to refer to dichotomous and polytomous nominal variables as well as to dichotomous and polytomous ordinal variables. In contrast to metric variables discrete variables take values only in a limited set of categories. For more details see Roberts (1979).

Models and methods for dealing with discrete data are important in the spatial sciences, just as in other social and economic sciences. Therefore, it is no surprise that in recent years there has been an increasing interest in handling discrete data. Many methods and models widely used in the spatial sciences have been originally worked out in quantitatively more advanced social and economic science disciplines such as psychometrics, sociometrics and econometrics. Under certain circumstances however, specific problems can arise, especially when such methods and models have not been sufficiently adapted to the specific needs of the spatial sciences. One well known problem of this kind is the spatial dependence among spatial data which often requires developing a specific spatial methodology.

In discrete data analysis two subfields can be distinguished: exploratory discrete data analysis and explanatory discrete data analysis. Exploratory discrete data analysis already has a long tradition especially in psychometrics. Exploratory data analysis tends to suggest rather than to test hypotheses and basically aims at identifying structures in complex phenomena and at understanding complex data structures. In contrast to explanatory discrete data analysis no definite statistical model is imposed and tested. There is a wide variety of exploratory statistical procedures such as ordinal and nominal principal component analysis, factor analysis and cluster analysis, correspondence analysis, homogeneous scaling and log-linear modelling.

Interest in explanatory discrete data analysis has rapidly grown during the past two decades. Research on explanatory methods and models with discrete variables has primarily two sources:

- explanatory discrete data analysis (in a strict sense) aiming to analyse cause-effect relationships between a set of independent variables and one or more dependent variables, where at least the dependent variable(s) are discrete. Section 2 will be devoted to a further discussion of explanatory discrete data analysis.

and

- explanatory discrete choice analysis aiming to analyse the behaviour of populations of individuals in a given choice context (such as e.g. residential migration, industrial location, travel mode, urban labour supply). This will be discussed in greater detail in Section 3.

The relationship between explanatory discrete data analysis and choice analysis may be seen from the fact that probabilistic choice models are simply multinomial response models which are stripped of their behavioural interpretation. Both explanatory discrete data and choice analysis have most recently received a great deal of attention in modern spatial research, witness the increased popularity of asymmetric log-linear models, linear logit/logistic regression models, generalised linear models, various classes of discrete choice models and so forth. The purpose of this paper is to identify and classify some major lines of recent methodological developments in this field.

2. DISCRETE SPATIAL DATA ANALYSIS

2.1 Introduction

Until recently, the great potential offered by discrete (spatial) data methods has too often been neglected or underestimated. Fortunately, in recent years, significant progress has been made in the treatment of discrete variables. In the field of both parametric and non-parametric statistics and econometrics, a great variety of techniques and models for explanatory data analysis based on discrete data has been developed. Many of these new approaches are also increasingly being applied in geography, regional science and related disciplines (for instance, in the analysis of consumer choice behaviour, locational perceptions and preferences, contingency table analysis, scenario and qualitative impact analysis, plan evaluation and conflict analysis, and so forth).

While disciplines such as psychology and sociology have been extensive users of discrete data and have well developed experimental designs and statistical methods for dealing with such data (see De Leeuw et al. 1983, and Wegener 1982), many other fields (such as geography and regional science) have been late entrants to this new area and are still in an experimental phase. There is still a long way to go in the handling of qualitative spatial data (for instance, in the field of polyhedral dynamics, differential topology, discrete-to-continuous transforms, and qualitative spatial filter and aggregation analysis), though it has to be added that in recent years also remarkable advances have been made (for instance in the field of Generalised Linear Models, structural equation modelling with latent variables, correspondence analysis and other qualitative multivariate techniques, fuzzy set analysis, qualitative evaluation analysis and discrete choice analysis including panel data and time-event histories).

A bottleneck regarding the application of methodologies of discrete data analysis in empirical research practice is the lack of attention to <u>behavioural theories</u> supporting or underlying the mathematical/statistical methodologies and models. In general, much attention is paid to the formal aspects of mathematical and statistical theories, but unsatisfactory attention is paid to (spatial) behavioural processes being studied or to the structure of regions or cities at hand. Clearly, if various applications ignore the underlying behavioural processes, then the results obtained cannot be meaningfully interpreted.

The analysis of discrete data falls within the broader framework of multivariate data analysis. Table 1 shows a classification of the most important multivariate data analysis models. Even if the distinction between exploratory and explanatory models is not very strict, it is possible to identify members of these two subclasses. The first row of Table 1 refers to exploratory data analysis models and the second to explanatory data analysis models. A further distinction in Table 1 is made with respect to manifest and latent variables involved. In contrast to manifest (measurable) variables, latent variables do not correspond directly to anything which is measurable. They can only be indirectly observed by means of observables (indicators) in a more or less accurate way. Examples of latent variables are permanent income, economic expectations, economic growth, socioeconomic status, quality of life, social aspiration, motivation and attidues. Furthermore, multivariate models can be classified by the types of variables involved. In Table 1 a distinction is made only between metric and discrete variables.

In the present section, the emphasis will mainly be put on statistical problems inherent in explanatory models for discrete variables. There is a wide variety of models for dealing with such problems. Those which have become most widely used in the case of manifest variables are linear logistic regression/logit models and asymmetric log-linear models. Whereas the latter class is restricted to cross-classified discrete data problems, logistic regression and logit models are also related to problems in which the independent variables are metric in nature. These models may be considered as a special case of the more general quantal response models to which also the probit models belong. Truncated regression and censored regression (tobit) models are examples of limited-dependent variable discrete models in which the dependent variables are limited to their range due to some underlying stochastic choice mechanism.

Finally, (restricted and unrestricted) latent class models analyse the relationships among a set of discrete manifest and latent variables. In the most general case, the latent variables are related to each other (structural relationships) and to a set of manifest variables (measurement relationships). It is noteworthy that latent class models can be regarded as a discrete analogue to structural equation models with latent variables in the metric case.

Table 1. Classes of Multivariate Data Analysis Models

	Manifest Variables		Latent and Manifest Variables	
	metric	discrete	metric	discrete
Exploratory Data Analysis Models	cluster analysis, multidimensional scaling	nominal and ordinal cluster analysis, symmetric log-linear modelling, correspondence analysis, homogeneous scaling	factor analysis	nominal and ordinal factor analysis
Explanatory Data Analysis Models	conventional regression models	quantal response models (including linear logistic/logit and probit regression models), asymmetric log-linear models, censored and truncated regression models	structural equation models with latent variables (LISREL and PLS approaches), confirmatory factor analysis	(restricted and unrestricted latent class models

In recent years, much progress has been made in integrating different models for discrete data analysis into a more general framework, the so-called generalised linear models (GLM) approach designed by Nelder and Wedderburn (1972) and implemented in the GLIM (Generalised Linear Interactive Modelling) computer package. The class of GLM's is obtained by extending conventional regression models by allowing a distribution function from an exponential family and a link function relating mean and linear predictor. This approach provides a unifying methodology that fits all members of the GLM class by means of a common unified estimation procedure, the so-called iterative weighted least squares procedure. GLM's for discrete data include linear logit models, probit models and log-linear models as special cases. By enlarging the generalised linear model methodology by adding an additional composite link function defined by Thomson and Baker (1981), one may develop a model in which the expected value of an observation depends on more than one linear predictor. New models for ordinal data (such as the proportional odds and the proportional hazards models as suggested by McCullagh (1980), as well as latent class models) can also be embraced by the GLM framework (see also Arminger 1984). In the next section, the family of log-linear models will first be discussed, with special emphasis on the asymmetric case in which an explicit distinction is made between dependent and independent variables.

2.2 Log-linear Models

In fact, log-linear models closely resemble the regression and analysis of variance models for normal metric data: they are linear in the logarithms of the expected cell frequencies of a p-dimensional contingency table. A treatment of the theory of log-linear models can be found in Bishop et al. (1976), Haberman (1978, 1979) and Fienberg (1981), where also procedures for handling structural zeros in incomplete contingency tables and contingency tables with ordered categories are discussed. When all variables in a cause-effect relationship are discrete in nature, the sample data can be displayed in the form of an asymmetric contingency table where one dimension is treated as a dependent variable.

Let us now consider the general form of a (symmetric) log-linear model for a 3-dimensional I x J x K table. Suppose that the total of the counts is n. Let n_{ijk} be the observation for entry (i,j,k) in the contingency table and let m_{ijk} denote the corresponding expected value for cell (i,j,k), given some parametric model.

Then the most general log-linear model takes the form

$$\ln m_{ijk} = \ln E(n_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)} \quad (1)$$

subject to the usual ANOVA-like constraints

$$\sum_{i=1}^{I} u_{1(i)} = \sum_{j=1}^{J} u_{2(j)} = \sum_{k=1}^{K} u_{3(k)} = 0 \quad (2)$$

and

$$\begin{aligned} &\sum_{i=1}^{I} u_{12(ij)} = \sum_{j=1}^{J} u_{12(ij)} = \sum_{i=1}^{I} u_{13(ik)} = \\ &\sum_{k=1}^{K} u_{13(ik)} = \sum_{j=1}^{J} u_{23(jk)} = \sum_{k=1}^{K} u_{23(jk)} = \\ &\sum_{i=1}^{I} u_{123(ijk)} = \sum_{j=1}^{J} u_{123(ijk)} = \sum_{k=1}^{K} u_{123(ijk)} = 0 \end{aligned} \quad (3)$$

The usual convention for log-linear models implies that u represents the overall mean effect, $u_{1(i)}$, $u_{2(j)}$ and $u_{3(k)}$ main effects, $u_{12(ij)}$, $u_{13(ik)}$ and $u_{23(jk)}$ first-order interaction effects and $u_{123(ijk)}$ a second-order interaction effect. Clearly, extensions of (1) - (3) to higher-order contingency tables is straightforward. It should be added that the GLIM computer package uses different constraints leading to cornered effects instead of centralised ones.

If the general model (1) - (3) imposes no restrictions on (m_{ijk}), it is called saturated. Then it has as many independent parameters as there are entries in the table. If some restrictions are imposed on the parameters (i.e., by equating different u-terms zero or by assigning them some a priori specified values), a wide range of unsaturated and hybrid log-linear models describing the relationship between the dependent and independent variables can be specified. Hybrid models may also be particularly useful in the case of underidentification problems, for instance, age-period-cohort analysis (see Fienberg and Mason 1978). The number of possible log-linear models increases as the number of dimensions of a contingency table increases. The question should be answered on the basis of the principle of parsimony and its goodness-of-fit to the data at hand. In this case, there are different approaches which are based, e.g. on F statistics, partitions of of the likelihood-ratio chi-square, standardised values of the parameter estimates, residual analysis and so forth.

When log-linear models are used for explanatory purposes, no additional analytical problems will arise. In that case, the marginals of the contingency table corresponding to the independent variables have to be treated as fixed, so that the product-multinomial sampling scheme with independent multinomial samples for each independent by independent variable combination has to be taken. Then log-linear models can be employed to assess the effects of the independent variables on the dependent ones. Moreover, also the interrelationships between the dependent variables can be determined by log-linear models. It should be added that there are also alternative strategies for estimating the model parameters and the expected cell frequencies: the iterative proportional fitting procedure, combined with the linear constraints on the design method (see Aufhauser 1984, among others), the non-iterative weighted least squares procedure described by Grizzle et al. (1969) and the iterative weighted least squares procedure suggested by Nelder and Wedderburn (1972).

