Transpn Res.-B Vol. 16B, No. 3, pp. 167–219, 1982
Printed in Great Britain.

0191-2615/82/030167–53$03.00/0
Pergamon Press Ltd.

BEHAVIOURAL THEORIES OF DISPERSION AND THE MIS-SPECIFICATION OF TRAVEL DEMAND MODELS†

H. C. W. L. WILLIAMS
School of Geography, University of Leeds, Leeds LS2 9JT, England.

and

J. D. ORTUZAR
Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, England

(*Received* 15 *June* 1980)

Abstract—Conventional or first generation transport models have for some time been heavily criticised for their lack of behavioural content and inefficient use of data; more recently second generation or disaggregate travel demand models based on a theory of choice between discrete alternatives have also been viewed critically. First, it has been argued that implemented structures—and particularly the Multinomial Logit model—have not been sufficiently general to accommodate the "interaction" between alternatives; and second, and perhaps more importantly, that the underpinning theory, involving a perfectly discriminating rational man (*homo economicus*), endowed with complete information is an unacceptable starting point for the analysis of behaviour. In this paper the potential errors in forecasting travel response arising from theoretical misrepresentation are investigated; more generally, the problems of inference and hypothesis testing in conjunction with cross-sectional models are noted.

A framework is developed to examine the *consequences* of the divergence between the behaviour of individuals in a system, the observed, and that description of their behaviour (which is embedded in a forecasting model) imputed by an observer, the modeller. The extent of this divergence in the context of response to particular policy stimuli is examined using Monte Carlo simulation for the following examples: (i) alternative assumptions relating to the structure of models reflecting substitution between similar alternatives; (ii) alternative decision-making processes; (iii) limited information and "satisficing" behaviour; and (iv) existence of habit in choice modelling.

The method has allowed particular conclusions to be made about the importance of theoretical misrepresentation in the four examples. More generally, it highlights the problems of forecasting response with cross-sectional models and draws attention to the problem of validation which is all too often associated solely with the goodness of statistical fit of analytic functions to data patterns.

1. INTRODUCTION

In the absence of specific information on travel and locational response to transport system changes it has become almost universal practise to infer the propensity of individuals to modify their behaviour from trip patterns revealed at a single cross section. Indeed, the traditional use of the cross-sectional approach has transcended the significant differences between two generations of travel demand models. Explanations and theories of traveller behaviour become associated with the interpretation of dispersion—the variability exhibited when individuals are associated with different travel related options.

It is a fundamental assumption of the cross-sectional approach that a measure of the response to (incremental) change may be assessed from demand functions simply by determining their derivatives with respect to the policy variables in question. Thus, if we write the demand model as a functional relation

$$P_\rho = f_\rho(\mathbf{Z}_1, \ldots \mathbf{Z}_\rho, \ldots \mathbf{Z}_N; \boldsymbol{\theta}) \quad \rho = 1, \ldots N \tag{1}$$

between the probability P_ρ of occupying a particular state A_ρ, $\rho = 1, \ldots N$; the vectors of attributes $\mathbf{Z}$, and the parameter set $\boldsymbol{\theta}$, then the response δP_ρ to a policy stimulus, identified by the combination of incremental changes in the components $\{\delta Z^\mu_{\rho'}\}$ may be written

$$\delta P_\rho = \sum_{\rho'} \sum_\mu \frac{\delta f_\rho}{\delta Z^\mu_{\rho'}} \cdot \delta Z^\mu_{\rho'} \quad \rho = 1, \ldots N. \tag{2}$$

†An initial version of this paper was presented at the P.T.R.C. proceedings, Annual Summer Meeting, University of Warwick, 9–11 July, 1979. It is an extended version of the paper "Behavioural Travel Theories, Model Specification and the Response Error Problem".

P_ρ (and δP_ρ) might, for example, refer to the probability (and its change) that an individual will select a particular travel related option. We shall write this equation in the matrix notation

$$\delta \boldsymbol{P} = \boldsymbol{f}'(\boldsymbol{Z}, \boldsymbol{\theta}) \cdot \delta \boldsymbol{Z} \tag{3}$$

in which f' contains all partial derivatives of the demand function.

The assumption that a realistic *stimulus-response* relation may thus be derived from the model (1), with its parameters estimated from cross sectional observations, has been elevated to such a level of faith that the notion of validity of a travel *response* forecasting model is often interpreted exclusively in terms of the statistical goodness-of-fit achieved between the base predictions and observed travel behaviour. If this view were not so prevalent in transportation studies it would be unnecessary to reassert that "goodness of statistical fit" might be a necessary but is not a sufficient condition for a valid response model. This has been demonstrated both in numerical tests on demand models, and in the increasing number of studies in which it has been possible to compare the results of demand predictions with observed behaviour in before-and-after studies (see, e.g. Senior and Williams, 1977; Williams and Senior, 1977; the Kent County Council, 1976; Papaoulias and Heggie, 1976; and Train, 1978).

The popularity of the so-called second generation demand models conceived in terms of probabilistic consumer choice concepts has, in part, been founded on the successful marriage of an explicit theory of behaviour with a micro-representation permitting an efficient statistical analysis of dispersion at the level of the individual traveller. One of the powerful motivations behind their development was the feeling that appropriately specified models which captured the essence of human behaviour would be transferable in both space and time. While the merits of the "conventional" micro-approach have been widely recognised, there is a broad and growing body of literature inspired by what are considered to be the deficiencies of the commonly adopted model representations, and in particular the theoretical basis underpinning them (see, e.g. Stopher and Meyburg, 1976; Burnett and Thrift, 1979; Heggie, 1978; Heggie and Jones, 1978; Banister, 1978; Burnett and Hanson, 1979). While these criticisms may be discussed under the generic heading of behavioural misrepresentation, there are two particularly important classes of comment which are worthy of note, relating to: the definition and characterization of travel related options which are deemed relevant to the choice and response contexts; and secondly, the nature and characterization of the decision process attributed to an individual. In the latter context it is argued that the notion of a perfectly discriminating "rational man" endowed with complete information is an unrealistic starting point for the analysis and prediction of travel behaviour.

Criticism against a prevailing orthodoxy is of course essential to prevent stagnation as the state-of-the-art evolves, but critical assessments can be, and often are, made in the absolute with little regard to their practical implications, and in particular, the way in which theoretical statements assert themselves in empirical demand analyses. Some criticisms are testable through numerical experimentation, and it is by this means that we shall seek to assess some of the consequences of behavioural misrepresentation as manifested through the mis-specification of micro-models.

Out of a very broad class of possible mis-specification problems we have singled out four particular areas for illustration and numerical experimentation, relating to: the selection of a demand model inappropriately structured in relation to its accommodation of the perceived similarity between travel related substitutes; alternative models of the decision process by which individuals are considered to identify a preferred alternative; the relaxation of the assumption of perfect information and discrimination in the choice process; and the existence of habit in choice contexts. As we are not in the possession of information on actual *response* behaviour we have simulated data pertaining to choice contexts on the basis of particular specifications of the behavioural process. A model(s) has then been estimated with this data and its predictions compared with the simulated response.

Although we have distinguished the various simulation experiments for the purposes of presentation, a number of themes relating to the issue of theory representation will weave their way through the discussion. It is a central concern to determine whether two or more competing models of behaviour can each result in acceptable "statistical fits" to cross sectional data, and differ significantly in their response predictions. By this means we attempt to determine the extent to which an observer (the modeller) can be indifferent to the many sources of variability giving rise to dispersion.

First we must address the general problem of specification and the role of theory in model formation. The particular aspects of decision contexts and frames of reference within which the experimental tests on behavioural misrepresentation are conducted will then be examined.

2. THEORETICAL PERSPECTIVES: MIS-SPECIFICATION AND RESPONSE ERRORS

The *specification* of a cross-sectional micro-model (1) occupies a central place in the development of analytic forecasting procedures. While it is recognised that in any particular study data limitations and resource restrictions may well have a vital—and possibly overwhelming influence on this process, it is important to establish how theoretical considerations influence the construction of a demand function such as (1). In this paper we shall in the main be concerned with the relationship between the form of the model and certain behavioural premises assumed to underpin it.

Because a considerable amount of criticism of the current generation of micro-models is directed at precisely this relationship it is necessary to make a brief excursion into the nature of the theoretical base, in so far as this is relevant to the issues raised later in the paper. A natural starting point is the theory of rational choice behaviour and its formal adaptation. This asserts that

> "...a decision maker can rank possible alternatives in order of preference, and will always choose from available alternatives the option which he considers most desirable, given his tastes and the relative constraints placed on his decision making, such as his level of income or time availability. Suitably modified to take account of the psychological phenomena of learning and perception errors, this theory...forms the foundation of modern economic analysis". (Domencich and McFadden, 1975.)

Two questions may immediately be posed here. Is the theory reasonable, and if it is, is it of any use? Many critics of the theory of rational choice, *adopting a particular interpretation of what it implies*, tend to oppose its strong normative flavour in utility analyses and point especially to the assumption of perfect information and discrimination in the decision process as being particular offenders against reality.

The concept of *satisficing* and "bounded rationality" proposed by Simon (1955) is perhaps the most widely discussed alternative to the above approach. As Eilon (1972) has remarked

> "...optimisation is the science of the ultimate: satisficing is the art of the feasible. The optimiser sets off in a single minded fashion to determine the best solution to a given problem in given circumstances...the satisficer on the other hand, acquiesces to the proposition that it is seldom possible to define the ultimate in unambiguous terms and that it is sufficient to do well enough".

The notion of satisficing is often cited as having some support from the study of human information processing and psychological theories in which decision making is conceptualized in terms of attitude formation, triggers and responses within a dynamic "environment" of the decision maker. While the economic and pyschological perspectives tend to examine consumer behaviour from different vantage points, they are in one sense complementary; the former tends to take preferences as given and does not attempt to explain them, the latter goes one stage further back to provide such explanations.

When the cost and effort of acquiring information and indeed making decisions are duly accounted for, many of the distinctions between the so-called "optimising" and "satisficing" approaches become moot points. Indeed, as McFadden (1975) has further remarked

> "...classical economic analysis makes the assumption of homo economicus virtually tautological, if an object is chosen then it must maximise "utility" *as the chooser perceives it*" (our emphasis).

In this sense we may perhaps regard *any* aspects of *observed* traveller behaviour which are deemed "irrational" or "idiosyncratic" as but apparent and simply the consequences of inaccurate or inappropriate descriptions of the behaviour in the *frame of reference* selected by the *observer.* This aspect of analysis will be further considered below.

We must first return to the question raised above and ask whether the (extended) framework of rational choice theory interpreted through utility maximisation is of any use. A central consideration is, whether the theory places any meaningful restrictions on the demand functions. We can provide a partial answer to this question by referring to those empirical studies of travel behaviour which have employed demand functions unfettered by the restrictions of behavioural theories. It has been demonstrated (Senior and Williams, 1977; Williams and Senior, 1977) that the properties (e.g. the signs of elasticity parameters) and resultant response predictions of a number of models widely used in Transportation Studies are badly at variance with what are generally regarded as acceptable.† An important point to note is that the unacceptability of certain structures can be diagnosed *a priori*, while others may be ruled out after estimation (see Williams, 1977). Thus we may appeal to a theory of demand—based for example, on the rational choice paradigm—in order to organize our *a priori* assumptions and ensure that the demand functions which embody them will be free from some of the absurdities and inconsistencies (interpreted within that paradigm!) which may, and indeed do, arise from the indiscriminte use of "pragmatic" models.

In order to determine what restrictions, if any, the theoretical postulates (be they drawn from economics, information science, physics or psychology) impose on a demand model we can ask which functional forms $f(\mathbf{z})$ are consistent with those postulates, and conversely we can enquire about which theories or hypotheses, are consistent with a given functional form. Two strategies have been adopted in addressing these issues. First, we can start with specific hypotheses and constructively derive choice probabilities from them. As McFadden (1978) notes, the primary drawback of this approach is that it often leads to analytic intractibilities, or results in functional forms impractical for empirical analyses. (In later sections where tests on mis-specification demand that complex behavioural models be formed by this procedure we resort to Monte Carlo simulation for the generation of choice probabilities.) A second approach is to attempt to verify directly or indirectly that a manageable demand function(s) is consistent with a given hypothesis. Such an approach has been adopted elsewhere in applied consumer demand studies and is exemplified in the formation of an acceptable linear expenditure model system (Deaton, 1975, Chap. 3).

Rejection of a particular model may take place on a range of criteria varying considerably in their formal status. Inconsistency with a theoretical postulate is a rigorous and unambiguous basis for rejection. In contrast the adoption of statistical goodness-of-fit criteria and parameter comparability with previously determined values are often subject to a possible range of such measures. While we may screen for and purge a model from the more dramatic or pathological aspects of mis-specification, the functional form and included variables may still be inappropriate and unrepresentative of the true process, which of course is not known *a priori*! Specification errors will no doubt invariably exist, and the realistic goal of the demand analyst is the generation of a model which is *plausible* and at the same time workable. From a theoretical viewpoint it is of interest and perhaps vital to test a model against less restrictive or simply different theoretical forms. In particular we might want to enquire whether large specification errors occur when the theoretical asumptions adopted in the generation of a particular model are violated. This is a prominent theme of the paper.

It is convenient, and useful for later considerations, to refer to the *mother logit homily* introduced by McFadden (1975) which states that any continuous qualitative choice model can be written in multinomial logit form without loss of empirical generality. Thus, a normalized choice model of the form

$$P_\rho = f_\rho(\mathbf{Z}, \boldsymbol{\theta}) \tag{4}$$

satisfying

$$\sum_\rho f_\rho = 1 \tag{5}$$

†This point should not be confused with the imputation by the modeller to *observed* behaviour discussed above. Here we are discussing the properties of given models used to predict behaviour.

may be written as

$$P_\rho = \frac{\exp(G_\rho(\mathbf{Z}, \boldsymbol{\theta}))}{\sum_\rho \exp(G_\rho(\mathbf{Z}, \boldsymbol{\theta}))} \tag{6}$$

with

$$G_\rho = \log f_\rho. \tag{7}$$

G may of course be a very complicated non-linear function of the attributes and parameters. Any choice model may thus be represented whether underpinned by utility theory or not. This simple transformation allows mis-specification issues to be discussed and in certain cases numerically assessed within the logit framework.

If we wish to statistically assess or "test" a model,

$$P_\rho = \frac{\exp(G^0_\rho(\mathbf{Z}, \boldsymbol{\theta}))}{\sum_\rho \exp(G^0_\rho(\mathbf{Z}, \boldsymbol{\theta}))} \tag{8}$$

consistent with a set of premises T° against a theoretically or empirically less restrictive form f^*_ρ consistent with T^*, then by writing

$$P^*_\rho = \frac{\exp(G^*_\rho)}{\sum_\rho \exp(G^*_\rho)} \tag{9}$$

with

$$G^*_\rho = G^\circ_\rho + \alpha \Delta G_\rho \tag{10}$$

and

$$\Delta G_\rho = g_\rho(\mathbf{Z}', \boldsymbol{\phi}'), \tag{11}$$

a test which determines whether α is significantly different from zero would serve to statistically assess whether one or more of the assumptions upon which the model (8) had been constructed are violated. If α is significant then the function g_ρ, expressed in terms of the sets $\mathbf{Z}'$ and ϕ' and reflecting the difference between T^* and T, will assert itself in the exponent.

Now this process is considerably more straightforward to state than to execute. While all discrete choice models will conform to the mother logit homily, it may simply not be useful to express them in this form—as in the case of the intractible multinomial probit function. There do however exist well known applications of the above procedure. These include tests for the existence of "interaction terms" in the case of linear specifications against the so-called universal logit form by McFadden *et al.* (1976); and the test of a linear multinomial logit model against nested logit forms, to which we shall return.

The rejection of a model against a more general specification as described above is achieved by appealing to *observed* choice behaviour. Alternatively we might conduct a set of numerical experiments to investigate the extent of mis-specification when applying a model (8) consistent with T° in circumstances were the form (9) underpinned by T^* is considered to represent the true process. The latter may for example correspond to a relaxation of one of the theoretical restrictions or assumptions imposed on the former. To emphasise a possible distinction between the attribute values and parameters in the two models we write

$$P^*_\rho = \frac{\exp(G^*_\rho(\mathbf{W}^*, \boldsymbol{\phi}^*))}{\sum_\rho \exp(G^*_\rho(\mathbf{W}^*, \boldsymbol{\phi}^*))} \tag{12}$$

and

$$P_\rho = \frac{\exp(G_\rho(\mathbf{Z}, \boldsymbol{\theta}))}{\sum_\rho \exp(G_\rho(\mathbf{Z}, \boldsymbol{\theta}))}. \tag{13}$$

By generating synthetic choice data according to the model (12) for given sets of $\mathbf{W}^*$ and $\boldsymbol{\phi}^*$, and estimating the parameters $\boldsymbol{\theta}$ of the model (13) applied to that data, according to a set of statistical criteria, it is possible to assess: first, whether the latter provides a good description of the pattern $\{P_\rho^*\}$; and secondly, how the response to change, obtained by varying a set of attributes, differs in the two systems—that is how the *estimated* elasticities, which involve parameters from the set $\boldsymbol{\theta}$, differ from the true elasticities derived from (12).

