DISAGGREGATE MODELS OF CHOICE
IN A SPATIAL CONTEXT 1)

Wal van Lierop
Peter Nijkamp

Research memorandum 1981-14

Paper presented at the IIASA-seminar on 'Urban Systems Modelling',
Moscow, October 1980.

1) The authors thank Wietze Boomsma, Franz Palm, Piet Rietveld and Koos Sneek for their helpful comments on an earlier version of this paper.
1. **INTRODUCTION**

The analysis of spatial allocation and choice patterns in urban and regional systems is a central issue in regional science and geography. In many traditional analyses, cost-minimizing or utility-maximizing principles have been used to explain and predict spatial behaviour of people; economic decision criteria were pivotal elements in analyzing regional and urban mobility patterns.

Since, however, many western industrialized countries have reached a full maturity, the locational and mobility choices are guided by less purely economic-oriented criteria. Many qualitative aspects (such as environmental conditions and the quality of the housing stock) are increasingly influencing spatial choice behaviour. This evokes the need for a broader analysis of spatial choice mechanisms.

Furthermore, the allocation and choice patterns demonstrate an increasing heterogeneity and diversity among spatial actors, so that also the question as to the *scale* of analysis (aggregate versus disaggregate) becomes crucial.

This paper is devoted to a critical survey of modern choice models, which are being used nowadays, among others, in attempts to approach spatial allocation and choice patterns in a more realistic way, by means of taking into account the abovementioned qualitative aspects and elements of scale in addition to economic decision criteria. After a methodological introduction, several categories of such choice models are reviewed in a spatial context. Having evaluated these classes of choice models, we will pay more specific attention to two disaggregate choice models, viz. the multinomial logit and the multinomial probit model. The features of the latter pair of models will be illustrated by means of a numerical exercise.

2. **METHODOLOGICAL REMARKS**

The methodology of spatial choice analysis can be based on several theoretical frameworks. Two main categories can be distinguished, viz. the traditional (mainly neoclassical) and the behavioural theories.

The traditional approach to spatial choice analysis takes for granted the notions of utility and indifference. Usually, the choice criteria in microeconomic decision-making are assumed to be the same for all individuals, though the shape of the individual utility functions is not necessarily equal. Consequently, the same set of attributes of a commodity or of an alternative or the same set of commodities will normally not lead to the same utility for all subjects, while also interpersonal utility compar-
isons are often not possible. It is clear that the application of this traditional theory to spatial groups at a more aggregate level (e.g., social classes, income groups) has until now assumed a uniform utility function for all members of that group.

This traditional approach aims at explaining and forecasting spatial interaction patterns on the basis of classical assumptions of rational behaviour and perfect information. There are several limitations inherent in this approach: consumer interactions (e.g., bandwagon effects) are normally neglected, learning effects are left aside, rational decision-making under perfect information is usually an illusion for spatial choice behaviour (such as migration, commuting and shopping), complementarity of individual choices (such as multi-purpose trips) are usually abandoned and often no insight is given into the distribution of the several alternative choices over the groups, i.e. within the population.

The behavioural theories are based on motives and attitudes of decision-makers (or groups); see among others Burnett (1973), Clark and Cadwallader (1973), Downs (1970), Colledge and Brown (1967), Gould (1973), Rushton (1967), Saarinen (1976) and Simon (1957). In these theories, such notions as 'satisficer' principles, bounded rationality, behavioural environment and the distribution of choices over the population play an important role.

The behavioural theories can be subdivided into 2 main classes, viz. the revealed preference approach and the direct preference approach.

The first class aims at analyzing realized, single and complex choices, by means of information on past behaviour (see among others, Pirie (1976) and Rushton (1969, 1971)). The basic assumption is that human preferences and aims can be inferred a posteriori from the results of decisions, i.e. from actual behaviour. If one takes for granted consistency of individual choices (transitivity), a similar ranking of alternatives among individuals, and a sufficiently long time period to define indifference, such an ex post analysis can also be used to forecast future behaviour.

The revealed preference approach also has several limitations. Mental processes of consumers are neglected, uncertainties and constraints in choices are mainly left aside, and interaction effects are not taken into account, so that the strong parallel between actual behaviour and internal preference structures is illusive. Furthermore, due to lack of reliable information on actual individual behaviour, the revealed preference approach is often macroscopic (aggregate) in nature.
The direct preference approach analyzes choices and choice-processes *ex ante*, by means of (mainly individual or micro-economic) information on preferences and/or perceptions of choice-makers and alternatives, i.e. it concentrates directly on the process of decision making. The information, needed to estimate direct preference models, should be based on questionnaires about attributes of the relevant alternatives and choice-makers with clearly measurable values and preference and/or perception dummies. These real and dummy-values define together the 'expected utility' of the choice decisions.

Due to its ability to treat individual preference and perception data this direct approach is usually microscopic (disaggregate). It has the potential to take into account mental images and processes, such as the cognitive perception of commodities and external effects. Furthermore, the direct preference approach may take account of different sets of attributes for individuals and a multiplicity of aims. It is compatible with related areas of knowledge (such as psychology). Consequently, this approach leads to a more integrated economic-psychological framework for the explanation of choice behaviour in general and in socio-geographical space in particular.

Clearly, the direct preference approach is in this respect not yet entirely perfect: constraints on spatial choices emerging from the (behavioural) environment are often still hard to integrate, learning processes and future anticipations are hardly touched upon, while also interaction effects are difficult to integrate.

A more unambiguous evaluation of the various behavioural approaches to analyze choices and choice-processes, than by means of the already mentioned judgements, is very hard to give. This is caused by the fact that the aim of the analysis is not entirely the same for each approach individually. Clearly, one may use methodological, theoretical, logical and empirical criteria such as credibility, plausibility, soundness, or empirical verifiability to judge the various approaches (cf. Nooteboom (1980)), but the operationalization of such concepts depends on the specific aim of the analysis and on the reliability and nature of available data. In Table 1 we have tried to give a simple illustration of the difference between the revealed preference approach and the direct preference approach.
Table 1. Difference between revealed preference and direct preference approach

In Table 1 the position of the revealed preference approach in block II is clear: it analyzes realized choices by means of information on past decisions. The same holds for the position of the direct preference approach in block IV: it analyzes choice processes, based on interviews about present behaviour. The filling up of the fields I and III is, however, not directly obvious. For instance, those practical analyses of choices and choice-processes which in theory should preferably be done by means of the direct preference approach, are in the estimation phase often forced to use revealed preference data. When these data consist mainly of recent ex post data on actual behaviour (for example, in case of repetitive decisions) the influence of the direct preference approach will still be significant. When there exists, however, a bigger time-lag between the depending and explanatory variables in such an estimation model, the revealed preference approach will become more important. The conclusion arises that in blocks I and III of Table 1 interactions take place between the revealed and direct preference *theories* and their more practical model applications.
Given the above mentioned features of the various approaches, the behavioural theories and especially the direct preference approach appear to us - from a methodological point of view - to be fairly rich in scope (at least compared to the alternative approaches). Further, it has to be stressed that the direct preference approach is rather flexible: it is not at variance with the traditional approach, while some of its elements can also be used in an ex post revealed preference approach. However, in order to take better advantage of the direct preference approach, more future research should be done into its mainly practical imperfections. For the improvement of such imperfections recently already attempts have been undertaken to apply so-called 'longitudinal' analysis of the choice-makers in their decision process (see, among others, Baanders and Slootman, 1980, and Clark and Smith, 1981).

Although the empirical content of behavioural theories of choice, and more specific of direct preference theory compared to revealed preference theory, can only be judged after real world application, it should be noted that there are several techniques (such as multinomial logit and probit models and within their framework multidimensional scaling techniques) which can help to make this approach potentially useful as an explanatory and forecasting method for choice behaviour.

The rest of the paper will, in a spatial context, concentrate on a whole set of such techniques. The starting point will always be the direct preference theory. It has to be remarked, however, that many of the described models have - mainly due to data problems - regularly been applied as revealed preference models.

3. DIRECT PREFERENCE APPROACH AND DISAGGREGATE MODELS OF CHOICE

The direct preference approach assumes that when a certain individual has to choose one alternative from a given set, he will make his choice on the basis of an ex ante judgement process about which alternative will presumably give him ex post a maximum utility. In psycho-economic modelling this judgement process takes the form of a utility function which may include measured or reported, perceived and preferred values of all kinds of attributes defining the alternative concerned as well as of socio-economic variables characterizing the choice-maker. Such utility functions are used within the framework of a series of choice models that have been developed during the last decade. Our survey study is concerned with a systematic description of such quantitative methods for the analysis of individual, micro or disaggregate, spatial choice behaviour in order to get a better understanding of the spatial interactions on a macro or aggregate level.
It has to be noted that, although the philosophy of these methods is entirely concentrated on the choice process of any random individual, sometimes - for reasons of efficiency in the model phases of estimating or forecasting spatial behaviour - the relevant population is subdivided into significant different groups or classes. Such a classification can be based on features of groups in the population and/or availability of alternatives. In each class the scope of research will then be: what choices will any person - typified as belonging to that class - make? Aggregation over all classes will give again the choice patterns for the entire population.

