SERIE RESEARCHMEMORANDA

FACTS AND FIGURES
IN REGIONAL SCIENCE

Peter Nijkamp
Neil Wrigley

Researchmemorandum 1984-4

VRIJE UNIVERSITEIT
EKONOMISCHE FAKULTEIT
AMSTERDAM

Paper presented at the
European Meeting of the
Regional Science Association,
Poitiers, August 1983.

1. INTRODUCTION

Ever since its emergence in the fifties, regional science has made a vigorous attempt at providing an explanation for real-world spatial configurations and development patterns. Despite a formal mathematical and statistical orientation, its major aim has always been to offer an analytical basis for understanding empirical facts from a complex spatial reality. In doing so, regional science methodology has been strongly affected by natural science approaches. Location-allocation models, optimization models, entropy models, transportation and traffic models, regional growth models and the like reflected the (implicit) assumption that facts in a spatial world could be adequately measured by means of figures based on a cardinal metric system. Despite sometimes invalid or weak underlying assumptions, the main stream in regional science methodology has often made heroic attempts to measure attributes of phenomena on a cardinal scale. Verification of facts and validation of theories had to pass - in many cases - the narrow filter of methods and techniques characterized by cardinal measurability (cf. Roberts, 1979).

Until recently, the great potential offered by qualitative methods of analysis has too often been neglected or underestimated, and much (sometimes radical) criticism of the conventional methodology of various social sciences (including regional science) has resulted from inappropriate choices of the level of measurement in statistical and econometric analyses. 'Measuring the unmeasurable' (see Nijkamp et al., 1983) is indeed a paradoxical issue of great significance in social science research.

In recent years, significant progress has been made in the treatment of qualitative (or categorical) variables (i.e., variables measured on an ordinal, binary, or nominal scale). In the field of both parametric and non-parametric statistics and econometrics a wide variety of techniques and models have been designed which treat qualitative variables in an appropriate manner, and have resulted in terms such as 'soft econometrics', soft modeling', 'qualitative statistics', 'non-metric data analysis', etc. Many of these new techniques are also increasingly being applied in regional science and related desciplines (for instance, in the area of consumer choice behaviour, locational perceptions and preferences, contingency table analysis, scenario and qualitative impact analysis, plan evaluation and conflict analysis, and so forth).

The recently developed methods of qualitative data analysis aim at taking into account the limitations caused by measuring variables on a non-metric scale. They also aim at avoiding the use of non-permissible numerical operations on qualitative variables (for instance, a summation or multiplication of ordinal figures). In this respect, these methods may broaden the scope of conventional regional science research and lead to many new perspectives for an appropriate manipulation of qualitative information.

Fortunately, research in the fields of geography, economics (particularly regional and urban), and planning has become more reliant on qualitative data. Whilst disciplines such as psychology and sociology have been extensive users of qualitative data, and have well developed experimental designs and statistical methods for dealing with such data (see De Leeuw et al., 1983, and Wegener, 1982), the spatial analysis disciplines in general have been late entrants to this new area and are still in an experimental phase. No doubt, 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 Generalized Linear Models, latent variables methods, correspondence analysis and other qualitative multivariate techniques, fuzzy set analysis, qualitative evaluation analysis and discrete choice analysis including panel data and time-event histories).

In regard to qualitative data analysis, it is also of utmost importance to pay more attention to data collection (such as quality of data, sampling, survey design). In many situations, there is an overemphasis on completeness for data collection rather than concentration on data which are analytically the most important - especially in the context of forecasting and updating.

Unfortunately, the strong orientation of regional science research towards a quantitative methodology has not kept pace with the necessary quality and level of measurement of the data needed to operationalize models and methods. In fact, the majority of models developed in spatial research takes for granted a reliability and accuracy of data which is not fulfilled in practice (cf. James et al., 1982). Thus the theoretical developments in quantitative spatial analysis have created a gap between data needs for analytical models and data availability (see Figure 1).

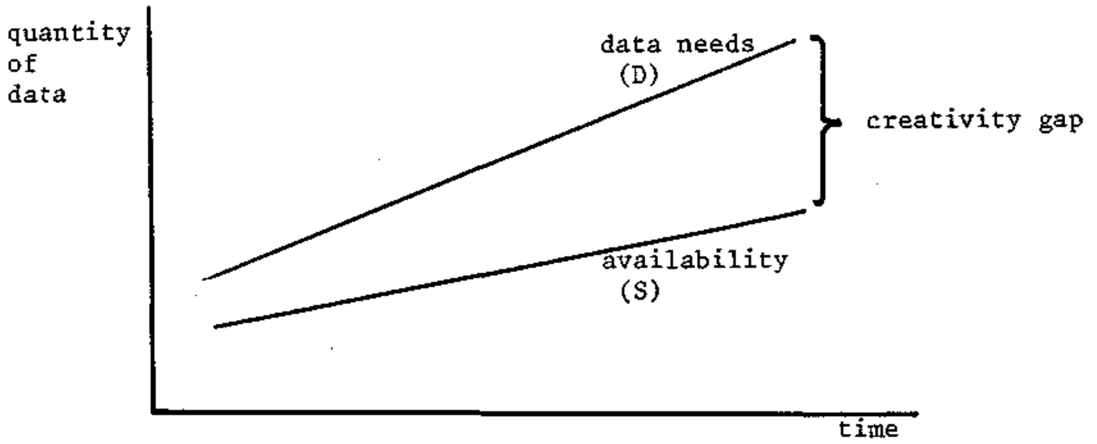


Fig. 1. Gap between supply and demand of data

This figure also illustrates the role of qualitative modeling in bridging the gap between the demand for data for analytical models (D) and the supply of available data (S). Qualitative data analysis takes for granted the nature of existing data and attempts to design appropriate models and techniques for a consistent transformation of these data by means of methodologically sound operations in order to derive metric inferences regarding the phenomena at hand. In this respect, it is a challenge for qualitative data analysis to fill the above mentioned creativity gap.