In recent years, symmetric log-linear models and - to some extent - also asymmetric log-linear models have become increasingly popular among geographers (see for instance, Aufhauser and Fischer 1984, Bahrenberg et al. 1984, Fingleton 1981, 1983, Willekens 1982, Willekens and Güvenç 1984 and Wrigley 1979, 1981, 1984a, b). In addition, Willekens (1984) could show that log-linear modelling has also a great potential for analysing spatial dyad structures (see also Aufhauser 1984). It is worth noting that conventional spatial interaction models can be reformulated as log-linear models. When log-linear models are applied in a straightforward manner to spatially dependent problems, standard model selection procedures (such as Brown's (1976) screening procedure and Aitkin's (1979) simultaneous test procedure) may erroneously detect interaction effects between variables which are spurious due to the spatial dependence of the measurements. An important step forward in adopting spatial dependence in the inferential process of choosing an adequate log-linear model has been undertaken by Fingleton (1983). Altogether, the conclusion can be drawn that log-linear models are becoming an increasingly powerful element in a geographer's toolbox.

2.3 Linear Logistic Regression and Linear Logit Regression Models

The present section will give a concise survey of linear logistic and linear logit models. The simplest form of these models is one in which the dependent variable is binary. Let $y_1, \ldots, y_I$ be statistically independent measurements of a binary random variable which may adopt values 0 or 1 with

$$\text{Prob}(y_i=1) = p(1/i) = p_{i1} \qquad i=1, \ldots, I \tag{4}$$

$$\text{Prob}(y_i=0) = p(0/i) = p_{i0} \qquad i=1, \ldots, I \tag{5}$$

Furthermore, let x_{ik} ($i=1, \ldots, I$; $k=1, \ldots, K$) represent the values of K independent variables for y_i. Now the <u>general linear logistic regression model for a binary dependent variable</u> has the following form

$$p(1/i) = \frac{\exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_{ik}\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_{ik}\right)} \tag{6}$$

and

$$p(0/i) = \frac{1}{1 + \exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_{ik}\right)} = 1 - p(1/i) \tag{7}$$

The parameters β_k($k=0, 1, \ldots, K$) are unknown while the independent variables may be discrete and/or metric. Independent discrete variables are included in (6) - (7), for instance, by adopting the principle of dummy coding.

The maximum likelihood (ML) equations which are necessary to estimate the parameters from (6) - (7) can be found by setting the minimum sufficient statistics $\sum_{i=1}^{I} y_i$ and $\sum_{i=1}^{I} x_{ik} y_i$ equal to their expected values:

$$\sum_{i=1}^{I} y_i = \sum_{i=1}^{I} \hat{p}(1/i) \tag{8}$$

and

$$\sum_{i=1}^{I} x_{ik} y_i = \sum_{i=1}^{I} x_{ik} \hat{p}(1/i) \qquad k=1, \ldots, K \tag{9}$$

The solutions for the (K+1) β's in (8) and (9) can be found by means of an iterative weighted least squares approach (see Nelder and Wedderburn 1972, Nerlove and Press 1973, Fienberg 1981). When confronted with a situation in which some or all independent variables are metric, it is not possible to carry out an overall goodness-of-fit or to use an easily interpretable criterion (such as R^2) for assessing the predictive power of the estimated model. By categorising the metric variables, one may then construct a corresponding logit regression model whose goodness-of-fit can be evaluated.

The logistic regression model (6) - (7) can also be reformulated and then shown to be equivalent to a model that is linear in the logits:

$$\ln \frac{p(1/i)}{1-p(1/i)} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} \qquad (10)$$

The left hand side in (10) is essentially a transformation of probabilities that is known as the logit transformation. Thus, this model is termed a linear logit model. This linear logit model is a discrete data representation which closely resembles the regression and analysis of variance models for metric variables and may be considered as their discrete variant.

ML-estimates for the logistic/logit models can be obtained through the iterative proportional fitting procedure or through an iterative weighted least squares procedure based on the Newton-Raphson or the David-Powell algorithm. In the case of large samples, also the non-iterative weighted least squares (GSK) approach may be utilised. It can be shown that the GSK-estimator is consistent and asymptotically normal, with the same asymptotic covariance matrix as the ML-estimator in sufficiently large samples. For determining whether the logit models provide an adequate fit to the data at hand, a goodness-of-fit measure based on the likelihood ratio statistics may be taken. For details on this issue see Maddala (1983, pp. 39); other goodness-of-fit measures are discussed in Amemiya (1981).

Superficially, linear logit and log-linear models look quite different. It is however, possible to show that every logit model with discrete independent variables can be represented as an equivalent log-linear model (cf. Fienberg 1981, p. 78). Thus, all results achieved for log-linear models also apply to linear logit models. Nevertheless, logit models when they are formulated in terms of log adds or a logit scale are often easier to interpret than log-linear models. In addition, it would often be rather inefficient to use estimation algorithms which are appropriate for the general log-linear model instead of those especially developed for linear logit models because of the presence of sampling constraints in the logit formulation (see for details also Fienberg and Meyer 1983).

Logit models are also members of a more general class of models which are known as quantal response models. An extensive treatment can be found in Finney (1971) (see also Haberman 1978). A quantal response model can be characterised by its distribution function F with an associated inverse function. If F is the logistic function, the logit model is obtained:

$$F(-\beta_o - \sum_{k=1}^{K} \beta_k x_{ik}) = \exp(-\beta_o - \sum_{k=1}^{K} \beta_k x_{ik})/(1+\exp(-\beta_o - \sum_{k=1}^{K} \beta_k x_{ik}) \quad (11)$$

The advantage of this distribution function is that it has a closed form expression in contrast to, for example, the normal one:

$$F(-\beta_o - \sum_{k=1}^{K} \beta_k x_{ik}) = \int_{-\infty}^{-\beta_o - \sum_{k=1} \beta_k x_{ik}} \frac{1}{(2\pi)^{\frac{1}{2}}} \exp(-\tfrac{1}{2}t^2)\, dt \quad (12).$$

which leads to the probit model, another example of a quantal response model. Unless the samples are rather large and the observations at the tails of the distribution have a large influence, probit and logit analysis will yield similar results. Due to its computational tractability and its mathematical simplicity, logit analysis has been preferred to probit analysis, for example in travel demand analysis. It is worth noting that when F corresponds to a continuous distribution, the results obtained for logit analysis can be readily applied to general quantal response models.

The models discussed so far may easily be generalised to handle polytomous dependent variables. In the case of an l'-polytomous dependent variable the linear logistic regression model takes the form:

$$p(l/i) = p_{il} = \frac{\exp \quad (\beta_{0l} + \sum_{k=1}^{K} \beta_{kl} x_{ik})}{1+\sum_{\tilde{l}=1}^{L} \exp\ (\beta_{0\tilde{l}} + \sum_{k=1}^{K} \beta_{k\tilde{l}} x_{ik})} \quad l = 1, \ldots, L \quad (13)$$

and

$$p(l'/i) = p_{il'} = \frac{1}{1+\sum_{\tilde{l}=1}^{L} \exp\ (\beta_{0\tilde{l}} + \sum_{k=1}^{K} \beta_{k\tilde{l}} x_{ik})} \quad l' = L+1 \quad (14)$$

where p_{il} represents the probability that the l-th category of the dependent variable with l' categories will be selected, given the values of K independent variables.

The dichotomous logit model (10) can also be extended to the polytomous case in a straightforward manner:

$$\ln \frac{p_{il}}{p_{il'}} = \beta_{0l} + \sum_{k=1}^{K} \beta_{kl} \, x_{ik} \qquad (15)$$

where L logits are taken into account. Both (13) - (14) and (15) are models that are only appropriate for polytomous nominal dependent variables. In case of an ordinal dependent variable it is convenient to use a set of logits based on continuation ratios (Fienberg 1981, p. 110):

$$\ln \frac{p_{il}}{\sum_{\tilde{l}>l} p_{i\tilde{l}}} \qquad (16)$$

It should be noted that this continuation ratio logit approach exhibits the asymptotic independence property which is not shared by the logits defined in (15). This means that if the ML-procedure is used for estimating the parameters in the set of L logit models, the estimation can be pursued separately for each member of the set concerned. The individual chi-square statistics may then be added up to the overall goodness-of-fit statistic for the whole set of models. Furthermore, the L models can be assessed independently from each other.

In recent years, dichotomous and polytomous (single equation) linear logit models have been applied in various cases, particularly in transportation analysis (for instance, modal choice problems in work and shopping trips); examples can be found in Bahrenberg et al. (1984). Special formulations of the linear logit model approach which are of particular interest for geography and regional science can be obtained by specifying discrete variable analogues to both space-time forecasting models and to the traditional trend surface model which became known under the name probability surface model (see Wrigley 1977).

Finally, it has been demonstrated by Goodman (1973), Heckman (1978) and Arminger (1983) that <u>discrete recursive simultaneous system models with discrete dependent variables</u> which allow a simultaneous assessment of all the interrelationships between a set of variables, are a natural extension of the discrete single-equation models described above.

In addition, if data are collected which are related to a number of discrete time points, some variables may sometimes be simultaneously both dependent and independent variables within the system at hand, as they may be considered as dependent on the temporally preceding variables and as independent with respect to the variables which temporally succeed them (see Goodman 1973). A more detailed analysis of recent developments in discrete non-recursive simultaneous system modelling can be found in Manski and McFadden (1981a) where, for example, special attention is given to mixed discrete linear systems which do not have a linear reduced form and thus yield specific difficulties for parameter estimations. The development of dynamic discrete simultaneous system modelling which would imply a significant progress compared to the static modelling of present discrete multivariate data analysis, is a logical follow-up of promising research efforts made in the recent past.

2.4 Censored and Truncated Regression Models

The abovementioned explanatory models were based on discrete values for the variables at hand. A specific class of models is made up by regression models in which the dependent variable is observed in only some of the ranges. This situation may lead to censored and truncated regression models (see Maddala 1983).

A censored sample of observations $y_1, \ldots, y_I$ emerges if we only record the values of the observations that are larger than a predetermined constant c, i.e.,

$$\left.\begin{array}{ll} y_i = y_i , & \text{if } y_i > c \\ y_i = c , & \text{if } y_i \leq c \end{array}\right\} \tag{17}$$

A truncated sample is obtained if the distribution of the y_i's $(i=1, \ldots, I)$ is cut off at point c, so that no observations are drawn for $y \leq c$.

Clearly, we may also have doubly truncated, doubly censored and mixed truncated-censored samples as well. A special class of censored models is the so-called tobit model, which is defined as follows:

$$\left.\begin{array}{ll} y_i = \beta' x_i + \varepsilon_i , & \text{if } \beta' x_i + \varepsilon_i > 0 \\ y_i = 0 \quad \text{otherwise} & \end{array}\right\} \tag{18}$$

where y_i is a known value of a dependent variable, β a vector of unknown regression coefficients, x_i a vector of unknown observations and ε_i a vector of disturbances.

The estimation of such models can be based on (iterative) maximum likelihood procedures, a reparametrisation of the tobit model, a two-stage estimation procedure, or a dummy variable approach. These models can also be generalised for a situation of stochastic and unobserved thresholds.

Truncated regression models have also been studied quite extensively in the literature, for instance, in the case of an earnings equation estimated from data for the negative-income-tax experiment. In this case, adjusted maximum likelihood procedures can be used to estimate the parameters of a truncated regression model. Such truncated models can be extended toward cases with endogenous stratification.

2.5 Latent Class Models

Latent class models aim at analysing the relationships between a set of discrete manifest and latent variables. Latent class models can be considered as a discrete analogue to structural equation models with latent variables. More extensive discussion of this topic can be found in Goodman (1978), Haberman (1979) and Bartholomew (1983), among others.