The description of these disaggregate spatial choice methods is not entirely exhaustive in this paper, but in brief the most recent important developments will be reviewed, with the main emphasis on model specification, but occasionally also on model estimation, model evaluation and testing, and aggregation and forecasting, while also some attention will be paid to the coherence of the various model developments.

It has to be noted that the reader is assumed to be familiar with the general foundations, properties and use of disaggregate choice theory and of traditional spatial interaction methods in the form of gravity and entropy-information models (see, for instance, Stopher and Meyburg (1975, 1976), Hensher and Stopher (1979), Manheim (1979), Nijkamp (1979A) and Van Lierop and Nijkamp (1980).

Disaggregate models of choice can be divided into two main groups:

A. **Deterministic Models**

These are models in which the utility functions are supposed to give an exact description of the alternatives and lagging attributes. Formally, the derivation of the preferred alternative is essentially a mathematically logical conclusion, from which uncertainties - inherent in any choice situation - are eliminated. An aggregation to class or population behaviour presupposes identical utility functions (or identical attributes in the utility functions) concerning the available alternatives for all relevant individuals.

B. **Probabilistic Models**

In probabilistic choice theory the choice process of individuals can be described by a straightforward mathematical programming procedure, which defines the probability of a specific choice. This probability decision will partly depend on the observable attributes. In some models also non-directly measurable elements play an important role.

In other ones, however, the choice probabilities may even be influenced
by the mutual relationships between alternatives. In other words, this approach does not, by definition, imply that similar attributes lead to identical choices. The aggregation phase of a probabilistic model requires the definition or approximation of probability density functions for the choice of the available alternatives in the relevant class or in the population.

After a descriptive survey of several models of both groups, we will synthesize all these models based on their main specification features in section 11 in an integrating table form.

4. **DETERMINISTIC MODELS**

A. **Logit models**

If it is assumed that the relevant utility functions are all exhaustive (i.e., all relevant factors are included and exactly measured), we will find among the group of deterministic disaggregated models the well-known logit model, defining in this case the 'choice-ratio' of any individual for any alternative compared to all other alternatives:

\[
p_{in} = \frac{\exp u_{in}}{\sum_{i' = 1}^{I} \exp u_{i'n}} ; \quad i, i' = 1, \ldots, I; \quad n = 1, \ldots, N
\]

with:

- \( p_{in} \) = 'choice'-ratio of individual for an alternative compared to all other alternatives (i.e., the ratio in which individual will 'choose' alternative out of a set of alternatives, he has at his availability),
- \( u_{in} \) = the utility of alternative for individual .

\( u_{in} \) is based on \( u(z_{in}) \), a utility function of a vector \( z_{in} \) of explanatory, non-stochastic variables \( z_{jin} \) (\( j = 1, \ldots, J \)).

\( J \) is the number of variables \( j \) that define alternative for individual .

In the deterministic logit model two special assumptions are made concerning \( u(z_{in}) \):

1. \( u(z_{in}) \) is exhaustive, which means that \( z_{in} \) exists of exactly measured values for all the lagging - both alternative-bounded, and relevant socio-economic - attributes \( z_{jin} \), which are assumed to be mutually (functionally) independent.

2. Mostly \( u(z_{in}) \) is defined by the following straightforward linear form:

\[
u(z_{in}) = \sum_{j=1}^{J} z_{jin}
\]

\( J \) is the number of variables \( j \) that define alternative for individual .
By construction of (4.1), the following condition holds:
\[
\sum_{i=1}^{k} p_{i} = 1
\]
(4.3)
i.e., the additivity constraints should be satisfied.

A very important hypothesis, which makes the deterministic logit model possible, is the 'independence from irrelevant alternatives axiom'. This property states that the relative ratio of 'choices' among two particular alternatives is not influenced by introducing other choice options. The advantage of the 'independence from irrelevant alternatives' is obviously that the model is easy to handle in terms of estimating and forecasting. A disadvantage is however that when, for instance, a third alternative almost similar to one of the already existing alternatives is taken into consideration as a relevant one, it is very difficult to give a good representation of the various 'choice' ratios. This problem may be illustrated by the well-known red-bus blue-bus example of Debreu (1960).

Assume that a certain individual can take the car or a red bus to go to work. The ratios of these traffic modes are 2/3 and 1/3, in other words a ratio of 2:1. Suppose a third alternative, a blue bus, is introduced, which is exactly the same as the first bus except for its colour, but this aspect may be regarded as not of any direct value for the choice between the several modes of travel. Intuitively one would expect as new proportions: 2/3, 1/6, and 1/6. According to the axiom of independence from irrelevant alternatives, however, the ratio of 2:1 between the car and the red bus has to remain the same; consequently, the new choice proportions will become 1/2, 1/4 and 1/4, which gives a wrong idea of reality. Of course this example is rather extreme, but still it illustrates how the calculation of 'choice' ratios is influenced in a negative way, when the available alternatives are quite similar. In case the new alternative has features which are completely dissimilar to the initial ones, this disadvantage will not occur.

It should already be mentioned here that it is very difficult - if not impossible - to solve the above described problem completely within the framework of the logit model, because the axiom of independence from irrelevant alternatives is one of the main assumptions of the logit model. Only a step-wise or 'nested' logit approach may get round part of the difficulty.
B. Gravity and Entropy models

Utility based (and in a micro scope developed) gravity and entropy models - which are traditionally more applied to macro situations - can also be regarded as deterministic models of choice (see among others, Nijkamp (1979A)). This is no surprise as there exists a direct relationship between the entropy concept from information theory and the logit model (see for more detailed description, Van Lierop and Nijkamp (1979)).

The basic feature of gravity and entropy models is the assumption that any spatial interaction is the result of three forces: attractiveness at a point of origin, attractiveness at a point of destination and distance friction. On the basis of information on these three forces (and of their related parameters), the spatial flow pattern at the macro level can be assessed in a deterministic way. With the aid of that, micro 'choice' ratios can be derived.

During the last decade these models have gained much popularity in geography and urban and regional science, especially thanks to their relatively simple framework. They have, however, also encountered much criticism for their restrictive underlying assumptions, particularly due to their rather physical concept, with which it is not really possible to explain spatial choice behaviour and interaction processes.

5. PROBABILISTIC MODELS

Two main reasons can be formulated supporting the probabilistic approach, (see also Andersson and Philipov (1980)). This approach is plausible:

(1) when the attributes are properly observed - i.e., the measurement and/or perception of all attributes resulted in 'realistic', exactly defined, corresponding values -, but the decision-making process is stochastic, or the individual does not consistently maximize his utility in terms of neo-classical micro-economic theory (i.e., repeatedly confronted with the same set of exactly described alternatives he will not permanently make the same choices);

(2) when the individual is acting rationally in terms of the neo-classical theory, but when some of the attributes are missing, either just unobserved and/or are observed with an error of measurement. In this case the probability approach can be defended on the basis of observation problems.

The consequence of these two reasons is the emergence of two directions of theoretical research

1. the constant utility model,
2. the random utility models.
These approaches regularly take the form of a variation on the basic logit concept. The derivation of the logit model in the constant utility method is, however, different from the one in the more general random utility method. They will therefore both be discussed separately in section 6 and 7 - 10, respectively.

6. THE CONSTANT UTILITY MODEL

The basic assumption for this model is the so-called choice axiom or IIA problem as defined by Luce (1959). This is the probabilistic version of the hypothesis of independence of irrelevant alternatives in deterministic choice theory. The axiom states that the presence or absence of an alternative is irrelevant to the relative probabilities of choice between any two other alternatives, although, of course, the absolute values of these probabilities will generally be affected. The same advantages and disadvantages which count for the 'independence of irrelevant alternatives rule' are relevant for the choice axiom.

Based on this fundamental assumption Luce (1959) proved that it is possible to define:

\[ P_{in} = \frac{M_{in}}{\sum_{i=1}^{I} M_{i'n}} \quad ; \quad i,i' = 1,\ldots,I ; \quad n = 1,\ldots,N \]  

where:

- \( P_{in} \) is the choice-probability that individual \( n \) will choose alternative \( i \) from a set of \( I \) available alternatives. In the constant utility model this choice probability equals the proportion of the population whose choice for alternative \( i \) is determined by the same vector of explanatory attributes as the one which is relevant for \( n \). \( P_{in} \) is - in contrast with section 4 - written with a capital, because of its stochastic character.

- \( M_{in} \) is a ratio-number, a cardinal utility figure measured on a scale defined by all available alternatives \( I \), and based on:
  1. a function of the vector of all the relevant and perfectly described explanatory variables;
  2. the analytical way in which the (measured, expected or perceived utility, resulting from the explanatory variables, is being linked to the choice-probability at the interval 0 - 1.