Now the value of $\boldsymbol{\theta}$ estimated by this process of applying the model (13) to the simulated data generated by (12) will in general be a function

$$\hat{\theta} = \hat{\theta}(\mathbf{W}^*, \boldsymbol{\phi}^*) \tag{14}$$

of both the given set of attributes $\mathbf{W}^*$ and parameters ϕ^*. We may now write

$$P_\rho^* = \frac{\exp\{G_\rho(\mathbf{Z}, \hat{\boldsymbol{\theta}}) + \Delta G_\rho\}}{\sum_\rho \exp\{G_\rho(\mathbf{Z}, \hat{\boldsymbol{\theta}}) + \Delta G_\rho\}} \tag{15}$$

with

$$\Delta G_\rho = G_\rho^*(\mathbf{W}^*, \boldsymbol{\phi}^*) - G_\rho(\mathbf{Z}, \hat{\theta}(\mathbf{W}^*, \boldsymbol{\phi}^*)). \tag{16}$$

If the estimated model $\{P_\rho;\ \rho = 1, \ldots N\}$ provides a good statistical fit to the data it will be because the difference $\{\Delta G\}$ fails to assert itself in the fit criterion. Interest will centre on how the vector of differences $\Delta\mathbf{P}$ defined by

$$\Delta P_\rho = P_\rho^* - P_\rho(\mathbf{Z}, \hat{\boldsymbol{\theta}}) \quad \rho = 1, \ldots N \tag{17}$$

and response error $\delta\Delta\mathbf{P}$, whose components are (see eqn 2)

$$\delta\Delta P_\rho = \delta P_\rho^* - \delta P_\rho(\mathbf{Z}, \hat{\boldsymbol{\theta}}) \quad \rho = 1, \ldots N \tag{18}$$

vary with the input values $\mathbf{W}$ and $\boldsymbol{\phi}$, or the theoretical differences between T and T^*. It should be noted that the stimulus giving rise to the response vectors $\delta\mathbf{P}^*$ and $\delta\mathbf{P}$ in both the synthetic data and applied model may or may not correspond to the same attribute set—that is, attribute *changes* may themselves be subject to mis-specification. Now, if the differences $\{\Delta P_\rho\}$ and $\{\delta\Delta P_\rho\}$, or some composite function appropriate to a statistical fit criterion, proved to be significant we would presumably conclude that the specification error was important under the conditions of the experiment. However, it might well be the differences $\Delta\mathbf{P}$ are in fact small—because $\Delta\mathbf{G}$ evaluated in the "base system" is small—yet $\delta\Delta\mathbf{P}$ is significantly larger as the *changes in* $\Delta\mathbf{G}$ and importantly the differences in the parameter sets start to assert themselves in the stimulus-response relation. The relative sizes of these differences $\Delta\mathbf{P}$ and $\delta\Delta\mathbf{P}$ will figure prominently in a number of mis-specification experiments which we shall describe. It is one of the objects of these tests to identify the conditions under which we are prepared to accept a model on statistical grounds which may lead to large response errors.

It should be emphasised here that while we have formally expressed models in terms of their exponential transforms according to the mother logit representation, the numerical experiments in later sections will not employ this feature. It has been used to bring out certain features of the mis-specification and response error problems. It may be useful later to bear in mind that the adoption of different sets of hypotheses as the basis for model formation, can ultimately be reflected in the differences in the exponents of a logit-type expression (whether or not the models are easily transformed to an explicit analytic form in practise).

We now turn to discuss the problems of mis-specification and theoretical misrepresentation in a way which draws out the role of the modeller as an *observer* and the importance of the framework within which data are interpreted, theoretical propositions organized and demand models formulated.

3. DECISION CONTEXTS, FRAMES OF REFERENCE AND THE SIMULATION OF BEHAVIOUR

Theoretical mis-representation, manifested in functional mis-specification, results partly from our ignorance of the "true" process governing revealed behaviour, and importantly from the strong desire to produce a workable model with the information available. As we indicated above, tests on potential mis-specification issues may be performed by examining the disparities which exists between alternative functional expressions reflecting differing descriptions of assumptions about behaviour. We must now relate these considerations to the more general problem of how an *observer* of behaviour can impose particular and inappropriate prescriptions of behaviour on the modelling process. In essence we wish to assess the consequences of inappropriate descriptions for forecasting travel response.

We introduce two *perspectives* Λ and Λ^* each of which involves a description of behaviour within a selected frame of reference. The former Λ entails a description adopted by an observer (the modeller) endowed with the information in traditional cross sectional surveys—namely choices and certain measured attribute values. Λ^* on the other hand, is a perspective of relative privilege and will involve a more "realistic" description of the individual decision process (it does not concern us *yet* whether this information is relevant and how it is obtained). In the examples to follow this section Λ^* will in fact involve simple refinements to the assumptions normally employed in the generation of choice models. In terms of the mis-specification tests described in the previous section the information in Λ and Λ^* will be used to underpin choice models $\{\mathbf{P}\}$ and $\{\mathbf{P}^*\}$.

We shall now assume that Λ and Λ^* are characterized by the following aspects of a choice context for each individual i:

$\Lambda^*: \{\mathbf{d}^*; \mathbf{A}^*(s^*, \mathbf{Z}^*); \psi^*; Q^*\}_i$	$\Lambda: \{\mathbf{d}; \mathbf{A}(s, \mathbf{Z}); \psi; Q\}_i$

in which we denote $\mathbf{d}$, $\mathbf{d}^*$ as the descriptors (attributes) of individual decision makers i; $\mathbf{A}$, $\mathbf{A}^*$ as the sets of alternatives out of which a selection is considered by individual i. Each alternative is characterized by state descriptors s and s^*, respectively (and might relate, for example, to a formal description of an individual travel tour or journey), and each state will be characterized by sets of attributes $\mathbf{Z}$ and $\mathbf{Z}^*$; ψ, ψ^* as the set of constraints to which the individual is subject. These may include: travel time, cost, family interdependence constraints, search time constraints, etc.; and Q, Q^* as the set of objectives motivating the choice process. These formal descriptors are introduced simply to characterize the two perspectives. In the formation of mathematical expressions for predicting response these aspects are incorporated into explicit models D_i^* and D_i of the decision process.

The residual dispersion which will in general characterize an observable trip pattern is described through probabilistic concepts in the modelling process. The notion of probability itself might vary according to the theory of behaviour selected. For example, the observer (modeller) might consider each individual to act rationally and consistently when repeatedly confronted by the same choice. In this case he might interpret the probability P_p^t that an individual t selects alternative A_p in terms of the proportion of a fictitious population π of individuals with *observable* attributes identical to t selecting A_p. Dispersion is attributed to the observer's uncertainty of the true subjective utility values which were taken to be probabilistically distributed over π. This is the usual interpretation adopted by choice modellers who apply random utility theory. Alternatively, the observer might regard the decision rule of any individual t to be intrinsically uncertain and the notion of probability is interpreted in terms of the relative frequency of choice of A_p in repeated trials due to variability in the state of mind. In this paper we shall generate models according to the first interpretation. The probabilistic model will involve an aggregation process over the decision contexts and resultant choices of members belonging to π, as we describe below.

The discrepancies between Λ and Λ^* characterize a whole family of mis-specification problems associated with: the description of the choice contexts; the specification of the decision models D and D^*; and the process of aggregation over individuals in π and π^*. Let us examine the scope of potential mis-representation by considering the assumptions embedded in the process by which conventional choice models, and the multinomial logit model in particular, are generated within random utility theory. An explicit statement of these assumptions will be used later for comparative purposes.

(a) Choice making populations are identified with individual segments of the transport markets. Constraints are handled *explicitly* or implicitly through the use of proxy variables.

(b) All individuals I_i in a given market segment s have the same deterministic choice set $\mathbf{A}^s$ containing alternative $A_1, \ldots A_\rho, \ldots A_{N_s}$. (We assume that certain obvious constraints are catered for, for example the possession of a driving licence in order to be a car driver, or of car availability as a pre-requisite for private travel.)

(c) The objectives of each individual are resolved in the formation of utility functions $U_\rho(\mathbf{Z}, \boldsymbol{\theta})$, $\rho = 1 \ldots N_s$ which are used to record preferences.

(d) The decision model is simply one of utility maximisation.

if

$$D_i\text{: Individual } i \text{ will select } A_\rho \text{ if}$$
$$U^i_\rho > U^i_{\rho'};\ \forall A_{\rho'} \epsilon \mathbf{A}^s. \tag{19}$$

(e) Within each market segment, dispersion is considered to arise from the *unobserved* attributes both of individuals and those associated with the description of choice alternatives. Formally the utility function U^i_ρ is decomposed into a "representative" component $\bar{U}_\rho$ and a residual ϵ^i_ρ accounting for a deviation from the 'group average' which absorbs non-observed attributes including taste variation in the utility function $U_\rho(\mathbf{Z}, \theta)$. Thus,

$$U^i_\rho = \bar{U}_\rho + \epsilon^i_\rho \qquad \forall A_\rho \epsilon \mathbf{A}. \tag{20}$$

(f) The aggregation process over the population comprising s is performed by assuming the "residuals" $\boldsymbol{\epsilon} = (\epsilon_1, \ldots \epsilon_\rho, \ldots \epsilon_N)$ to be distributed randomly over π^s. Specification of the joint probability density function $F(\boldsymbol{\epsilon})$ allows the choice probabilities P_ρ to be determined by integrating over that portion of utility space R_ρ for which the condition (19) holds; that is,

$$P_\rho = \int_{R_\rho} \mathrm{d}\boldsymbol{\epsilon}\, F(\boldsymbol{\epsilon}). \tag{21}$$

(g) If the random components $(\epsilon_1, \ldots \epsilon_\rho, \ldots \epsilon_N)$ are identically and independently distributed (uncorrelated with each other and the attributes $\mathbf{Z}$) according to Weibull functions $W(O, \sigma)$ with standard deviation σ, the aggregation process (21) may be resolved analytically to yield the multinomial logit model (Domencich and McFadden, 1975; Cochrane, 1975)

$$P_\rho = \frac{\exp(\theta \bar{U}_\rho)}{\sum_\rho \exp(\theta \bar{U}_\rho)} \tag{22}$$

in which

$$\theta = \frac{\pi}{\sigma\sqrt{(6)}}. \tag{23}$$

(h) The representative utilities $\bar{U}_\rho$ are expressed as linear functions of the attributes $\mathbf{Z}$, and the parameter set $\boldsymbol{\phi}$ (to be determined in the estimation process).

$$\bar{U}_\rho = \sum_\mu \phi_\mu Z^\mu_\rho. \tag{24†}$$

(i) The response $\delta\mathbf{P}$ arising from a change $\delta\mathbf{Z}$ is traced directly through eqns (24) and (22) by assuming constant parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$.

It can be seen that the assumptions involved in the formation of the linear logit model are

†The parameter θ in eqn (22) is related, through eqn (23) to the dispersion in utilities *measured in specific units.* θ may be absorbed into the utility function for the purposes of estimation, as in the expression

$$P_\rho = \frac{\exp(\bar{U}'_\rho(\mathbf{Z}, \boldsymbol{\phi}'))}{\sum_\rho \exp(\bar{U}'_\rho(\mathbf{Z}, \boldsymbol{\phi}'))}.$$

If "utility" is considered to be measured in, say, time units, then θ will be equal to the coefficient of the attribute *time* in the linear function $\bar{U}'_\rho$. In this paper we shall adopt both representations, and the specific interpretation of the parameters in any context will be apparent in the argument of the exponent.

many, and vary considerably in their formal content, from the postulation of a scalar utility function used to record preferences, to very specific analytic assumptions—e.g. Weibull distributed residuals—invoked for computational tractibility.

Now the adoption of the Multinomial Logit (MNL) model does not *necessarily* imply a commitment to the theory outlined above, as it is well known that the model can be derived from other theoretical stand-points (see, e.g. Thrift and Williams, 1981). There may also be other explanations of dispersion which are *statistically* consistent with the above model. Through a series of simulation experiments we seek such explanations.

In order to investigate the implications of the differences between the behavioural perspectives Λ^* and Λ it is necessary to generate a data base $\mathbf{P}^*$. If it were possible to aggregate analytically over the "individual" decisions of the members of π then it would be unnecessary to resort to simulation. In this paper we shall be dealing with behavioural processes which typically require numerical analysis for their resolution, and we have appealed to Monte Carlo simulation to generate the resultant model. The resolution of choice models using this method has also been considered by Albright *et al.*, 1977; Manski and Lerman, 1978; Ortuzar, 1978, 1979; Robertson, 1978; Robertson and Kennedy, 1979; among others. Each member of the population π of size M associated with a given market segment, is sampled and assigned to a particular alternative $A_1, \ldots A_\rho, \ldots . A_n$ according to the outcome of a simulated decision rule D^* associated with the perspective Λ^*. When M is large, the proportion M_ρ/M which "selects" A_ρ then approximates P^*_ρ, a member of a discrete probability distribution $P^*_1, \ldots, P^*_\rho, \ldots P^*_N$. What was considered sufficiently "large" in each experiment had to be determined in the process of the investigation. The vectors $\mathbf{P}^*$ associated with population $\pi(\mathbf{Z})$, corresponding to different observable attributes, are now considered to be the travel related choice data available to the modeller.

The model $\mathbf{P}$ underpinned by the perspective Λ is then estimated with the synthetic data and used to predict population *response* under conditions of change. These modelled forecasts are then compared with those *separately* generated using simulation according to D^*, in the manner described in the previous section, and the deviation or response error between the "true" and forecast probability measures is recorded. The experimental scheme is depicted in Fig. 1.

The model $\mathbf{P}$ adopted by the observer will often correspond to the multinomial logit form for which we shall provisionally accept the assumptions in perspective Λ with the stages (a)–(i) outlined above. Λ^* will involve the relaxation of one or more of the assumptions employed in these steps. The simulation tests involve relaxation of the assumptions associated with steps (g), (h), (d), (e) and (i) in the above list.

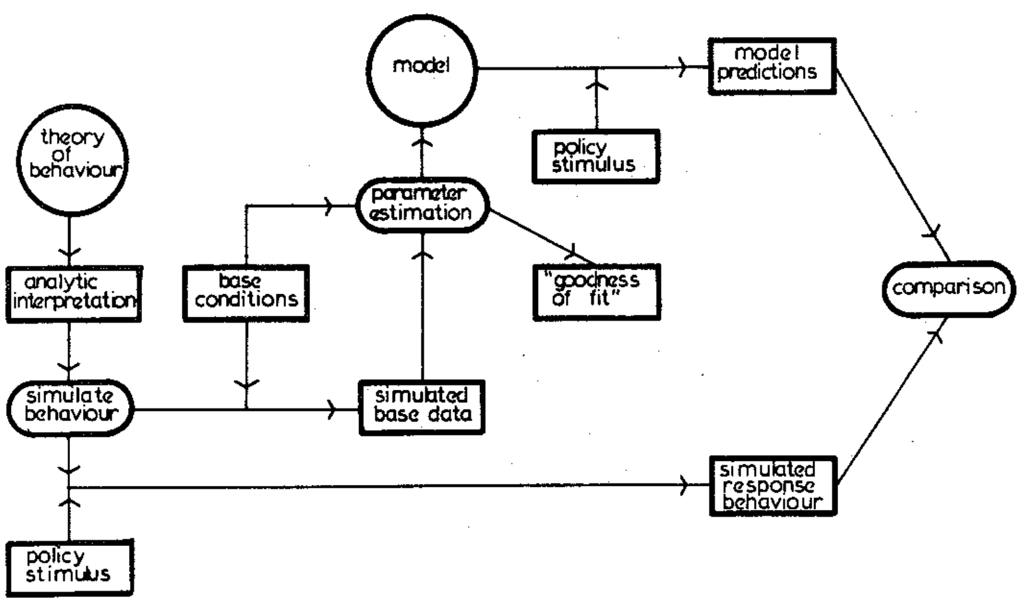


Fig. 1. Testing models of behaviour using simulation.

To repeat, we seek, in each case, alternative explanations and models which are statistically consistent with given patterns of choice behaviour, and which result in different response forecasts. The difference, referred to as the response error, is a measure, as we later discuss, of the consequences of failing to discriminate between alternative representations of behaviour at the cross section.

Now, *in practise* data consists of 1,0 information on whether an alternative in a set is accepted or not by an individual, and the existence of dispersion due to behavioural variations in decision making is imputed from the observed dispersion in the model $\{P_\rho\}$ estimated from data points corresponding to different **Z** values. Because details of the decision process are known in the Λ^* perspective the discrete probability distribution—the data–$\mathbf{P}^*_\rho \equiv \{P^*_\rho(\mathbf{Z})\}$ may be generated directly. The modeller might seek to attribute the source of the dispersion in $\mathbf{P}^*$ to the distribution in utility values according to eqn (23). If then a logit model

$$P_\rho = \frac{\exp(\bar{U}_\rho(\mathbf{Z}, \boldsymbol{\phi}'))}{\sum_\rho \exp(\bar{U}_\rho(\mathbf{Z}, \boldsymbol{\phi}'))}$$

is adopted to account for the variation in the data $\mathbf{P}^*$, and found to provide a good statistical fit, the estimated parameters will *be consistent with* an interpretation of $\mathbf{P}^*$ having been generated by utility maximisers whose utilities are Weibull distributed with means $\{\bar{U}_1, \ldots \bar{U}_\rho, \ldots \bar{U}_N\}$ and common standard deviation σ given by eqn (23), that is

$$\sigma = \frac{\pi}{\sqrt{(6)}\theta}. \tag{25}$$

As $\mathbf{P}^*$ will, in fact, have been generated according to an alternative set of rules or assumptions, it is clear that a good fit to the base pattern $\mathbf{P}^*$ is in this case consistent *in a statistical sense*, with at least two interpretations of the observed dispersion or variation in behaviour. It then remains to see whether the response predicted by the model adopted by the observer (the multinomial logit function in the above case) provides an accurate approximation to the simulated response recorded by $\mathbf{P}^*$ for a revised set of external conditions—interpreted as a policy stimulus.

We now proceed to the mis-specification experiments. Out of the whole family of possible tests introduced above we have chosen four. The first is, in nature, the most conventional and involves an assessment of mis-specification arising from the use of a model insufficiently refined to accommodate the perceived similarity between alternative choices. The test is designed not only to shed some light on the limitations of the multinomial and nested logit models, but more importantly from a theoretical standpoint, provides an important link with later considerations of the generation of choice probabilities from alternative decision models.

4. MODEL STRUCTURES AND THE SIMILARITY OF TRAVEL RELATED SUBSTITUTES

4.1 *Correlation and model structures*

The adoption of simple models such as the multinomial logit model for applications in which the choice alternatives are considered to be endowed with degrees of "similarity" has resulted in a series of ambiguities and inconsistencies which have recently been resolved within the framework of random utility theory (Williams, 1977; Daly and Zachary, 1978; McFadden, 1978; Hausman and Wise, 1978; Daganzo *et al.*, 1977). It is now well known that if the multinomial logit model is indiscriminately applied to choice contexts, involving: multiple modes; multiple routes; mode-route or location-mode combinations, etc. the cross elasticities obtained and resultant response properties of the models will often be unacceptable. In fact the multinomial logit model has frequently been applied in conjunction with utility functions which are inconsistent with its formulation (Williams 1977, 1980). To overcome these difficulties, encountered in the use of the simple logit function, a broader class of models with less restrictive properties of cross-substitution has been derived within the framework of random utility theory.