There is a great variety of latent class models. In most models the assumption is made that the manifest variables are conditionally independent given the latent variable(s). In principle, two broad categories of latent class models may be distinguished, viz. unrestricted and restricted models. We will first consider the unrestricted case.

Assume without loss of generality a general restricted latent class model with three discrete manifest variables and a discrete latent variable. Let $A = (A_i)$ with $i=1, \ldots, I$; $B = (B_j)$ with $j=1, \ldots, J$ and $C = (C_k)$ with $k=1, \ldots, K$ denote three (dichotomous or polytomous) discrete manifest variables. These 3 variables constitute a 3-way contingency table which cross-classifies a sample of n individuals with respect to A, B and C. Suppose that there is a (polytomous) discrete latent variable X with T categories (termed latent classes) which is able to explain the relationships among the variables in the 3-way contingency table. Assume also that each individual in the population concerned belongs to one and only one of the T latent classes. Then:

$$p_{ijk}^{ABC} = \sum_{t=1}^{T} p_{ijkt}^{ABCX} \tag{19}$$

where $p_{ijk}^{ABC} = p(A=i,\ B=j,\ C=k)$ is the probability of an event being in cell (i,j,k) and p_{ijkt}^{ABCX} the (unknown) probability of an observation belonging to cell (i,j,k,t) of the four-way table.

Next, the assumption of local independency (i.e., the manifest variables are conditionally independent given the latent variable X) leads to the following latent class model:

$$p_{ijkt}^{ABCX} = p_{it}^{A|X}\ p_{jt}^{B|X}\ p_{kt}^{C|X}\ p_{t}^{X} \tag{20}$$

where

$$p_{t}^{X} = p(X=t) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} p_{ijkt}^{ABCX} \tag{21}$$

is the probability of an event being in latent class t and $p_{it}^{A|X}$ the conditional probability of category i of A, given the value t of X . $p_{jt}^{B|X}$ and $p_{kt}^{C|X}$ are defined analogously. The usual restrictions such as

$$\sum_{t=1}^{T} p_{t}^{X} = 1;\ \sum_{i=1}^{I} p_{it}^{A|X} = 1;\ \sum_{j=1}^{J} p_{jt}^{B|X} = 1;\ \sum_{k=1}^{K} p_{kt}^{C|X} = 1 \tag{22}$$

are assumed to hold. Then $(p_{t}^{X},\ p_{it}^{A|X},\ p_{jt}^{B|X},\ p_{kt}^{C|X})$ denotes the vector of parameters in the general latent class model. When the number T of latent classes is unknown, T has to be determined within the analysis by considering various possible values for T.

In the simple latent class model (20) - (22) the manifest variables A, B and C are affected by the latent variable X, but the manifest variables do not have any direct effects upon each other. In other words it can be said that the latent variable X explains the relationships among the manifest variables. Clearly, (20) defines a log-linear model for the unknown probabilities p_{ijkt}^{ABCX} . Since the X_t's are unknown, the frequencies n_{ijkt} are not known either, so that p_{ijkt}^{ABCX} cannot be estimated by means of the cell frequencies (like in the ordinary log-linear model). The ML-estimates satisfy the following conditions:

$$n\ \hat{p}_t^X = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} n_{ijk}\ \hat{p}_{tijk}^{X|ABC} \tag{23}$$

$$n\ \hat{p}_{it}^{A|X} = \frac{\sum_{j=1}^{J} \sum_{k=1}^{K} n_{ijk}\ \hat{p}_{tijk}^{X|ABC}}{\hat{p}_t^X} \tag{24}$$

$$n\ \hat{p}_{jt}^{B|X} = \frac{\sum_{i=1}^{I} \sum_{k=1}^{K} n_{ijk}\ \hat{p}_{tijk}^{X|ABC}}{\hat{p}_t^X} \tag{25}$$

$$n\ \hat{p}_{kt}^{C|X} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} n_{ijk}\ \hat{p}_{tijk}^{X|ABC}}{\hat{p}_t^X} \tag{26}$$

Now the conditional probabilities $\hat{p}_{tijk}^{X|ABC}$ can be expressed in terms of the $\hat{p}_{ijk}^{ABC}$ and $\hat{p}_{ijkt}^{ABCX}$ as follows:

$$\hat{p}_{tijk}^{X|ABC} = \hat{p}_{ijkt}^{ABCX} \ / \ \hat{p}_{ijk}^{ABC} \tag{27}$$

with

$$\hat{p}_{ijk}^{ABC} = \sum_{t=1}^{T} \hat{p}_{ijkt}^{ABCX} \tag{28}$$

Here n_{ijk} is the observed cell frequency and the circumflex denotes the ML-estimates of the corresponding probabilities given the observed marginal totals. Because of condition (22) only T-1 of the p_t^X, I-1 of the $p_{it}^{A|X}$, J-1 of the $p_{jt}^{B|X}$ and the K-1 of the $p_{kt}^{C|X}$ have to be estimated. There are various possibilities to estimate these probabilities. For instance, the iterative proportional fitting procedure can easily be adapted for solving the ML-equations. In addition, Fisher's scoring procedure, a variant of the Newton-Raphson procedure, can be applied to any latent class model based on a log-linear model. In this latter case, latent class modelling can be considered as a special case of GLM's with a composite link function defined by Thomson and Baker (1981) (see also Arminger 1984). More details on estimation problems can be found in Goodman (1978, p. 542). Despite a general resemblance of the ML-equations of this general latent class to those for ordinary log-linear models, there are many difficulties in contrast to ordinary log-linear modelling, especially because ML-estimates are not uniquely determined (leading to multiple solutions or to solutions which are not ML-estimates). Thus, special attention must be paid to identification problems of an ML-estimate (cf. Haberman 1979, p. 542).

Like in confirmatory factor analysis, restricted latent class models are defined either by specific patterns of a priori fixed values of the conditional probabilities, or by equality restrictions of conditional probabilities in the same latent class (such as e.g. $p_{11}^{A|X} = p_{21}^{B|X}$), or by equality restrictions of conditional probabilities in different latent classes (such as e.g. $p_{11}^{A|X} = p_{12}^{A|X} = p_{23}^{B|X}$). The estimation procedures mentioned above can be modified in a straightforward manner for such restricted latent class models.

The most general (restricted and unrestricted) latent class models include several latent variables where the latent variables are related to each other (structural relationships) and to a set of manifest variables (measurement relationships). In this case specific care must be taken that the parameters are identifiable. Such latent class models are discussed in Goodman (1978). Muthén (1979) presents a simultaneous equation system with dichotomous manifest variables of metric latent variables which makes use of the probit transformation. Furthermore, it is worth noting that the LISREL model approach originally designed for metric variables may be extended by means of tetrachoric, polychoric and polyserial correlations in order to cover dichotomous, ordered polytomous and metric manifest variables of latent variables.

2.6 Generalised Linear Models as a Unifying Framework

The 'generalised linear model' (GLM) approach (see Nelder and Wedderburn 1972) provides a framework for integrating the discrete data models discussed in 2.2 - 2.4 and also links them with conventional linear metric data models. A detailed discussion of this GLM approach can be found in Nelder and Wedderburn (1972), Nelder (1974), Arminger (1982), McCullagh and Nelder (1983), and O'Brien and Wrigley (1984), among others. Here only a concise presentation will be given.

Assume that $y = (y_1, \ldots, y_I)$ is a dependent variable whose distribution is one of the exponential family of probability distributions with density function:

$$f(y;\ \Theta,\Phi) = \exp(|y\ \Theta - b(\Theta)|/\ a(\Phi) + c(y,\Phi)) \qquad (29)$$

where $a(\Phi)$, $b(\Theta)$ and $c(y,\Phi)$ are suitable monotonic functions. Θ is the canonical parameter of the exponential family and Φ is the (fixed, non-negative) scale parameter. Table 2 lists some examples of specific members of the exponential class of probability distributions.

Now a <u>generalised linear model</u> may be expressed as follows:

$$y_i = g^{-1}\ (\eta_i) + \varepsilon_i \qquad i=1,\ \ldots,\ I \qquad (30)$$

Here $\varepsilon_i (i=1,\ \ldots,\ I)$ denotes a <u>randomly distributed error</u>, while

$$\eta_i = \sum_{k=1}^{K} \beta_k\ x_{ik} \qquad (31)$$

is the <u>linear predictor</u> (linear-additive hypothesis), with parameter vector $\beta = (\beta_k)$, $k=1,\ \ldots,\ K$ and observations x_{ik} $(i=1,\ \ldots,\ I;\ k=1,\ \ldots,\ K)$ on the K (metric and/or discrete) independent variables. $g^{-1}\ (\eta_i)$ is the inverse of the following (monotonic twice differentiable) <u>link function</u>:

$$\eta_i = g(\mu_i) \qquad (32)$$

which relates the linear predictor to the theoretical mean

$$\mu = (\mu_1,\ \ldots,\ \mu_I) = E(y_1,\ \ldots,\ y_I) = b'(\Theta_1,\ \ldots,\ \Theta_I), \qquad (33)$$

where b' is the derivative of b with regard to the parameter vector.

<u>Table 2</u>: Characteristics of some specific probability distributions of the exponential family (see Arminger 1982, p. 20)

Characteristics	Normal	Poisson	Binomial
range of y	$(-\infty,\ \infty)$	0, 1, 2, ...	(0, 1)
$a(\Phi)$	Φ	1	$\frac{1}{I}$
$b(\Theta)$	$\frac{1}{2}\ \Theta^2$	e^Θ	$\ln(1+e^\Theta)$
$c(y,\Phi)$	$-\frac{1}{2}\left(\frac{y^2}{\Phi} +\ \ln\ 2\pi\Phi\right)$	$-\ln\ y\ !$	$\ln\binom{I}{Iy}$
$\mu=E(y) = b'(\Theta)$	Θ	e^Θ	$\frac{e^\Theta}{1+e^\Theta}$
Variance function $b''(\mu)$	1	μ	$\mu(1-\mu)$

In order to specify a certain GLM, it is necessary to define explicitly the following three components: linear predictor, link function and error distribution.

The specification of the linear predictor is determined by the process of data generation and the experimental design of the analysis. Table 3 outlines some standard members of the GLM family characterised by its specific link function and error distribution. There is at present a wide variety of link functions. In the case of discrete data models, appropriate link functions are provided by the logit transformation (such as in the case of the asymmetric log-linear model and of the linear logit regression model), by the probit transformation (such as in the case of the probit regression model) or the logarithmic transformation (such as in the case of the symmetric log-linear model). It is noteworthy that other variants of link functions are possible (such as e.g. the identity function in the case of linear metric data regression models, or the complementary log-log transformation), while they may also be specified in the GLM context.

The exponential family of possible density functions may include many continuous and discrete probability functions. The Poisson distribution plays the same central role in discrete data analysis as the normal distribution in metric data data analysis. Generalisations of the Poisson distribution lead to the binomial distribution (if the number of categories of the dependent variable is two) and to the multinomial distribution which can be considered as constrained Poisson distributions. The error components of the asymmetric log-linear, the logit regression and the probit regression models are either defined by the binomial or by the multinomial distribution (see Table 3).