The utility related to point (1) under \( M_{in} \) is denoted by \( u_{in} \), and is more or less the same as the comparable figure from the deterministic logit model. Like in this model, the related utility function, \( u(Z_{in}) \), is assumed to
be exhaustive. In the constant utility model one usually assumes, however, the following - less rigid - linear form for $u(z_{in})$:

$$u(z_{in}) = \sum_{j=1}^{J} \alpha_j z_{jin}$$

with $\alpha_j$ = a coefficient, a weighing factor.

The relation between the model which describes the stochastic process of defining $P_{in}$ and the deterministic utility $u_{in}$, is made by an analytic expression of point (2) under $M_{in}$; for instance by:

$$M_{in} = e^{u_{in}}$$

Substitution of this into (6.1) provides:

$$P_{in} = \frac{u_{in}}{\sum_{i'=1}^{I} e^{u_{i'n}}}$$

This description shows that only the process of choosing alternative $i$ can be stochastic in the constant utility model, while the utility, $u_{in}$, itself is fixed and exactly defined.

To illustrate the connection with the logit model framework in a better way, the following statement is made: in fact the above-mentioned means assuming that the probability $P_{in}$ is proportional to an arbitrary function $G$ of the vector of the explanatory - alternative related and socio-economic - variables. In formula:

$$P_{in} = \beta_n G(z_{in})$$

with:

$$\beta_n = \text{a coefficient of proportionality.}$$

The result of this proportionality assumption is, that the utility function (6.2) does not depend on the actual choice made, as the utility itself is constant! It is predetermined by the exhaustive (fixed) vector of explanatory variables. Taking into account the constraint that:

$$\sum_{i=1}^{I} P_{in} = \sum_{i=1}^{I} \beta_n G(z_{in}) = 1$$
it is easily seen that:

\[ p_n = \frac{1}{\sum_{i=1}^{I} G(z_{in})} \] \hspace{1cm} (6.7)

or:

\[ p_{in} = \frac{G(z_{in})}{\sum_{i=1}^{I} G(z_{in})} \] \hspace{1cm} (6.8)

This latter expression is mathematically equivalent to (6.1) and (4.1). Consequently, formula (6.8) is to be interpreted as a logit model, and the introduction of the assumption of proportionality leads towards the fulfilment of the independence from irrelevant alternatives axiom. So, the proportionality assumption should be equivalent to the assumption that the choice axiom holds.

7. **RANDOM UTILITY APPROACH**

Two special features of trying to model individual choice processes by means of a random utility approach are the following:

1. It takes into account that some of the alternative related and/or socio-economic attributes are often missing or unobservable for the researcher. So, these attributes have to be treated as stochastic variables.

2. It assumes that the individual decision-maker chooses an alternative under the condition of bounded rationality. That means, each individual is maximizing his utility rationally, but only in the framework of his personal criteria and his own well defined structure of preferences which is reflected in his personal utility function. Some attributes, which are theoretically assumed to be very important, may not be included into a personal utility function, as they are not preferred by the specific individual or are not perceived as being present. Also the values - personal weighing factors - attached to the attributes can be very different. Consequently, these individual observation elements also lead to a stochastic part in the individual utility function.

A result of these observation problems is that different choices may be made by people having exactly the same set of alternatives, described by exactly the same set of attributes. Consequently, the choice has to be looked upon as a random decision.
Theoretically, the two abovementioned observation problems can then be tackled by assuming that the individual utility or attractiveness, on which the choice for a specific alternative \( i \) is based, is a random variable, based on a set of arguments of a fixed vector of explanatory variables. This variable is denoted by \( U_{in} \), written with a capital because of its stochastic character. The connected utility function, \( U(\mathbf{z}_{in}) \), will be a random function of these explanatory, alternative related and socio-economic, attributes. In formula, \( U(\mathbf{z}_{in}) \) can be represented by:

\[
U(\mathbf{z}_{in}) = v(\mathbf{z}_{in}) + \eta(\mathbf{z}_{in}) + \xi(\mathbf{z}_{in}) ,
\]

in which:

\( v(\mathbf{z}_{in}) = \) a deterministic (non-exhaustive) function of the \( J' \) elements of the attribute vector \( \mathbf{z}_{in} \);

in which: \( J' \leq J \), the latter representing the maximum number of attributes present. \( v(\mathbf{z}_{in}) \) defines the so-called strict (or mean) utility of \( i \) for individual \( n \) and can be seen as the mathematical expectation of \( U(\mathbf{z}_{in}) \). It is the non-exhaustive version of \( u(\mathbf{z}_{in}) \) in (6.2). Usually \( v(\mathbf{z}_{in}) \) is, equally to \( u(\mathbf{z}_{in}) \), assumed to be a linear combination of the perceived or measured values of the fixed set of observed attributes. For all available alternatives, these values may take on any real number. They are assumed to be normally distributed and not to be related to each other in any way. As a result \( v(\mathbf{z}_{in}) \) can also be assumed to have a normal distribution. It can be written as:

\[
v(\mathbf{z}_{in}) = \sum_{j=1}^{J'} \delta_j z_{jin} \quad \text{(7.2)}
\]

with: \( \delta_j \) = a coefficient, a weight given to the \( z_{jin} \) by the choosing individual \( n \) \( J' \leq J \) (from the deterministic and constant utility models).

\( \eta(\mathbf{z}_{in}) = \) a stochastic function of the following elements:

a. individual 'taste variations' over some observed attributes;

b. individual measurement errors, or perception or preference disturbances of the weights, \( \delta_j \), given to one or more attributes, \( z_{jin} \), in (7.2);

c. possible inconsistencies in the individual's choice behaviour, based on \( v(\mathbf{z}_{in}) \);

d. the influences of the (restricted) assumption of linearity of the strict utility function.
\( \zeta(z_{in}) \) = a stochastic function of the influence on \( U(z_{in}) \) of missing, omitted or purposely unobserved attributes, (i.e., the impact of the \( J-J' \) attributes which might play an important role in the decision process).

As these last two factors are mostly difficult to separate in empirical research, they are usually taken together into one stochastic disturbance function \( \xi(z_{in}) \), representing the total individual deviation from the strict utility. Hence, (7.1) should be rewritten into:

\[
U(z_{in}) = v(z_{in}) + \xi(z_{in}). \tag{7.3}
\]

The at the start of section 3 mentioned maximizing process can in the random utility approach be described as the evaluation process by individual \( n \) of the utility \( U_i \) of all the available alternatives \( i \). In this framework the following general probability statement may be made for any chosen alternative \( i \) compared to all other alternatives:

\[
P_{in} = \Pr \left\{ U_{in} \geq \max[U_{i-1,n}, U_{i+1,n}, \ldots, U_{in}] \right\}. \tag{7.4}
\]

This means in fact:

\[
P_{in} = \Pr \left\{ [U_{in} > U_{in}] \land \ldots \land [U_{in} > U_{i-1,n}] \land [U_{in} > U_{i+1,n}] \land \ldots \right. \\
\left. \ldots \land [U_{in} > U_{in}] \right\}, \tag{7.5}
\]

or:

\[
P_{in} = \Pr \{ U_{in} > U_{i'n} \}, \forall i'. \tag{7.6}
\]

In words: the probability that a random alternative \( i \) will be chosen by individual \( n \) equals the probability that the utility (or attractiveness) of \( i \) exceeds or equals the utility of any other alternative \( i' \) for \( n \). So it is assumed that the choice maker behaves like a 'homo economicus'. He maximizes his utility, restricted by his social and economical possibilities.
By rewriting this last formula with the aid of (7.3), one obtains the fundamental equation of the random utility approach:

\[ P_{in} = \Pr \left\{ [v(z_{in}) + \xi(z_{in})] \geq [v(z_{i'n}) + \xi(z_{i'n})] \mid i,i'=1,\ldots,I; \right\} , \quad (7.7) \]

with the property again, of course, that:

\[ \sum_{i=1}^{I} P_{in} = 1 \]

i.e., the sum of all choice probabilities should exactly be equal to 1.

The actual calculation of the choice probabilities depends heavily on the form that will be chosen for the fundamental equation (7.7). In this respect, the distribution of the error term function, \( \xi(z_{in}) \), plays a crucial role. Several assumptions about this distribution can be made, each leading to different models. In order to present these models in a systematic way, we will follow a classification moving gradually from less to more general models (see also Daganzo (1979)). The following models will be discussed:

1. models with independent identically distributed error terms;
2. closed-form models without independent identically distributed error terms;
3. the multinomial probit model.

These models and their subdivisions will successively be discussed in sections 8-10.