Another bottleneck in terms of the application of the methodologies of qualitative data analysis in empirical research practice is the lack of attention to behavioral theories 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 not enough attention is paid to the behavioral processes being studied or to the structure of regions or cities at hand. In particular, if various applications ignore the underlying behavioral processes, then the results obtained cannot be meaningfully interpreted. In such cases, unsatisfactory results might lead 'practitioners' to be reluctant to use new models and techniques (including qualitative models). Thus the latter must be carefully applied by their developers to real world problems, in order for 'practitioners' to become comfortable in adopting the new techniques.

As mentioned already before, in recent years various techniques and models in the area of qualitative data analysis have found interesting applications in spatial research. Table 1 contains a concise survey based on Nijkamp (1983). The following main new developments are included in this Table:

- indirect methods: cope with the qualitative nature of data used in statistical or econometric techniques by employing allied metric concepts or indicators that reflect the information embodied by the original qualitative variables.
- statistical methods for multivariate qualitative data: aim at detecting a structure in an ordinal or binary data set so as to draw inferences from this data set in cardinal terms.
- statistical methods for multidimensional nominal data: try to provide appropriate solutions for treating variables characterized by dichotomous (binary) or polychotomous classifications, by employing information on the frequency of items in each nominal class.
- explanatory models with qualitative data: aim at developing statistical/econometric models, in which the predetermined variables are either cardinal or qualitative (or mixed), and the response variables cardinal or qualitative.
- qualitative structure analysis of complex systems: tries to infer statements regarding the qualitative structure of models without knowing the precise cardinal values of the variables described in those models.
- qualitative multiple criteria analysis: aims at providing an evaluation of discrete alternatives marked by qualitative impact assessment and/or preference weights.

Having briefly summarized now the current state-of-the-art in qualitative (spatial) data analysis, it may be interesting to point out some new promising directions in this field. This will be the subject of some subsequent sections.

indirect methods	statistical methods for multivariate qualitative data	statistical methods for multidimensional nominal data	explanatory models with qualitative data	qualitative structure analysis of complex systems	qualitative multiple criteria analysis
proxy variables	rank correlation analysis	contingency table analysis	logit analysis	graph theory	expected value method
dummy variables	concordance analysis	log-linear analysis	probit analysis	qualitative calculus	lexicographic method
order statistics	ordinal principal component analysis		scaling analysis	bifurcation theory	frequency method
latent variables	qualitative cluster analysis		regime analysis		ordinal concordance method
path models	multidimensional scaling analysis		generalized linear models		permutation method
partial least squares	homogeneous scaling analysis		linear probability models		metagame method
Lisrel models	correspondence analysis				eigenvalue method
fuzzy set analysis					scaling method
					regime method

Table 1. A summary of some qualitative data methods

2. CATEGORICAL DATA ANALYSIS AND GENERALIZED LINEAR MODELS

The models and techniques of categorial data analysis (logistic/logit models, probit models, log-linear models, screening and simultaneous test procedures, and so on) are now a firmly established part of research methodology within the fields of geography, economics, and planning. Moreover they are increasingly being viewed within the framework provided by Nelder and Wedderburn's (1972) family of 'generalized linear models' (GLMs) and implemented through the common computing environment provided by the GLIM package. These trends must inevitably continue and strengthen within the next few years. Thus one may look forward to an increasingly sophisticated understanding and use of these procedures in spatial analysis aided by the diffusion of a new generation of text-books (e.g. McCullagh and Nelder, 1983; Wrigley, 1984).

Developing outwards from this now established base we see the most promising directions for research as lying within two broad areas:

(i) In the exploitation within spatial analysis of some of the most recent extensions of GLMs including composite link function models, mixture models, model with compound dispersion parameters, and quasi-likelihood procedures.

(ii) In attempts to integrate the still rather separate literatures and research traditions of the analysis of spatially dependent data on the one hand and categorical data analysis on the other.

2.1 Conventional GIMs and the potential of recent extensions

At the heart of Nelder and Wedderburn's 'generalized linear models' (GLMs) approach lies the concept that many apparently separate statistical models and modelling traditions can be reconciled in a single system with a common notation and a unified estimation procedure. The potential benefits of this perspective for the analysis of qualitative spatial data are threefold. First, the algebraic differences which have characterized categorical data models become largely redundant. Second, the consistent notation of GLMs allows users to more readily perceive the interrelatedness of models and encourages an exploratory approach to data analysis. Third, the GLM perspective not only simplifies the various categorical data methods to a single family of models, but also links these models in a unified manner with conventional continuous (quantitative) data linear models.

Details of the GLM framework are discussed by Nelder (1983), Arminger (1983), McCullagh and Nelder (1983), and in the spatial analysis literature by O'Brien (1983) and O'Brien and Wrigley (1983) amongst others. Very briefly, an generalized linear model can be written in the form:

$$Y_i = g^{-1}(\eta_i) + \varepsilon_i \tag{1}$$

where Y_i is a response variable which is assumed to come from the exponential family of probability distributions, $\eta_i = \sum_k \beta_k X_{ik}$ is the 'linear predictor', ε_i is an error term, and g^{-1} is the inverse of the 'link' function

$$\mu_i = g^{-1}(\eta_i) \tag{2}$$

by which the linear predictor η_i, is related to μ_i, the expected value of Y_i [i.e. $\mu_i = E(Y_i)$].

Many spatial analysts are now familiar with the standard combinations of error distributions and link functions shown in Table 2, and the conventional continuous and categorical data GLMs which the combinations produce. Less familiar to spatial analysts, however, are some of the recent extensions of GLMs, yet such extensions are now beginning to play a significant role in the specialized literature which surrounds the GLIM package. (See, for example, recent issues of the twice-yearly GLIM Newsletter which contains valuable discussions of these developments and user-supplied GLIM 'macros').

Model	Error Distribution	Link function
Linear regression	Normal	Identity
ANOVA (fixed effects)	Normal	Identity
ANOVA (random effects)	Gamma	Identity
Logit regression	Binomial or Multinomial	Logit
Probit regression	Binomial or Multinomial	Identity
Linear logit model for cell f problems	Binomial or Multinomial	Logit
General log-linear model for cell g problems	Poisson	Logarithmic
Log-linear model for cell f problems	Poisson	Logarithmic

Table 2. Examples of conventional GLMs (cell references are to Wrigley, 1983, Table 1).