Table 3: Examples of generalised linear models

Model	Link function	Error Distribution
classical linear regression	identify: $\eta = \mu$	normal
symmetric log-linear	logarithmic: $\eta = \ln \mu$	Poisson
asymmetric log-linear	logit: $\eta = \ln \frac{\mu}{1-\mu}$	binomial or multinomial
logit regression	logit: $\eta = \ln \frac{\mu}{1-\mu}$	binomial or multinomial
probit regression	probit: $\eta = F^{-1}(\mu)$ where F is the cumulative normal distribution function	binomial or multinomial

It should be added that Nelder and Wedderburn (1972) have shown that ML-estimates can be obtained for all subclasses of GLM's by using the iterative weighted least squares procedure (based upon the Newton-Raphson algorithm) in order to maximise the log-likelihood function L of (29) which can be expressed in canonical form as:

$$L(\Theta,\Phi|y) = \sum_{i=1}^{I} L_i(\Theta_i,\Phi|y) = \sum_{i=1}^{I} \frac{(y_i\Theta_i - b(\Theta_i))}{a_i(\Phi)} + c(y_i,\Phi) \tag{34}$$

From (34) the maximum likelihood equations for β can be derived as:

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{I} \frac{\partial L_i}{\partial \beta} = \sum_{i=1}^{I} \left(\frac{\partial\, L_i}{\partial\, \beta_k}\right) \tag{35}$$

$$= \sum_{i=1}^{I} \frac{d\, L_i}{d\, \Theta_i} \cdot \frac{d\, \Theta_i}{d\, \mu_i} \cdot \frac{d\, \mu_i}{d\, \eta_i} \left(\frac{\partial\, \eta_i}{\partial\, \beta_k}\right)$$

$$= \sum_{i=1} \left(\frac{y_i - b'(\Theta_i)}{a_i(\Phi)}\right) \cdot \frac{a_i(\Phi)}{\text{var}(y_i)} \cdot \frac{d\, \mu_i}{d\, \eta_i} x_{ik} =$$

$$= \sum_{i=1}^{I} \frac{y_i - \mu_i}{\text{var}(y_i)} \frac{d\, \mu_i}{d\, \eta_i} x_{ik} = 0$$

with

$$\text{var}(y_i) = E(y_i - \mu_i)^2 = b''(\Theta_i)\; a_i(\Phi) \tag{36}$$

Note that the solution of the ML-equations is equivalent to the iterative weighted least squares approach with a weighting function

$$w_i = \left(\frac{d\mu_i}{d\eta_i}\right)^2 \cdot \frac{1}{\text{var}(y_i)} \tag{37}$$

Expression (35) can be simplified for canonical links. If:

$$\Theta_i = \eta_i \tag{38}$$

then:

$$\frac{\partial\, L}{\partial\, \beta_k} = \sum_{i=1}^{I} \frac{y_i - \mu_i}{a_i(\Phi)} x_{ik} = 0 \tag{39}$$

and consequently:

$$\sum_{i=1}^{I} y_i\, x_{ik} = \sum_{i=1}^{I} \mu_i\, x_{ik} \tag{40}$$

has to be valid for the ML-solution. Furthermore, it can be shown that:

$$\sum_{i=1}^{I} \frac{1}{a_i(\Phi)} \; y_i \; x_{ik} \tag{41}$$

are sufficient statistics for β_k. More details on estimating problems of GLM's can be found in Nelder and Wedderburn (1972) and Arminger (1982).

Recently some progress has been made to extend the GLM approach by integrating composite link function models, quasi-likelihood models and mixture models. In contrast to the simple form of the link function by a (1,1) relationship between μ_i and η_i (see equation (32)), composite link functions allow each μ_i to be a linear combination of some intermediate quantities which are themselves functions of η_i. Uses of composite link functions include latent class modelling as well as the proportional odds and the proportional hazards models suggested by McCullagh (1980) for ordinal data (see Arminger 1984 and Nelder 1984). Uses of comlikelihood models the distribution is not fully defined. It is only assumed that the value of the variance is related to the mean.

The quasi-likelihood function $D(y_i, \mu_i)$ for the i-th observation is defined as

$$\frac{\partial D(y_i,\mu_i)}{\partial \mu_i} = \frac{y_i-\mu_i}{var(\mu_i)} \tag{42}$$

Wedderburn (1974) shows that $D(y_i,\mu_i)$ has many of the properties of a log-likelihood function. If the distribution of y is a one-parameter exponential one, then D and L are identical. In mixture models it is assumed that the error distribution is a mixture of several components rather than being homogeneous. The mixture may be continuous or discrete. In the latter case the number of components may be known or unknown (Nelder 1984). Flowerdew and Aitkin's (1982) compound Poisson migration model may serve as an example of a mixture model in a spatial context.

2.7 Conclusion

The area of discrete data analysis is marked by fast dynamics. A major issue in the context of geographical research is the integration of categorical data and spatiotemporal autocorrelation. Recently, some interesting developments have taken place in this field. For example, Odland and Barff (1982) have linked the logic of existing space-time interaction tests to categorical data models in order to model the space-time patterns of urban housing deterioration. Moreover, Fingleton (1983) has explored some fundamental features of complex categorical data sample survey designs in order to include spatial dependence effects in the context of log-linear modelling. He could also show that considerable care is necessary when log-linear models are applied to spatially dependent data, as then standard model selection procedures may under certain circumstances erroneously detect interaction effects between variables which are spurious as a consequence of the spatial dependence of the measurements. So far, the number of applications of categorical data in a spatial dependence context is limited, although the need for combining discrete data models with spatially dependent data is increasing. Hence, this is an important field of future spatial data research.

3. DISCRETE CHOICE ANALYSIS

3.1 Introduction

Spatial choice analysis has already a long tradition in human geography and regional science in so far as these disciplines focussed attention on the location of firms, the structure of migration patterns, etc. Location-allocation models based on programming theory have played an important role in aggregate spatial choice and interaction analysis. In the seventies, discrete choice models have come to the fore (see e.g. Bahrenberg et al. 1984). In the context of discrete choices - defined as decisions among a finite set of discrete alternatives such as transportation modes or routes, labour force participation, residential and work location, industrial location or relocation, recreational trips, etc. - conventional marginalist microeconomic consumer theory is not applicable.

Discrete choice models have been highly inspired by the work of Thurston (1927) and Luce (1959). They specify choice probabilities for each element from a finite set of discrete alternatives among which an actor can choose. The choice probabilities are assumed to conform to the basic rules of probability; they depend only on measured attributes of alternatives and on some socioeconomic characteristics of the choice maker. Spatial discrete choice models differ from general discrete choice models only with respect to the fact that alternatives and decision makers (choice makers, individuals) are distributed - and thus interrelated - over space. Thus a spatial choice context is characterised by spatially dependent alternatives (for instance, housing units in choosing a residential location or jobs in choosing workplace).

In recent years an extensive body of methodological research on discrete choice problems came into being with special emphasis on developing utility based discrete choice models. Such models emphasise that the decision is made at the individual level on the basis of a utility-maximising principle. The decision is based on an evaluation of foreseeable utilities attached to each alternative which constitute a basis for estimating the choice probabilities. It is assumed that the alternative with the highest utility is chosen. A considerable part of the theoretical underpinnings of utility based discrete choice models is rooted in the Lancastrian version of consumer choice theory and in psychological theories on choice behaviour.

Individual discrete choice analysis may be either deterministic or stochastic (random). In contrast to deterministic approaches, random utility models consider the choice of an alternative by an individual actor as a random decision. They are able to take into account two aspects: (i) a decision maker chooses normally an alternative in a situation of bounded rationality and (ii) certain choice relevant attributes are unobserved or omitted by the analyst.

Subsection 3.2 will outline the basic idea of random utility modelling within a discrete choice context. In subsection 3.3 some classic random utility based discrete choice models will be discussed, while subsection 3.4 is devoted to some more recent modelling developments. These subsections are mainly based on Fischer and Maier (1984).

3.2 Discrete Choice Analysis and Random Utility Modelling

In general, a discrete choice problem can be identified by an exhaustive finite set of disjoint choice alternatives or possibilities:

$$A = (1, \ldots, A'), \qquad (43)$$

a population of individuals (actors or choice makers)

$$I = (1, \ldots, I'), \qquad (44)$$

a space Z of observable attributes including measured attributes of alternatives and individual characteristics, a probability density p(z) with

$z \in Z$ characterising the distribution of attributes in I and a choice probability function $P(a|Z)$ specifying the conditional probability of choosing alternative $a \in A$, given the attributes $z \in Z$. Usually $P(a|z)$ is a priori specified up to a parameter vector Θ. Assuming that there is a priori knowledge of a structure relation from z to a, the frequency distribution $f(a,z)$ in I can be described as

$$f(a,z) = P(a|z,\Theta) \cdot p(z) \qquad (a,z) \in (A \times Z) \qquad (45)$$

where the choice probability function has to satisfy the following conditions

$$0 \leq P(a|z,\Theta) \leq 1 \qquad (a,z) \in (A \times Z) \qquad (46)$$

and

$$\sum_{a \in A} P(a|z,\Theta) = 1 \qquad z \in Z \qquad (47)$$

It is noteworthy that it is not necessary to assume that all choice makers face the same choice set A, because the fact that an alternative is not available to one of the individuals can be taken into acount by means of the z-value, so that consequently the corresponding choice probability is equal to zero.

Utility based discrete choice models take for granted that decisions are made at the individual level on the basis of the principle of a maximisation of a utility function. In general, it is assumed that there exists a distribution of utility functions from which the utility function U associated with $(i,a) \in (I \times A)$ is fixed but latent. In a formal sense, this utility function

$$u_{ia} = U(x_i, y_a) , \qquad (i,a) \in (I \times A) \qquad (48)$$

denotes the (subjective) utility of alternative a to the i-th choice maker where x_i is a vector of attributes characterising individual i and y_a a vector of attributes of alternative a with

$$z_{ia} = (x_i, y_a) \qquad (i,a) \in (I \times A) \qquad (49)$$

Note that y_a may include an alternative-specific dummy variable. Next, we define $u_{i.}$ as

$$u_{i.} := (u_{ia}, a \in A) \qquad i \in I \qquad (50)$$

Then $u_{i.}$ represents the set of preferences of choice maker i over all relevant alternatives. Alternative a will be selected if and only if:

$$u_{ia} \geq u_{ia'} \qquad (a, a') \in (AxA) \quad a' \neq a \tag{51}$$

In most random utility discrete choice models, it is assumed that the utility u_{ia} of an alternative a for a choice maker i can be partitioned into two components (additive utility hypothesis), viz. a systematic (or deterministic) component:

$$v_{ia} = V(z_{ia}, \beta) \qquad (i, a) \in (IxA) \tag{52}$$

and a random component:

$$\varepsilon_{ia} = \mathcal{E}(z_{ia}, \beta) \qquad (i, a) \in (IxA) \tag{53}$$

The deterministic component accounts for the effects of the measured attributes of both the alternatives and the individuals. The random component (a set of random disturbances) represents the impacts of neglected characteristics affecting the i-th individual's decision as well as measurement errors in the data regarding the attributes.