8. MODELS WITH INDEPENDENT IDENTICALLY DISTRIBUTED ERROR TERMS

(A). Rational Model

In the rational model (Manheim, 1979), the error terms in formula (7.7) are assumed to be equal to zero. This is useful when the variability of the \( \xi \)'s is assumed to be small across the given alternatives. A problem is that such a kind of model without error terms is rather unstable: small specification errors may lead to large prediction errors; a change in the attributes of one alternative may lead to a complete shift of the choices.
(B). Multinomial Logit Model

If it is assumed, that the error terms in formula (7.7) are modelled by a set of variates which:
- are mutually independent identically extreme-value distributed;\(^1\)
- have zero mean and are uncorrelated with the perceived or directly measured attribute-values and related parameters of the alternatives (i.e., with the parts \(v(z_i)\) in (7.7));
- are consistent with respect to maximization (i.e., in case two disturbances have the same distribution (not necessarily with the same parameters), then also their combined maximum, \(\max(\xi(z_i), \xi(z_{i'}))\), must have that distribution),

then the so-called multinomial logit (M.N.L.) model results.

These requirements are met by a Gumbel (or Gnedenko or Weibull) distribution, which is a skewed distribution that can be almost normalized by taking logarithms. The Gumbel distribution has the following form:

\[
\Pr\left\{\xi(z_{in}) \leq \xi^*\right\} = e^{-e^{-\xi^* + \gamma}};
\]

in which: \(\xi^* = v(z_{i'n}) - v(z_{in}) + \xi(z_{i'n})\),

and: \(\gamma = \text{Euler's constant}, \; \gamma \approx 0.577\).

When formula (8.1) is assumed to hold it is not difficult to show (for the exact mathematical derivation see McFadden (1973)) that equation (7.7) reduces to the multinomial logit formula:

\(^1\) Hensher and Johnson (1981, page 105), write in this respect: 'Independence' indicates that the correlation between the unobserved attributes associated with each and every pair of alternatives in a choice set and across choice sets is zero.

'Identically distributed' says that taste variation exists over the observed attributes (and is allowed for in the random component), yet it is neutral between alternatives, having the same distribution (i.e., equal variance) around the mean (or representative) utility level.
This equation (8.2) is without doubt up till now the most widely used disaggregate demand model to analyse spatial interactions. The reasons for this seem to be that the model was relatively easy to calibrate and that its properties were generally well understood.

One of the most important properties of the M.N.L. model is that, with the assumptions concerning the probability distribution of the error terms, again the same elements are introduced into the model, as defined by the 'independence of irrelevant alternatives' hypothesis ('choice axiom') or 'IIa-property'). As Holman and Marley proved (see Luce et al. (1965)), the choice axiom is theoretically equivalent to the assumptions about the distribution of the error terms in the fundamental equation of the random utility approach. That means that also in the M.N.L. model, the relative probability of choice of two alternatives depends only on their measured attractiveness. (The measurement in this respect might take place direct or indirect by means of perception or even preference weights, which are transformed into cardinal values.) In cases where the unobserved components of alternatives (the error or rest terms) are correlated, introduction of a new alternative that is highly correlated with another one but is only marginally inferior to it, has hardly any effect on the choice probabilities of all other available alternatives. It is namely highly unlikely that the new alternative will be chosen. When such relations exist, the error terms assumptions of the M.N.L. model cause obvious problems.

Under such circumstances a solution might be to try to capture these interdependences between alternatives by defining adjusted specifications for the functions \( v(z_{in}) \). That is, however, mostly very difficult, as it might mean the introduction of a non-linear function. A few authors have also tried to solve this problem by developing \( ad hoc \) corrections for the logit model (see among others Domencinch and McFadden (1975)); others created new models with interdependent error terms and error terms with different variances. These last mentioned two kinds of models will be reviewed in section 9 after the discussion of a special case of the M.N.L. model in (C), a presentation of a group of different disaggregated sequential choice models in (D), and an overview of a practical \( ad hoc \) attempt to deal with the choice axiom within the framework of the M.N.L. model in (E) below.
(C). Binary Logit Model

A special case of the multinomial logit model - widely used in practice - is the binary logit model. In this choice analysis, it is convenient to express (8.2) as:

\[
\ln \left( \frac{P_{1n}}{P_{2n}} \right) = v(z_{1n}) - v(z_{2n}) \tag{8.3}
\]

or, to write:

\[
v(z_{1n}) - v(z_{2n}) = v(z_n) \tag{8.4}
\]

so that:

\[
\ln \left( \frac{P_{1n}}{1 - P_{1n}} \right) = v(z_n) \tag{8.5}
\]

The latter expression is traditionally the most common way of representing the binary logit model.

(D). Elimination By Aspects Method

Tversky (1972a, 1972b) has developed a disaggregated choice method which makes use of the random utility concept, but is completely different from the methods described sofar. It has mainly been applied in psychology, and does not concern error terms or even assumptions about possible error terms.

The general assumption of this Elimination By Aspects (E.B.A.) Method is that the choice maker selects an alternative in a sequential process based only on the known, identified, explanatory attributes or aspects (as they are always called in Tversky's approach). These aspects are scaled in order of importance and are to be interpreted as desirable features. The selection of a particular aspect leads to an elimination of all alternatives which do not contain this desired aspect. The process terminates with the decision based on the last relevant aspect. With that decision the final alternative choice is made.

The analytical description of this procedure yields the Elimination By Aspects model. In the original form of this model, the choice of the aspects, which will be crucial in the successive selection steps, is made at random. The choice probabilities in the model can be defined as an increasing function of the importance of the relevant aspects.
Special features of the E.B.A. model are:

- Aspects which are shared by all alternatives do not affect the final choice probability. This might be a restriction in analyzing interactions, as the relative total values of alternatives cannot exert any influence. The selection process only takes into account the presence of aspects. That means that gradual differences of aspects do not have an influence on the choice probabilities. Only completely dissimilar aspects play a role in the E.B.A. model. In analyzing actual spatial interactions this is usually a rigid restriction, as it might exclude inter alia relevant socio-economic variables from the decision process. A consequence of this feature is also that the choice between alternatives which are very similar can hardly be explained by a sequential process. Of course this last point is a weak element in most other decision explaining (or forecasting) methods. But in the E.B.A. model there are even in theory no possibilities at all to solve this problem.

- A technical disadvantage of the method is the computational burden which is usually necessary to estimate the outcome of the (assumed) sequential process, and which increases very fast when the number of alternatives and different aspects increase.

Variations on the above mentioned method in which attempts are made at better covering these problems, mainly by a better structuring of the choice process, are:

a. The Elimination By Tree Method

This method (see Tversky and Sattath (1979)), assumes a clear tree-structure in its analysis of a choice decision process. All aspects in the process are divided into the different branches of a tree, which lead to the available alternatives. There exists, however, no strict hierarchy between the successive choice steps. This means: whenever an aspect is accepted as a starting point for the evaluation procedure, then one continues the process only in the branches which have been selected by that aspect. The rest of the tree, with all its aspects, is assumed to be not relevant any longer.
b. **The Hierarchical Elimination Method**

This method (see Tversky and Sattath (1979)) is a refinement of the elimination by tree method, in as far as it assumes an hierarchy in the choice process. One always starts at the root of the tree and then selects in successive steps the relevant branches via the related aspects.

The technical model description of this method shows a recursive approach in the estimation phase. To be more precisely: in the estimation phase of the model one starts at the level of the realized alternative choices and returns step by step, until the most elementary aspect choice level is reached again. The latter means, until the choice level is reached from which the aspects are the least specific for the ultimate choice, i.e., at the root of the tree.

(E). **Sequential or Multilevel Logit Model**

A practical way of dealing with the choice axiom problem, inherent in the use of the M.N.L. model, is the introduction of sequential or multilevel logit models (see McFadden (1978)). This approach is in line with the above mentioned ideas of assuming a certain hierarchy in choice and decision processes. By splitting the choice problem into several process stages, one will, in the mathematical approach of that process be confronted with conditional choice probabilities. The systematics that is introduced in that way into the model to analyze the choice problem can solve part of the difficulties caused by the choice axiom.

This is because, by using such an sequential approach, the number of real choice situations, available alternatives (or sub-alternatives, represented by sets of important explanatory attributes) and the number of parameters in each successive stage, is declining very fast. On the other hand, there will be only a limited loss of efficiency compared to a direct estimation of the model.

A good illustration of a sequential logit model can be given by a spatial distribution problem in which a combined choice has to be made for destination x and mode y.

Assume: $P_{y|x|n}$ is the probability that a certain mode will be chosen by an individual n, when his destination is already given, as well as the marginal choice probability $P_{x|n}$ for that destination.

So it is supposed that at first x is chosen and only then y.

Furthermore, the assumption should be made that:
defines the total utility of the sequentially approached choice problem. The terms at the right hand side of this usually linear function have the following meaning:

\[ U(z_{xyn}) = A(z_{xyn}) + B(z_{xyn}) \] (8.6)

\[ A(z_{xyn}) \] = a random utility subfunction based on a vector of directly observed variables or variables which are perceived by person \( n \) as important; these variables vary in regard to both destination \( x \) and mode choice \( y \);

\[ B(z_{xyn}) \] = a random utility subfunction based on a vector of, for person \( n \), important explanatory variables, which vary only in regard to destination \( x \).