Some of the simplest, but potentially most useful, extensions of the conventional GLMs of Table 2 can be derived using composite link functions (Thompson and Baker, 1981). Instead of the one-to-one relationship shown in (2) between the expected values, μ_i , and the linear predictors, η_i , composite link functions allow each μ_i to be a function of several η's. In practice, this is done by defining each μ to be a linear combination of some intermediate quantities (say τ_i) which are themselves functions of the corresponding η_i (see Arminger, 1983). Uses of composite link function models in qualitative data analysis currently include latent class analysis (see Arminger, 1983) and the analysis of fused cells in contingency tables (see Thompson and Baker, 1981). In addition, Nelder (1983) suggests that McCullagh's (1980) logistic/logit models for ordered categories can be cast in this framwork, and many other possibilities have recently been suggested (e.g. Roger, 1983).

Other interesting extensions of conventional GLMs occur when the error term in (1) is assumed to be a mixture of several components rather than being homogeneous. One example of such extensions which already exists in the spatial analysis literature is Flowerdew and Aitkin's (1982) compound Poisson migration model, and the SSRC in Britain are currently supporting additional research (Flowerdew, 1983) in this area of spatial analysis. Nelder (1983) notes that a quasi-likelihood procedure may often be useful for deriving satisfactory parameter estimates for such models, and it is clear that this procedure holds considerable potential for many other areas of qualitative spatial data analysis.

The increasingly wide use of the GLMs approach is, of course, in no small part due to the power and the flexibility of the operational statistical modelling environment provided by the associated GLIM computer package. The new version of the package, GLIM 4, will shortly be released and will form part of a new statistical data analysis system which has provisionally been known as PRISM (see Wrigly, 1983). The new features of GLIM 4, e.g. the extended range of default options of link functions and error distributions, including the important multinomial error distribution (see Clarke, 1982; O'Brien, 1983), and the proposed link in the PRISM system to GRAPH 1 - a set of graphics subroutines for simple data displays and more sophisticated uses (generalized text drawing, picture segments, and graphical input-output)-can only advance its use in spatial analysis.

As such, there is clearly an urgent need for existing quantitative geography/spatial analysis textbooks to be updated to reflect the increasing use of GLIM and the GLMs framework, and thus to help bridge the gap between research practice and undergraduate/graduate teaching.

2.2 Integrating Techniques for Handling Spatial Dependence and Categorical Data

Despite the fact that categorical data analysis is now a firmly established part of research methodology within geography, regional/urban economics, and planning, there have, so far, been suprisingly few attempts to integrate the spatial autocorrelation/dependence and categorical data literatures. Here we see a major challenge which spatial analysts must take up and from which they could make a distinctive contribution.

Over the past fifteen years, considerable attention has been paid to the problems which derive from applying classical statistical models and inferential procedures which assume independent observations to geographical data which typically exhibit systematic ordering over space. During this period, the analysis of spatial data has been developed from a stage of relatively uncritical application of standard inferential tests and statistical models, to a stage of wide appreciation of the effects of spatially dependent data. Tests for spatial autocorrelation have been developed, classical statistical models and tests have been modified to handle spatially dependent data and extensive research on spatial process modelling and spatial time-series analysis has been conducted (e.g. Cliff and Ord, 1973, 1981; Hepple, 1974; Hordijk and Nijkamp, 1978; Bennett, 1979; Griffith, 1980, Haining, 1980; Folmer, 1983).

To some extent, it is possible to view the categorical data methods and models now used in spatial analysis as being in an analogous position to the continuous data methods/models used a decade ago. As such, integration of the categorical data and spatial dependence literatures is likely to proceed along the same path as described above, but at a much accelerated rate. Indeed, as Wrigley (1983a) shows, the first steps along this path have already been taken. For example, Fingleton (1983a, 1983b) has shown that considerable care is necessary when attempting to apply the now familiar log-linear models for categorical data to spatially dependent data. In these circumstances the standard model selection procedures of log-linear modelling may erroneously detect interaction effects between variables (or even main effects) which are, in fact, spurious and a consequence of the spatially dependent observations. Fingleton's solution to this problem involves modifying the standard calculation of Pearson's chi-square statistic, for - in the presence of positive spatial dependence - it can be shown that this will assume an inflated value. Similarly, Odland and Barff (1982) have attempted to combine the logic of existing space-time interaction tests with categorical data models in order to model the space-time patterns of housing deterioration in an American city. Further research in both these areas is clearly possible. For example, in the case of log-linear models with spatially dependent data, Fingleton has shown how knowledge of the precise nature of the spatial autocorrelation might ultimately be used to estimate variance-covariance matrices which include terms representing the structure of the spatial dependence whilst in the case of Odland and Barff's categorical space-time process models there is clearly considerable potential in attempting to develop linkages to Besag's (1974,1975) important class of auto-logistic models (see Haining, 1982, 1983).

We, therefore, essentially see the integration of the spatial autocorrelation/dependence and categorical data literatures as following an accelerated version of the path established in the late 1960s and 1970s for continuous spatial data. However, the wider backcloth of statistical methodology is clearly not the same as in that period, and these broader trends must influence the nature of the integration. In particular, we foresee two wider trends in statistical methodology as likely to affect research in this area within the next five years.

First, it is clear that spatial analysis is not immune to the wider debate on robustness and resistance which has taken place within statistics and to the still wider issue of exploratory versus confirmatory approaches to data analysis (see the commentaries in the geographical literature by Cox and Jones (1981), Jones (1984)). At the very least, the next few years are likely to see the adoption of a much wider range of diagnostic tools and a more prevailing interactive and exploratory approach to statistical modelling in spatial analysis (Wrigley, 1983b). Pregibon (1981, 1982; see also Wrigley, 1983a) has shown how diagnostic methods for continuous data regression models can be extended to some of the centrally important models of categorical data analysis, and there seems to be great potential for adopting these methods as standard tools and attempting to extend their current scope to handle issues of spatial dependence. In this context there is a valuable analogy in, and perhaps some interesting potential linkages to, the recursive residual diagnostic tests used to assess the instability of regression relationships over time (see Hepple, 1979; Martin, 1979; Dunn, 1982).