The most widely used statistical specification of the systematic component in practice is the linear-in-parameters multiattribute model based on the theory of conjoint measurement (see Timmermanns 1984) leading to

$$u_{ia} = z_{ia}\, \beta + \varepsilon_{ia} \qquad (i, a) \in (IxA) \tag{54}$$

The parameter vector β (being constant across the individuals) reflects the tastes of the individuals. The distribution of

$$\varepsilon_{i.} = (\varepsilon_{ia}, a \in A) \qquad i \in I \tag{55}$$

conditional to $z_{i.}$ is specified as a member of a parametric distributional family

$$F(\varepsilon_{i.} \mid z_{i.}, \gamma) \tag{56}$$

As a consequence, the parameter vector Θ is equal to

$$\Theta = (\beta, \gamma) \tag{57}$$

The utility vector (u_{ia}, $a \in A$) has a multivariate probability distribution conditioned on $z_{i.}$ and β. If this distribution or that of V is known, the choice probability that choice maker i with associated attributes

$$z_{i.} := (z_{ia},\ a \in A) \qquad i \in I \tag{58}$$

will choose alternative a among all relevant alternatives is then equal to the probability of drawing a utility vector from this multivariate probability distribution in such a manner that

$$u_{ia} \geq u_{ia'} \qquad \bigwedge_{a' \in A} a' \neq a \in A \tag{59}$$

Thus,

$$\begin{aligned} P(a|z_{i.},\ \beta) &= \mathrm{Prob}(V(z_{ia},\ \beta) + \epsilon(z_{ia},\ \beta) \geq \\ &\geq V(z_{ia'},\ \beta) + \epsilon(z_{ia'},\ \beta);\ a' \in A \bigwedge_{a' \in A} a' \neq a \in A) \\ &= p_{ia} \qquad (i,\ a) \in (I,A) \end{aligned} \tag{60}$$

is the fundamental equation of random utility models. The choice probabilities as defined in (60) depend on the distribution of the error terms vector $\epsilon(z_{ia},\ \beta)$. Taking the linear-in-parameters additive specification of the systematic component (see (54)), (60) leads to

$$\begin{aligned} P(a|z_{i.},\ \beta) &= \mathrm{Prob}(z_{ia}\ \beta + \varepsilon_{ia} \geq z_{ia'}\ \beta + \varepsilon_{ia'}\ , \\ &\qquad \bigwedge_{a' \in A} a' \neq a \in A) = \\ &= \mathrm{Prob}((z_{ia} - z_{ia'})\beta \geq (\varepsilon_{ia} - \varepsilon_{ia'}), \\ &\qquad \bigwedge_{a' \in A} a' \neq a \in A) \qquad (i,\ a) \in (I \times A) \end{aligned} \tag{61}$$

The estimation of parameters in the choice probability function depends very much upon the type of sampling procedure used. Stratified samples have the great potential of reducing sampling costs. In general, two kinds of stratifications may be distinguished: exogenous and endogenous sampling. In exogenous stratified sampling, subsets of individuals with exogenously given attributes are selected (e.g. geographically clustered samples). Endogenous sampling refers to situations in which the sampling process is founded on an endogenous stratification. In mode choice modelling, for example, it is common practice to use data from samples of individuals at car parks, train stations and bus-stops instead of household surveys in order to reduce sampling costs. Such samples are endogenously stratified because the choice of mode of transport is an endogenous variable.

ML- estimation from exogenously stratified samples does not yield new problems in comparison with estimation from random samples (see Cosslett 1981). It is however, more difficult to get consistent and asymptotically efficient estimates of the choice model parameters for choice-based (endogenously stratified) samples, especially in the case of more complicated types of sampling processes involving stratification on exogenous and endogenous variables at the same time. Only recently, progress has been made in developing statistically sound estimators in such a context. Interesting surveys on the state-of-the-art are given in Cosslett (1981), and in Manski and McFadden (1981a); see also Hausman and Wise (1978) and Lerman and Manski (1979).

3.3 Classic Random Utility Models

The specific functional form of a random utility model is particularly based on the distribution of the random disturbances for (60). It has been an important objective in discrete choice analysis to find appropriate distribution functions which lead to computationally attractive choice probabilities and which also assume a sufficiently realistic behavioural basis. In the sequel to this section, the most important classic random utility models will be further discussed. These models are:

(i) the family of conditional or multinomial logit (MNL) models,
(ii) the family of generalised extreme value (GEV) models (and as a special case the subclass of nested multinomial logit (NMNL) models)
and
(iii) the family of multinomial probit (MNP) models.

(i) The family of conditional or multinomial logit models

The hypothesis of independently and identically distributed (IID) random disturbances with the double exponential (Weibull, Gumbel or Gnedenko) distribution leads directly to the family of conditional or multinomial logit models. The cumulative distribution function is equal to

$$F(\varepsilon_{i.}|z_{i.}) = \prod_{a\varepsilon A} \exp(-\exp(-\varepsilon_{ia})) \quad i\varepsilon I \tag{62}$$

implying that choice makers with identically measured attributes have identical tastes and that the correlation between unobservable (or omitted) attributes of alternatives (or choice makers across alternatives) and individuals is zero (see Horowitz 1981, and Hensher and Johnson 1981, p. 105). The choice probabilities defining the family of MNL models can be expressed as:

$$P(a|z_{i.}, \beta) = \exp(V(z_{ia}\ \beta))/\sum_{a'\varepsilon A} \exp(V(z_{ia'}, \beta) \qquad (63)$$

$$(i, a) \ \varepsilon \ (IxA)$$

Given the linear-additive specification of the systematic component, (63) can be formulated as

$$P(a|z_{i.}, \beta) = \exp(z_{ia}\ \beta)/\sum_{a'\varepsilon A} \exp(z_{ia'}, \beta) \qquad (64)$$

$$(i, a) \ \varepsilon \ (IxA)$$

Evidently, the MNL models are obviously in accordance with Luce's (1959) choice axiom of the independence of irrelevant alternatives (IIA). This IIA-axiom states that

$$P(a|z_{i.}, \beta)/P(a'|z_{i.}, \beta) = \exp(V(z_{ia}, \beta)/\exp(V(z_{ia'}, \beta)$$

$$(a, a') \ \varepsilon \ (AxA) \qquad (65)$$

meaning that the relative probabilities of any two alternatives depend only on their systematic components of utility and are independent of other alternatives of the choice set. It can be shown that the IIA-axiom and the IID-hypothesis are theoretically equivalent. Due to the IIA-property, MNL models have an attractive feature, as they permit the introduction and/or elimination of alternatives in the choice set without reestimating the utility function parameters and the choice probabilities. Evidently, this feature greatly facilitates estimation and forecasting.

Because of its computational tractability and its mathematical simplicity the MNL model approach has been preferred to other choice models and has been widely used in a variety of applications, notably in the field of travel demand analysis (see e.g. Hensher and Stopher 1979, Stopher and Meyburg 1976b, Domencich and McFadden 1975). For many discrete choice problems the MNL model approach is quite attractive and appropriate, though it has to be admitted that the underlying assumptions are quite restrictive and may lead to counter-intuitive behavioural predictions, especially in cases when there are alternatives which are close substitutes for each other. It is not difficult to construct hypothetical examples, such as the well known 'red bus-blue bus' problem that violates the IIA-assumption. In this example, the only distinguishing attribute between two of the alternatives is the colour, all the other attributes being the same. Violations of the premises of the MNL model, however, may lead to inconsistent estimates of β and the choice probabilities (see Horowitz 1981). Procedures to determine violations of the IIA-assumption are discussed in Hensher and Johnson (1981, p. 150 pp), among others.

Due to the computational tractibility of the MNL model approach many attempts have been made to relax the IIA-assumption and thus to overcome the problem of similarities between alternatives. Such research efforts resulted in generalisations of the family of MNL models. The simplest class of random utility models generalising the MNL model approach is the family of generalised extreme value models.

(ii) <u>The family of generalised extreme value models</u>

The family of generalised extreme value (GEV) models can be derived from the <u>hypothesis of a broader class of multivariate extreme value distributions of the random disturbances</u> (see McFadden 1978, 1981, Manski 1981, and Smith 1982, among others).

$$F(\varepsilon_{i.} | z_{i.}) = \exp(-G((\exp(-\varepsilon_{ia}), a\varepsilon A), z_{i.})) \qquad i \varepsilon I \tag{66}$$

where G is a nonnegative, linear homogeneous in $(\exp(-\varepsilon_{ia}), a\varepsilon A)$ function with certain properties regarding its behaviour such as:

$$\lim_{\exp(-\varepsilon_{ia})\to\infty} G((\exp -\varepsilon_{ia}), a\varepsilon A), z_{i.}) = \infty \qquad i \varepsilon I \tag{67}$$

and regarding its partial derivatives, viz. that the continuous mixed partial derivatives exist and are nonnegative odd and nonpositive even (McFadden 1981, p. 227).

Then the GEV choice probabilities can be obtained as:

$$P(a|z_{i.}, \beta) = \exp(z_i \beta) \frac{G^a (\exp(z_{ia} \beta, a\varepsilon A))}{G (\exp(z_{ia} \beta, a\varepsilon A))} \qquad (i, a) \varepsilon (IxA) \tag{68}$$

with G^a denoting the partial derivative of G with respect to $z_{ia} \beta$. Different specifications of G lead to different GEV models. The family of GEV models allows the random terms to be correlated, but - in contrast to the probit model approach (see later on) - they do not have unequal variances. Furthermore, GEV models exhibit the appealing feature of leading to a closed form for the choice probabilities (see (67), in contrast with the family of multinomial probit models. Evidently, the family of GEV models can be considered as a generalisation of the MNL models because of the following specification:

$$G(\exp(-\varepsilon_{ia}), a \varepsilon A) = \sum_{a' \varepsilon A} \exp z_{ia} \beta \qquad (i,a) \varepsilon (I x A) \tag{69}$$

It is clear that (63) may be regarded as a special case of (66). Another very important special case is the <u>hierarchically structured</u> or <u>nested multinomial logit</u> (NMNL) <u>model</u> (see McFadden 1978, 1981). In this model approach a nested form of the decision structure is assumed, i.e., an individual actor first takes a decision independent of any other choice problem while he subsequently takes decisions conditional to the previous ones. Note that this assumption implies that the utility function has to have an additive separable form.

For illustration purposes let us consider the most simple case of a two-level nested decision structure. For example, consider a situation in which a combined choice has to be made for a travel destination and a transportation mode. Suppose that the destination is selected first and then the mode. Let B denote the set of possible destinations specific for the first decision level in the hierarchy and let C_b represent the set of travel modes with respect to $b \varepsilon B$ available at the second level. Thus,

$$A = ((b,c) \ , b \varepsilon B \wedge c \varepsilon C_b \tag{70}$$

Following Manski (1981) let us specify the function G as

$$G = \sum_{b \varepsilon B} \exp(\lambda_b \log \sum_{c \varepsilon C_b} \exp(z_{bc}(\beta | \lambda_b))) \tag{71}$$

with

$$0 < \lambda_b \leq 1 \qquad b \varepsilon B \tag{72}$$

Then the probability of choosing $(b,c) \varepsilon (B,C_b)$ shows the following nested logit form:

$$P(b,c) = \frac{\exp(z_{bc}\ (\beta|\lambda_b) + (\lambda_b - 1)\ \log \sum_{c'\varepsilon C_b} \exp(z_{bc'}\ (\beta|\lambda_b)))}{\sum_{b'\varepsilon B} \exp(\lambda_{b'}\ \log \sum_{c'\varepsilon C_{b'}} \exp(z_{b'c'}\ (\beta|\lambda_{b'})))} \qquad (73)$$

When λ_b, the parameter measuring the common correlation among the random disturbances ε_{bc}, equals 1, then the result is the MNL model approach. Note that choice makers taking decisions at the first level will take into account the expected maximum utility of choices at the second level. More general G functions leading to GEV models nested to multiple levels are discussed in McFadden (1981).

It is worthwhile noting that GEV models can be interpreted as elimination-by-strategy (EBS) models - a general family of random preference maximising models suggested by Tversky (1972a, b) which assume a hierarchy in the choice process and which also allow very general and flexible patterns of similarities between alternatives where the transition probabilities have a multinomial logit functional form as in the case of Tversky's elimination-by-aspect (EBA) models.