Within the framework of the M.N.L. model the result is:

\[
P_{y|x_n} = e^{U(z_{xyn})} \frac{\sum_{y'=1}^{Y_x} e^{U(z_{xyn'})}}{\sum_{y'=1}^{Y_x} e^{U(z_{xyn'})}}
\]

\[
A(z_{xyn}) \frac{\sum_{y'=1}^{Y_x} e^{A(z_{xyn'})}}{\sum_{y'=1}^{Y_x} e^{A(z_{xyn'})}}
\]

(8.7)

where:

\( y, y' = 1, \ldots, Y_x \), are the possible mode choices for a destination \( x = 1, \ldots, X \), for individual \( n = 1, \ldots, N \).

Further:

\[
P_{x|x_n} = \frac{\sum_{y=1}^{Y_x} e^{U(z_{xyn})}}{\sum_{x=1}^{X} \sum_{y=1}^{Y_x} e^{U(z_{xyn})}}
\]

\[ B(z_{xyn}) \left\{ e^{ \sum_{y=1}^{Y_x} e^{A(z_{xyn'})}} \right\} / \left\{ e^{ \sum_{x=1}^{X} e^{B(z_{xyn'})}} \right\} \]

(8.8)

in which:

\( x, x' = 1, \ldots, X \), are all possible destinations for individual \( n = 1, \ldots, N \).
When so-called inclusive values \( W_{x^n} \) are defined for formula's (8.7) and (8.8) as:

\[
W_{x^n} = \ln \left\{ \frac{\sum_{y=1}^{Y_X} A(z_{x^n})}{e^{x_n}} \right\}, \tag{8.9}
\]

then they, (8.7) and (8.8) can be written as:

\[
P_y|x_n = \frac{A(z_{x^n})}{e^{x_n}} W_{x^n}, \tag{8.10}
\]

and

\[
P_{x^n} = \frac{B(z_{x^n}) + W_{x^n}}{\sum_{x'=1}^{X} e^{x_n}} \frac{X}{X' \sum_{x'=1}^{X} e^{x_n'}}, \tag{8.11}
\]

These last 2 formula's represent the sequential or multilevel logit model.

A method for estimating the joint model, i.e.:

\[
P_{x^n} = \frac{U(z_{x^n})}{\sum_{x'=1}^{X} \sum_{y'=1}^{Y_X} e^{x_n'}}, \tag{8.12}
\]

is to first estimate the conditional choice model (8.7), then to use that result to define \( W_{x^n} \) in (8.9), and - after substituting that inclusive value into (8.11) - to estimate finally this last marginal probability model.

A problem with this practically-oriented sequential or multilevel logit model is that, in general, it may be inconsistent with utility maximization (see McFadden (1978) ), although it has a good potential to explain real choice behaviour.
9. CLOSED-FORM MODELS WITHOUT INDEPENDENT IDENTICALLY DISTRIBUTED ERROR TERMS

9.1. Models with Positive Correlation between Error Terms with the Same Variance

(A). Nested Logit Model

An empirical generalization of the sequential or multilevel logit model is the so-called nested logit model. It is defined by assuming that the inclusive values $w_{x_n}$ (from 8.9) have coefficients which are unequal to 1, viz. $(1-\sigma)$; with $\sigma \neq 0$.

The nested logit model then exists of formula (8.10) plus the adjustment of formula (8.11) into

$$P_{x_n} = \frac{e^{B(z_{x_n}) + (1-\sigma)w_{x_n}} \prod_{k=1}^{x_n} e^{B(z_{x_k}')} + (1-\sigma)w_{x_k}'}{\sum_{x_n} e^{B(z_{x_n}) + (1-\sigma)w_{x_n}}}. \quad (9.1)$$

The special feature of the nested logit model is that it permits pairwise correlation between unobserved attributes. For instance, in the example of the physical distribution problem this allows the error term of the step of the destination choice to be more or less correlated with the error term of the phase of the conditional mode choice. So it is possible to take into account special relations between destination and specific mode choices. Yet, at the same time, this 'taste variation' is in total still assumed to remain neutral within and between all choice levels of the hierarchically defined process. So all complete alternatives are still assumed to have the same utility distribution for the error terms. In other words: the essential elements of the choice axiom still hold.

The nested logit model has been used quite often in practice (see among others: Ben-Akiva (1973), McFadden (1975, 1978) and Ameniya (1976)).

(B). General Extreme Value Model

McFadden (1978) proved that both the multinominal, the binary, the sequential or multilevel, and the nested logit model are special cases of a family of general extreme value (G.E.V.) choice models. This family is derived from stochastic utility maximization, and allows a general pattern of positive correlation among the error terms.

Above all, in the conditional (nested) structure of the models of this
family no equal variance of the disturbances is assumed to exist between the several levels of the choice process. On the other hand within a level the total variances still have to be equal. So within a level the effect of taste variations over the observed attributes in the disturbance term should be neutral in the G.E.V. model.

The Central Theorem of the G.E.V. model (see McFadden (1978, page 80)) is:

Suppose that

1. $D(c_1, \ldots, c_I)$ is a non-negative homogeneous-of-degree-one function of $(C_j, \ldots, C_I) \geq 0$.

2. $\lim_{c_i \to \infty} D(c_1, \ldots, c_I) = +\infty$ for $i = 1, \ldots, I$.

3. for any distinct $(i_1, \ldots, i_k)$ from $\{1, \ldots, I\}$, should yield that:

   $\delta^k D/\delta c_{i_1} \ldots, \delta c_{i_k}$ is non-negative if $k$ is odd, and non-positive if $k$ is even.

Then,

$$P_{in} = e^{\mathbf{v}(\mathbf{z}_{in})} \frac{D_i(e^{\mathbf{v}(\mathbf{z}_{In})}, \ldots, e^{\mathbf{v}(\mathbf{z}_{In})})}{D(e^{\mathbf{v}(\mathbf{z}_{in})}, \ldots, e^{\mathbf{v}(\mathbf{z}_{In})})}$$

(9.2)

defines a probabilistic choice model of alternatives $i = 1, \ldots, I$, which is consistent with utility maximization (see also Daly and Zacharay (1976)).

This formula (9.2) is the basic model of the family of G.E.V. models.

The special case in which:

$$D(c_1, \ldots, C_I) = \sum_{i=1}^{I} c_i$$

(9.3)

forms the basis for the M.N.L. model.

From a more general $D$ function satisfying the central theorem of the G.E.V. model:
\[ D(\hat{c}) = \sum_{h=1}^{H} a_h \left\{ \sum_{i \in L_h} \frac{1}{1-\sigma_h} c_i \right\}^{1-\sigma_h} \]

where:

- \( h (= 1, \ldots, H) \) represents the classes in which the alternatives might be grouped,
- \( L_h \subseteq \{1, \ldots, I\} \),
- \( \bigcup_{h=1}^{H} L_h = \{1, \ldots, I\} \),
- \( a_h > 0 \),
- \( 0 \leq \sigma_h < 1 \),

it is possible to derive the nested logit model with a single class \( h \) of the form of formula (8.10) and (9.1) (see McFadden (1978)).

Because of the fact that in (9.4) it is a sufficient condition to satisfy the central theorem when \( \sigma \) lies between zero and one, it is also a sufficient condition for the nested logit model to be consistent with the basic concept of random utility theory, i.e. with stochastic utility maximization. This means that the coefficient \((1-\sigma)\) of the inclusive value in (9.1) has to fall into the unit interval. This coefficient of the inclusive value provides in that case an index of the correlation between the unobserved attributes in the first step of the nested logit model. When \( \sigma = 0 \) there exists independency, and when \( \sigma \) goes towards 1 the dependency grows. In other words: it can be seen as an indicator of the validity of the choice axiom for the nested logit model.

Functions of the form of (9.4) can also be used to deal with multilevel nested problems, which result into models which can better be described as tree models (see among others: Ben-Akiva and Lerman (1977), McFadden (1978), and especially Van Lierop (1981)).

Regarding the empirical possibilities of the G.E.V. method, it has to be mentioned that the qualities of the estimates of the choice probabilities appear to improve compared to describing the actual behaviour with a M.N.L. model. Up till now it is, however, uncertain how efficient this more general model is in cases with more than 3 alternatives.
The idea behind this theory, developed by Wann Yu (1979), is that choice alternatives should not be regarded only in terms of their inherent attractiveness, defined by the lagging attributes, but also in terms of their degree of similarity to one another. This means: the relative 'prominence' of an alternative becomes a new and relevant choice attribute in any situation. One might regard the figure, which can be introduced as a measure for the prominence of alternatives into the model, describing the choice process, to be a substitute for allowing positive correlation between error terms.

Yu proves empirically, for a case with no more than 3 alternatives, that by taking into account such a new 'prominence' attribute, a modified behavioural choice theory is defined within the framework of the logit model and without the strict hypothesis of the choice axioms (IIa assumption).