Second, there has recently been growing interest in statistics in the intrinsic and centrally important relationship between sample survey design and statistical analysis, in particular the effects of complex survey design on classical statistical models and methods (e.g. Holt et al., 1980a, 1980b; Holt and Scott, 1981). Some of this (e.g. Holt et al., 1980a; Brier, 1980) addresses the impact of complex survey design on categorical data techniques and, in this context, it should be noted that spatial sampling produces similar problems (i.e. positive dependence between observations from similar origins which in the survey analysis context is said to produce a 'design effect' exceeding 1). As a result, we foresee considerable potential in exploiting some of the central features of this categorical data/complex sampling scheme literature to accommodate spatial dependence effects in categorical data models, and we note that Fingleton (1983a, 1983b) has already begun to explore these possibilities in the context of log-linear modelling.

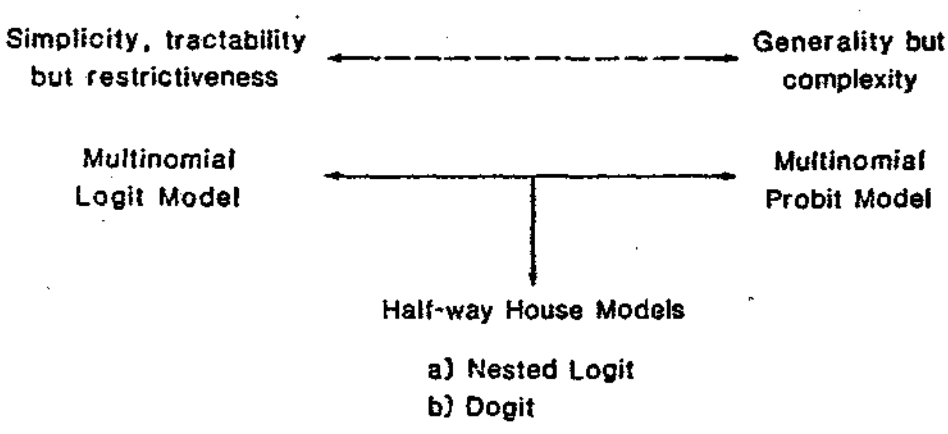
3. DISCRETE CHOICE MODELLING

The area of research which has come to be known as discrete choice modelling represents the interlinkage of developments in micro-economic theory, psychological choice theory, and the statistical analysis of categorical data. As such it is merely a convenience to separate developments in this area from those in categorical data analysis/generalized linear modelling, and the trends discussed above must inevitably influence discrete choice modelling. Nevertheless, there are certain debates and developments which are currently restricted to the discrete choice literature and these can be expected, reciprocally, to influence the previously identified trends.

3.1 The search for less restrictive discrete choice models

For a considerable time, examples of situations in which the 'independence from irrelevant alternatives' (IIA) property of the multinomial logit (MNL) model will yield counter-intuitive behavioural forecasts have been known (see Wrigley, 1984, for illustrations). This has given rise to a search for alternative discrete choice models based upon less demanding assumptions and with less restrictive properties of cross-substitution embodied in their structure.

Figure 2 Alternative discrete choice models



A summary of the nature of this search is provided by Figure 1 (see for more details also Wrigley, 1983). The most general least restrictive, of the alternative discrete choice models is the multinomial probit (MNP) model (Daganzo, 1979). The MNP allows the random components of utility of the choice alternatives to be correlated and to have unequal variances. It also permits random taste variation across individuals. By restricting the form of the variance-covariance structure of the MNP model, a family of special cases of the MNP can be specified, and these can provide (e.g. Miller and Lerman, 1981) considerable insight into the nature of spatial choice problems. Unfortunately, the MNP model is conceptually complex and computationally unwieldy, and the most popular estimation procedure (which uses the Clark approximation to reduce the estimation problem to one of sequential univariate integration) remains controversial (Sheffi, Hall and Daganzo, 1982). As a result, there has been a search for what might be termed 'half-way house' models, which lie somewhere between the generality and complexity of the MNP model and the restrictiveness but tractability of the MNL model.

The two most popular 'half-way house' models are the dogit (Gaudry and Dagenais, 1979) and the nested logit (see Sobel, 1980, for an introductory review). In particular, the nested logit, which is a special case of the so-called generalized extreme value (GEV) model, has recently held a prominent theoretical and empirical position. The work of Williams (1977), Daly and Zachary (1978), Ben-Akiva and Lerman (1979), and McFadden (1979) established the consistency of nested logit models with random utility maximization. It also established the correct form and significance of the composite utility terms and the associated parameters within such models.

The nested logit model can readily handle correlated random components of utility, and it thus embodies more general properties of cross-substitution than the MNL without sacrifice of computational tractability.

The discrete choice models in Figure 2 belong to a general set of choice models which are termed 'compensatory' models. In addition, most of them assume that the parameters of the utility function are 'context free' and that there is no shifting in the valuation (weight) given to an attribute as the choice set changes in composition. Both of these characteristics of existing discrete choice models have recently been questioned. Timmermans (1983a), for example, has considered the potential of non-compensatory choice models, though as yet the only members of this class of models which have assumed significance are the elimination-by-aspects (EBA) models proposed by Tversky (1972a, 1972b) and to some extent these are a special case being 'globally' compensatory in nature though non-compensatory 'locally' (Tversky, 1972a, p.296). Similarly, Meyer and Eagle (1981, 1982) have attempted to incorporate context induced parameter instability into discrete choice models, and have proposed a set of adjusted logit models which seek, simultaneously, to overcome problems caused by correlated random components of utility and 'context dependence/weight shifting'.