(iii) The family of multinomial probit models

The most general class of random utility models which circumvents the IIA-problem can be obtained by assuming the random disturbances $\varepsilon_{i.} = (\varepsilon_{i1}, \ldots, \varepsilon_{iA'})$ to be multivariate normally distributed with zero mean and an arbitrary variance-covariance matrix. This assumption directly leads to the multinomial probit (MNP) models which show the attractive feature of allowing the random terms of utility to be correlated and to have unequal variances. Furthermore, they permit variation in tastes among choice makers with identical observed attributes (see Daganzo 1979, among others). The random coefficients MNP model specification suggested by Hausman and Wise (1978) among others, may illustrate these properties.

$$u_{ia} = z_{ia}\,\tilde{\beta} + z_{ia}\,\delta_i + \kappa_{ia} \qquad (i,a) \in (I \times A) \qquad (74)$$

with

$$\delta_i \sim N(0, \Sigma_\delta) \qquad i \in I \qquad (75)$$

and

$$\kappa_{i.} := (\kappa_{i1}, \ldots, \kappa_{iA'}) \sim N(0, \Sigma_\kappa) \qquad i \in I \qquad (76)$$

where $\tilde{\beta}$ denotes the mean vector parameter vector describing the average taste of the choice makers in the population and δ_i characterises the deviations of the i-th choice maker's tastes from those which are represented in $\tilde{\beta}$. It should be noted that δ_i is constant for all alternatives in the choice set faced by the i-th choice maker. κ_{ia} is an additive error term. δ_i and $\kappa_{i.}$ are assumed to be multivariate normally distributed meaning that, for any $(i,i') \in (I \times I)$ with $i \neq i'$, δ_i and $\delta_{i'}$ as well as $\kappa_{i.}$ and $\kappa_{i'.}$ are independent realisations of the corresponding random vectors.

$$\tilde{\varepsilon}_{ia} := z_{ia}\,\delta_i + \kappa_{ia} \qquad (i,a) \in (I \times A) \qquad (77)$$

is the total random disturbance term for choice maker i and alternative a, where κ_{ia} may be considered to represent unobserved or omitted attributes of alternatives as well as purely random behaviour with respect to choice makers. Following Fischer and Nagin (1981) the parameters $\tilde{\beta}$, Σ_δ and Σ_κ are not jointly identified. Usually necessary identification conditions have to be imposed on the components of Σ_κ.

In contrast to MNL and GEV models the functional relationship between the choice probabilities and the measured attributes cannot be computed in an analytically closed form (except the binary case, i.e., for two choice alternatives). They can only be expressed as multiple integrals in which the domain of integration is bounded from below:

$$P(a|z_{i.}, \beta, \Sigma_\varepsilon) = \int_{\varepsilon_{ia}=-\infty}^{\infty} \Big(\prod_{\substack{a'=1 \\ a' \neq a}}^{A'} \int_{\varepsilon_{ia'}=-\infty}^{V(z_{ia},\beta)-V(z_{ia'},\beta)+\varepsilon_{ia}} N(\varepsilon_{i.}|0,\Sigma_\varepsilon)\,d\varepsilon_{ia'} \Big)\,d\varepsilon_{ia}$$

$$(i,a) \in (I \times A) \qquad (78)$$

where $N(\varepsilon_{i.}|0, \Sigma_\varepsilon)$ is a multivariate normal density function with zero mean and a variance-covariance matrix Σ_ε. The number of integrals equals the number A' of the choice alternatives.

Because of the considerable computational complexity of (78), MNP models have been considered as theoretically appealing and flexible but practically less manageable specifications in discrete choice analysis (except the binary case). Only quite recently some progress has been made in providing some more efficient and accurate procedures such as direct numerical integration methods (see Hausman and Wise 1978), iterative approximation procedures (see Daganzo et al. 1977) or the simulated frequency method (see Lerman and Manski 1981). Some examples of the probit model approach used in a spatial context can be found in van Lierop and Nijkamp (1984), van Lierop and Rima (1984) and Fischer and Maier (1984).

Clearly, the tension between the generality of the MNP and the tractability of the MNL model is hard to solve. Horowitz (1980) has tried to assess the effects of using the relatively simple MNL model in situations where the assumptions of the MNP modelling are satisfied but those of MNL modelling are not by evaluating the accuracy of the MNL model as an approximation to a variety of MNP models. He was able to show that MNL models give asymptotically accurate estimates of the ratios of the coefficients of the deterministic components of MNP utility functions, even when the MNL choice probabilities differ greatly from the MNP ones. Alternative and complementary contributions to the identification and a diagnostic checking of specification errors in random utility models can be found in Horowitz (1981, 1982).

3.4 Generalisations of the Classic Discrete Choice Models

The random utility based discrete choice models discussed so far have been termed classic due to the fact that they describe the behaviour of choice makers within an a-temporal framework. Many discrete choice situations however, such as for example, the labour force participation of middle-aged women and residential location decisions of young people, are recurrent. In such cases, choices made by an individual are not static in time but exhibit a dynamic pattern. An increasing interest in developing discrete choice models which explicitly incorporate dynamic aspects of behaviour can be observed in the last few years (see e.g. Heckman 1981b, Manski 1981, Daganzo and Sheffi 1982, de Palma and Leferre 1982, Leonardi 1983, Sonis 1983, Halperin 1984).

Especially stochastic panel data discrete choice approaches look quite appealing. They have been developed for analysing the structure of discrete choices within an intertemporal setting and represent a more genuinely behavioural approach than conventional discrete choice models. In such models, two main effects are taken into account: serial correlation in the unobserved atttributes known to the decision maker, but unknown to the analyst (termed spurious state dependence by Heckman 1981b), and (structural) state dependence. State dependence results from the fact that the individuals' current choices are to some extent dependent upon their previous choices, or in other words that past experience affects the range of alternatives and preferences relevant to choices made in subsequent time periods. Serial correlation is the result of omitted and unmeasured attributes which do not (or only marginally) change from time period to time period (see Daganzo and Sheffi 1982, p. 1378). Halperin (1984) shows that considerable progress has recently been made in developing panel data discrete choice models.

In his pioneering paper, Heckman (1981b) illustrates that serial correlation and state dependence can be represented within the framework of a more general panel data discrete choice model. This general model (including serial correlation models as well as state dependence models among others as special cases) is based on an MNP formulation for analysing the effects of conditional probability relationships between the occurrence of a discrete choice in one time period and its occurrence in previous time periods.

In order to focus the discussion on the most essential aspects, a simplified version of Heckman's model will be described in the sequel, a version which is sufficiently general and flexible to take account of dynamic aspects (see Manski 1981).

Let t=1, ..., T denote an exogenously given sequence of time periods. Furthermore, assume that at any period t=1, ..., T a choice maker $i \varepsilon I$ has to choose an alternative, say a_{it}, from A where the decisions made are based on the principle of maximising the random utility function:

$$u_{iat} = z_{iat}\, \beta + h_{iat}\, \gamma + \alpha_{ia} + \varepsilon_{iat} \qquad a \varepsilon A \tag{79}$$

The first term on the right-hand side of (79) represents the effects of known time-varying exogenous variables. Thus, the z-vector may incorporate past, current and expected values of future exogenous variables which determine the choice at period t. The second term represents state dependent effects where h_{iat} denotes a state variable with

$$h_{iat} = \begin{cases} 1 & \text{if } i \text{ chooses } a \text{ at } t-1 \ (t > 1) \\ 0 & \text{otherwise} \end{cases} \tag{80}$$

The disturbance in (79) is decomposed into two components: one component (third term on the right-hand side of (79)) which allows for time invariant individual specific effects of unmeasured attributes (with a general serial correlation structure) known to the decision maker but unknown to the analyst, and one component (fourth term on the right-hand side of (79)) which permits unanticipated purely random effects denoted by ε_{iat} which are unknown to both the choice maker and the analyst. α_{ia} with $a \varepsilon A$ is assumed to be multivariate normally distributed and independent of the measured attributes. The term (ε_{iat}, $a \varepsilon A$) being fully characterised by a multivariate normal distribution for any i and t represents white noise. Thus, recurrent choice (t=1, ..., T) is modelled as a sequence of static random utility maximising decisions. Following Manski (1981), four types of models based on (79)-(81) may be distinguished according to whether neither of, one of, or both of the serial correlations and state dependent effects are present.

In <u>temporally independent models</u> it is assumed that both γ and the α's degenerate to zero. Thus, neither state dependent nor serial correlation effects are present. This leads to the following expression of the probability of a sequence of choices for t=1, ..., T:

$$P(a_{it}, t=1, \ldots, T | z_{i..}, \beta) = \prod_{t=1}^{T} P(a_{it} | z_{i.t}, \beta) \quad i \varepsilon I \tag{81}$$

with

$$z_{i..} := (z_{iat}, a \varepsilon A, t=1, \ldots, T) \qquad i \varepsilon I \tag{82}$$

and

$$z_{i.t} := (z_{iat}, a \varepsilon A) \qquad i \varepsilon I, t=1, \ldots, T \tag{83}$$

Note that in the case of temporal independence a time series of choices made by a single decision maker cannot be distinguished from a set of choices made by a cross-section of individuals at a single time point.

Thus, estimation of temporally independent models is identical to estimation of the static models described in 3.3.

State dependence models, the second class, assume that non-trivial state dependent effects (i.e. $\gamma \neq 0$) exist, but serial correlation is absent. Then the probability of a sequence of choices for t=1, ..., T can be written as

$$P(a_{it}, t=1, \ldots, T|z_{i..}, \beta, \gamma) = \prod_{t=1}^{T} P(a_{it}|z_{i.t}, a_{i(t-1)}, \beta, \gamma) \qquad i\varepsilon I \qquad (84)$$

where $a_{i(t-1)}$ is an initial condition which has to be assumed to be exogenously specified. It is important to stress that the h-variables are statistically predetermined and that thus estimation theory developed for classic models can be applied without difficulty.

In contrast to models with temporal independence and models with state dependence, the next two model classes do not possess the Markov property. Then the probability of a sequence of choices for t=1, ..., T does not have a meaningful product decomposition (Manski 1981). The third class of models consisting of serial correlation models is defined by the absence of state dependence and by non-trivial unmeasured individual specific α-effects. Since all choices at t=1, ..., T depend on the same unmeasured time-invariant effects $\alpha_{ia}(a\varepsilon A)$, the choices $a_{it}(t=1, \ldots, T)$ are mutually dependent.

According to Manski (1984) this phenomenon can be considered as the choice model manifestation of the mover-stayer problem. Because the choice probabilities $P(a_{it}|z_{i.t}, \beta, \alpha)$ with t=1, ..., T have a multinomial probit form, the choice probability function parameters β and α can be estimated from a cross-section of measurements by means of estimation procedures for static discrete choice models.

Unrestricted models containing both serial correlation and state dependent effects form the most complete class of models. In this case, h_{iat} is a function of the utilities $(u_{ia(t-1)}, a\varepsilon A)$ which themselves depend on the unmeasured α-effects. As a consequence, the h_{iat}'s are correlated with the disturbance term $(\alpha_{ia} + \varepsilon_{iat})$. Obviously, the joint probability of the choice sequence has to be directly considered. Estimation of unrestricted models leads to specific difficulties. A comprehensive discussion of this issue can be found in Heckman (1981a).

While the generalisation of conventional random utility models discussed so far refers to a situation where the utility of an alternative to an individual only depends on his experience of past choices, de Palma and Lefevre (1982) deal with a choice context in which decision makers interact in their decision process. As a consequence, the attributes describing a choice problem of a decision maker also depend on the behaviour of other members of I. They suggest a continuous-time Markov model which allows individuals to interact in their decision process. Alternative and partly complementary contributions to simultaneous multiple person choice problems can be found in Margolis (1980), who placed the individualistic rationality axiom in the context of the 'economics of altruism', and in Miyao and Shapiro (1981), who developed a disaggregate random utility choice model for actors confronted with congestion effects due to simultaneous similar choices of competing actors.