The model is a variant of McFadden's (1975) elimination by strategy model, which is in turn a generalization of the elimination by aspects method and which introduces into the model - instead of a prominence weight - a similarity factor of alternatives, defined in terms of the degree of overlap of their aspects.

In formula, the prominence theory of choice can be represented by:

\[
P_{in} = \frac{v(z_{in}) + \log Q_{in}}{\sum_{i'} e^{v(z_{i'n}) + \log Q_{i'n}}} ,
\]

with:

\[
Q_{in} = I \left( \sum_{i'=1}^{I} r_{ii'n} \right)^{-1} .
\]

This term \( Q_{in} \) can be regarded as the measure of the average dissimilarity between alternative \( i \) and all the other available alternatives, as:

\[
r_{ii'n} = \text{an index of similarity between alternatives } i \text{ and } i' \text{ for individual } n .
\]
This similarity measure $r_{i' i''}$ can be operationalized among others by defining it as the cosine of the angle ($\theta$) between the attribute vector $\vec{z}_{i''}$ and $\vec{z}_{i'}$, i.e.:

$$r_{i' i''} = r(\vec{z}_{i''}, \vec{z}_{i'}) = \cos \theta_{\vec{z}_{i''}, \vec{z}_{i'}} = \frac{\vec{z}_{i''} \cdot \vec{z}_{i'}}{||\vec{z}_{i''}|| ||\vec{z}_{i'}||} \quad (9.7)$$

A problem with this most common similarity measure is, however, that it is not necessarily unique. Many other similarity measures can be defined. Maybe this problem can be solved by adding some efficiency criteria.

Yu showed that it is possible to see his prominence theory to be an empirical generalization of both the multinomial logit model and the nested logit model. Especially because of the latter characteristic, this method might also be considered to belong to the family of general extreme value models.

The benefit of the prominence theory of choice-approach, compared to the other so far mentioned models, is that it provides a simple operational variant in cases where attribute dependencies among choice alternatives are relevant, in other words when the choice axiom has to be avoided. However, more study into the similarity measure is necessary in order to resolve the uniqueness problem according to a behaviouristic analysis, and to give the approach also a more firm theoretical foundation.

9.2. Model with Independent Exponential Distributed Error Terms

(D). Negative Exponential Distribution Model

In this model (see also Daganzo (1979, pp. 14-15)) the error terms are assumed to be independently, exponentially distributed with a zero mean and different standard deviation $\sigma(\vec{z}_{i''})$. So the assumption of 'identically' is dropped, compared with (8.1).

In formula:
According to Daganzo (1979), with these assumptions it is possible to obtain the probability of choice after a few algebraic manipulations. Before presenting the choice function, it should be noted that the utility (attractiveness) of any alternative \( i \) cannot exceed an upper bound \( t(z_{in}) \), defined by the strict utility and the standard deviation of that alternative for \( n \):

\[
t(z_{in}) = v(z_{in}) + \sigma(z_{in}) .
\]  

(9.9)

The subscript \( i \) indicates the alternative with the \( i \)th largest upper bound on the utility as measured and/or perceived by person \( n \), \( \hat{T}(z_{in}) \).

It has to be assumed that \( T(z_{0n}) = \infty \) and \( T(z_{I+1,n}) = -\infty \). Furthermore should hold that \( P_{I+1,n} = 0 \).

The choice function of the negative exponential distribution (N.E.D.) model can, in the most general case with \( k \) classes, then be written as:

\[
P_{in} = \sum_{k=i}^{I} \left\{ \frac{\sigma(z_{kn})^{-1}}{k} \sum_{i'=1}^{k} \frac{t(z_{i'n}) - t(z_{kn})}{\sigma(z_{i'n})} \right\} \cdot \left\{ \exp \left( - \sum_{i'=1}^{k} \frac{t(z_{i'n}) - t(z_{kn})}{\sigma(z_{i'n})} \right) \right. 
\]

\[
- \exp \left( - \sum_{i'=1}^{k} \frac{t(z_{i'n}) - t(z_{kn})}{\sigma(z_{i'n})} \right) \right\} .
\]  

(9.10)

By means of a recursive calibration procedure one can calculate from (9.10) the choice functions for each alternative.

A difficult problem may arise in a model with many alternatives, from which several have approximately the same estimated utility. When in such a case the variance of a distribution term in (9.10) is increased, then at the same time the choice probability of the alternative with the enlarged variance tends to increase.
It might be possible to generalize the N.E.D. model by a model with Weibull error terms. This is a subject of further research.

10. **MULTINOMIAL PROBIT MODEL**

A random utility model in the form of a multinomial probit (M.N.P.) model is characterized by error terms with a joint multivariate normal distribution, with a zero mean and an arbitrary variance-covariance matrix. Thus, in an M.N.P. model the variances of the error terms are allowed to be different and also the error terms themselves are permitted to be mutually correlated. This means that the M.N.P. model can be seen as a generalization of both G.E.V. and N.E.D. models by incorporating the elements of both models.

The starting formula of the M.N.P. model choice function is again the fundamental equation of random utility models as described in chapter 7. By defining a joint multivariate normal density function for the disturbances, $\xi(z^*)$, with a related distribution function $F_{\xi(z^*)}$, and characterizing:

$$E[z_n^*] = 0$$

we may formulate the following variance-covariance matrix:

$$E [\xi(z^*)] = \Sigma_n$$

where:

$$\xi(z^*_n) = [\xi(z_{1n}^*), \ldots, \xi(z_{m,n}^*), \ldots, \xi(z_{1n}^*)]'$$

we may formulate the following variance-covariance matrix:

$$E [\xi(z^*_n) \cdot \xi'(z^*_m)] = \Sigma_n$$

This is equal to:

$$\Sigma_n = \begin{bmatrix}
\sigma^2 & \sigma_{1n} & \ldots & \ldots \\
\sigma_{1n} & \sigma^2 & \sigma_{2n} & \ldots \\
\sigma_{2n} & \sigma_{2n} & \sigma^2 & \ldots \\
\ldots & \ldots & \ldots & \ddots \\
\sigma_{1n} & \ldots & \sigma_{1n} & \sigma^2
\end{bmatrix}.$$
with:

\[ \sigma_{ii'n} = E [\xi(\hat{z}_{i'n}) \cdot \xi(\hat{z}_{i'n})] \]  

and:

\[ \sigma_{in}^2 = E [\xi^2(\hat{z}_{in})] \]  

Now, it is possible to estimate the probabilities \( P_{in} \).

For the distribution of the continuous random error term of the \( i \)-th argument, \( \xi(\hat{z}_{in}) \), the following characterization is relevant (see, for instance, Mood et al. (1974) pp. 138):

\[ \Pr [\xi(\hat{z}_{in}) \leq \xi^*] = F_{\xi}(\hat{z}_{in}; \xi^*) \]  

Introduction of the equivalence of (10.7), viz.: \[ \int_{-\infty}^{\xi^*} f_{\xi}(\hat{z}_{in}; \xi) \, d\xi \] into the fundamental equation of the random utility approach (7.8) yields the well defined probability function of the M.N.P. model, i.e.:

\[ P_{in} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} v(\hat{z}_{in}) - v(\hat{z}_{i'n}) + \xi(\hat{z}_{in}) \]

\[ \cdots v(\hat{z}_{in}) - v(\hat{z}_{i'n}) + \xi(\hat{z}_{in}) \]

\[ \cdot \int_{-\infty}^{\xi^*} f_{\xi}(\hat{z}_{in}; \xi) \left( \xi(\hat{z}_{in}), \ldots, \xi(\hat{z}_{in}), \xi(\hat{z}_{in}) \right) \]

\[ \cdot \{ d\xi(\hat{z}_{in}) \cdots d\xi(\hat{z}_{i'n}) \} \]

\[ = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_i \left( v(\hat{z}_{in}) - v(\hat{z}_{i'n}) + \xi(\hat{z}_{in}), \ldots, v(\hat{z}_{in}) - v(\hat{z}_{i'n}) + \xi(\hat{z}_{in}) \right) \]

\[ \cdot \xi(\hat{z}_{in}), \ldots, v(\hat{z}_{in}) - v(\hat{z}_{i'n}) + \xi(\hat{z}_{in}) \]

\[ + \xi(\hat{z}_{in}) \right) \, d\xi(\hat{z}_{in}) \]  

(10.8)
Because of the non-convenient closed form representation of the multiple integrals in (10.8), computation of the choice probabilities of the M.N.P. model was until recently hardly possible. After a first success by Hausman and Wise (1978), in solving the computational problems of the M.N.P. model, Daganzo (1979) and his associates rediscovered, a single numerical approximation method by Clark (1961), that is surprisingly accurate in estimating M.N.P. models.

This method has the potential to calculate very quickly the choice probabilities for a reasonably large number of alternatives within the framework of a M.N.P. model.