The next five years are likely to see a continuation of this search for less restrictive but computationally tractable discrete choice models. However, this will probably occur at a slower rate as there is now an urgent need to gain wider empirical experience of the performance of the current generation of models. In this context, there is considerable potential for cross-fertilization with the categorical data/generalized linear modelling literature. For example, MNP models are little known outside the specialized discrete choice modelling literature, and it would be of considerable benefit to see their structure and computational problems debated within the context of this wider literature. Similarly, assessment of the empirical performance of competing discrete choice models can only be aided by a sensitivity to the potential of the logistic/logit model diagnostics suggested by Pregibon (1981, 1982).

3.2 Choice experiment data, decompositional multiattribute preference models, and discrete choice models

Discrete choice models are calibrated using survey data on observed choice behaviour. Such data are often subject to many sources of confoundment, and are often available for only a limited range of attribute levels: a range which may be insufficient to permit extrapolation of choice predictions to the unobserved attribute mixes which are likely to define any 'new' alternatives which might enter the choice set. In these circumstances, preference data obtained from controlled choice experiments involving hypothetical alternatives may have a particularly valuable role to play. Such experiments allow the separation out of error components which become confounded in normal sample survey data, and allow decision makers to formulate preferences over new or unattainable alternatives.

In practice, choice experiments of this kind form the basis of a class of choice models known as 'decompositional multiattribute preference models' (see Timmermans, 1983a, Timmermans et al, 1983).

Such models are sometimes regarded as competitive alternatives to discrete choice models; the issue being closely related to the relative merits of revealed versus expressed preferences. However, there is a growing consensus (e.g. Timmermans, 1983b; Hensher and Louviere, 1981) that the two types of models may be used in a complementary way depending on the type of spatial choice problem. Timmermans (1983b), for example, has noted that:

> 'Discrete choice models might prove to be most valuable for predicting discretionary behaviour which is characterised by relatively homogeneous and monotonic underlying preferences and compensatory composition rules as a response to relatively minor and continuous changes in the attributes of the choice alternatives. In contrast, decompositional multiattribute preference models seem to be better for predicting constrained behaviour with large variations in underlying preferences and type of composition rules as a response to substantial and discontinuous changes in the attributes of the choice alternatives or to the introduction of new alternatives'.

Over the next five years, we foresee a closer integration of discrete choice models and decompositional multiattribute preference models, and a wider sensitivity to the appropriate roles of choice experiment and sample survey data. However, once again, there is currently an urgent need for comparative empirical studies to assess the performance of both types of models.

3.3 Sample design effects

A majority of applications of discrete choice models assume that parameter estimation will be conducted using standard maximum likelihood techniques on sample survey data collected from simple or exogenously stratified random sample designs. However, the search for refined cost effective sampling procedures in discrete choice modelling has frequently led to the adoption of an endogenous (choice-based) approach to sample stratification. This approach presents additional estimation difficulties but it has generated important research (e.g. Lerman and Manski, 1979; Manski and McFadden, 1981; Coslett, 1981) which serves to stress the intrinsic and centrally important relationship between sample survey design and statistical analysis. As noted above, recent attempts to accommodate spatial dependence effects into log-linear modelling (Fingleton, 1983a; 1983b) have drawn upon the growing literature in statistics on the question of categorical data and complex (clustered, multistage, etc.) sampling schemes (e.g. Holt et al, 1980a). As a result, we foresee considerable potential in attempts to link the discrete choice modelling, categorical data and sampling survey literatures through a wider and deeper understanding of the effects of sample design on the most commonly employed statistical methods for qualitative data.

3.4 Panel data and dynamic modelling

Discrete choice models have traditionally been estimated using cross-sectional sample survey data. Recently, however, there has been increased interest in the potential of longitudinal or panel data, for it is well known that in the analogous context of conventional continuous data linear models the use of panel data often results in greater efficiency, in both statistical and behavioural terms, than the use of separate cross-sectional samples.

Panel data discrete choice models must attempt to capture two main effects. The so-called state dependence (or 'structural state dependence') effect results from the fact that individuals' choices are to some extent dependent on their previous choices, whereas the serial correlation (or 'spurious state dependence') effect results from neglected or unobserved attributes which do not change (or change only slowly) from one time period to the next. This latter effect usually arises from individual differences in the propensity to experience an event and it, therefore, usually reflects the effect of population heterogeneity. The problem, of course, is to distinguish between the effects of spurious and structural state dependence. Halperin (this volume) reviews the work of Heckman (1981), Daganzo and Sheffi (1979, 1982), Tardiff (1980), Johnson and Hensher (1982) and shows that considerable progress has recently been made in this context, particularly involving use of the MNP model. In fact, Daganzo and Sheffi (1982) have recently suggested that MNP panel data problems with either structural or spurious state dependence are merely special cases of a more general model, and that choice of a structural state dependence model, serial correlation (spurious state dependence) model, or any combination of the two, is simply a specification issue to be decided by the modeller.

The analysis of longitudinal/panel/event history data is currently a major area of advance in statistical methodology and we foresee continued theoretical development in this area of discrete choice modelling. Recently, a number of major panel surveys have been undertaken by transport planners, economists and geographers, some of which contain much richer locational information than typically available in previous panel data sets. As such, we look forward to empirical studies in which the recently developed discrete choice models for panel data are assessed, and to the development and empirical testing of alternative approaches to modelling the time-space activity data recorded by such surveys.

4. QUALITATIVE MULTIVARIATE MODELS

The area of qualitative multivariate analysis is very heterogeneous. It varies from path analysis and linear structural models to correspondence analysis and scaling techniques. Some elements in this field have already received attention from econometricians and geographers for many years (for instance, latent variables models), while other elements are still underdeveloped (for instance, correspondence analysis). In the present section some major issues are selected for further discussion notably linear structural equation models and partial least squares models.

4.1 Linear structural equation models (Lisrel)

Linear structural equation (Lisrel) models are based on latent variable methods and may be regarded as a synthesis of path analysis, factor analysis and simultaneous equation models. Thanks to the path breaking work of Jöreskog and Goldberger (1975), Lisrel models have received increasing attention in econometrics (see for a discussion and survey among others Bentler, 1983, and Folmer, 1983). Lisrel is now fully operational, and has been applied to many empirical problems.