3.5 Alternative Approaches

In recent years, various new distinct contributions to discrete spatial choice analysis have been made. The models generally used for discrete choice analysis are often marked by specific constraints with regard to their choice (for instance, linear response models). Such constraints on representational formalisms and on search procedures may be relaxed however in order to arrive at a greater degree of generality than that provided by the conventional statistical approaches (see Smith 1984). Examples of inductive formalisms that are more expressive in their ability to represent implicational statements involving categorical variables than the restricted conventional forms are:

- the introduction of variables that can assume general relationships with respect to sets of categorical variables and expressions that involve conjunctions and disjunctions of such relationships (see (Smith 1984),
- the use of general variable-valued logic in order to express the inductive inferential forms employed by human experts in solving complex tasks (see Michalski and Chilanski 1980),
- the use of an augmented predicate calculus in which each predicate, variable or function possesses an assigned annotation containing

relevant problem-oriented information including, for instance, the definition of the concept represented by the descriptor (see Michalski 1983).

In this regard, also new procedures for seeking inductive generalisations expressed in these formalisms may be imagined, for instance, algorithms based on training procedures for human decision making. Applications of such procedures concern especially housing market search problems (see Smith 1984). It is also increasingly being argued that conventional statistical models provide a very restrictive framework for choice analysis, as they do not leave open possibilities for inductive search behaviour guided by heuristic knowledge. Therefore, it may be important to stress the relevance of the manner in which an individual organises, retrieves and updates a representation of the world as a determinant of decision making behaviour (see also Smith and Lundberg 1984).

Heuristic search rules may be particularly appropriate if the set of relevant alternatives is non-exhaustive or if efficient search algorithms for complex choice problems are lacking. The viewpoint that decision making is a concurrent and heuristically-guided search of a physical space and its mental representation has become more prominent, since recent advances in computer science and cognitive psychology have opened the possibility of constructing computational process models to represent complex choice behaviour of individuals with constrained computational capacity (see also Tversky and Kahneman 1981). Such approaches may imply radical departure from models of rational decision making, as they are based on pattern-matching methods, similarity analysis, and context-dependent modular design. Examples of related heuristic rules are: the elimination-by-aspects rule and the greatest attractiveness difference rule (see also Wallsten 1980).

It follows from the abovementioned propositions that research into human spatial decision making should focus attention both upon the way individuals structure their knowledge bases and upon the processes by which these knowledge bases are accessed and modified. Altogether, it may be concluded that learning mechanisms may be a useful complement to traditional statistical spatial choice models.

The relevance of psychological perception for spatial choice processes has also been emphasised by Timmermans (1984), who tried to link objective attributes of spatial alternatives to their perceived values and to analyse next the cognitive process of combining separate judgements of attributes into an overall judgement. The author paid particular attention to such combination rules from the viewpoint of non-compensatory preference statements (dominance rules, conjunctive rules, disjunctive rules, lexicographic rules, elimination-by-aspects rules, maximin rules, etc.) and compensatory preference statements (linear additive rules, multiplicative rules, complex multilinear rules, etc.).

In conclusion, the area of psychological and heuristic decision analysis provides a rich framework for progress in human spatial decision making based on discrete data.

4. PROSPECTS

The field of explanatory discrete spatial data and choice analysis appears to be an extremely useful and fruitful research area. Despite the considerable progress which has been made especially in the last decade, there are still various open methodological problems and questions, such as:

- the estimation of random utility based models in the context of more complicated types of sampling processes involving exogenous and endogenous stratification at the same time,
- the estimation of unrestricted panel data discrete choice models, i.e. panel data models including both serial correlation and state dependence effects,
- the development of methods for forecasting aggregate population behaviour given an estimated discrete choice model as well as a description of the environment in which future decisions are made,
- the identification and inclusion of spatial auto- and cross-correlation in the case of discrete spatially dependent data,
- the nature and validity of explanatory discrete analysis in the case of forecasting models,
- simultaneous equation discrete choice modelling in which one or more of the attributes affecting choice are dealt with exogenously.

Finally, it should be emphasised again that a spatial choice model simply arises if the alternatives are distributed over space. Even if there is generally no need to develop a specific methodology for dealing with spatial discrete choice problems, it would be fruitful in some cases to combine the methodology of discrete choice analysis with both the research tradition of the analysis of spatially or temporal-spatially dependent data on the one hand and with the activity-constraint/time-budget framework on the other hand. By means of the time geographic approach it is possible to take into account far-reaching aggregate socio-spatial constraints upon individual choice which may strain the assumption of a rational decision making process independent of a retraining choice context (see Longley 1984). Up to now, surprisingly few attempts have been undertaken to integrate the spatial auto-correlation methodology mainly developed in geography in the early 1970s and the discrete data methodology. Here quantitative geographers might make a refreshing contribution. Wrigley (1984b) shows that the first steps in this direction have already been made. For example, Fingleton (1983) has attempted to explore some of the central features of complex discrete data sample survey designs in order to accommodate spatial dependence effects in the context of log-linear modelling. Naturally, there is a continuing need to combine these two kinds of methodologies.

REFERENCES

Aitkin, M., 'A simultaneous test procedure for contingency table models', Applied Statistics, vol. 28, 1979, pp. 233-242.

Amemiya, T., 'Specification and estimation of a multinomial logit model', Technical Report 211, Institute of Mathematical Studies in the Social Sciences, Stanford University, Stanford, Ca., 1976.

Arminger, G., 'Klassische Anwendungen verallgemeinerter, linearer Modelle in der empirischen Sozialforschung', Working Paper, Universität-Gesamthochschule Wuppertal, 1982 (mimeographed).

Arminger, G., 'Regressionsmodelle in der exponentiellen Familie', Working Paper, Universität-Gesamthochschule Wuppertal, 1983 (mimeographed).

Arminger, G., 'Analysis of qualitative individual data of latent class models with generalised linear models', in Nijkamp, P., Leitner, H. and Wrigley, N. (eds), Measuring the Unmeasurable, Martinus Nijhoff, The Hague, 1984 (forthcoming).

Aufhauser, E., 'Log-lineare Modelle und ihre Anwendung zur Analyse raümlichter Interaktionsmatrizen', Master's Thesis, Department of Geography, University of Vienna, 1984.

Aufhauser, E. and Fischer, M.M., 'Loglinear modelling and spatial analysis', paper presented at the 24th European Regional Science Conference', Milan 1984.

Bahrenberg, G., Fischer, M.M. and Nijkamp, P. (eds), Recent Developments in Spatial Data Analysis: Methodology, Measurement, Models, Gower, Aldershot, 1984.

Bartholomew, P.J., 'Latent variable models for ordered categorical data', Journal of Econometrics, vol. 22, 1983, pp. 229-243.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W., Discrete Multivariate Analysis: Theory and Practice, 2nd edition, MIT Press, Cambridge, Mass., 1976.

Brown, M.B., 'Screening effects in multidimensional contingency tables', Applied Statistics, vol. 25, 1976, pp. 37-46.

Cosslett, S.R., 'Efficient estimation of discrete-choice models', in: Manski, C.F., McFadden, D. (eds), Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, Mass., 1981, pp. 51-111.

Daganzo, C., Multinomial Probit. The Theory and its Application to Demand Forecasting, Academic Press, New York, 1979.

Daganzo, C. and Sheffi, Y., 'Multinomial probit with time-series data: Unifying state dependence and serial correlation models', Environment and Planning A, vol. 14, 1982, pp. 1377-1388.

Daganzo, C., Bouthelier, F. and Sheffi, Y., 'Multinomial probit and qualitative choice: A computationally efficient algorithm', Transportation Science, vol. 11, 1977, pp. 338-358.

Domencich, T.A. and McFadden, D., Urban Travel Demand: A Behavioral Analysis, North Holland Publ. Co., Amsterdam, 1975.

Fienberg, S.R., The Analysis of Cross-Classified Categorical Data, 2nd edition, MIT Press, Cambridge, Mass., 1981.

Fienberg, S.E. and Mason, 'Identification and estimation of age-period-cohort models in the analysis of discrete archival data', in: Schluessler, K.F.(ed), Sociological Methodology, Jossey-Bass, San Francisco, 1978, pp. 1-67.

Fienberg, S.E. and Meyer, M.M., 'Loglinear models and categorical data analysis with psychometric and econometric applications', Journal of Econometrics, vol. 22, 1983, pp. 191-214.

Fingleton, B.,'Log-linear modelling of geographical contingency tables', Environment and Planning A, vol. 13, 1981, pp. 1539-1551.

Fingleton, B.,'Log-linear models with dependent spatial data', Environment and Planning A, vol. 15, 1983, pp. 801-813.

Finney, D.J., Probit Analysis, 3rd edition, Cambridge University Press, London, 1971.

Fischer, M.M. and Maier, G., 'Discrete choice and labour supply modelling', Paper presented at the Besançon Symposium of the IGU Working Group on Systems Analysis and Mathematical Models, August 21-23, 1984.

Fischer, G.W. and Nagin, D., 'Random versus fixed coefficient quantal choice models', in: Manski, C.F. and McFadden, D. (eds), Structural Analysis of Discrete Data with Econometric Application, MIT Press, Cambridge, Mass., 1981, pp. 273-304.

Flowerdew, R. and Aitkin, M., 'A method for fitting the gravity model based on the Poisson distribution', Journal of Regional Science, vol.22, 1982, pp. 191-202.

Goodman, L.A., 'The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach', Biometrika, vol. 60, 1973, pp. 179-192.

Goodman, L.A., Analysing Qualitative/Categorical Data, Log-Linear Models and Latent Structure Analysis, Addison-Wesley, London, 1978.

Grizzle, J.E., Starmer, C.F. and Koch, G.G., 'Analysis of categorical data by linear models', Biometrics, vol. 25, 1969, pp. 489-504.

Haberman, S.J., Analysis of Qualitative Data. Vol. 1: Introductory Topics, Academic Press, New York, 1978.

Haberman, S.J., Analysis of Qualitative Data. Vol. 2: New Developments, Academic Press, New York, 1979.

Halperin, W.C., 'The analysis of panel data for discrete choices', in: Nijkamp, P., Leitner, H. and Wrigley, N. (eds), Measuring the Unmeasurable, Martinus Nijhoff, The Hague, 1984 (forthcoming).

Hannan, M.T. and Tuma, N.B., 'Dynamic analysis of qualitative variables: Applications to organizational demography', in: Nijkamp, P., Leitner, H. and Wrigley, N. (eds), Measuring the Unmeasurable, Martinus Nijhoff, The Hague, 1984 (forthcoming).

Hausman, J.A. and Wise, D.A., 'A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences', Econometrica, vol. 46, 1978, pp. 403-426.

Hausman, J.A. and Wise, D.A., 'Stratification on endogenous variables and estimation: The Gary income maintenance experiment', in: Manski, C.F. and McFadden, D. (eds), Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, Mass., 1981, pp. 365-391.

Heckman, J.J., 'Dummy endogenous variables in a simultaneous equation system', Econometrica, vol. 46, 1978, pp. 931-959.

Heckman, J.J., 'Statistical models for discrete panel data', in: Manski, C.F. and McFadden, D. (eds), Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, Mass., 1981a, pp. 114-178.

Heckman, J.J., 'The incidental parameters problems and the problem of initial conditions in estimating a discrete time-discrete data stochastic process', in Manski, C.F. and McFadden, D. (eds), Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, Mass., 1981b, pp. 179-195.

Hensher, D.A. and Johnson, L.W., Applied Discrete Choice Modelling, Croom Helm, London, 1981.

Hensher, D.A. and Stopher, P.R. (eds), Behavioural Travel Modelling, Croom Helm, London, 1979.

Horowitz, J.L.,'The accuracy of the multinomial logit model as an approximation to the multinomial probit model to travel demand', Transportation Research, vol. 14B, 1980, pp. 331-341.