In the remainder of this section we shall use the term "elasticity function" and "cross-substitution" to refer to variation of a choice probability P_ρ with respect to the change in a utility component $\bar{U}_{\rho'}$ which will in general be a function of the set of attributes $\{Z^{\mu}_{\rho'};\ \mu = 1,2,\ldots\ m\}$. These elasticities

$$\mathscr{E}_{\rho\rho'} = \frac{\partial P_\rho}{\partial U_{\rho'}} \cdot \frac{\bar{U}_{\rho'}}{P_\rho} \tag{26}$$

are related to those defined in terms of the attributes $\mathscr{E}^{\mu}_{\rho\rho'}$ in the following straightforward manner

$$\mathscr{E}^{\mu}_{\rho\rho'} = \mathscr{E}_{\rho\rho'} \cdot \frac{Z^{\mu}_{\rho'}}{\bar{U}_{\rho'}} \cdot \frac{\partial \bar{U}_{\rho'}}{\partial Z^{\mu}_{\rho'}} \tag{27}$$

In the considerations on alternative model structures and mis-specification we shall be interested in the *response* properties of models, and in particular in the number of parameters which characterize the matrix $\mathscr{E}$ with elements $\{\mathscr{E}_{\rho\rho'}\}$.

The structure of random utility models generated by eqn (21)

$$P_\rho = \int_{R_\rho} \mathrm{d}\boldsymbol{\epsilon} F(\boldsymbol{\epsilon})$$

is directly determined by the distribution function $F(\boldsymbol{\epsilon})$ and a number of parameters which characterize the resultant function $f(\mathbf{Z}, \boldsymbol{\theta})$ in eqn (1) will be embodied in the variance-covariance matrix $\underset{\sim}{\Sigma}$. The elements of this matrix are defined by

$$\Sigma_{\rho\rho'} = E(\epsilon_\rho \epsilon_{\rho'}) \tag{28}$$

in which $E(\cdot)$ denotes the expectation value. In general the number of parameters in the matrix $\mathscr{E}$ will be the same as that entering $\underset{\sim}{\Sigma}$.

The variance-covariance matrix corresponding to the multinomial logit model, generated from identical and independent distributions, is given by

$$\underset{\sim}{\Sigma} = \sigma^2 \boldsymbol{I} \tag{29}$$

$$= \frac{\pi^2}{6\theta^2} I \tag{30}$$

in which $\boldsymbol{I}$ is the unit matrix, and the elastic matrix $\mathscr{E}$ is characterized by the single parameter θ.

A natural point of enquiry is the potential mis-specification arising from the application of models characterized by diagonal (or other restricted forms of) variance-covariance matrices in choice contexts for which more general structures are appropriate. This is a question we shall take up later.

A number of model structures have now been derived which accommodate varying degrees of "similarity" as expressed through the correlation between the stochastic residuals in eqn (21). These range from: the nested or hierarchical logit model (HL) (Williams, 1977; Daly and Zachary, 1978; McFadden, 1979);† the cross-correlated logit (CCL) model (Williams, 1977); the general extreme value (GEV) class of models proposed by McFadden (1978) which contains the hierarchical and multinomial logit functions as special cases; to the general but computationally unwieldy multinomial probit (MNP) model (Domencich and McFadden, 1975; Daganzo *et al.*, 1977).

The existence of a broad class of models constructed on less restrictive principles than the multinomial logit model has allowed mis-specification tests to be performed on the latter. As we noted above, McFadden *et al.* (1976) have tested the MNL model against a universal logit

†An excellent discussion of the nested logit model can be found in the paper by Sobel (1980).

specification, Hausman and Wise (1978) and Horowitz (1979a, 1979b) have examined the multinomial logit against multinomial probit forms—while Ben-Akiva (1974) and others have in turn tested the MNL model against alternative nested (HL) specifications.

In this section we shall examine the performance of restricted members of the logit family—the MNL and alternative HL varieties—against a 3 parameter cross-correlated representation which contains these specifications as special cases. We shall examine the potential mis-specification in contexts involving a combination of choice "dimensions" X and Y–say location and mode. If X_μ, $\mu = 1, 2, \ldots$ represents the set of alternatives available in the X-dimension and Y_ν, $\nu = 1, 2, \ldots$ those in the Y-dimension, then the total set of available alternatives $\{\ldots A_\rho \ldots\}$ is composed of the combination $\{\ldots X_\mu Y_\nu \ldots\}$.

The utility function governing choice between these alternatives will be taken to be of traditional form (see Ben-Akiva, 1974; Williams, 1977)

$$U(X, Y) = U_X + U_Y + U_{XY}. \tag{31}$$

The components of eqn (31) may be written

$$U(\mu, \nu) = U_\mu + U_\nu + U_{\mu\nu} \qquad \forall X_\mu \epsilon X,\ Y_\nu \epsilon Y, \tag{32}$$

in which U_μ and U_ν are themselves components specific to the choice dimensions X and Y respectively, while $U_{\mu\nu}$ is an interaction term. Thus in a location (X) and mode (Y) choice context, U_X would refer to components which vary over locations but not modes, U_Y vary over modes but not locations, while U_{XY} are components, perhaps transportation costs which vary over location and modes.

If now $U(\mu, \nu)$ is written in terms of the representative utilities and residuals

$$U(\mu, \nu) = \bar{U}(\mu, \nu) + \epsilon(\mu, \nu) \tag{33}$$

in which

$$\bar{U}(\mu, \nu) = \bar{U}_\mu + \bar{U}_\nu + \bar{U}_{\mu\nu} \tag{34}$$

and

$$\boldsymbol{\epsilon}(\mu, \nu) = \boldsymbol{\epsilon}_\mu + \boldsymbol{\epsilon}_\nu + \boldsymbol{\epsilon}_{\mu\nu} \tag{35}$$

the variance-covariance matrix can be expressed in terms of the following expectation value

$$\sum\nolimits_{\mu\nu,\mu'\nu'} = E(\epsilon_\mu + \epsilon_\nu + \epsilon_{\mu\nu}, \epsilon_{\mu'} + \epsilon_{\nu'} + \epsilon_{\mu'\nu'}). \tag{36}$$

We shall now assume that ϵ_μ, ϵ_ν and $\epsilon_{\mu\nu}$ are *separately* identically and independently distributed, with

$$E(\epsilon_\mu \epsilon_{\mu'}) = \sigma_X^2\, \delta_{\mu\mu'} \tag{37}$$

$$E(\epsilon_\nu \epsilon_{\nu'}) = \sigma_Y^2 \delta_{\nu\nu'} \tag{38}$$

$$E(\epsilon_{\mu\nu}, \epsilon_{\mu'\nu'}) = \sigma_{XY}^2\, \delta_{\mu\mu'} \delta_{\nu\nu'} \tag{39}$$

with all cross terms (such as $E(\epsilon_\mu \epsilon_\nu)$ vanishing). The Kronecker delta is defined in the usual way

$$\begin{aligned} \delta_{\alpha\beta} &= 1 \quad \text{if } \alpha = \beta \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{40}$$

The matrix elements of $\boldsymbol{\Sigma}$ now become

$$\Sigma_{\mu\nu,\mu'\nu'} = \sigma_X^2 \delta_{\mu\mu'} + \sigma_Y^2\, \delta_{\nu\nu'} + \delta_{XY}^2\, \delta_{\mu\mu'} \cdot \delta_{\nu\nu'}. \tag{41}$$

We wish to determine the conditions under which the application of MNL and HL models—themselves underpinned by $\boldsymbol{\Sigma}$ matrices which are special cases of eqn (41), will give

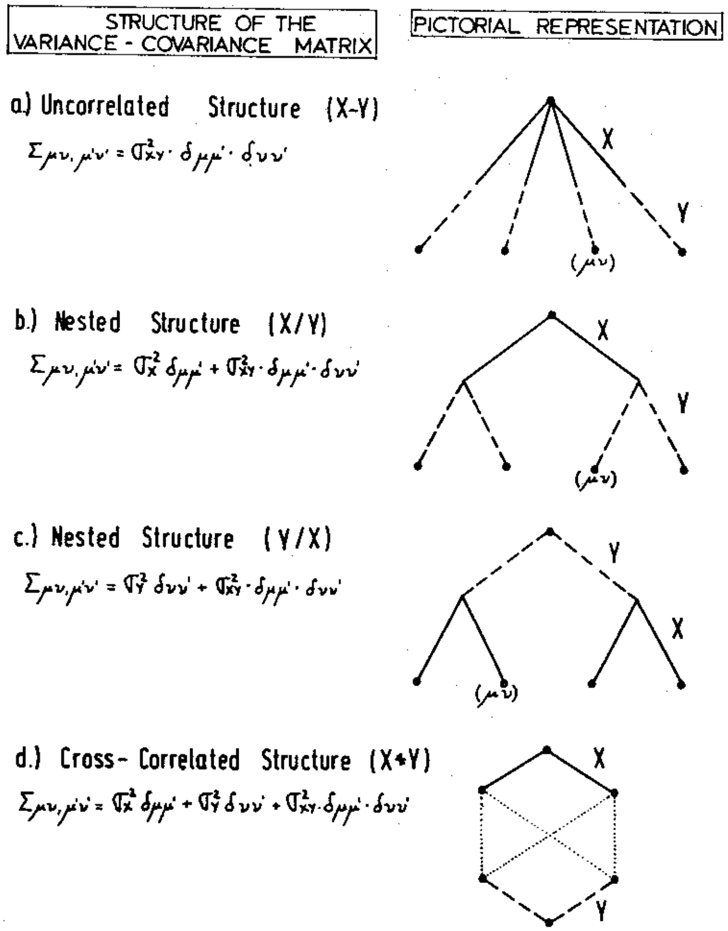


Fig. 2. Representation of the structure of choice models.

Table 1. Characteristics of selected members of the logit family

MODEL	"STANDARD" FORM	MOTHER-LOGIT REPRESENTATION "G-FUNCTION" (see Eq. 7)	PARAMETER RELATIONS AND OTHER RESTRICTIONS
Multinomial Logit Model	$P_{\mu\nu} = \dfrac{\exp\{\Delta(\bar{U}_\mu + \bar{U}_\nu + \bar{U}_{\mu\nu})\}}{\sum_{\mu\nu} \exp\{\Delta(\bar{U}_\mu + \bar{U}_\nu + \bar{U}_{\mu\nu})\}}$	$G_{\mu\nu} = \Delta(\bar{U}_\mu + \bar{U}_\nu + \bar{U}_{\mu\nu})$	$\Delta = \dfrac{\pi}{\sigma_{XY} \cdot \sqrt{6}}$
'Nested' or 'Hierarchical' Logit Model (X/Y)	$P_{\mu\nu} = \dfrac{\exp\{\beta(\bar{U}_\mu + \tilde{U}_{\mu*})\}}{\sum_\mu \exp\{\beta(\bar{U}_\mu + \tilde{U}_{\mu*})\}} \dfrac{\exp\{\Delta(\bar{U}_\nu + \bar{U}_{\mu\nu})\}}{\sum_\nu \exp\{\Delta(\bar{U}_\nu + \bar{U}_{\mu\nu})\}}$	$G_{\mu\nu} = \beta\bar{U}_\mu + (\beta - \Delta)\tilde{U}_{\mu*} + \Delta(\bar{U}_\nu + \bar{U}_{\mu\nu})$	$\Delta = \dfrac{\pi}{\sigma_{XY} \cdot \sqrt{6}}$ $\beta = \Delta(1 + \frac{6}{\pi^2} \cdot \sigma_X^2 \Delta^2)^{-\frac{1}{2}}$ $\tilde{U}_{\mu*} = \frac{1}{\Delta} \log \sum_\nu \exp\{\Delta(\bar{U}_\nu + \bar{U}_{\mu\nu})\}$
'Nested' or 'Hierarchical' Logit Model (Y/X)	$P_{\mu\nu} = \dfrac{\exp\{\lambda(\bar{U}_\nu + \tilde{U}_{\nu*})\}}{\sum_\nu \exp\{\lambda(\bar{U}_\nu + \tilde{U}_{\nu*})\}} \dfrac{\exp\{\Delta(\bar{U}_\mu + \bar{U}_{\mu\nu})\}}{\sum_\mu \exp\{\Delta(\bar{U}_\mu + \bar{U}_{\mu\nu})\}}$	$G_{\mu\nu} = \lambda\bar{U}_\nu + (\lambda - \Delta)\tilde{U}_{\nu*} + \Delta(\bar{U}_\mu + \bar{U}_{\mu\nu})$	$\Delta = \dfrac{\pi}{\sigma_{XY} \cdot \sqrt{6}}$ $\lambda = \Delta(1 + \frac{6}{\pi^2} \cdot \sigma_Y^2 \Delta^2)^{-\frac{1}{2}}$ $\tilde{U}_{\nu*} = \frac{1}{\lambda} \log \sum_\mu \exp\{\Delta(\bar{U}_\mu + \bar{U}_{\mu\nu})\}$
'Cross-Correlated' Logit Model	For specification see Williams (1977)		

rise to serious mis-specification errors when the true choice process is characterised by a 3 parameter model with non-zero σ_X, σ_Y, and σ_{XY}.

Some characteristics of the 1-parameter MNL and 2-parameter NL specifications together with a pictorial representation of the structures of correlation between the choice alternatives are given in Fig. 2. We have expressed the specification of these models in terms of their G-functions (in Table 1). Thus the traditional representation of the

Nested Logit (X/Y) model (see Williams, 1977),

$$P_{\mu\nu} = \frac{\exp\{\beta(\bar{U}_\mu + \tilde{U}_{\mu*})\}}{\sum_\mu \exp\{\beta(\bar{U}_\mu + \tilde{U}_{\mu*})\}} \cdot \frac{\exp\{\Delta(\bar{U}_\nu + \bar{U}_{\mu\nu})\}}{\sum_\nu \exp\{\Delta(\bar{U}_\nu + \bar{U}_{\mu\nu})\}} \tag{42}$$

with

$$\tilde{U}_{\mu*} = \frac{1}{\Delta} \log \sum_\nu \exp\{\Delta(\bar{U}_\nu + \bar{U}_{\mu\nu})\} \tag{43}$$

can, by simple manipulation, be written in mother logit form

$$P_{\mu\nu} = \frac{\exp\{\beta\bar{U}_\mu + (\beta-\Delta)\tilde{U}_{\mu*} + \Delta(\bar{U}_y + \bar{U}_{\mu\nu})\}}{\sum_{\mu\nu} \exp\{\beta\bar{U}_\mu + (\beta-\Delta)\tilde{U}_{\mu*} + \Delta(\bar{U}_\nu + \bar{U}_{\mu\nu})\}} \tag{44}$$

thus revealing the structure of the G-function. This makes transparent the fact that when off-diagonal elements of the correlation matrix Σ vanish ($\sigma_X \to 0$), β tends to Δ and the model collapses to the multinomial logit expression. Correspondingly, the 2-parameter elasticity matrix elements

$$\mathscr{E}_{\mu\nu,\mu'\nu'} = \Delta\bar{U}_{\mu'\nu'}\{\delta_{\mu\mu'} \cdot \delta_{\nu\nu'} - P_{\mu'\nu'} + \frac{(\beta-\Delta)}{\Delta}(\delta_{\mu\mu'}P_{\nu'|\mu} + P_{\mu'\nu'} \cdot P_{\nu'|\mu'})\} \tag{45}$$

with

$$P_{\nu|\mu} = \frac{\exp\{\Delta(\bar{U}_\nu + \bar{U}_{\mu\nu}))\}}{\sum_\nu \exp\{\Delta(\bar{U}_\nu + \bar{U}_{\mu\nu})\}} \tag{46}$$

reduce in this limit to their 1-parameter forms.

Note that the model will only be consistent with utility maximisation if the estimated elasticity parameters β and Δ satisfy the inequality

$$\beta - \Delta \leqslant 0. \tag{47}$$

A violation of this condition can result in the pathological condition in which certain elasticity elements attain the wrong sign—and increase in the desirability of an alternative may result in a decrease of its share. This anomaly has in fact been associated with a class of nested logit models employed in British Transport Studies (Williams and Senior, 1977). We shall be particularly anxious to identify the conditions in which this pathological condition arises in the mis-specification tests below.

In order to assess the extent of mis-specification errors we seek a model(s) which accommodates the full degree of cross-substitution implied by the variance-covariance matrix (41) with $\sigma_X \neq 0$, $\sigma_Y \neq 0$, $\sigma_{XY} \neq 0$ and which collapses in appropriate limits to the Nested and Multinomial logit models.

One possibility is to appeal to the GEV class for the generation of a suitably structured model model with 3 parameter degrees of freedom. McFadden (1978) has shown that if $\mathscr{H}(V_1, \ldots V_p, \ldots V_N)$ is a non-negative linear homogeneous function of non-negative representative utilities $V_1, \dots V_\rho \dots V_n$, and subject to certain additional restrictions, the model

$$P_\rho(V_1, \dots V_\rho, \dots V_N) = V_\rho \cdot \frac{\partial \mathcal{H}}{\partial V_\rho} \cdot \mathcal{H}^{-1} \tag{48}$$

will be consistent with utility maximisation. For the choice contexts considered in this section the $\mathcal{H}$-functions required to form the multinomial and nested models are given in Table 2. Although the following three parameter generalization of these forms

$$\mathcal{H}(\gamma = \beta/\Delta, \delta = \lambda/\Delta) = \sum_\mu \left\{ \sum_\nu V_{\mu\nu}{}^{1/\gamma} \right\}^\gamma + \sum_\nu \left\{ \sum_\mu V_{\mu\nu}{}^{1/\delta} \right\}^\delta - \sum_{\mu\nu} V_{\mu\nu} \tag{49}$$

suggests itself, this does not lead to a simple expression whose parameters can be readily related to the elements of the Σ matrix The search for a *simple exact* closed form analytic expression was thus abandoned† in favour of the intuitively more appealing method of direct numerical solution of eqn (21) by Monte Carlo simulation using Weibull distributions for *each* component in the U_X, U_Y, and U_{XY} dimensions.‡ In the process it was possible to test the accuracy of the cross-correlated logit (CCL) model proposed by Williams (1977) as a 3-parameter generalization of the MNL and HL models to which it reduces in appropriate limits. This model is not however consistent with utility maximisation. Its characteristics and pictorial representation are given in Fig. 2, and its full specification is given in Williams (1977).