A new direction in Lisrel modeling has recently been provided by Muthén (1983), who demonstrated that categorical data can also be included in conventional Lisrel models, by applying either limited information (univariate and bivariate) multi-stage least squares or full information maximum likelihood procedures. Furthermore, it can be shown that longitudinal data can also be taken into account in this qualitative Lisrel approach. This new research direction means that qualitative data analysis may now be fully extended to Lisrel models, so that latent variables methods with qualitative data can be cast in an explanatory framework.

Another important research direction in Lisrel modeling has been pointed out by Folmer (1983), who has dealt with spatial auto- and cross-correlation in the framework of Lisrel models for spatial impact analysis. By regarding the impacts of spatial auto- and cross-correlation as latent impacts, he was able to identify spatial auto- and cross-correlation and to respecify regional models suffering from spatial spillover effects. Clearly, this research issue is highly relevant for spatial analysis and deserves full attention in future regional and urban analyses.

4.2 Partial least squares (PLS)

Partial least squares (PLS) is now a well established method in quantitative research and systems analysis thanks to the innovative work of Herman Wold (see for a review also Dijkstra, 1983). PLS models are - unlike maximum likelihood methods - not based on the stringent 'hard' assumption that the observables are joinly characterized by a specified multivariate distribution subject to independent observations. Another advantage of PLS methods over conventional maximum likelihood techniques is that these methods also provide predictive inferences in case of qualitative variables.

PLS methods aim at bridging the gap between substantive theory and statistical techniques by providing a general scope for causal-predictive analysis in causal chain models. As PLS models stem from path analysis with latent variables and as the PLS approach is distribution-free and not impaired by interdependence among the observations, it is a very general and flexible technique for causal-predictive inferences. In this respect, PLS can also treat qualitative (notably categorical) data in a straightforward way.

New directions in qualitative PLS modeling can be found inter alia in contingency table analysis with categorical data, in qualitative canonical correlation analysis, in nested qualitative causal models, and in mixed (quantitative and qualitative) path analysis. Especially in the area of large-scale and complex models based on qualitative information regarding the variables describing this system PLS models offer a potentially important tool.

Unfortunately, very little attention has so far been given to spatially dependent variables in PLS models. As such, this represents a potentially important research area which deserves development. Similarly, closer attention should also be given to the Stone-Geisser test for predictive relevance and to the jackknife statistic in case of spatially dependent qualitative data (see Folmer, 1983). This would also be a rich field of future research in regional science.

5. CONCLUSION

The field of qualitative data analysis represents an extremely rich research area which is continuing to develop at a rapid pace. There is at present a wide variety of methods available that may be helpful tools for classification and reduction problems, statistical tests, econometric modeling and plan evaluation problems. Despite many promising advances, there are still many methodological problems which remain. Some of these have been noted above. Others include:

- the design of methods especially appropriate for consistent linking of qualitative and quantitative data in individual models
- the design of modules for a consistent integration of models in a multiregional context, some of which may be based on precise information and others on qualitative data
- the aggregation and scale problems regarding space, time and model dimensions (as well as the linkages between these three items), when some or all of the models are based on qualitative data
- the nature and validity of qualitative analysis in case of forecasting problems
- the identification of spatial auto- and cross-correlation in case of qualitative spatial data
- the treatment of qualitative data in optimization models (e.g., in a spatial interaction context).

Thus the final conclusion from this essay is that qualitative data analysis has the potential to provide one of the focal points for future research in spatial analysis.

References

Arminger, G., Analysis of Qualitative Individual Data and of Latent Class Models with Generalized Linear Models, Measuring the Unmeasurable (P. Nijkamp, H. Leitner and N. Wrigley eds.), Martinus Nijhoff, The Hague, 1983 (forthcoming).

Ben-Akiva, M.E. and Lerman, S.R. 1979, Disaggregate Travel and Mobility Choice Models and Measures of Accessibility, in D.A. Hensher and P.R. Stopher, eds., Behavioural Travel Modelling, (Croom Helm, London), 654-79.

Bennet, R.J., 1979, Spatial Time Series: Analysis, Forecasting and Control, (Pion, London).

Bentler, P.M., 1983, Simultaneous Equation Systems as Moment Structure Models, Journal of Econometrics, vol. 22, no. 1/2, pp. 13-42.

Besag, J.E., 1974, Spatial Interaction and the Statistical Analysis of Lattice Systems, Journal of the Royal Statistical Society B 36, 192-236

Besag, J.E., 1975, The Statistical Analysis of Non-Lattice Data, The Statistician 24, 179-95

Brier, S.S., 1980, Analysis of Contingency Tables Under Cluster Sampling, Biometrika 67, 591-96

Clarke, M.R.B., 1982, GLIM 4 - The New Facilities, in R. Gilchrist, ed., GLIM 82: Proceedings of the International Conference on Generalized Linear Models, (Springer, New York) 25-35

Cliff, A.D. and Ord, J.K., 1973, Spatial Autocorrelation, (Pion, London)

Cliff, A.D. and Ord, J.K., 1981, Spatial Processes: Models and Applications, (Pion, London)

Coslett, S.R., 1981, Efficient Estimation of Discrete Choice Models, in: C.F. Manski and D. McFadden (eds.) Structural Analysis of Discrete Data With Econometric Applications, (M.I.T. Press, Cambridge, Massachusetts)

Cox, N.J. and Jones, K., 1981, Exploratory Data Analysis, in: N. Wrigley and R.J. Bennett eds. Quantitative Geography: A British View, (Routledge and Kegan Paul, London), 135-43

Daganzo, C.F., 1979, Multinomial Probit: The Theory and Its Application to Demand Forecasting, (Academic Press, New York)

Daganzo, C.F. and Sheffi, Y., 1979, Estimation of Choice Models from Panel Data, Paper presented at 1979 Meeting of the Regional Science Association, Los Angeles, California

Daganzo, C.F. and Sheffi, Y., 1982, Multinomial Probit with Time-Series Data: Unifying State Dependence and Serial Correlation Models, Environment and Planning A, 14, 1377-88

Daly, A.J. and Zachary, S., 1978, Improved Multiple Choice Models, in: D.A. Hensher and M.Q. Dalvi eds., Determinants of Travel Choice,(Teakfield, Farnborough, Hants.), 335-57