Horowitz, J., 'Identification and diagnosis of specification errors in the multinomial logit model', Transportation Research, vol. 15B, 1981, pp. 345-360.

Horowitz, J., 'Specification tests for probabilistic choice models', Transportation Research A, vol. 16, 1982, pp. 383-394.

Horowitz, J., 'Random utility models as practical tools of travel demand analysis', in: Jansen, G.R.M., Nijkamp, P. and Ruijgrok, C.J. (eds), Transportation and Mobility in an Era of Transition, 1984 (forthcoming)

Issaev, B., Nijkamp, P., Rietveld, P. and Snickars, F. (eds), Multi-Regional Economic Modelling: Practice and Prospect, North-Holland Publ. Co., Amsterdam, 1982.

Leeuw, J. de, Keller, W.J. and Wansbeek, T. (eds), 'Introduction to the Special Issue on 'Interface between Econometrics and Psychometrics', Journal of Econometrics, vol. 22, 1983, pp. i - vi.

Leonardi, G., 'An optimal control representation of a stochastic multistage-multiactor choice process', in: Griffith, D.A. and Lea, A.C. (eds), Evolving Geographical Structures, Martinus Nijhoff, The Hague, 1983, pp. 62-72.

Lerman, S.R. and Manski, C.F., 'Sampling design for discrete choice: State of the art', Transportation Research, col. 13A, 1979, pp. 29-44.

Lerman, S.R. and Manski, C.F., 'On the use of simulated frequencies to approximate choice probabilities', in: Manski, C.F. and McFadden, D. (eds), Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, Mass., 1981, pp. 305-319.

Lierop, W.F.J. van, and Nijkamp, P., 'Perspectives of disaggregate choice models on the housing market', in: Pitfield, D.A. (ed), Discrete Choice Modelling, Pion, London, 1984 (forthcoming).

Lierop, W.F.J. van and Rima, A., 'Residential mobility and probit analysis', in: Bahrenberg, G., Fischer, M.M. and Nijkamp, P. (eds), Recent Developments in Spatial Data Analysis: Methodology, Measurement, Models, Gower, Aldershot, 1984, pp. 393-407.

Longley, P.A., 'Discrete choice modelling and complex spatial choice: An overview', in Bahrenberg, G., Fischer, M.M. and Nijkamp, P. (eds), Recent Developments in Spatial Data Analysis: Methodology, Measurement, Models, Gower, Aldershot, 1984, pp. 375-391.

Luce, R.D., Individual Choice Behavior, Wiley, New York, 1959.

Maddala, G.S., Limited-Dependent and Qualitative Variables in Econometrics, Cambridge University Press, Cambridge, Mass., 1983.

Manski, C.F., 'Structural models for discrete data: The analysis of discrete choice', in: Leinhart, S. (ed), Sociological Methodology Jossey-Bass, San Francisco, Washington and London, 1981, pp. 58-109.

Manski, C.F. and McFadden, D., 'Alternative estimators and sample designs for discrete choice analysis', in: Manski, C.F. and McFadden, D. (eds), Structural Models for Discrete Data with Econometric Applications, MIT Press, Cambridge, Mass., 1981a, pp. 2-50.

Manski, C.F. and McFadden, D. (eds), Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, Mass., 1981b.

Manski, C.F. and Lerman, S., 'The Estimation of choice probabilities from choice-based samples', Econometrica, vol. 45, 1981, pp. 1977-1988.

Margolis, H., Selfishness, Altruism and Rationality, Cambridge University Press, Cambridge, Mass., 1982.

McCullagh, P., 'Regression models for ordinal data', Journal of the Royal Statistical Society, vol. 42B, 1980, pp. 109-142.

McCullagh, P. and Nelder, J.A., Generalized Linear Models, Chapman and Hall, London, 1983.

McFadden, D., 'Conditional logit analysis of qualitative choice behaviour', in Zarembka, P. (ed), Frontiers in Econometrics, Academic Press, New York, 1974, pp. 105-142.

McFadden, D., 'Modelling the choice of residential location', in: Karlqvist, A., Lundqvist, L., Snickars, F. and Weibull, J.W. (eds), Spatial Interaction Theory and Planning Models, North-Holland Amsterdam, 1978, pp. 75-96.

McFadden, D., 'Qualitative methods for analysing travel behaviour of individuals: Some recent developments', in: Hensher, D.A. and Stopher, P.R. (eds), Behavioural Travel Modelling, Croom-Helm, London, 1979, pp. 279-318.

McFadden, D., 'Econometric models of probabilistic choice', in: Manski, C.F. and McFadden, D. (eds), Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, Mass., 1981, pp. 198-272.

Michalski, R.S., 'A theory and a methodology of inductive learning', Report no. UIUCDCS-R-83-1122, Department of Computer Science, University of Illinois, Urbana, 1983 (mimeographed).

Michalski, R.S. and Chilansky, R.L., 'Learning by doing told and learning from examples', International Journal of Policy Analysis and Information Systems, vol. 4, 1980, pp. 125-161.

Miyao, T. and Shapiro, P., 'Discrete choice and variable returns to scale', International Economic Review, vol. 22, 1981, pp. 257-273.

Muthén, B., 'A structural probit model with latent variables', Journal of the American Statistical Association, vol. 74, 1979, pp. 807-811.

Muthén, B., 'Latent variable structural equation modeling with categorical data', Journal of Econometrics, vol. 21, 1983, pp. 43-65.

Nelder, J.A., 'Log-linear models for contingency tables: A generalization of classical least squares', Applied Statistics, vol. 23, 1974, pp. 323-329.

Nelder, J.A. and Wedderburn, R.W.M., 'General linear models', Journal of the Royal Statistical Society, vol. 135A, 1972, pp. 370-384.

Nerlove, A. and Press, S.J., 'Univariate and multivariate log-linear and logistic models', Technical Report R-1306-EDA/NIH, Rand Corporation, Santa Monica, Cal., 1973.

Nijkamp, P., Leitner, H. and Wrigley, N. (eds), Measuring the Unmeasurable, Martinus Nijhoff, The Hague, 1984.

O'Brien, L.G. and Wrigley, N. 'A generalized linear models approach to categorical data analysis', in: Bahrenberg, G., Fischer, M.M. and Nijkamp, P. (eds), Recent Developments in Spatial Data Analysis: Methodology, Measurement, Models, Gower, Aldershot, 1984, pp. 231-251.

Odland, J. and Barff, R., 'A statistical model for the development of spatial problems: Applications to the spread of housing deterioration', Geographical Analysis, vol. 14, 1982, pp. 327-339.

Palma, A.de and Lefevre, C., 'Individual decision-making in dynamic collective systems', Paper presented at the IIASA-Workshop on Spatial Choice Models in Housing, Transportation and Land Use Analysis: Towards a Unifying Effort, Laxenburg, 29th March-1st April, 1982.

Roberts, F.S., Measurement Theory with Applications to Decisionmaking, Addison-Wesley, Reading, Mass., 1979.

Smith, T.R., 'General representional formalisms and search procedures for inferring models from categorical data', in: Nijkamp, P., Leitner, H. and Wrigley, N. (eds), Measuring the Unmeasurable, Martinus Nijhoff, The Hague, 1984. (forthcoming).

Smith, T.E., 'Random utility models of spatial choice: A structural analysis', Paper presented at the IIASA Workshop on Spatial Choice Models in Housing, Transportation and Land Use Analysis: Towards a Unifying Effort, Laxenburg, 29th March-1st April, 1982.

Smith, T.R. and Lundberg, C.G., 'Psychological foundations of individual choice behaviour and a new class of decision making units', in: Bahrenberg, G., Fischer, M.M. and Nijkamp, P., Recent Developments in Spatial Data Analysis: Methodology, Measurement, Models, Gower, Aldershot, 1984, pp. 355-373.

Sonis, M.,'Dynamic discrete time multinomial logit and dogit models, Paper presented at the European Conference of the Regional Science Association, Poitiers, France, August 1983 (mimeographed).

Stopher, P.R. and Meyburg, A.H. (eds), Behavioral Travel Demand Models, Lexington Books, Lexington, Mass., 1976.

Tardiff, T., 'Definition of alternatives and representations of dynamic behaviour in spatial choice models', Transportation Research Record, no. 723, 1980, pp. 25-34.

Thomson, R. and Baker, R.J., 'Composite link functions in generalized linear models', Applied Statistics, vol. 30, 1981, pp. 125-131.

Thurstone, L., 'A law of comparative judgement', Psychological Review, vol. 34, 1927, pp. 273-286.

Timmermans, H.J.P., 'Decision models for predicting preferences among multiattribute choice alternatives', in: Bahrenberg, G., Fischer, M.M. and Nijkamp, P. (eds), Recent Developments in Spatial Data Analysis: Methodology, Measurement, Models, Gower, Aldershot, 1984, pp. 337-354.

Tversky, A., 'Elimination-by-aspects: A theory of choice', Psychological Review, vol. 79, 1972a, pp. 281-299.

Tversky, A., 'Choice-by-elimination', Journal of Mathematical Psychology, vol. 9, 1972b, pp. 341-367.

Tversky, A. and Kahneman, D., 'The framing of decisions and the psychology of choice', Science, vol. 211, 1981, pp. 453-458.

Tversky, A. and Sattath, S., 'Preference trees', Psychological Review, vol. 86, 1979, pp. 542-573.

Wallsten, T.S., 'Processes and models to describe choice and inference behavior', in: Wallsten, T.S. (ed), Cognitive Processes in Choice and Inference Behavior, Lawrence Erlbaum Ass., Hillsdale, N.J., 1980, pp. 215-237.

Wedderburn, R.W.M., 'Quasi-likelihood functions, generalized linear models and the Gauss-Newton method', Biometrika, vol. 61, 1974, pp. 439-447.

Wegener, B. (ed), Social Attitudes and Psychophysical Measurement, Lawrence Erlbaum Ass., Hillsdale, N.J., 1982.

Willekens, F., 'Log-linear modelling of spatial interaction', Papers of the Regional Science Association, vol. 52, 1984 (forthcoming).

Willekens, F. and Güvenç, N., 'Hybrid log-linear models', in: Nijkamp, P., Leitner, H. and Wrigley, N. (eds), Measuring the Unmeasurable, Martinus Nijhoff, The Hague, 1984 (forthcoming).

Wrigley, N., An Introduction to the Use of Logit Models in Geography', Concepts and Techniques in Modern Geography 10, Geo Abstracts, Norwich, 1976.

Wrigley, N., Probability Surface Mapping. An Introduction with Examples and Fortran Programmes, Concepts and Techniques in Modern Geography 16, Geo Abstracts, Norwich, 1977.

Wrigley, N., 'Developments in the statistical analysis of categorical data', Progress in Human Geography, vol. 3 (3), 1979, pp. 315-355.

Wrigley, N., 'Categorical data analysis', in: Wrigley, N. and Bennett, R.J. (eds), Quantitative Geography: A British View, Routledge and Kegan Paul, London, 1981, pp. 111-122.

Wrigley, N., 'Statistical models for discrete choice analysis', in: Bahrenberg, G., Fischer, M.M. (eds), Proceedings of the 3rd European Conference on Theoretical and Quantitative Geography, Bremer Beiträge zur Geographie und Raumplanung 5, University of Bremen, Bremen, 1984a (forthcoming).

Wrigley, N., 'Categorical data methods and discrete choice modelling in spatial analysis: Some directions for the 1980's', in: Nijkamp, P., Leitner, H. and Wrigley, N. (eds), Measuring the Unmeasurable, Martinus Nijhoff, The Hague, 1984b (forthcoming).