The above mentioned means that the dilemma of the choice axiom (or independence of irrelevant alternatives or IIa hypothesis) can be avoided when one tries to estimate choice probabilities of individual spatial interaction decisions. The possibility to allow a full variance-covariance structure for the random utilities of alternatives in the empirical model-approach of choice processes, which structure permits things like:

- general taste variations between individuals,
- dependence between different alternatives, and
- methods to treat errors in the data in a straightforward manner,

leads to the conclusion that the M.N.P. model is a significant realistic disaggregate spatial interaction model. It has to be remarked, however, that some care should always still be taken with this model in case of many comparable alternatives.

11. EVALUATION

In table 2 an integrated survey of the described disaggregate spatial choice models and their most important assumptions is given.

It is clearly seen from sections 3 to 10 and from this table 2 that the disaggregate spatial choice models can be split up into several classes, based on the way in which the relevant models deal with such topics as: dependence between different alternatives, taste variations among individuals and data-errors. The main, although not exclusive, selection criteria in this respect are:
### Table 2: Integrated survey of disaggregate spatial choice models and their most important assumptions.

<table>
<thead>
<tr>
<th>Features</th>
<th>Independence of irrelevant alternatives</th>
<th>Choice axiom</th>
<th>Error terms independently and identically distributed</th>
<th>Multinomial Independence problem</th>
<th>Error terms positive correlated</th>
<th>Error terms independent exponential distributed</th>
<th>Error terms independent assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deterministic Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Logit models</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>- Gravity and Entropy models</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Probabilistic Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Constant Utility Theory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Logit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Random Utility Theory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Models with Independent Identically Distributed Error Terms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Rational model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Multinomial logit model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Binary logit model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Elimination by aspects methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Elimination by tree model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. 'Hierarchical' Elimination model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Sequential or multilevel logit model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Closed Form models without Independent Identically Distributed Error Terms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 Models with positive correlation between error terms with the same variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Nested logit model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- General Extreme Value model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Prominence Theory of Choice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 Model with independent exponential distributed error terms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Negative exponential distribution model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Multinomial Probit Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**LEGENDA:**
- = explicit relevant
- = implicit relevant
0 = not relevant
Φ = implicit relevant, but more possibilities
I. Independence of Irrelevant Alternatives Assumptions

First one can distinguish the group of models based on the assumption of independence of irrelevant alternatives. The most important representative of this group is the deterministic logit model. This is the least general disaggregate spatial choice model. Implicitly also gravity and entropy models can be regarded as members of this group, at least when these kinds of models can - by means of information theory - be translated into a micro form of the logit model. Many probabilistic models, the ones which use the choice axiom hypothesis or have independently and identically Gumbel distributed error terms, are influenced by the same features as defined by the independence of irrelevant alternatives assumption.

II. Choice Axiom or IIa Hypothesis

Among the class of models with this assumption one finds the constant utility logit model, and implicitly almost all random utility models, except for some more general ones.

The same model characteristics which are introduced into models by the independence of irrelevant alternatives assumption are found again under the choice axiom or IIa hypothesis.

III. Assumption of Independently and Identically Gumbel Distributed Error Terms

The multinomial, binary, sequential or multilevel and nested logit models fall into this class. This group of models has widely been applied in practice. The assumption of independently and identically Gumbel distributed error terms leads implicitly again to the introduction of the features of the choice axiom or IIa hypothesis.

IV. Ad Hoc Solutions Independence Problem

It were mainly the shortcomings of the existing models concerning the treatment of such topics as correlation, number of alternatives and individual taste variations, found in the empirical applications, which led to modifications of the well-known disaggregate demand models and to the development of completely new approaches for individual choice analysis. Examples are to be found in the sequential or multilevel logit model, the elimination by aspects method, elimination by tree model and the hierarchical elimination model.
V. Absence of Independent Identically Distributed Error Terms in Closed Form Models

Other approaches try to tackle the problems of independence and taste variations among individuals in a more fundamental way. The general extreme value model may be regarded as the most well-known theory in this respect. It admits positive correlation among the error terms. The nested logit model and the prominence theory of choice, which can be derived as special cases of the general extreme value model, fall into this class. An alternative approach is offered by the negative exponential distribution model, in which framework independent exponential error terms are defined.

VI. Presence of Full Variance-Covariance Matrix for the Rest Terms

The most general theory up till now for real empirical analysis of spatial interaction choices in a disaggregate way is presented by the multinomial probit model. This model, which can intuitively be viewed as the generalization of most of the foregoing methods, allows - as described - the introduction of a full variance-covariance matrix for the rest terms into the individual utility functions without any restrictive assumptions. In this way, the M.N.P. model seems not only to be able to meet the required theoretical standards, but also to tackle practice in a very accurate way.

Whether the very high theoretical standards of the M.N.P. model can indeed practically be met and whether they are completely necessary, will depend on the problem at hand. For certain problems the estimation of variances and covariances will be of primary interest to the analyst. Then the model cannot be formulated usefully in an other way than as a M.N.P. model. In situations in which the set of alternatives is partly unknown or not exhaustively sampled it is, however, even not possible to formulate a M.N.P. model. A model based on the independence of irrelevant alternatives assumption or the choice axiom, like for instance a M.N.L. model, might in such cases still give useful solutions. Sequential and nested structures of spatial interaction choices can up till now in the easiest way be approached with sequential or multilevel and nested logit models.

In case of alternative sampling procedures, like choice-based sampling, the M.N.L. model offers many opportunities. When there are many alternatives available one might have to switch to an elimination by aspects model, etc.
Van Lierop and Nijkamp (1980) give a list of methodological, theoretical, logical and practical criteria, which can be helpful when choosing a specific model for a given research project.

Concluding we would like to remark that the M.N.P. model will not necessarily be the only exclusive model of choice for most disaggregate spatial demand studies, although it offers significant theoretical advantages over other disaggregate models of choice in a spatial context.

12. A NUMERICAL ILLUSTRATION 1)

As an illustration of the theoretical potential of the M.N.P. model, we simulated a spatial choice problem and tested 2 random utility models on it, viz. a M.N.L. and a M.N.P. model. The aim was to show that when one allows correlation between alternatives belonging to a finite set of spatial choice possibilities, probit gives in general better results than logit. The example is taken from a fictitious home-workplace commuting situation in the Netherlands. The answers to questions of a home inquiry about individual travel behaviour have been simulated for a representative group of 250 persons of the relevant population. They had 3 alternatives to go to work: car, bus or train.

The following model specifications (conform section 7) were used to calculate the choice probabilities:

\[ U_{In} = \gamma_{0In} + \gamma_{1In}z_{1In} + \gamma_{2In}z_{2In} + \gamma_{3In}z_{3In} + \xi(z_{1In}) \]  \quad (12.1)

in which:

- \( U_{In} \) is the individual utility or attractiveness a random person \( n \) attaches to (or expects to get from) alternative 1 of the set of alternatives he has at his availability, here: taking the car to travel from his home to his workplace and back.

- \( z_{1In} \) is car availability; (the precise meaning of the subscripts is: aspect 1 of alternative 1 for person \( n \)).

- \( z_{2In} \) is valuation of travel time

- \( z_{3In} \) is costs

- \( \gamma_{0In} \) is constant term

1) The authors thank Annemarie Rima for her computational assistance.
\[ U_{2n} = \gamma_0 2n + \gamma_1 2n z_1 2n + \gamma_2 2n z_2 2n + \gamma_3 2n z_3 2n + \xi(z_{2n}) \]  
\[ (12.2) \]

in which:

- \( U_{2n} \) = the utility of individual \( n \) in taking the bus to commute between home and work.
- \( z_{1,2n} \) = valuation travel time
- \( z_{2,2n} \) = valuation waiting time
- \( z_{3,2n} \) = costs
- \( \gamma_{0,2n} \) = constant term
- \( \gamma_{1,2n} \) = parameters
- \( \gamma_{2,2n} \) = parameters
- \( \gamma_{3,2n} \) = parameters
- \( \xi(z_{2n}) \) = disturbance factor.

\[ U_{3n} = \gamma_{1,3n} z_{1,3n} + \gamma_{2,3n} z_{2,3n} + \gamma_{3,3n} z_{3,3n} + \gamma_{4,3n} z_{4,3n} + \xi(z_{3n}) \]  
\[ (12.3) \]

in which:

- \( U_{3n} \) = the utility of individual \( n \) in taking the train to go to work
- \( z_{1,3n} \) = valuation travel time
- \( z_{2,3n} \) = valuation waiting time
- \( z_{3,3n} \) = costs
- \( z_{4,3n} \) = valuation comfort
\[ \left\{ \begin{array}{l} \gamma_{13n} \\ \gamma_{23n} \\ \gamma_{33n} \\ \gamma_{43n} \end{array} \right\} = \text{parameters} \\
\xi(z_{3n}) = \text{disturbance term}. \]

An important usual restriction that should be taken into account also with the model in this example is that the sum of the choice probabilities of all the alternatives is exactly equal to 1 and that each of them separately varies between 0 and 1.