Dunn, R., 1982, Parameter Instability in Models of Local Unemployment Responses, Environment and Planning A, 14, 75-94

Dijkstra, Th., 1983. Some Comments on Maximum Likelihood and Partial Least Squares Methods, Journal of Econometrics, vol. 22, no. 1/2, pp. 67-90

Fingleton, B., 1983a, Independence, Stationarity, Categorical Spatial Data, and the Chi-Squared Test, Environment and Planning A 15, 483-99

Fingleton , B., 1983b, Log-linear Models With Dependent Spatial Data, Environment and Planning A 15, 801-13

Flowerdew, R., 1983, Poisson Regression Models For the Analysis of Geographical Data, SSRC Research Grant D/00/23/0083

Flowerdew, R. and Aitkin, M., 1982, A Method of Fitting the Gravity Model Based on the Poisson Distribution, Journal of Regional Science, 22, 191-202

Folmer, H., 1983, Measurement of Effects of Regional Economic Policy, Ph.D. Diss., Dept. of Economics, University of Groningen, The Netherlands.

Gaudry, M.J.I. and Dagenais, M.G., 1979, The Dogit Model, Transportation Research B 13B, 105-11

Griffith, D.A., 1980, Towards a Theory of Spatial Statistics, Geographical Analysis 12, 325-39

Haining, R.P., 1980, Spatial Autocorrelation Problems, in: Herbert, D.T. and Johnston, R.J., (eds.) Geography and the Urban Environment Vol. 3, (John Wiley, Chichester), 1-44

Haining, R.P., 1982, Interaction Models and Spatial Diffusion Processes, Geographical Analysis 14, 95-108

Haining, R.P., 1983, Spatial and Spatial-Temporal Interaction Models and the Analysis of Patterns of Diffussion, Transactions Institute of British Geographers, New Series 8, 158-86

Heckman, J.J., 1981, Statistical Models for Discrete Panel Data, in: C.F. Manski and D. McFadden, eds., Structural Analysis of Discrete Data: With Econometric Applications (M.I.T. Press, Cambridge, Massachusetts), 114-78

Hensher, D.A. and Louviere, J.J., 1981, An Integration of Probabilistic Discrete Choice Models and Experimental Design Data: An Application in Cultural Economics, School of Economic and Financial Studies, Macquarie University, NSW 2113, Australia

Hepple, L.W., 1974, The Impact of Stochastic Process Theory Upon Spatial Analysis in Human Geography, in: Board, C., Chorley, R.J., Haggett, P. and Stoddart, D.R. eds., Progress in Geography 6, (Edward Arnold, London), 89-142

Hepple, L.W., 1979, Regional Dynamics in British Unemployment, and the Impact of Structural Change, in: N. Wrigley, ed., Statistical Applications in the Spatial Sciences, (Pion, London)

Holt, D. and Scott, A.J., 1981, Regression Analysis Using Survey Data, The Statisticia 30, 169-78

Holt, D., Scott, A.J. and Ewings, P.D., 1980a, Chi-Squared Tests with Survey Data, Journal of the Royal Statistical Society A, 143, 303-20

Holt, D., Smith, T.M.F. and Winter, P.D., 1980b, Regression Analysis of Data from Complex Surveys, Journal of the Royal Statistical Society A, 143, 474-87

Hordijk, L. and P. Nijkamp, 1977, Dynamic Models of Spatial Interaction. Environment & Planning A, vol. 9, pp. 505-519.

James, L.R., Mulaik, S.A., and J.M. Brett, Causal Analysis: Assumptions, Models and Data, Sage Publ., Beverly Hills, 1982.

Johnson, L.W. and Hensher, D.A., 1982, Application of Multinomial Probit to a Two-Period Panel Data Set, Transportation Research A, 16A, 457-64.

Jones, K., 1984, Confirmatory and Exploratory Data Analysis in Geography, (Croom Helm, London).

Jöreskog, K.G., and A.S. Goldberger, 1975, Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable, Journal of the American Statistical Association, vol. 70, pp. 631-639.

Leeuw, J. de, W.J. Keller, and T. Wansbeek (eds.), 1983, Introduction to the Special Issue on 'Interfaces between Econometrics and Psychometrics', Journal of Econometrics, vol. 22, no. 1/2, pp. 1-15.

Lerman, S.R. and Manski, C.F., 1979, Sample Design for Discrete Choice: State of the Art, Transportation Research A, 13A, 29-44

McCullagh, P, 1980, Regression Models for Ordinal Data (With Discussion), Journal of the Royal Statistical Society B, 42, 109-42

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

McFadden, D., 1979, Quantitative 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), 279-318

Manski, C.F. and McFadden, D., 1981, Alternative Estimators and Sample Design for Discrete Choice Analysis, in: C.F. Manski and D. McFadden, eds., Structural Analysis of Discrete Data With Econometric Applications, (M.I.T. Press, Cambridge, Massachusetts), 2-50

Martin, R.L., 1979, Subregional Phillips Curves, Inflationary Expectations, and the Intermarket Relative Wage Structure: Substance and Methodology, in: N. Wrigley, ed., Statistical Applications in the Spatial Sciences, (London, Pion), 64-110

Meyer, R.J. and Eagle, T.C., 1981, A Parsimonious Multinomial Choice Model Recognising Alternative Interdependence and Context-Depenedent Utilty Functions, Carnegie-Mellon University, Graduate School of Industrial Administration, Working Paper 26-80-81

Meyer, R.J. and Eagle, T.C., 1982, Context-Induced Parameter Instability in a Disaggregate Stochastic Model of Store Choice, Journal of Marketing Research 19, 62-71

Miller, E.J. and Lerman, S.R., 1981, Disaggregate Modelling and Decisions of Retail Firms: A Case Study of Clothing Retailers, Environment and Planning A, 13, 729-46

Muthén, B., 1983, Latent Variable Structural Equation Modeling with Categorical Data, Journal of Econometrics, vol. 22, no. 1/2, pp. 43-65.