4.2 *Simulation experiments with alternative model structures*

The models and their underlying utility functions will now be referred to in terms of their utility coordinates.

$$\{\bar{U}; \Sigma\} \equiv \{\bar{U}_X, \bar{U}_Y, \bar{U}_{XY}; \sigma_X, \sigma_Y, \sigma_{XY}\},$$

and the test on mis-specification by the juxtaposition

$$\boxed{\bar{\mathbf{U}}^*; \Sigma^*(\sigma_X^*, \sigma_Y^*, \sigma_{\mathbf{X}}^*)} \quad \boxed{\bar{\mathbf{U}}; \Sigma(\sigma_X, \sigma_Y, \sigma_{XY}).}$$

Table 2. The generation of members of the logit family from the GEV system (for notation please see text)

MODEL	$\mathcal{H}$-FUNCTION	$V_{\mu\nu}$
Multinomial Logit MNL	$\sum_\mu \sum_\nu V_{\mu\nu}$	$e^{\Delta(\bar{U}_\mu + \bar{U}_\nu + \bar{U}_{\mu\nu})}$
Hierarchical Logit HL(X/Y)	$\sum_\mu \{\sum_\mu V_{\mu\nu}^{\Delta/\beta}\}^{\beta/\Delta}$	$e^{\beta(\bar{U}_\mu + \bar{U}_\nu + \bar{U}_{\mu\nu})}$
Hierarchical Logit HL(Y/X)	$\sum_\nu \{\sum_\mu V_{\mu\nu}^{\Delta/\lambda}\}^{\lambda/\Delta}$	$e^{\lambda(\bar{U}_\mu + \bar{U}_\nu + \bar{U}_{\mu\nu})}$

†There are many possible homogeneous degree one functional expressions $\mathcal{H}(\mathbf{V})$ which reduce in appropriate limits to those shown in Table 1. Equation (49) is one such form. It appears that the existence of "cross-correlation" does considerably complicate the resultant model. We decided to record our rather negative experiences with the GEV system in the hope that other might improve on our attempts in the formation of a practical model from this potentially very useful method.

‡This method generates a model consistent with the variance-covariance matrix (41) but does introduce a slight approximation in the limits when either σ_X or σ_Y tend to zero, as an approximation must be invoked in order to form the nested logit model *from the sum of Weibull distributions* (Williams, 1977). This effect of approximation was tested and was found to be insignificant. It does not affect the conclusions drawn from the simulation experiments. It should however be noted that it is not necessary to invoke an approximation in the formation of the nested logit model from utility maximising behaviour.

Data are generated by direct simulation from the sum of Weibull functions distributed according to $\bar{\mathbf{U}}^*$, Σ^* $(\sigma_X^*, \sigma_Y^*, \sigma_{XY}^*)$. The four models (MNL, HL(X/Y), HL(Y/L), CCL) are then adopted for assessing mis-specification errors. In the multinomial logit model and alternative hierarchical forms the respective parameters were estimated by maximum likelihood. For the CCL model the parameters were theoretically determined (Williams, 1977). In all tests two alternatives were taken in each of the X and Y dimensions, allowing a four alternative choice model to be generated.

Two series of tests were performed. In the first the corresponding pairs of representative utilities $\bar{\mathbf{U}}^*$ and $\bar{\mathbf{U}}$ were taken to be identical and the simulation tests involved variation of the co-ordinates $(\sigma_X^*, \sigma_Y^*, \sigma_{XY}^*)$. A standardization or "normalization" condition is used to bound the joint variation of these quantities, and is of the form

$$\sigma_X^{*2} + \sigma_Y^{*2} + \sigma_{XY}^{*2} = \text{constant}. \tag{50}$$

A particular co-ordinate $(\sigma_X^*, \sigma_Y^*, \sigma_{XY}^*)$ corresponds to a simulation test. To illustrate the possible combinations of these three components we can appeal to a property of an equilateral triangle for which the sum of perpendicular distances to the three sides from an interior point is equal to the height of the triangle.

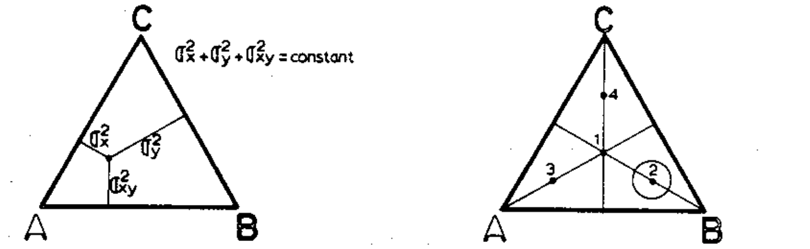
Four particular co-ordinate test points were taken, as shown in Fig. 3(b) *in addition* to test points randomly sampled within the triangle. Any test point may thus be identified with a point in or on the boundary of the triangle as shown in Fig. 3(a). At interior points a three elasticity parameter model is necessary to capture the full range of cross-substitution embodied in the choice probabilities derived from the utility function (31). On the boundaries CB and CA the alternative hierarchical logit models HL(X/Y) and HL(Y/X) for which $\sigma_Y^* = 0$ and $\sigma_X^* = 0$ respectively, are appropriate. It is only at the vertex $C(\sigma_X^* = \sigma_Y^* = 0)$ that the multinomial logit model is *strictly* an appropriate representation. The four members of the logit family (adopted in the Λ-perspective), which correspond to particular points lines or areas of this triangle are used to assess the extent of the response error.

A sample set of results for the four models used to fit data generated from the test points are shown in Fig. 3(c) and (d), in order to illustrate the existence of response errors and pathological response behaviour. The results of the simulations, which are described fully in Williams and Ortuzar (1979) and Ortuzar (1979), are consistent with the following conclusions:

(i) The cross-correlated logit (CCL) model is a good theoretical approximation to that generated by the three parameter utility function. It is, as expected, more flexible than the other three members, and this is particularly apparent when the three quantities σ_X^*, σ_Y^*, and σ_{XY}^* are rather different from zero, and from each other. Because of the interaction between utility components in the separate choice dimensions X- and Y- its estimation is however complex (Williams, 1977) and, for this reason, does not commend itself. It is thus rather important to assess the theoretical error in applying the MNL and HL forms;

(ii) when $\sigma_X^* > \sigma_Y^*$; the specification HL(Y/X), which corresponds to $\sigma_Y > \sigma_X = 0$, will usually result in pathological response behaviour—the change in behaviour predicted by the calibrated model, when the utility of one of the alternatives is modified, is opposite to that simulated. The exact conditions under which anomalous response occurs appear to depend on the values of the representative utility values. When, in the second series of tests, the restriction ($\bar{U}_X^* = \bar{U}_X$; etc.) was replaced by conditions in which alternative values were placed on components of $\bar{\mathbf{U}} \neq \bar{\mathbf{U}}^*$ all response errors were found to be worse than their counterparts in the first series in which $\bar{\mathbf{U}}^* = \bar{\mathbf{U}}$. The performance of the HL models deteriorated and pathological behaviour became more prevalent. This pathological behaviour could in all cases be diagnosed from the value of the estimated parameters of the HL-forms through the violation of condition (47).

(iii) The multinomial logit model performs reasonably well when $(\sigma_X^*, \sigma_Y^*, \sigma_{XY}^*)$ correspond to interior points of the triangle and this is particularly true when $\sigma_X^* \simeq \sigma_Y^*$. When $\sigma_X^*, \sigma_Y^* \to 0$ the model fit to simulated base data not surprisingly is excellent. Although the limited structure of cross substitution becomes apparent near the sides of the triangle (CA) and (CB) the model was found to be considerably more robust than the authors had anticipated.

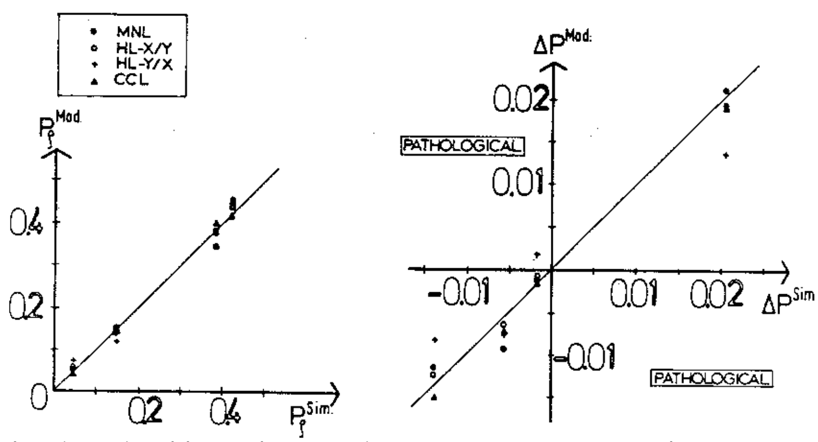


a.) The normalization condition.

b.) The test point coordinates ($\sigma_x, \sigma_y, \sigma_{xy}$)

POINT	MNL	HL-X/Y	HL-Y/X	CCL
1	good	good $\Delta \sim \beta$	good $\Delta \sim \beta$	very good
(2)	regular	very good	pathological $\Delta < \beta$	good
3	regular	pathological $\Delta < \beta$	very good	good
4	very good	very good $\Delta = \beta$	very good $\Delta = \beta$	very good

c.) Model performance at different test points.



d.) An illustration of base performance and response errors at test point 2.

Fig. 3. The design and results of simulation tests to investigate model structure variation.

(iv) Out of the three alternative structures (MNL and the two HL forms) the model which provides the best fit to base (simulated) data, *and is consistent with any restrictions appropriate to its structure* (e.g. eqn 47), provides a good estimate of the response to change.

Although the simulation tests have been confined to a choice context with a small number of alternatives (4), it is thought that the conclusions above would not be qualitatively modified when the number of alternatives is increased, because the focus of the mis-specification tests is the variance-covariance matrix itself the *structure* of which does not vary with problem size. The results summarized above and particularly (iv) seem to indicate that the choice between the MNL and HL-logit forms is not unduly restrictive—with respect to a less restrictive member of the logit family—and would seem to lend some numerical support for the suggestion of Ben-Akiva (1977a) and Williams and Senior (1977) that results of alternative HL and MNL models could be compared and an appropriate model selected according to the estimated value of the "similarity coefficient", which in the present notation is $1-(\beta/\Delta)$ for the HL(X/Y) form. It should however be emphasised that the above tests have not included variation over the choice population of the value of travel (or location) attributes which has been shown by Hausman and Wise (1978), Horowitz (1979) and Cardell and Reddy (1977) to be a significant source of mis-specification. We shall return to this point below.

In this section we have discussed the generation of alternative (and post hoc rationalization of existing) model forms within the framework of random utility maximisation which accommodate differing degrees of similarity between choice alternatives, and have examined the consequences of employing models with restricted properties of cross substitution in particular 2-dimensional choice contexts. We have emphasised the essential continuity between what have often been viewed as very distinct model structures. It is now time to turn to alternative derivations of models and interpretations of their structure based on different formulations of the decision process by which choices are considered to be made.

5. THE MULTICRITERION PROBLEM, DECISION RULE SETS AND THE GENERATION OF CHOICE MODELS

5.1 *Functional characteristics and decision processes*

In this section we shall discuss two issues: alternative conceptualisations of the decision process through which choices are considered to be made; and, the formation of choice models which are underpinned by such processes. This will provide the basis for numerical tests on behavioural mis-representation to be presented in the next section. The discussion will also form a link with the rationalization of alternative model structures developed above, and also on the significance of information limitations in the decision process dealt with later.

One strong focus of the growing literature on mis-specification is associated with the form of the utility function, and particularly the linear-in-parameters-linear-in-attributes (LPLA) varieties accompanying the vast majority of multinomial logit model applications. The body of criticism directed at LPLA forms has stimulated interest in the specification and estimation of non-linear formulations of varying designs (see, e.g. Lerman and Louviere, 1978; Louviere, 1979; Hensher, 1979; Gaudry and Wills, 1977).

Commentary on these *functional* characteristics has been intertwined with statements about alternative models of the *decision process* considered to underpin the choice models. Although a variety of sentiments have been expressed about this relationship, a not untypical view is that because LPLA forms are associated with an *unrealistic* compensatory decision making process (the trade-offs implied by eqn (24) allow a change in one or more of the attributes to be *compensated* by changes in others) the model cannot be appropriately specified.† It is often argued that the decision process is characterized by perception non-linearities, discontinuities, and is more plausibly of a non-compensatory nature‡, and that these features should naturally assert themselves in a model specification (see the discussions in Golob and Richardson, 1979; Recker and Golob, 1979; and Gensch and Svestka, 1978). There is then the view that the distinction between compensatory and non-compensatory models is a theoretically significant classification, which it is claimed has some empirical basis in market research studies.

Before considering these constructs in more detail, it should not be forgotten that one of the prime motivations behind the construction of alternative models of the decision process based, e.g. on elimination-by-aspects (Tversky, 1972), was precisely the limitations of simple scaleable

†Compensatory behaviour should not be associated solely with LPLA functions. Various transformations can generate model forms (e.g. multiplicative in attributes) which are consistent with trade-off principles in the normal sense of that term.

‡In individual decision models based on a non-compensatory mechanism some "good" aspects of an alternative may not be allowed to compensate for inferior aspects which are ranked higher in importance in the selection procedure, simply because that alternative may be eliminated on the basis of the latter in the process of "searching" for an option.

choice models typified by the multinomial logit form. Because they were endowed with the controversial "independence from irrelevant alternatives" property, and could not accommodate satisfactorily the inclusion of options characterized by degrees of similarity, it was recognised as long ago as the early 1960s that such simple models would often portray restrictive and unrealistic properties. Their defects have sometimes been attributed exclusively to an *optimisation* framework underpinning them, rather than the particular assumptions used in conjunction with this approach.

The development of more general model structures for eqn (21) with less restrictive joints distribution functions (as outlined in the previous section), which allow degrees of similarity to be accommodated, has *in a certain sense* removed some of the *original* justification for the construction of alternative decision models. This does not, of course, imply that the currently adopted models of the decision process, and the assumptions through which the more general random utility models have been achieved are necessarily appropriate or realistic, nor does it remove the need to examine competing frameworks.

In the following section we shall examine the consequences of adopting a model based on compensatory (LPLA) behaviour in a situation for which the data are generated by individuals conforming to alternative, including "non-compensatory", rules. That is, we shall juxtapose the two perspectives

$$\boxed{\Lambda^*\text{: various } D^*} \qquad \boxed{\Lambda\text{:} D = \text{linear compensation utility model}}$$

Before this may be achieved however we shall draw out some of the characteristics of "compensatory" and "non-compensatory" decision making rules and discuss their joint membership of a *set* of decision rules. For this purpose we shall examine the decision process in terms of the formal solution of a *multicriterion choice problem*, which is assumed to confront a population of decision makers. This will allow us also to discuss some of the distinctions which are sometimes associated with the notions of "optimisation" and "satisficing" discussed above.

5.2 *The multicriterion problem and decision rule sets*

Let us consider each individual confronted by the decision to be endowed with a *set* of goals or objectives Q and a *set* of constraints ψ. In terms of these we shall formally state the multicriterion problem as follows:

$$\begin{gathered}\max_{\{\text{options}\}} \{\zeta_1(Z_1^1)\ldots.\zeta_1(Z_N^1)\}\\ \vdots \\ \max_{\{\text{options}\}} \{\zeta_\mu(Z_l^\mu)\ldots.\zeta_\mu(Z_N^\mu)\}\\ \vdots \\ \max_{\{\text{options}\}} \{\zeta_M(Z_1^M)\ldots\zeta_M(Z_N^M)\}\end{gathered} \tag{51}$$

subject to the vector of constraints

$$\mathbf{g}(\mathbf{Z}) \leqslant \mathbf{b} \tag{52}$$

in which $\zeta_\mu(z_p^\mu)$ is the value of a criterion function associated with the attribute Z_p^μ accompanying alternative A_p. For example, we might be interested in finding an option (say a mode) in an N-membered set $\{A_1\ldots A_N\}$ which minimises travel time, minimises cost, maximises comfort and safety, etc. These attributes associated with any particular alternative might, in addition, be required to satisfy absolute constraints such as (52).

If a single alternative is found which simultaneously satisfies these optimality criteria (i.e. optimises the M functions in (51)) and whose attributes are feasible in terms of (52) then an unambiguous optimal solution is obtained. There will however, in general, be conflicts between objectives—that is, options will be superior in some respects and inferior in other—and this of course gives the multicriterion problem its flavour.

There are a number of important questions we must address before a choice model based on this multicriterion problem may be constructed:

(i) What strategies might be adopted for the resolution of this problem?

(ii) Are there differences in the stragegies adopted by different individuals in the population π?

(iii) How are these strategies to be formally represented?

(iv) How do we aggregate over the population π to produce a model to be estimated with individual data?

It is especially important to emphasise that the probabilistic choice models which we wish to discuss are derived by aggregating over the actions of the individuals within the population π, and that while any or all individuals may indulge in a "non-compensatory" decision process, it may or may not be appropriate to characterize the "sum-total" of these decisions and the resultant choice model in these terms. We shall encounter this point again.

Let us now consider the first of these issues—how may an individual confronted by a hypothetical decision context resolve the multicriterion problem. There is a wide literature dispersed over several fields which involves the application of decision theory to problems of this kind. In certain disciplines one will meet a distinction between *optimising* and *satisficing* approaches (Eilon, 1972) and because it is a so-called vector optimisation problem, some notion of what optimisation means must be supplied in this context. For our purposes it is unnecessary to be more specific on this issue, and our discussion will centre on the formulation of *compensatory, non-compensatory* and *hybrid* decision rules.