It was assumed that the choice utilities for car and bus influence each other and that the choice of the train is independent of the choice of any other alternative. This gave the following variance-covariance matrix:

\[ \Sigma_n = \theta \begin{bmatrix} T_{1n} & \rho \sqrt{T_{1n}T_{2n}} & 0 \\ \rho \sqrt{T_{1n}T_{2n}} & T_{2n} & 0 \\ 0 & 0 & T_{33n} \end{bmatrix} \]

(12.4)

In which:

\[ T_{1n} = \text{generalized costs of alternative } i \ (i = 1, 2, 3) \text{ for individual } n \ (n = 1, \ldots, n, \ldots, 250). \]

\[ \theta, \rho = \text{parameters which are the same for all alternatives with ranges } -1 \leq \rho \leq 1 \text{ and } \theta > 0, \text{ which guarantees } \Sigma_n \text{ to be positive semi-definite.} \]

In the logit model this variance-covariance matrix is of course irrelevant. A direct confrontation of multinomial probit and multinomial logit in order to see which one fits better - by means of, for instance, a generalized likelihood ratio test - is not possible, because there is no direct connection between these models: they are not nested, i.e. the one is not a specific case of the other. For a generalized likelihood test it is however a condition sine qua non that the relevant models are nested 1).

1). See Mood et al. (1974), page 419.
To solve this problem a third model should be added as an intermediate one: an independent probit model. This is a special case of the multinomial or covariance probit model and has the following variance-covariance matrix:

\[ \Sigma_n = R \times I_n, \quad (12.5) \]

in which:

\[ R = \text{the identity matrix with dimension } I_n \times I_n \quad (i.e. \text{defined by the number of alternatives available to individual } n). \]

As a result it is possible to test independent probit and covariance probit against each other.

Hausman and Wise (1978, blz. 415) showed that, after normalization of the variances, the independent normal distribution and the extreme value distribution are almost the same. This means that the independent probit model and the logit model are based on approximately the same theoretical aspects (both assume e.g. independence) and also that in practice they will provide about the same results. As a consequence, testing of multinomial probit against independence probit is almost equal to testing multinomial probit against multinomial logit.

The likelihood ratio test itself is based on:

- \( H_0 = \text{independent probit specification with loglikelihood } L_0 \),
- \( H_1 = \text{multinomial probit specification with loglikelihood } L_1 \),

and can be defined as:

reject \( H_0 \) if

\[ \lambda = -2 (L_0 - L_1) > \chi^2_{1-\alpha} (q) \quad (12.6) \]

with: \( q = \text{number of degrees of freedom} \)
\( \alpha = \text{probability of an error of type I} \).
The relevant loglikelihood values are calculated with the choice modelling computer program CHOMP, as developed by Daganzo and Schoenfeld (1978). The comparison of logit and independent probit is not necessary in this model, because the measurable part of the utility function is identical for both these specifications, so that also the loglikelihood values will be identical. We found: $L_0 = -305,15370$ and $L_1 = -277,20543$. This means that the $\lambda$ from (13.6) becomes: 55,89654. As a result the $H_0$-hypothesis can be rejected for $\alpha = .005$ and $q < 31^{1)}$. In our example, the number of degrees of freedom is 4 and thus independent probit and also logit will be rejected in favour of covariance or multinomial probit.

From the generated individual choices, one can derive the theoretical estimations of the individual probabilities to choose alternative 1, 2 or 3. Putting the mean utility and the variance-covariance into the computer program CONFID, as developed by Sparmann and Daganzo (1979), provides next predictions for the multinomial probit case of the probabilities of choosing one specific alternative from the three available options. Also the predictions in case of a logit specification can be calculated by means of that computer program. The prediction results of both are given in Table 3.

<table>
<thead>
<tr>
<th>Probabilities</th>
<th>M.N. Probit</th>
<th>Independent Probit</th>
<th>M.N. Logit</th>
<th>Theoretical</th>
</tr>
</thead>
<tbody>
<tr>
<td>alt. 1</td>
<td>.32311</td>
<td>.28592</td>
<td>.38367</td>
<td>.328</td>
</tr>
<tr>
<td>alt. 2</td>
<td>.35132</td>
<td>.39079</td>
<td>.35982</td>
<td>.356</td>
</tr>
<tr>
<td>alt. 3</td>
<td>.32059</td>
<td>.31858</td>
<td>.25651</td>
<td>.316</td>
</tr>
</tbody>
</table>

Table 3. Individual choice probability prediction results of the CONFID computer program.

In this short example the differences between M.N.L. and M.N.P. are not yet very significant. It seems however, justifiable to conclude that on the whole multinomial probit gives better results. It is obvious that especially the ratio between alternatives 1 and 2 is very well predicted by multinomial probit, due to the existing correlation between these alternatives.

How this kind of probit analysis will work in more complex practical cases, is a new field of research. Results of work recently done in this field are very promising. As a final remark of this illustrative chapter and as a provisional conclusion of this paper we would therefore like to state that in case of correlation among alternatives in a spatial choice analysis problem, multinomial probit models are worthwhile.

1) See Daganzo (1979), page 208.
REFERENCES


Lierop, W. van, Micro-Economisch Consumentengedrag en Personenvervoer, Det. of Regional Economics, Free University, Amsterdam, 1978, (mimeographed).


Manski, C.F., Recent Advance in and New Directions for Behavioural Travel Modelling, Report of the Workshop on Research Issues in Mathematical Techniques for Behavioural Travel Modelling, Fourth International Conference on Behavioural Travel Modelling, held at Elbsee, Germany, 1979 (mimeographed).


Rima, A., Analyse van keuzeprocessen met behulp van Logit en Probit, working-paper, Free University, Amsterdam, 1980.


Stopher, P.R., and Meyburg, A.H. (eds.), *Workshop-papers for the Fourth International Conference on Behavioural Travel Modelling, held at Elbsee, Germany*, 1979 (mimeographed).


Williams, H.C.W.L. and Ortuzar, J.D., Behavioural Theories of Dispersion and the Mis-specification of Travel Demand Models, School of Geography and Institute of Transportation Studies, University of Leeds, 1980.


Serie Research Memoranda:

1977-1 L. Hordijk, P. Nijkamp, Estimation of Spatiotemporal Models
- New directions via distributed lags
and Markov schemes.

1977-2 P. Nijkamp, Gravity and entropy models: The state of the art.

1978-1 J. Klaassen, Valuta problemen in de Jaarrekening.

1978-2 P. Nijkamp, M.W. van Psychometric scaling and preference
Veenendaal, methods in spatial analysis.

1978-3 P. Nijkamp, J. Spronk, Interactive Multiple Goal Programming.

1978-4 P. Nijkamp, P. Rietveld, New Multi Objective Techniques in
Physical Planning.

1978-5 P. Nijkamp, Decision Models for Planning against Stagnation.


1978-7 P. Nijkamp, An Analysis of Interdependent Decision via
Non-Linear Multi-Objective Optimization.

1978-8 J.D.P. Kasper, Conflict Patterns and Compromise Solutions
in Fuzzy Choice Theory.

1978-9 H.J. Blommestein, P. De Bijdrage van Detailhandelsmarketing aan
Nijkamp, P. Rietveld, Maatschappelijke Welvaart en Welzijn.

1978-10 P. Palm and A. Zellner, A Multivariate Analysis of Spatial
Large Sample Estimation and Testing Proce­
dures for Dynamic Equation Systems.

P.A. Cornelis, H. van der Groepseffectiviteit.
Flier,

1978-12 W. van Lierop and Een onderzoek bij deelnemers aan een
P. Nijkamp, bedrijfsspel.

1978-13 R.J. Joseph, A utility framework for interaction models
Beschermingsconstructies in het huidige en for spatial processes.
komende vennootschapsrecht.

1978-14 P. Nijkamp en A multi-attribute analysis of spatial
W. van Veenendaal, choice Behaviour.

Frans Blommaert and Hein Schreuder,

1979-2 H. Schreuder, De maatschappelijke verantwoordelijkheid van
Ondernemingen.

1979-3 P. Nijkamp and P. Rietveld, Multilevel Multi-objective Models in a
Regional System.

1979-4 J. Arntzen, G. Bornmalm-
Jardelsw and P. Nijkamp, Duality, Segmentation and Dynamics on a
Regional Labour Market.

1979-5 P. Nijkamp, A. Soffer Soft Multicriteria Decision Models for
Urban Renewal Plans.

1979-6 drs. A.J. Mathot A Model of choosing a car with or without
a credit.

1979-7 H. Blommestein, P. Nijkamp Shopping Perceptions and preferences: A multi-
en W. van Veenendaal, dimensional Attractiveness Analysis of
Consumer and Entrepreneurial Attitudes.

1979-8 H.J. Blommestein and The Aggregate Demand for Money in the Netherlands-
F.C. Palm a new look at a study of the Bank of the
Netherlands.