Nelder, J.A. and Wedderburn, R.W.M, 1972, Generalized Linear Models, Journal of the Royal Statistical Society A, 135, 370-84.

Nelder, J.A., Statistical Models for Qualitative Data, (P. Nijkamp, H. Leitner and N. Wrigley, eds.), Martinus Nijhoff, The Hague, 1983 (forthcoming).

Nijkamp, P., A Survey of New Methods in Qualitative Spatial Data Analysis, Research Paper, Dept. of Economics, Free University, Amsterdam, 1983 (mimeographed).

Nijkamp, P., H. Leitner, and N. Wrigley (eds.), Measuring the Unmeasurable, Analysis of Qualitative Spatial Data, Martinus Nijhoff, The Hague, 1983.

O'Brien, L.G., 1983, Generalized Linear Modelling Using the GLIM Sytem, Area 15,

O'Brien, L.G. and Wrigley, N., 1983, A Generalized Linear Models Approach to Categorical Data Analysis, in: G. Bahrenberg, M.M. Fischer and P. Nijkamp, eds., Recent Developments in Spatial Analysis: Methodology, Measurement, Models, (Gower, Aldershot)

Odland, J. and Barff, R., 1982, A Statistical Model for the Development of Spatial Patterns: Applications to the Spread of Housing Deterioration, Geographical Analysis 14, 327-39

Pregibon, D., 1981, Logistic Regression Diagnostics, The Annals of Statistics 9, 705-24

Pregibon, D., 1982, Resistant Fits for Some Commonly Used Logistic Models with Medical Applications, Biometrics 38, 485-98

Roberts, F.S., 1979, Measurement Theory, with Applications to Decisionmaking, Utility, and the Social Sciences, Addison Wesley, Reading, Mass.

Roger, J.H., 1983, Composite Link Functions With Linear Log Link and Poisson Error, GLIM Newsletter 7, 15-21

Sheffi, Y., Hall, R. and Daganzo, C.F., 1982, On the Estimation of the Multinomial Probit Model, Transportation Research A, 16A, 447-56

Sobel, K., 1980, Travel Demand Forecasting with the Nested Multinomial Logit Model, Transportation Research Record 775, 48-55

Tardiff, T.J., 1980, Definition of Alternatives and Representation of Dynamic Behaviour in Spatial Choice Models, Transportation Research Record 723, 25-30

Thompson, R. and Baker, R.J., 1981, Composite Link Functions in Generalized Linear Models, Applied Statistics, 30, 125-31,

Timmermans, H.J. P., 1983a, Decision Models for Predicting Preferences Among Multiattribute Choice Alternatives, in: G. Bahrenberg, M.M. Fischer and P. Nijkam eds., Recent Developments in Spatial Analysis: Methodology, Measurement, Models, (Gower, Aldershot)

Timmermans, H.J.P., 1983b, Decompositional Multiattribute Preference Models in Spatia Choice Analysis: A Review of Some Recent Developments, Progress in Human Geography (forthcoming)

Timmermans, H.J. P., van der Heijden, R. and Westerveld, H., 1983, Decision-Making Among Multiattribute Choice Alternatives: A Model of Spatial Shopping Behaviour Using Conjoint Measurements, Environment and Planning A, 15,

Tversky, A., 1972a, Elimination-by-Aspects: A Theory of Choice, Psychological Review 79, 281-99

Tversky, A., 1972b, Choice-by-Elimination, Journal of Mathematical Psychology 9, 341-

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

Williams, H.W.C.L., 1977, On the Formation of Travel Demand Models and Economic Evaluation Deasures of User Benefit, Environment and Planning A, 9, 285-344.

Wrigley, N., 1983b, Quantitative Methods: On Data and Diagnostics, Progress in Human Geography, 7, pp. 565-575.

Wrigley, N., 1984, Categorical Data Analysis for Geographers and Environmental Scientists, (Longman, London).

Wrigley, N., Categorical Data Analysis and Discrete Choice Modeling in Spatial Analysis, Measuring the Unmeasurable; Analysis of Qualitative Spatial Data. P. Nijkamp, H. Leitner and N. Wrigley (eds.), Martinus Nijhoff, The Hague, 1983 (forthcoming).

Serie Research Memoranda (1)

1980-1	P. Nijkamp and H. Voogd	New Multicriteria Methods for Physical Planning by Means of Multidimensional Scaling Techniques
1980-2	H.P. Smit	Medium- and Long-Term Models for the ESCAP-Region - A Review of exisging models and a proposal for a new model system
1980-3	P.v. Dijck en H. Verbruggen	Productive Employment in Developing Countries' Exporting Industries
1980-4	P. Nijkamp and L. Hordijk	Integrated Approaches to Regional Development Models; A survey of some Western European Models
1980-5	P. Nijkamp	Soft Econometric Models; An Analysis of Regional Income Determinants
1980-6	P. Nijkamp and F.v. Dijk	Analysis of Conflicts in Dynamical Environmental Systems via Catastrophe Theory
1980-7	E. Vogelvang	A short term econometric Model for the Consumer demand of Roasted Coffee in The Netherlands
1980-8	N.van Hulst	De effectiviteit van de geleide loonpolitiek in Nederland
1980-9	P. Nijkamp	A survey of Dutch integrated Energy-Environmental-Economic Policy Models
1980-10	P. Nijkamp	Perspectives for Urban Analyses and Policies
1980-11	P. Nijkamp	New developments in Multidimensional Geographical data and Policy Analysis
1980-12	F.C. Palm, E. Vogelvang and D.A. Kodde	Efficient Estimation of the Geometric Distributed Lag Model; some Monte Carlo Results on Small Sample Properties

---

1981-1	E. Vogelvang	A quarterly econometric model for the Price Formation of Coffee on the World Market
1981-2	H.P. Smit	Demand and Supply of Natural Rubber, Part I
1981-3	R. Vos	The political Economy of the Republic of Korea; A proposal for a model framework of an open economy in the ESCAP-region, with emphasis on the Role of the State
1981-4	F.C. Palm	Structural Econometric Modeling and Time Series Analysis - Towards an Integrated Approach