Perhaps the best known and most widely adopted approach to the multiple objective problem is the trade-off strategy which forms the basis for compensatory decision models, in which a *single* objective function

$$\mathcal{J} = \mathcal{J}(\zeta_1, \zeta_2, \ldots \zeta_M) \tag{53}$$

is formed and the appropriate option is extracted. If the ζ_μ functions are simply the attributes Z^μ themselves, or linear transformations on them, $\mathcal{J}$ may be written

$$\mathcal{J} = \mathcal{J}\{\sum_\mu \alpha_\mu Z_1^\mu; \ldots ; \sum_\mu \alpha_\mu Z_\rho^\mu; \ldots ; \sum_\mu \alpha_\mu Z_N^\mu\} \tag{54}$$

and the conventional type of linear "trade-off" problem is addressed. The "trade-off" parameters α are determined from either the stated or revealed preferences of the *individual* decision maker.

One of the characteristics of the trade-off approach is its symmetric treatment of the objective functions. An alternative general approach to the problem is to treat these functions asymmetrically, by either ranking them or converting some or all to constraints by introducing "norms" or thresholds. That is we might require, e.g. that any acceptable alternative has an associated travel time less than a particular amount. Formally, the restriction is imposed that

$$Z_1^\mu, \ldots, Z_\rho^\mu, \ldots Z_N^\mu \leq \bar{Z}^\mu \tag{55}$$

in which $\bar{Z}^\mu$ is a maximum (or minimum when the inequality sign is reversed) satisfactory value for Z^μ.

The creation of norms or thresholds restricts the range of feasible alternatives which individuals are considered to impose on their decision process. Various forms of satisficing model are generated by converting some (or all) of the constraints into norms and establishing a *structured search* for the desired alternative in conjunction with an eliminating strategy.

There are a great many ways in which this "resolution" or "search" strategy *may be considered to be* organized. It might be that a complex cyclic process is used by an individual in which the thresholds become sequentially modified until a unique alternative is found. Equally an individual might be prepared to curtail the search at any point according to a pre-specified decision rule in which case some or all of the attributes or alternatives may not be considered. Indeed, when the notion of satisficing is applied to travel related decisions and particularly those involving location (Heggie, 1978; Young and Richardson, 1978; Thrift and Williams, 1981) the decision model is closely bound up with the acquisition of information in the search process. As Young and Richardson (1978) have remarked, a search may be characterized by an elimination process based on *attributes* or one based on *alternatives*. In the former, attributed are selected in turn and alternatives are "processed" and maintained or rejected in the search depending on the value of these attributes; while in the latter, *alternatives* are selected in turn and their "bundle of attributes" examined. At any stage of the process, alternatives which do not satisfy norms or other constraints are eliminated. In a complex decision process of selecting a house say, both strategies of appraisal may well exist simultaneously. A more detailed consideration of alternative decision strategies is given by Foerster (1979).

The important point to note is that there exists a *set* of decision rules $\mathscr{D}$ from which an individual *may be considered to refer to* in order to resolve a complex choice problem, and that what strategy *is* adopted may well be context dependent. Both the so-called "compensatory" (e.g. linear trade-off) and "non-compensatory" (e.g. lexicographic order rules) refer to particular members of this set. Other members will include hydrid rules which are combinations of "non-compensatory" and "compensatory" elements.

While attempts may be made to determine empirically how individuals do resolve travel related problems, it would appear that any model which emphasises one strategy to the total exclusion of others would appear on *a priori* grounds to be theoretically restrictive. We are not suggesting that it is an easy task to construct (let alone estimate) a computationally tractable model which embraces a broad range of decision rules, but to the extent that *current* frameworks are deemed to be overly simplistic, however, it remains to assess the numerical consequences of their inherent restrictions.

We may now decompose the probability P_ρ of selecting an alternative A_ρ from the set **A** in terms of the outcome of all possible ways by which A_ρ might be selected. That is we can write

$$P_\rho = \sum_{D^* \in \mathscr{D}} P(A_\rho | D^*)\, P(D^* | \mathscr{D}) \tag{56}$$

in which $P(D^*|\mathscr{D})$ is the probability that the decision rule D^* is selected from a finite and non-empty set $\mathscr{D}$; $P(A_\rho|D^*)$ is the probability that A_ρ is chosen on the basis of the selected decision rule D^*; and $\sum_{D^* \in \mathscr{D}}$ represents summation over all decision rules in the set $\mathscr{D}$. It should be noted here that the effect of policy measures may in principle appear in both models for $P(A_\rho|D^*)$ and $P(D^*|\mathscr{D})$.

We can trivially identify the so-called "compensatory" and "non-compensatory" models as special cases of this general structure. For example, the LPLA multinomial logit model *may* be considered to arise as the following special case: $P(D^*|\mathscr{D}) = 1$ if D^* is a linear trade-off construction (see equation (24)). That is, all members belonging to π are endowed with the identical rule D^*, while for all other strategies $P(D^*|\mathscr{D}) = 0$; and $P(A_\rho|D^*) \equiv$ MNL model, constructed on the basis of the usual source of dispersion due to distributed utility values. We shall encounter other special cases below.

The existence of a possibly large set of rules $\mathscr{D}$ from which individuals in the population π may be considered to refer to, has enriched the possible sources of dispersion which may underpin observed behaviour. The decomposition (56) itself emphasises that individuals—if the probabilities are interpreted in terms of proportions of π—may not only differ in the rules they select and thus be associated with different behaviour, but may differ in the value of "parameters" (e.g. thresholds, trade-off valuations, etc.) associated with any given rule.

The process of model formation may be regarded as one of summing over the contribution to P_ρ from the decision rules associated with all members of the population π. This may proceed *explicitly* or *implicitly*. We can for example specify the set of relevant decision rules and parameterize the probability distribution determining their frequency of selection, generate the choice probability $P(A_\rho|D^*)$ and finally perform the summation to accumulate the contributions to each share. Alternatively, indirect attempts may be made to show that a constructed model is consistent with eqn (56).

Before proceeding to the mis-specification tests in the next section we shall consider Tversky's elimination-by-aspects (EBA) model (Tversky, 1972) which is one of the best known to be based on "non-compensatory" constructs. This will allow us to classify the model with the general framework provided by eqn (56), and, importantly to reconsider the basis for alternative model structures which accommodate degrees of similarity between alternatives.

5.3 *Some comments on Tversky's Elimination-By-Aspects* (EBA) *Model*

In Tversky's model individuals are considered to move from choice set to choice set eliminating alternatives in the process, according to a particular *transition probability*. Formally the model may be derived from the following *recursive* relation

$$P(A_p|\mathbf{A}) = \sum_{\mathbf{B}} P(\mathbf{B}:\mathbf{A})P(A_p|\mathbf{B}) \tag{57}$$

with the transition probability $P(\mathbf{B}:\mathbf{A})$ given by

$$P(\mathbf{B}:\mathbf{A}) = \frac{V_B}{\sum_{\mathbf{B}'\epsilon Q_A} V_{B'}} \tag{58}$$

in which $\mathbf{B}$ is a non-empty subset of alternatives $\mathbf{A}$; $P(A_p|\mathbf{A})$, $P(A_p|\mathbf{B})$ are the probabilities of selecting alternative A_p from the choice sets $\mathbf{A}$ and $\mathbf{B}$ respectively; while V_B is the scale value of the collection of aspects which are *unique to and common within* members of set B. V_B may be regarded as a measure of the unique advantages of the alternatives in set B; $\sum_{\mathbf{B}'\epsilon Q_A}$ denotes summation over all subsets belonging to the set $\mathbf{A}$. In Tversky's model then thresholds are so defined as to render alternatives satisfactory or not satisfactory with respect to each particular attribute or "aspect". Because the scales V (which are variously referred to by the terms "utility", "weight" or "attractiveness") relate only to those aspects unique and common to particular sets, it is clear from eqn (58) that the relative probability of selecting two alternatives will not depend on the scale value common to them. In this way the model accommodates the similarity between the alternatives and introduces differential substitutability between different members of the set of alternatives.

It is interesting to note Tversky's own comments on the behavioural interpretation of the model

> "The EBA model accounts for choice in terms of a covert elimination process based on sequential selection of aspects. Any such sequence of aspects can be regarded as a particular state of mind which leads to a unique choice. In light of this interpretation, the choice mechanism at any given moment in time is entirely deterministic; the probabilities merely reflect the fact that at different moments in time different states of mind (leading to different choices) may prevail. According to the present theory, choice probability is an increasing function of the values of the relevant aspects. Indeed, the elimination by aspects model is compensatory in nature despite the fact that at any given instant in time, the choice is assumed to follow a conjunctive (or a lexicographic) strategy. Thus the present model is compensatory "globally" with respect to choice probability but not "locally" with regard to any particular state of mind". (Tversky, 1972, p. 296).

As McFadden and others have remarked, the EBA model is also consistent with the behaviour of a population of preference maximisers, each with lexicographic preferences over aspects in which

> "...the transition probabilities can be interpreted as the result of a process in which an individual drawn randomly from the population has a ranking of all aspects of alternatives, and moves serially down the ranking, eliminating alternatives which fail to have the desired aspect, until a single choice remains". (McFadden, 1978, p. 14).

Tversky's remarks on the characterization of the decision process may thus be reinterpreted as follows: non-compensatory behaviour may be taken to characterize each individual in the population π, while that group as a whole, conforming to the EBA model appear to act in a compensatory fashion.

The model is consistent with eqn (56) in which $\mathscr{D}$ represents a set of lexicographical orderings. Once a strategy D^* is drawn, all but one alternative, say A_ρ, will be eliminated requiring that $P(A_\rho|D^*) = \delta_{\rho\rho'}$ if D^* eliminates all alternatives other than $A_{\rho'}$. (Note, that there will in general be more than one strategy which will result in a given alternative being selected.) The probability $P(D^*|\mathscr{D})$ of drawing the ordering D^*, (to be consistent with eqn 58), is related to the scale values $\mathbf{V}$ defined above.

It should be emphasised that in the use of and fitting of the EBA model a knowledge of the internal decision process of a person and an explicit statement of the lexicographical orderings is not presupposed. We have found it useful to discuss the EBA model in this explicit fashion in order to compare other models consistent with eqn (56) to be described below.

In the Tversky model the quantities $\mathbf{V}$ are the unknowns and these must be determined in an application. They may be obtained *directly* from the various probabilities and conditional probabilities which characterize the results of a choice process (see, e.g., Tversky, 1972; Makowski *et al.*, 1978) or parameterised in terms of the values of attributes $\mathbf{Z}_B$ unique and common to the various sets (see McFadden, 1978)

$$V_B = V_B(\mathbf{Z}_B, \boldsymbol{\phi}). \tag{59}$$

Because the intersection of the sets of aspects in the general case will be very complex, the number of parameters to be estimated will proliferate rapidly with the number of alternatives. In certain important cases, however, in which the structure of similarity (interpreted in terms of the commonality between the aspects of alternatives) is relatively simple, the resultant models are of more manageable form.

Consider again the "two dimensional" choice context involving location mode combinations, and the 4-alternative model in particular, with A_ρ, $\rho = 1 \ldots 4$. These will be taken to correspond to $A_{\mu\nu}$, $u = 1,2$; $\nu = 1,2$ as before, with $\mu \epsilon X$ and $\nu \epsilon Y$ referring to the location and mode, respectively. The similarity between the alternatives embodied in the variance-covariance matrix (41) may now be discussed in terms of commonality or overlap in the sets of aspects characterizing each alternative. These sets and the structure of their overlap for the four cases encountered in Fig. 2 are shown in Fig. 4. If we let $V_{\mu\nu}$ denote the scale value for aspects unique to $A_{\mu\nu}$, while $V_{\mu *}$ and $V_{\nu *}$ signify those common to alternatives in dimensions X and Y respectively, the EBA models are readily constructed. We shall write the model for the "cross-correlation" case from which the others may be determined as special cases.

Directly from eqns (57) and (58) we have

$$P(A_{\mu\nu}|\mathbf{A}) = \frac{V_{\mu\nu} + V_{\mu *}P(A_{\mu\nu}|A_{\mu 1}, A_{\mu 2}) + V_{*\nu}P(A_{\mu\nu}|A_{1\nu}, A_{2\nu})}{\sum\limits_{\mu\nu} V_{\mu\nu} + \sum\limits_{\mu} V_{\mu *} + \sum\limits_{\nu} V_{*\nu}} \tag{60}$$

with

$$P(A_{\mu\nu}|A_{\mu 1}, A_{\mu 2}) = \frac{V_{\mu\nu} + V_{*\nu}}{\sum\limits_{\nu} V_{\mu\nu} + \sum\limits_{\nu} V_{*\nu}} \tag{61}$$

and

$$P(A_{\mu\nu}|A_{1\nu}, A_{2\nu}) = \frac{V_{\mu\nu} + V_{\mu *}}{\sum\limits_{\mu} V_{\mu\nu} + \sum\limits_{\mu} V_{\mu *}}. \tag{62}$$

A given alternative, say A_{11} (location 1 and mode 1) may be chosen in three "ways" corresponding to the different lexicographic orders in the set $\mathscr{D}$. Most directly, it will be selected if an aspect unique to A_{11} is considered most important. Secondly, a contribution to P_{11} will come from the selection of an aspect common to the set of alternatives containing A_{11} and A_{12}

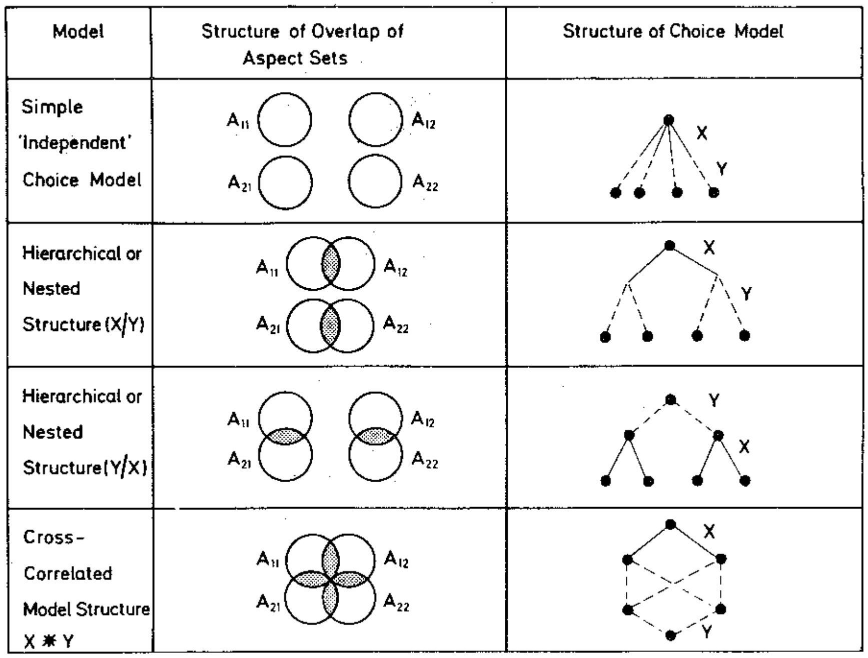


Fig. 4. Pictorial representation of the structure of selected Elimination-by-aspects models for the XY choice context.

with the subsequent elimination of A_{12}. The final contribution will be achieved from the selection of an aspect common and exclusive to the set (A_{11} and A_{12}) with the subsequent elimination of A_{21}. These three contributions correspond to those terms in the numerator of (60). The two "tree like" structures (X/Y and Y/X) and the simple scaleable choice model will be obtained when $V_{\mu *}$ or $V_{*\nu}$ or both are zero, respectively (see Fig. 4).

It appears that there is a great deal of similarity between the *structures* of the corresponding models of Figs. 2 and 4. This will also be reflected in a comparison between their elasticity matrices. Indeed, McFadden (1980) has suggested that the GEV system and EBA models are roughly comparible in flexibility and complexity. He adds however that

> "...one drawback of EBA for econometric applications is that the motivation for the model provides little guidance for parametric specification of the scale functions V" (McFadden, 1980).

Indeed, one of the criticisms raised against Tversky's model is its formulation in terms of 0–1 aspects rather than intervally scaled variables (see, e.g. Gensch and Svestka, 1978). In the next section we shall discuss a class of models which include elimination criteria based on the latter.

6. TESTS ON COMPENSATORY, NON-COMPENSATORY AND HYBRID DECISION MODELS

6.1. *Introduction*

In this section we consider the formulation of models, characterized by eqn (56), which involve an explicit specification of the decision rules D^* and parametrization of the probability distributions determining their frequency of selection. Having generated choice models by accumulating contributions to $\{P_\rho^*\}$ from all different strategies belonging to $\mathscr{D}$ we perform a series of experiments and estimate the parameters of a test model from the simulated data in the usual way. We shall discuss and present the results of tests using two different types of model based on distinct characterizations of $\mathscr{D}$, $P(D^*|\mathscr{D})$ and $P(A_\rho|D^*)$. In the first series of tests (T_1) the contributions to P_ρ^* are decomposed in terms of the numbers of attributes considered in the choice process. The resultant model will then be characterized by specific distributions of the sets of attributes considered by individuals. In the second series (T_2) we directly apply the explicit decomposition (56) with a set of lexicographical (rank order) rules, and a probabilistic choice model $P(A_\rho|D^*)$ involving distributed thresholds which are used to eliminate unsatisfactory options. In this process we draw on elements of the models discussed recently by Recker and Golob (1979) and Gensch and Svetska (1978).

Let us recapitulate on the motivation for the experiments—what is it we are looking for? Basically we wish to discover whether any variants in the decision process which give rise to observed patterns of dispersion will result in serious response errors when that variability is interpreted and fitted by the LPLA multinomial logit model, deemed to be underpinned by "trade-off" behaviour. The essential features we wish to bring out are the interpretational and numerical consequences of estimating the parameters of a particular choice model with data underpinned by the notion of individuals who might formulate their decision process in terms of *priorities*, who might scrutinize a limited number of attributes associated with alternatives, and generally fail to make a "global" assessment of alternatives. As we pointed out above this does not imply that they behave irrationally within their own frame of reference. In approaching these issues it is essential to bear in mind the comments of the last section, namely that in general it is the *composite* of different decision processes which will be responsible for the observed dispersion, and the discontinuities and sequences which one might attribute to any "individual" may well become "smoothed" on aggregation over members of the population π.

Let us reconsider two extremes of the spectrum of possible decision processes:

(i) an individual makes a decision on the basis of a complete set of M attributes in the traditionally discussed trade-off fashion;

(ii) the decision is made on the basis of what is considered the most important attribute (strong lexicographic behaviour).

Between these two extremes there is a range of possibilities derived from rules consistent with the active consideration of $1 \leq m \leq M$ attributes. In the series of tests to be described in this section we shall continuously span this range of possibilities using alternative models of the decision process.

There are many ways in which a model can be decomposed and the contributions from various sources explicitly recognized. The usefulness of a particular decomposition will depend on the problem or hypothesis at hand. The above considerations suggest that a decomposition in terms of the distribution over the population π of the sets of attributes actively considered in the choice process, is a useful starting point in the formulation of a model, as this feature is one of the *outcomes* of "non-compensatory" behaviour.

6.2 *Models embodying Distributed Attribute Sets* (DAS).

We shall define the following: $\mathscr{S}$ is the set of all non-empty sets of attributes derived from the M attributes $\{Z^1, \ldots Z^\mu, \ldots Z^M\}$; $P(S|\mathscr{S})$ is the probability that a particular set of attributes S from $\mathscr{S}$ is used to characterize a choice process; $P(A_\rho|S)$ is the probability that an alternative A_ρ is selected on the basis of a decision rule which involves various attributes belonging to the set S. The probability P_ρ can now be synthesized from the following contributions

$$P_\rho = \sum_{S \in \mathscr{S}} P(A_\rho|S)\, P(S|\mathscr{S}) \tag{63}$$

in which $\sum_{S\in\mathscr{S}}$ denotes summation over all contributions to P_ρ from all sets belonging to $\mathscr{S}$.

The set $\mathscr{S}$ contains elements—which are sets of attributes—which may be ordered according to the number of attributes in each set. That is $\mathscr{S}$ contains the non-empty sub-sets:

$S_1 \colon \{Z^1\}, \ldots \{Z^\mu\}, \ldots \{Z^M\}$

.

.

$S_m \colon \{Z^1, Z^2, \ldots, Z^m\}, \ldots \{Z^{M-m+1}, \ldots Z^{M-1}, Z^M\}$

.

.

$S_M \colon \{Z^1, Z^2 \ldots Z^M\}$,

in which S_m, $m = 1, \ldots, M$ is the sub-set containing m-membered set of attributes. There are clearly ${}^MC_m$ possibilities of selecting m attributes from M and this is therefore the number of members in set S_m. The total number of non-empty sets in $\mathscr{S}$ is then simply $\sum_{m=1}^{M} {}^MC_m = 2^M - 1$, and these range as indicated above between the M sub-sets containing 1 attribute and the single set with M attributes. This decomposition is clarified in Fig. 5(a). We shall in the following interpret $P(S_m|\mathscr{S})$ as the proportion of individuals in π who select an alternative having considered m attributes in the decision process.

The decomposition (63) may now be further organized in terms of the various sets which are contained in the M sets $\{S_1, \ldots S_m, \ldots S_M\}$, as follows

$$P_\rho = \sum_{m=1}^{M} \sum_{S \epsilon S_m} P(A_\rho|S)\, P(S|S_m)\, P(S_m|\mathscr{S}) \tag{64}$$

in which $P(S|S_m)$ is the probability of drawing a *particular* m-membered set S from the set S_m; and $\sum_{S\epsilon S_m}$ denotes summation over contributions from all sets S which contain m members.

It is clear that the usual trade-off model involving M attributes can be taken to correspond to the special case

$$\begin{aligned} P(S_m|\mathscr{S}) &= 1 \quad \text{if } m = M \\ &= 0 \quad \text{otherwise,} \end{aligned} \tag{65†}$$

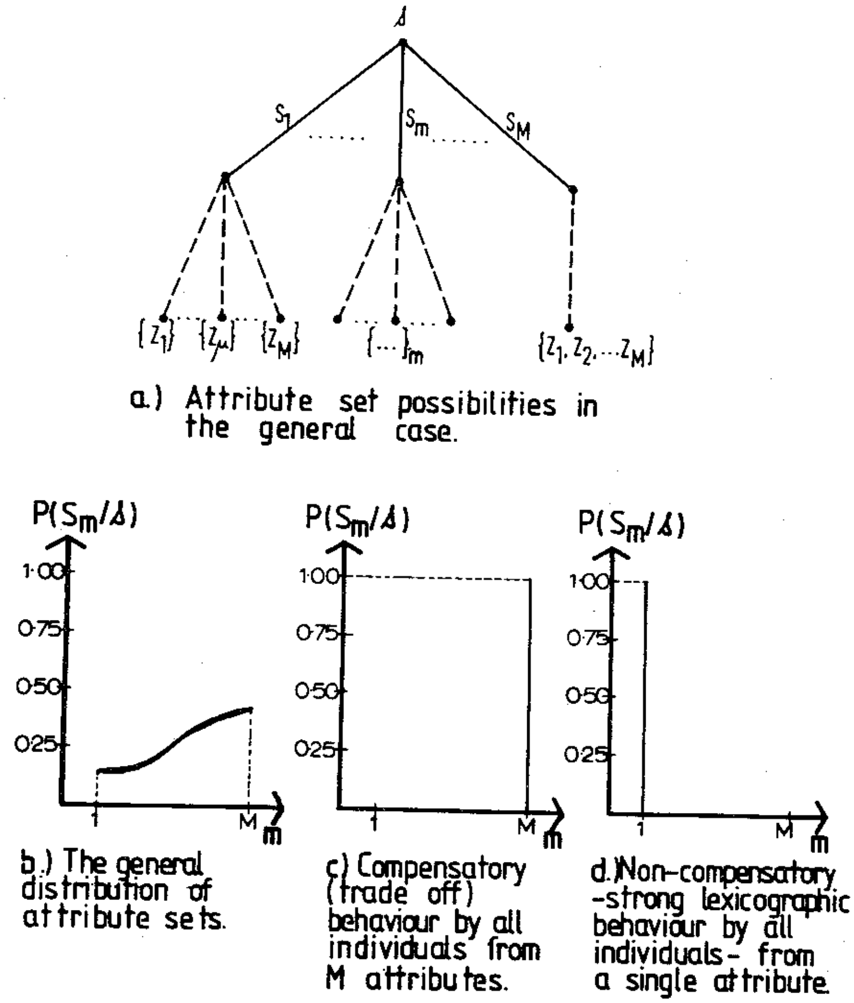


Fig. 5. Distribution of attribute set sizes in the general and special cases.

†It may be of course that this case is consistent with other decision rules.

and, at the other extreme, the strong lexicographic rule corresponds to

$$\begin{aligned} P(S_m|\mathscr{S}) &= 1 \qquad \text{if } m = 1 \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{66}$$

The spectra appropriate to these cases, together with the general distribution of attribute set sizes are shown in Figs. 5(b)–(d). (We have portrayed the *discrete* distribution $P(S_m|\mathscr{S})$ as a continuous curve in Fig. 5b.)

It should be emphasized that the decomposition (64) does not constitute a model until particular assumptions and specifications are associated with the various contributions $P(A_\rho/S)$, $P(S|S_m)$ and $P(S_m|\mathscr{S})$. The expression (64) is consistent with a continuum of decision processes and as we pointed out above, it may in fact not be convenient to decompose the quantity P_ρ in these terms. For our purposes such a decomposition is useful and we shall parameterize the distribution of attribute set sizes $P(S_m|\mathscr{S})$ in such a way that it is possible to span a range of possibilities between the two above extremes (65) and (66).

It is clear that the dispersion in the resultant data set {P*} will in general arise from the three sources associated with the distributions $P(S_m|\mathscr{S})$, $P(S|S_m)$ and $P(A_\rho|S)$. That is, variation in behaviour over members of π will depend on the different propensities to make decisions involving various sets of attributes, and on the distribution of the decision rule over those "individuals" who scrutinize the attributes in the particular set S.

In order to generate a model we need then to specify $P(S_m|\mathscr{S})$, $P(S|/S_m)$ and $P(A_\rho|S)$. Let us consider first the choice model $P(A_\rho|S)$ which is the probability of selecting A_ρ from the set **A** on the basis of a decision rule consistent with an examination of the m attributes in the set S, which is a member of S_m. To provide $P(A_\rho|S)$ with a behavioural basis we adopt a decision rule involving the m attributes. One possibility, and the one taken here, is to adopt a linear trade-off rule with the restricted set of attributes in S. In this way we shall attempt to simulate "non-global" or "satisficing" behaviour by introducing compensatory behaviour within the context of a *distribution* of attribute sets.

Within this formulation we may construct a decision rule in which an alternative $A_\rho \epsilon \mathbf{A}$ is selected on the basis of maximum utility, according to:

$$\sum_{\substack{\{Z^\mu\}\in S \\ S\in S_m}} \phi_\mu Z_\rho^\mu \geqslant \sum_{\substack{\{Z^\mu\}\in S \\ S\in S_m}} \phi_\mu Z_{\rho'}^\mu \quad \forall A_{\rho'} \in \mathbf{A}. \tag{67}$$

Here $\sum_{\{Z^\mu\}\epsilon S}$ denotes summation over the particular attributes in the m-membered set S.

We must now confront the issue of dispersion in these utility functions, and in particular whether all or some of the m attributes in the set are considered observable or not. If we considered all M attributes $\{Z^1 \dots Z^M\}$ to be the only ones relevant to the choice process and these were observable and measureable, then distributed utilities will result if the vector of parameters ϕ is distributed over the population π. We can express this "taste" variation in the usual way by writing each component ϕ_μ in the form of a representative (mean value) $\bar{\phi}_\mu$, common to all sets, and a random component η_μ. The utility function associated with the m-membered set can now be written

$$U_\rho^S = \sum_{\substack{\{Z^\mu\}\in S \\ S\in S_m}} \bar{\phi}_\mu Z_\rho^\mu + \sum_{\substack{\{Z^\mu\}\in S \\ S\in S_m}} \eta_\mu Z_\rho^\mu. \tag{68}$$

It is clear from (68) that the variance in the utility values will increase with the number of attributes in the set S. When the distributions of the various residuals are independent we have

$$\operatorname{Var}\{U_\rho^S\} = \sum_{\substack{\{Z^\mu\}\in S \\ S\in S_m}} \{Z_\rho^\mu\}^2 \operatorname{Var}(\eta_\mu). \tag{69}$$

A consequence of this specification would be that an increased propensity for individuals in π to make decisions on the basis of smaller attribute sets results in some of the dispersion from the utility maximization being "squeezed out". Under this formulation the two extremes shown in Figs. 5(c) and 5(d) correspond respectively to the maximum and minimum contributions to dispersion arising from variation accompanying the linear compensation rule.

From an experimental viewpoint it is an attractive feature to include the above model in a mis-specification test. However apart from any response error which might arise from the distribution of attribute sets we should have to contend with the additional complication of the attribute dependence of the residual in eqn (68) when applying a multinomial logit which is free from such dependence. Our initial tests confirmed the existence of mis-specification due to taste variation which has been thoroughly examined by Hausman and Wise (1978), Horowitz (1979), Cardell and Reddy (1977), among others. Instead of adopting a random parameter model for the model to be estimated (which was not available to us), and because of the desirability of constructing the experiments in such a way that data generated in the extreme case (65) would result in zero (or negligible) response error when a multinomial logit model is adopted by an observer, we resorted to the following compromise, in which the utility function (68) was expressed as

$$U_p^S = \sum_{\substack{\{Z^\mu\}\in S \\ S\in S_m}} \bar{\phi}_\mu Z_p^\mu + \sum_{\substack{\{Z^\mu\}\in S \\ S\in S_m}} \eta_\mu. \tag{70}$$

This retained the desired effect of the reduction in variability as the set size decreases. Further, if the distributions of the residuals are bell-shaped the model $P(A_p|S)$ resulting from this specification will be very well approximated by a multinomial logit function, as we confirmed in the tests with Weibull distributions. In the limit when all members of π are associated with the complete set S_M, a negligible response error will thus arise.†

It remains to specify the distributions $P(S_m|\mathscr{S})$ and $P(S|S_m)$. A particular parameterization which allows a continuous variation of the attribute set size distribution between two extremes $m = 1$ and $m = M$ is a binomial distribution truncated at $m = 1$ and appropriately normalized. That is $P(S_m|\mathscr{S})$ is taken as

$$\begin{aligned} P(S_m|\mathscr{S}) &= B(m|M, q) \\ &= \frac{{}^MC_m\, q^m(1-q)^{M-m}}{1-(1-q)^M} \end{aligned} \tag{71}$$

in which q, the binomial parameter, varies in the range $0 \leq q \leq 1$. The mean attribute set size is given by

$$\bar{\eta}_q = \frac{Mq}{1-(1-q)^M}, \tag{72}$$

and it can be seen from (71) and (72) that the special cases referred to in eqns (65) and (66) correspond to $q = 1$ and $q = 0$ respectively. In the former case the complete set of attributes is examined by all members of π, and in the latter only one-membered sets are involved in the decision process. As q decreases from unity we include a consideration of smaller attribute sets. sets.

We need finally to specify the probability $P(S|S_m)$ of selecting a *particular* set $(\ldots.)_m$ containing m members from all the ${}^MC_m$ possible m-membered sets in S_m. Our strategy here is to perform sensitivity analyses in two series of tests. In the first we shall assume that all the ${}^MC_m$ sets are equiprobable and that there is no propensity to prefer one attribute over another in the generation of attribute sets. As it might reasonably be argued that because individual attributes are perceived with different importance in the process of selection between alternatives in the choice model $P(A_p|S)$, we include a second series which involves *additional* biases in the consideration of attributes, leading to differential probabilities of selecting the various attribute sets in S_m.

†If the random component in eqn (70) is associated also with unobserved factors, then the set $\{Z^1, \ldots Z^M\}$ corresponds only to the observable attributes in the usual way.

We can now summarize the strategy in the two series of tests on the first model system described in this section which we shall refer to as a Distributed Attribute Set (DAS) model, as follows:

Λ^*: compensatory decision making involving attribute sets distributed over the population π.	Λ: LPLA multinomial logit model (compensatory decision making from a single full set of attributes).

We now describe the experiments and their results.

6.3 *Experiments with Distributed Attribute Set (DAS) Models*

It is our intention to estimate the parameters $\boldsymbol{\theta}$ of a multinomial logit model

$$P_\rho = \frac{\exp\{\sum_{\mu=1}^{M} \theta_\mu Z_\rho^\mu\}}{\sum_{\rho} \exp\{\sum_{\mu=1}^{M} \theta_\mu Z_\rho^\mu\}} \tag{73}$$

with data derived from the model (64). The characteristics of this data set and the results of the tests themselves will be a function of the parameters of the three distributions which are summarized here: $P(S_m|\mathscr{S})$: q—the binomial parameter; $P(S|S_m)$: $\boldsymbol{\alpha}$—the vector of relative weights associated with membership by each attribute of a set of attributes (see below); and $P(A_\rho|S)$: $\bar{\boldsymbol{\phi}}$, $\boldsymbol{\sigma}$ the parameters associated with the mean utility and distribution (standard deviation) of the independent residuals (see eqn 70). We may thus write the probability P_ρ^* as follows

$$P_\rho^* = f(\mathbf{Z};\, q,\, \boldsymbol{\alpha},\, \bar{\boldsymbol{\phi}},\, \boldsymbol{\sigma}) \tag{74}$$

and it is necessary to determine the salient characteristics of the response error as a function of these several parameters. Our strategy in all tests has been to retain the parameter set $\bar{\Phi}$ at a fixed value (which corresponded to realistic values derived from a mode choice context) and assess the consequence of varying q, $\boldsymbol{\alpha}$ and σ. We are interested in and will present results for the following aspects: the "goodness of statistical fit" accompanying the use of the model (73); the variation of the estimated logit parameters $\boldsymbol{\theta}$ with q and $\boldsymbol{\alpha}$, and the response error (which will be taken here as the difference between estimated and simulated elasticities) as a function of the parameters, q, $\boldsymbol{\alpha}$ and σ.

A mode choice context is used for the experiments in which the particular observed $\mathbf{Z}$ values correspond to 3 attributes (in-vehicle time; cost/income; and out-of-vehicle time) for each of three modes (car driver, car pool and bus).† At each data point, corresponding to a *particular* combination of attribute values $\mathbf{Z}$, the probability $P_\rho^*(\mathbf{Z})$ $\rho = 1, 2, 3$ of selecting the three alternatives is generated by Monte Carlo simulation. That is we take a sample π of "individuals" of a given size, and for each member perform the following operations:

(i) generate an attribute set size S_m from the discrete distribution (71);
(ii) select a particular attribute set $S \epsilon S_m$ by sampling from the discrete distribution $P(S|S_m)$;
(iii) sample values of residuals $\boldsymbol{\eta}$ from equal variance Weibull distributions associated with each attribute belonging to the set S,
(iv) allocate to the "chosen" alternative according to eqn (70).

The accumulated proportions of the population associated with each alternative will, as we mentioned in the Section 3, approximate well the probability distribution $\{P_1^*(\mathbf{Z}), P_2^*(\mathbf{Z}), P_3^*(\mathbf{Z})\}$ when the size of the population π is sufficiently large. Sample sizes varying from 500 to 25000 were used to assess the appropriate size. Note again that *in general* our generated data is in the form of fractional quantities unlike the 1, 0 information collected in actual studies.

†The variation of $\mathbf{Z}$ was derived from observations drawn from a sample collected in Washington D.C. We are grateful to S.R. Lerman for providing us with this information.

This procedure was repeated at a sufficient number of data points with different **Z** values to ensure that the estimated coefficients of the logit model (73) had fully converged. The model estimated in these tests and used for forecasting response contained a utility function of the following form

$$\bar{U}_p = \theta_1 \begin{pmatrix}\text{in-vehicle} \\ \text{time}\end{pmatrix} + \theta_2(\text{cost/income}) + \theta_3\begin{pmatrix}\text{out-of-} \\ \text{vehicle time}\end{pmatrix} + \theta_4 + \theta_5 \qquad (75)$$

in which θ_4 and θ_5 are car driver and bus specific constants respectively. The parameters $\theta_1, \dots \theta_5$ were estimated by the Berkson–Theil method (McFadden, 1976) and the goodness-of-fit expressed as the coefficient of determination (R^2) associated with the corresponding regression equations.

6.3.1 *The case of no bias in attribute set membership.* In this series of tests we stipulated that the probability of selecting an attribute set is a function of its size alone, and the q-variation was then examined. The results of this series are summarized in Fig. 6. The goodness of fit of the multinomial logit model (73) with the utility function (75) was impressive over the whole q-range ($0 \leq q \leq 1$) and only decreased from $R^2 = 0.996$ at $q = 1$ to $R^2 = 0.910$ at $q = 0.01$. At all tested points the parameters $\boldsymbol{\theta}$ were significant at the 95% level and in no case was a mode specific constant significant at the 90% level. The parameters θ_1, θ_2 and θ_3 all decreased in a very similar manner as increasingly smaller attribute sets were introduced, and as a consequence the "values" of in-vehicle and out-of-vehicle times θ_1/θ_2 and θ_3/θ_2 changed very little over the q-range. These characteristics are shown in Figs. 6(b) and (c).

It can be seen from Fig. 6(a) that the response error, measured by the aggregate direct elasticity, and derived by reducing the cost on the car pool mode, increases as q is reduced and achieves a maximum at $q = 0$. Although this deviation between the estimated and simulated elasticities is very significant in proportional terms at $q = 0$ and corresponds to a 40% error, it is debatable whether the resultant *absolute* discrepancy in the *aggregate* shares would be considered of *practical significance*.

Because of the particular assumptions used in eqn (70) to formulate the model $P(A_p|S)$ we have not retrieved in this range of tests the strong lexicographic limit characterized by *zero* dispersion in the choice model. There remains a residual dispersion in the utility function at $q = 0$. In order to "squeeze out" this residual variation in behaviour the vector of standard deviations $\boldsymbol{\sigma}$ of the residuals η was reduced in stages to zero, and the above statistics re-estimated. The error was found to increase, but importantly the statistical fit of the estimated multinomial logit model rapidly deteriorated with the parameters θ_1, θ_2 and θ_3 becoming increasingly badly specified. In fact, before the strong lexicographic limit (with zero dispersion) was achieved the logit model had been rejected.† In other words it was not acceptable to fit a LPLA logit model to data generated in this limit.

6.3.2 *Experiments with heterogeneously distributed attribute sets.* It is important to consider the consequences of relaxing the assumption

$$P(S|S_m) = \frac{1}{{}^{M}C_m} \qquad \forall S \epsilon S_m \qquad (76)$$

adopted in the above tests, and enquire whether the response error is a function of any bias in the formation of attribute sets scrutinized by individual decision makers. In order to examine this issue we stipulated that the probabilities that each member of the M attributes Z^1, $Z^2 \dots . Z^M$ will enter a given set are in proportion $\alpha_1, \alpha_2 : \dots : \alpha_M$, with $\sum_{\mu} \alpha_{\mu} = 1$. It is now straightforward to compute the relative frequency of the ${}^{M}C_m$ different sets in S_m—the above tests being characterized by $\boldsymbol{\alpha} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

The effect of variation with $\boldsymbol{\alpha}$ will be considered in two ways in order to bring out the variation of response error with q and $\boldsymbol{\alpha}$. Firstly, the above tests are repeated for variation with q with the vector ($\alpha_1 = 0.165$, $\alpha_2 = 0.4$, $\alpha_3 = 0.435$). This particular set was suggested by the *relative* weights of the attributes in the utility function (68). In other words, the same set of

†As σ was reduced one or more of the coefficients acquired an unacceptable sign.

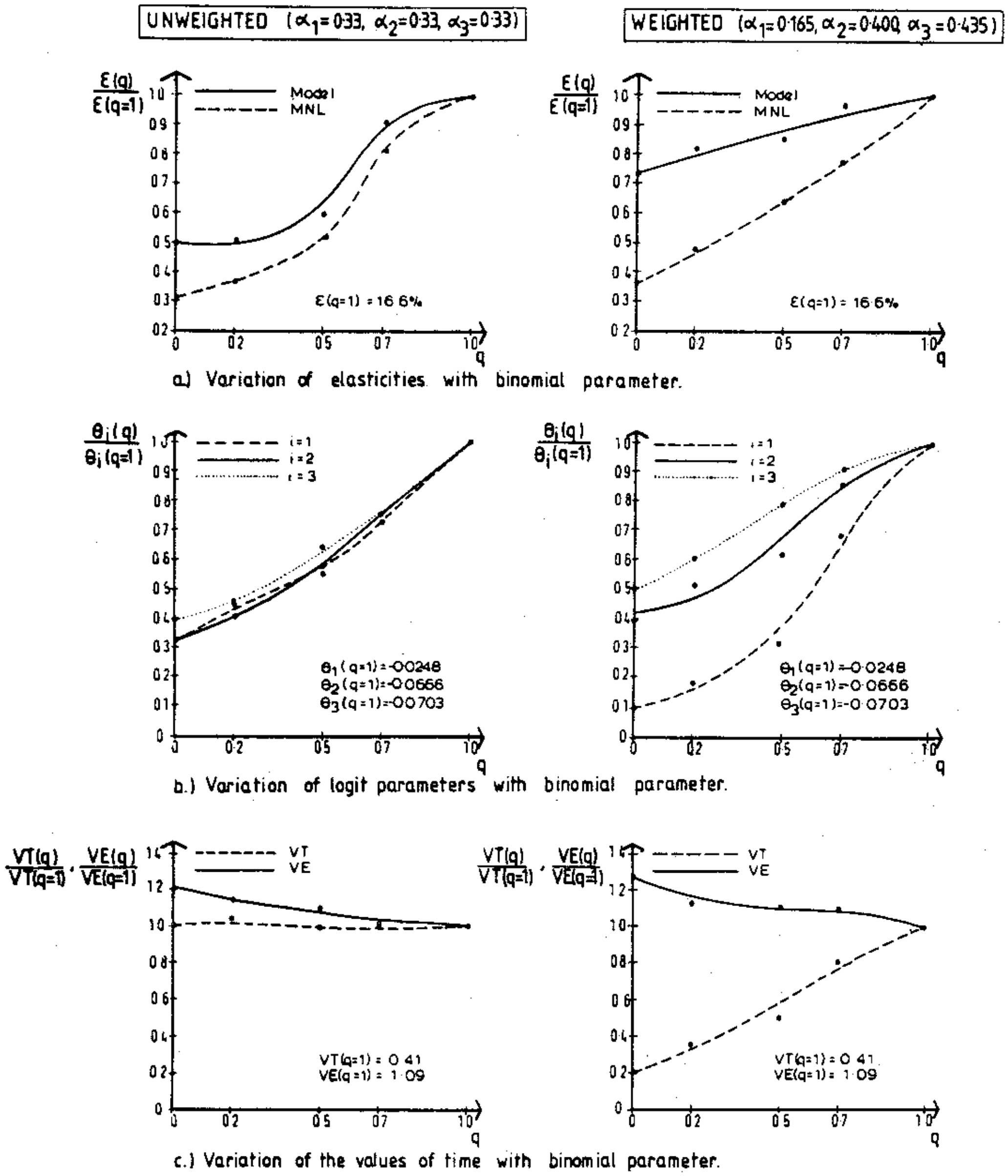


Fig. 6. Characteristics of simulation experiments with Distributed Attribute Set (DAS) model.

parameters is used to determine both the probability that a particular attribute will be considered, and also the relative "weights" or "values" accorded to the variables in the linear utility function. In a further series of tests a variety of vectors α were selected and the model statistics and response error determined at a particular distribution of attribute set sizes ($q = 0.5$).

The results of the series with ($\alpha_1 = 0.165$, $\alpha_2 = 0.4$, $\alpha_3 = 0.435$) are summarized in Fig. 6 and can be compared directly with the values for $\alpha = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. The multinomial logit model could again be very satisfactorily estimated at all values of q with R^2 values ranging from 0.996 at $q = 1$ to 0.956 at $q = 0.01$. The parameters θ_1, θ_2 and θ_3 were again significant at the 95% level with expected signs, and the mode-specific constants not significant at the 90% level.

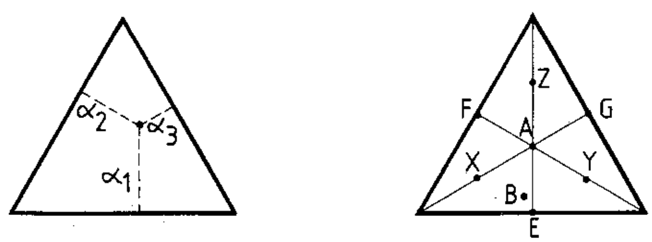
It can be seen from Figs 6(b) and (c) that the q-variation now induces significantly different behaviour in the ratios $\theta_i(q)/\theta_i(q = 1)$ with the result that the *measured* value of in-vehicle time decreases substantially over the q-range. A comparison of the Figs. in 6(a) reveals that the imposition of attribute set entry bias also has a significant influence on the response error which increases, relative to the homogeneous case, over the whole range. The maximum error is again recorded at $q = 0$ where the estimated direct elasticity of $\sim$12% substantially over-estimates the simulated value of $\sim$6%.

Having established that the effect of bias in the attribute sets has an influence on the response error at varying q-values, it was considered desirable to test for error over a wider range of $\boldsymbol{\alpha}$ values for a given value of q, taken as 0.5. The various $\boldsymbol{\alpha}$ test points are illustrated with reference to the triangle shown in Fig. 7, and the specific co-ordinates α_1, α_2, α_3 are also given in the figure.†

In all cases well specified logit models with highly significant values for θ_1, θ_2 and θ_3, and impressive R^2-values were obtained. Again the mode specific constants did not prove significant. The response error is very variable over the nine test points and ranges from zero at point X to 25% at point G. There is little doubt that this variability will increase for smaller q-values as we allow the bias to impress itself on a range of smaller attribute set sizes.

One of the interesting features of this experiment was the relationship between the estimated parameters $\boldsymbol{\theta}$, the input vector $\bar{\boldsymbol{\phi}}$ in the utility function (68) and the co-ordinates in the $\boldsymbol{\alpha}$-vector. If we form the ratios $\theta_i/(\phi_i)$, $i = 1, 2, 3$ and express these quantities as relative proportions ζ_i, $i = 1, 2, 3$, then it may be noted from the table in Fig. 7 that these quantities are ranked according to α_1: α_2: α_3. This is particularly apparent at test points E, F and G (where one attribute is omitted from consideration in the generation of data) where $\boldsymbol{\zeta}$ is approximately equal to $\boldsymbol{\alpha}$,

6.3.3 *A summary of the tests on the DAS models.* We can summarize the results of all the above tests as follows: the effect of introducing smaller attribute sets and concommitantly reducing variability associated with a distribution of utility values introduces a response error which varies both with the distribution of attribute set sizes and with the bias in the formulation of attribute sets—the maximum error being recorded when sets contain a single attribute.‡ The actual error obtained by varying the choice set bias is rather variable, but with the exception of two test points, and in the context of the whole series of tests, we found that the multinomial logit model *underestimated* the direct elasticities. This is a consequence of the relative weight α_2 given to the attribute Z^2 in the tests.



Coordinate points of tests.

	A			E			F			G			X			Y			Z		
Coordinates α	0·33	0·33	0·33	0·0	0·5	0·5	0·5	0·0	0·5	0·5	0·5	0·0	0·165	0·165	0·67	0·165	0·67	0·165	0·67	0·165	0·165
R^2	0·980			0·997			0·996			0·993			0·997			0·990			0·972		
Elasticity of model %	9·9			15·4			0·0			20·0			7·1			18·0			5·4		
Elasticity of MNL %	8·6			13·2			2·6			14·7			7·1			15·6			6·7		
ζ_i	0·31	0·33	0·36	0·06	0·46	0·48	0·50	0·06	0·44	0·53	0·43	0·04	0·28	0·21	0·51	0·26	0·50	0·24	0·53	0·21	0·26

Fig. 7. Characteristics of simulation experiments with the DAS model for selected attribute weightings ($q = 0.5$).

†In this case the position in the triangle is given for illustrative purposes only because the sum of the co-ordinates themselves, and not the sum of their squares, is normalised.

‡We have presented the response error in terms of the *direct* elasticities corresponding to a change in the value of a particular attribute (cost of the car pool mode). It was found that the results for cross elasticities produced few comments worthy of note.

With the above model we have attempted to simulate the effects of a "non-global" assessment of the attributes of alternatives—which is sometimes associated with "satisficing" behaviour—by characterizing the outcome of the decision process in terms of a distribution of attribute sets. The particular *behavioural* mechanism by which attributes are *excluded* is in this model latent and underpins the set size distribution. We considered it appropriate to examine the effects of including a specific "elimination mechanism" accompanying what is generally regarded as *non-compensatory* behaviour at the micro-level. It is to the formulation of this model and the corresponding experimental tests that we now turn.

6.4 *A joint compensatory-non-compensatory (hybrid) model*

Let us return to the decomposition of the choice model expressed by eqn (56) which we repeat here for convenience

$$P_p = \sum_{D^* \in \mathscr{D}} P(A_p|D^*)\, P(D^*|\mathscr{D}).$$

In Tversky's model which is underpinned by lexicographic or attribute rank order rules D^*, the probabilistic choice model $P(A_p|D^*)$ attains a value of either zero or one.

In this section we consider the formulation of a model which includes sources of behavioural variation arising from both $P(D^*|\mathscr{D})$ and $P(A_p|D^*)$. In the former we shall consider $\mathscr{D}$ to be composed of a fixed distribution of rank order rules, while the latter will be based on a decision mechanism belonging to the class of weak lexicographical processes (see the discussion by Foerster, 1979) and involves elements of "compensatory" and "non-compensatory" behaviour. It embodies a *critical tolerance principle* which relates to the psychological concept of *just noticeable differences.* Alternatives will be eliminated from further consideration in a decision process according to the perceived difference between attribute values in relation to *individually* defined thresholds or critical tolerances.

The general behaviour underpinning the formulation of the choice model is as follows:

(i) Individuals rank attributes in order of importance. For individual i this rank order will be denoted R_i.

(ii) The available alternatives are scrutinized with reference to the values of the attribute which is first in the list R_i, and alternatives are eliminated if a threshold constraint for that attribute is violated.

(iii) This process is repeated for the second and subsequent entries in R_i until either a single alternative remains or the set of attributes upon which the alternatives are assessed is exhausted—with more than one alternative remaining.

(iv) If the latter case results, a decision is made between the *remaining* alternatives on the basis of a compensatory utility maximising rule.

The rationale for this process is simply that a failure to discriminate easily between alternatives on the basis of thresholds is assumed to encourage a closer joint scrutiny of the several relevant characteristics in a compensatory fashion. Put another way, an individual will be considered to engage in "compensatory" behaviour unless particular alternatives notably distinguish themselves as inferior in terms of their individual attributes.

It is now necessary to specify the elimination process in detail. To do this we draw on elements of the Elimination-by-Aspects models recently discussed by Recker and Golob (1979) and by Gensch and Svestka (1978), in which the criterion for acceptance or rejection of a particular alternative *with respect to a given attribute,* is based on an individual's perception of the *best available alternative at any particular stage of the search.* We define the following quantities: $\mu^i(r)$ is the attribute associated with the rth rank of the order R_i for individual i; $\mathbf{A}(i, r)$ is the set of alternatives still under consideration by individual i when the rth attribute is considered; $T_{pi}^{\mu(r)}$ is the percentage difference between the individual evaluation of the alternative judged best with respect to attribute $\mu(r)$ and the evaluation of alternative A_p on the attribute $\mu(r)$, and $\mathscr{T}_\mu{}^i$ is the critical tolerance between the evaluation of any alternative on the attribute μ and an acceptable standard. This quantity will be considered distributed over the population π with probability $P(\bar{\tau}_\mu; \sigma_\mu)$. $\bar{\tau}_\mu$ is the mean, and σ_μ the standard deviation of this distribution.

We now define the elimination process in terms of the quantities $T_{pi}^{\mu(r)}$ determined as follows:

$$T = {}_{pi}^{\mu(r)} = \frac{\left| \underset{A_{p'}\in \mathbf{A}(i,r)}{\text{opt}} \{Z_{p'}^{\mu}\} - Z_p^{\mu} \right|}{\underset{A_{p'}\in \mathbf{A}(i,r)}{\text{opt}} \{Z_{p'}^{\mu}\}} \tag{77}$$

in which Opt denotes the maximum value associated with attribute μ over the set of alternatives $\mathbf{A}(i, r)$. Here the value is interpreted in terms of the measured quantities Z, and Opt is the appropriate maximization or minimization (e.g. maximum {comfort} or minimum {cost}, etc.). The elimination criterion is now simply defined as follows:

$$\begin{aligned} T_{pi}^{\mu} < \mathcal{T}_{\mu}^{i}: &\quad \text{Accept } A_p \\ T_{pi}^{\mu} > \mathcal{T}_{\mu}^{i}: &\quad \text{Reject } A_p. \end{aligned} \tag{78}$$

We summarize this aspect of the model as follows: at any stage of the search process determined by the rank of the attribute currently under scrutiny, a set of alternatives $\mathbf{A}(i, r)$ will be available to an individual i. For each alternative A_p in $\mathbf{A}(i, r)$ the deviation of the relevant attribute Z_p^{μ} from the best value *in this set* is expressed as a proportion and compared with the critical value $\mathcal{T}_{\mu}^{i}$ which is similarly expressed as a proportion or percentage. The alternatives remaining in the choice set are then examined with reference to the next attribute— and so on.

If when the list R_i is exhausted two or more alternatives remain in the set $A(i, r_M)$ then a compensatory rule is adopted in the usual way and expressed through the LPLA multinomial logit function in terms of the observable attributes $\mathbf{Z}$

$$P_p^i = \frac{\exp\left\{\sum_{\mu=1}^{M} \phi_\mu Z_p^\mu\right\}}{\sum_{A_p\in \mathbf{A}(i,r_M)} \exp\left\{\sum_{\mu=1}^{M} \phi_\mu Z_{p'}^\mu\right\}}, \qquad \forall A_p \in \mathbf{A}(i, r_M). \tag{79}$$

For simulation purposes in the context of the model as a whole the probabilistic choice model $P(A_p|D^*)$ was operationalized by sampling from independent Weibull variates.

In order to specify the distribution $P(D^*|\mathcal{D})$ of rank orders $D^* \equiv R$ over the set of all rank orders $\mathcal{D} \equiv \mathcal{R}$ we note that the number of possible rank orders derived from M attributes is $M!$ We shall consider the generation of individual orders through the sequential sampling of attributers from the set $\{Z^1, Z^2, \ldots Z^M\}$ according to the probabilities $\{\alpha_1: \alpha_2: \ldots \alpha_M\}$. After each attribute is "removed" from the set and placed in rank, the selection probabilities for the remaining attributes are appropriately normalized. For small numbers of attributes the distribution $P(R|\mathcal{R})$ may be written down in a straightforward manner and rank orders directly sampled in the usual way.

We can now express the composite model P_p as a function of the parameters of the above distributions. That is

$$P_p = f_p(\mathbf{Z}|\bar{\tau}, \boldsymbol{\sigma}, \boldsymbol{\phi}, \boldsymbol{\alpha}). \tag{80}$$

Consider for a moment the variation of P_p with the means $\bar{\tau}_\mu = 1 \ldots . M$ of the distributions of critical tolerance for fixed ϕ and α. If these components of the vector τ are large then the number of alternatives eliminated in the sequential analysis of attributes will be very small, and in the limit as $\bar{\tau} \to \infty$ all alternatives will be assessed in a *compensatory* manner. No response error will then be obtained if a multinomial logit model is used for prediction. On the other hand, in the limit as these components tend to zero only the "best" alternative defined with respect to the highest ranking attribute in the list R_i will be retained, and *strong lexicographical* behaviour results. If we set all the mean values equal to T we can, as in the above series of tests, span a range of behaviour between these two extemes.† As an *outcome* of this process will be a distribution over the population of the sets of attributes considered before a selection is made the model is clearly consistent with the general decomposition expressed in eqn (63).

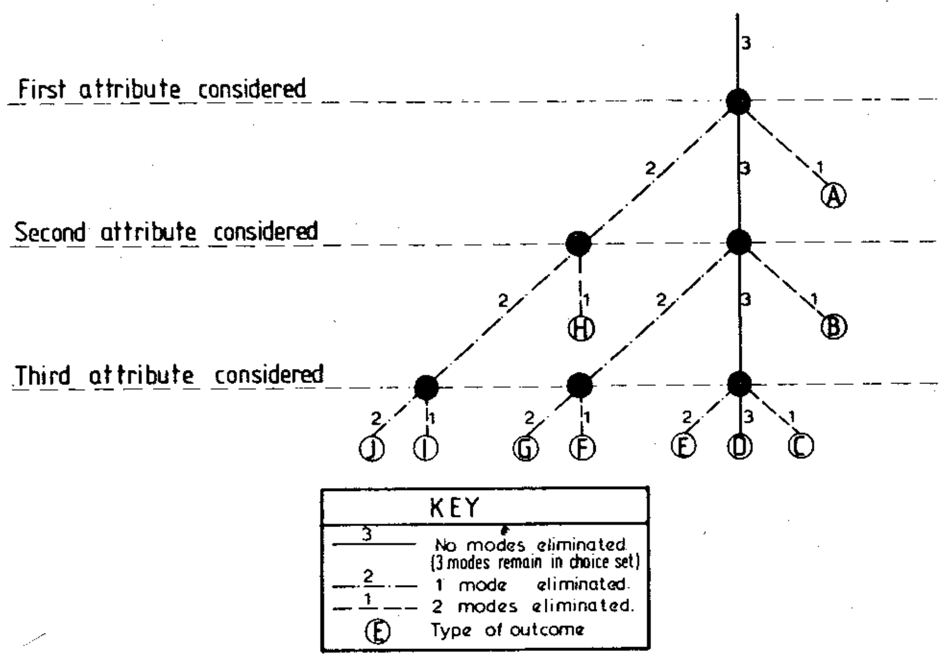
We now proceed to a description of tests on what we shall refer to as the "Hybrid Model".

6.5 Simulation tests with the Hybrid Model

The context and data used for this series of tests were those adopted in the experiments with the DAS model. In the three attribute cases the set $\mathcal{R}$ consists of the 6 rank orders $R_1 \ldots R_6$

$$\begin{Bmatrix} Z^1 \\ Z^2 \\ Z^3 \end{Bmatrix} \begin{Bmatrix} Z^1 \\ Z^3 \\ Z^2 \end{Bmatrix} \begin{Bmatrix} Z^2 \\ Z^1 \\ Z^3 \end{Bmatrix} \begin{Bmatrix} Z^2 \\ Z^3 \\ Z^1 \end{Bmatrix} \begin{Bmatrix} Z^3 \\ Z^1 \\ Z^2 \end{Bmatrix} \begin{Bmatrix} Z^3 \\ Z^2 \\ Z^1 \end{Bmatrix}$$

and these are generated in a straightforward fashion with the vector $\alpha = (\alpha_1, \alpha_2, \alpha_3)$. The various possible outcomes of the decision process upon which $P(A_p|R)$ is based are illustrated in Fig. 8.



a.) Possible consequences of the decision processes in the Hybrid Model.

Mean threshold (T)	10·00	7·50	5·00	2·50	2·00	1·50	1·25	1·00	0·75	0·50
R^2	0·994	0·995	0·942	0·923	0·929	0·912	0·915	0·881	0·747	0·565
Elimination on one attribute %	0·0	0·3	1·3	7·2	9·3	14·3	17·3	21·7	29·8	43·5
Elimination on two attributes %	0·0	0·0	0·2	3·6	7·0	14·1	19·4	25·5	31·6	34·9
Elimination on three attributes %	0·0	0·0	0·3	6·0	10·6	16·9	19·7	20·8	18·0	9·1
Trade-off on two attributes %	0·8	4·3	18·2	51·4	53·5	46·6	39·2	30·5	20·4	12·5
Trade-off on three attributes %	99·2	95·4	80·0	31·8	19·6	8·1	4·4	1·5	0·2	0·0

b.) Distribution of the decision process in the Hybrid Model as a function of mean threshold.

Fig. 8. Characteristics of the Hybrid model.

†It should be recalled that T is expressed as a proportion or percentage.

Again the tests were divided into two series corresponding to *equiprobable* rank orders, in which case the above ranks each enter with probability $\frac{1}{6}$, and "*non-flat*" or biased order distributions determined by the specific vector $\boldsymbol{\alpha}$. In both series of tests the thresholds were normally distributed with mean T and variance proportional to T.† Again this allows some of the dispersion in the data to be "squeezed out" as the tolerance is reduced.

The results of the simulation tests for $\boldsymbol{\alpha} = (0.33, 0.33, 0.33)$, including the goodness-of-fit measures, are shown in Figs. 8(b), 9(a)–(c). Only for the case $T = 0.5$ at the "lower end" of the tests range did significant estimation problems arise as judged by the "goodness-of-fit" and the significance of the parameters. It should be appreciated what this quantity $T = 0.5$ represents "on average" an alternative will be rejected at any stage of the assessment if its associated attribute μ is $>50\%$ worse than that corresponding to the best available alternative, as measured by the optimal Z^{μ} value. The significance of this tolerance level in terms of size must be viewed in the context of the size of the best available attribute. If this is small then a large T

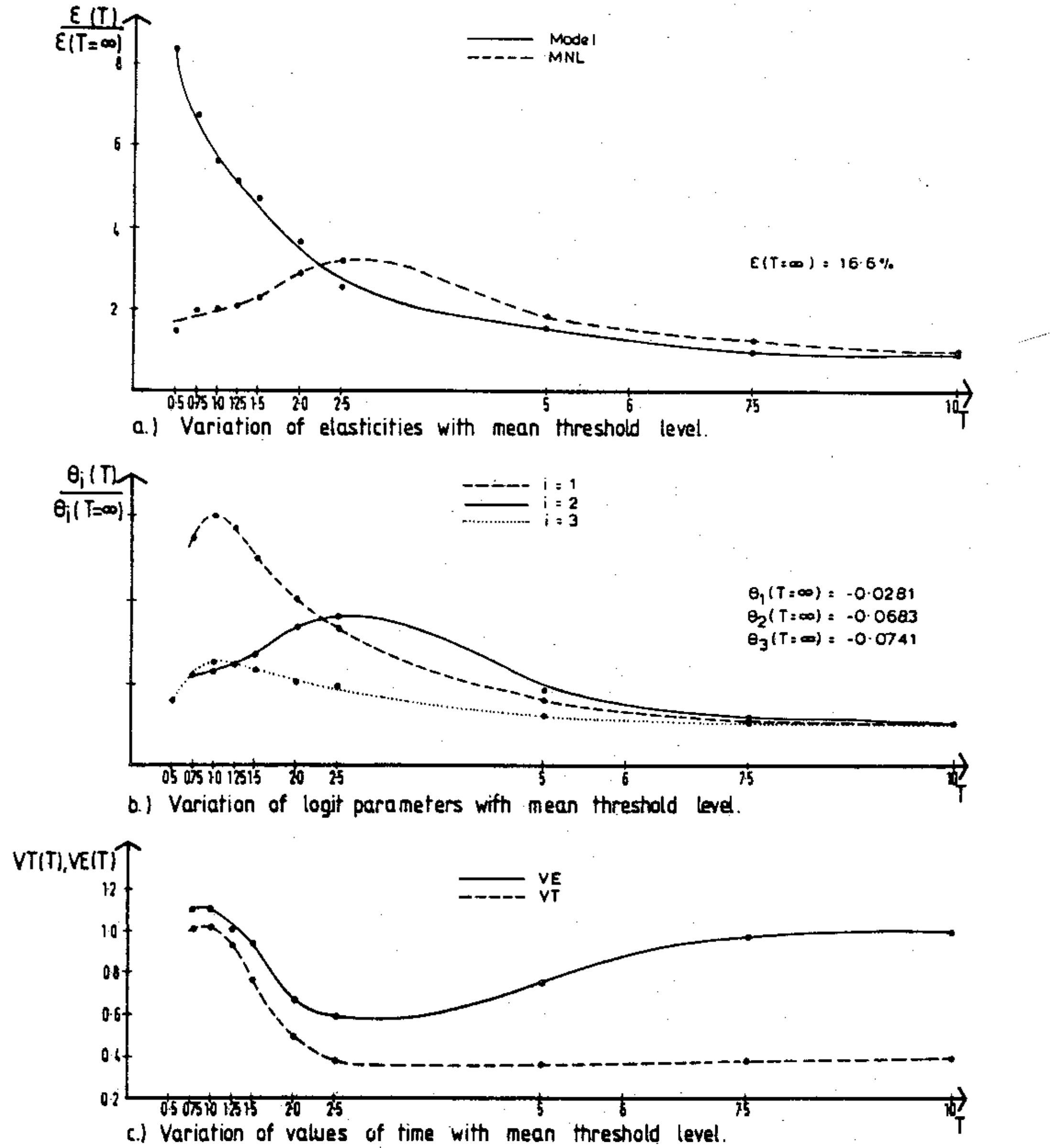


Fig. 9. Characteristics of simulation experiments with Hybrid model.

†At $T = 0.5$ a small number of negative tolerances were sampled with the adopted variation, and were simply discarded.

value might be required to retain the second best option. This is a consequence of basing an elimination process on proportions and not differences.

It is interesting to note that as T is reduced from a high value ($T = 10$) the entry of significant mode specific constants occurs first at $T = 2.0$ and it appears that this is responsible for the salient features of the curves in Figs. 9. As T is reduced θ_2 increases, the "values of time" decrease, and the model elasticity increases until the emergence of significant mode specific constant reverses the trend in θ_2 and produces the maxima and minimum shown. In contrast, the simulated response continues to rise as T decreases with the result that a large response error is apparent in the range below $T = 2.0$. The constants begin to play a prominent role as T is reduced further in this range, but it is not until $T = 0.5$ that the multinomial logit model begins to encounter serious difficulties in accommodating the dispersion in the synthesized data. In other words *we would not reject the logit model* until $T = 0.5$.

The unacceptability of the model as $T \to 0$ is consistent with the previous finding that the variation in behaviour appropriate to the strong lexicographical limit can not be satisfactorily accounted for by the multinomial logit function for the α values selected. Furthermore, we found that the fit statistics deteriorated at a given T as the standard deviations of the threshold distributions were jointly reduced.

In the range for which the logit model was not rejected we found perhaps surprisingly, little variation in the results portrayed in Fig. 9 when different rank order parameters α were selected. It appears that in the hybrid model the elimination mechanism itself is largely responsible for the response error and, as we have suggested, it is the influence of the constants entering the multinomial logit model, to improve the statistical fit measure, which are responsible for the "distortion" of the model elasticity. This is of course part of the occupational hazard of operating with an inappropriately specified model.

6.6 *Discussion*

In this section we have tried in a variety of ways to determine the significance of relaxing the conventional decision mechanism as a basis for model development, and in particular the implications for mis-specification of applying the LPLA logit model. As we introduce additional sources of behavioural variation and hence the parameters of the distributions representing them, the possible range of assumptions and experiments increase considerably. We have concentrated on what we feel have been significant parameter variations but have by no means exhausted the possibilities.

It is clear that response errors may be generated when distributed attribute set assumptions are introduced, whether these are accompanied by implicit or explicit behavioural mechanisms. The absolute size of the errors and general characteristics of the parameters of the logit model, are clearly dependent on the specific assumptions used to generate the the data as might have been expected.

Further general comments on these tests will made in the conclusion. We now proceed to an examination of the closely related issues associated with the role of information on the decision process and the existence of variation of choice sets over the population confronted by a choice.

7. INFORMATION, DISPERSION AND CHOICE SET GENERATION

7.1 *Introduction*

In the models considered above at least one attribute or aspect of *each* alternative in a choice set is scrutinized in the search process. There is a growing literature which confirms the widely held view that in location (of jobs and residence) and travel choice contexts individuals act under a restricted knowledge of alternatives and of their attribute values (see, e.g. the discussion in the papers by Kirby, 1979; Richardson, 1978; Thrift and Williams, 1981; and the references cited therein). Indeed, the geographical concept of a mental map is a recognition of the spatial heterogeneity of information (Gould and White, 1974; Young and Richardson, 1978).

In this section we shall concentrate on the relaxation of the assumption embodied in eqn (19), that each individual in the population π has the same choice set A available to them. While it is clear that this assumption of homogeneous and complete information, which is invoked to produce a workable model is an abstraction from reality, it is important to determine the extent to which it is a source of serious mis-specification.

Models which explicitly recognize the role of information in a choice process have tended to emphasise the dynamics of the search process in conjunction with aspiration levels and "satisficing" behaviour, and a well developed mathematical theory of such processes is available as detailed by Weibull (1978) and others. We can contemplate a series of simulation experiments based on an *explicit* model of the search process in which individuals sequentially consider alternatives and terminate that search when the net perceived utility of accepting an alternative is greater than a random "threshold" or satisfaction level (see also Richardson, 1978, for a discussion of such "alternative based" searches). An output of this process is a distribution of the sets of choice alternatives considered by individuals in the population.

We have instead resorted to a simulation framework in which the search process and imperfect knowledge are accommodated through the random generation of choice sets—which are simply collections of alternatives considered by an individual. The search process will be described in terms of its outcome, which is characterized by the distribution of number alternatives searched before a choice is made. The formulation of an appropriate model and organization of the simulation experiments will parallel those associated with the DAS model described above. We shall investigate the consequences of assuming complete information—identical and deterministic choice sets for all individuals (the perspective Λ), when the actual choice process involves a distribution of choice sets (the perspective Λ^*).

7.2 *Modéls embodying Distributed Choice Sets* (*DCS*)

In order to generate models involving distributed choice sets, the choice probability P_ρ is decomposed according to the various possible ways in which an alternative A_ρ may be selected from the various sets containing it. We can write, following Manski (1977)

$$P_\rho = \sum_{C} P(A_\rho|C)\,P(C|\mathscr{C}) \tag{81}$$

in which are defined the following: $\mathscr{C}$ is the set of all available choice sets; $P(C|\mathscr{C})$ is the probability of selecting choice set C from the set $\mathscr{C}$; and $P(A_\rho|C)$ is the probability of selecting alternative A_ρ from the choice set C.

We now consider the further decomposition in which the set of all choice sets $\mathscr{C}$ is partitioned according to the number of members in each subset C. Equation (81) can then be written

$$P_\rho = \sum_{n=1}^{N} \sum_{C\epsilon C_n} P(A_\rho|C)P(C|C_n)P(C_n|\mathscr{C}). \tag{82}$$

Here C_n defines the set of all choice sets within n members and $P(C|C_n)$ is the probability of drawing a *particular* choice set with n alternatives from C_n. $\sum_{C\epsilon C_n}$ denotes summation over all choice sets of size n.

Let us consider this decomposition for the N choices $A_1, \ldots, A_\rho, \ldots A_N$. There are clearly ${}^NC_n$ choice sets of size n in the set C_n and a maximum of $2^N - 1$ available choice sets in $\mathscr{C}$. These can be arranged as before in terms of the subsets $C_1, \ldots C_n, \ldots C_N$ as follows:

C_1: $\{A_1\}, \ldots \{A_\rho\} \ldots \{A_N\}$

.

.

.

C_N: $\{A_1 A_2 \ldots . A_N\} \ldots . \{A_{N-n+1}, \ldots A_{N-1}, A_N\}$

.

.

.

C_N: $\{A_1 A_2 \ldots . A_N\}$.

This decomposition of $\mathscr{C}$ is organized on the tree structure as shown in Fig. 10(a).

The conventional assumption in random utility models is, as is well known, *that all members of a market segment s* select an alternative from the single set C_{N^s} $(= \mathbf{A})$. That is

$$P(C_n|\mathscr{C}) = 1 \quad \text{if } n = N^s$$
$$= 0 \quad \text{otherwise.} \tag{83}$$

This will be a special case of a choice set distribution considered below.

In order to generate a choice model the probability $P(C|\mathscr{C})$, which has been decomposed into $P(C|C_n)$ and $P(C_n|\mathscr{C})$, must be specified in terms of the size of the sets and the attributes of their members. A decision model must also underpin the process of selecting A_p from the sets C containing it.

The two perspectives Λ and Λ^* will be formed based on these two facets of the choice process. Λ^* will involve a description of dispersion due to the heterogeneity of information interpreted analytically through the probability functions of choice set generation. In Λ it will be assumed that *all* members of π_r^s scrutinise one available choice set $\{A_1 \ldots A_N\}$ containing all alternatives. Both perspectives will involve a decision model based on utility maximisation—individuals select what they consider to be the best alternative *from the sample which each considers.*

We can now summarise the simulation experiments through the following description of the perspectives:

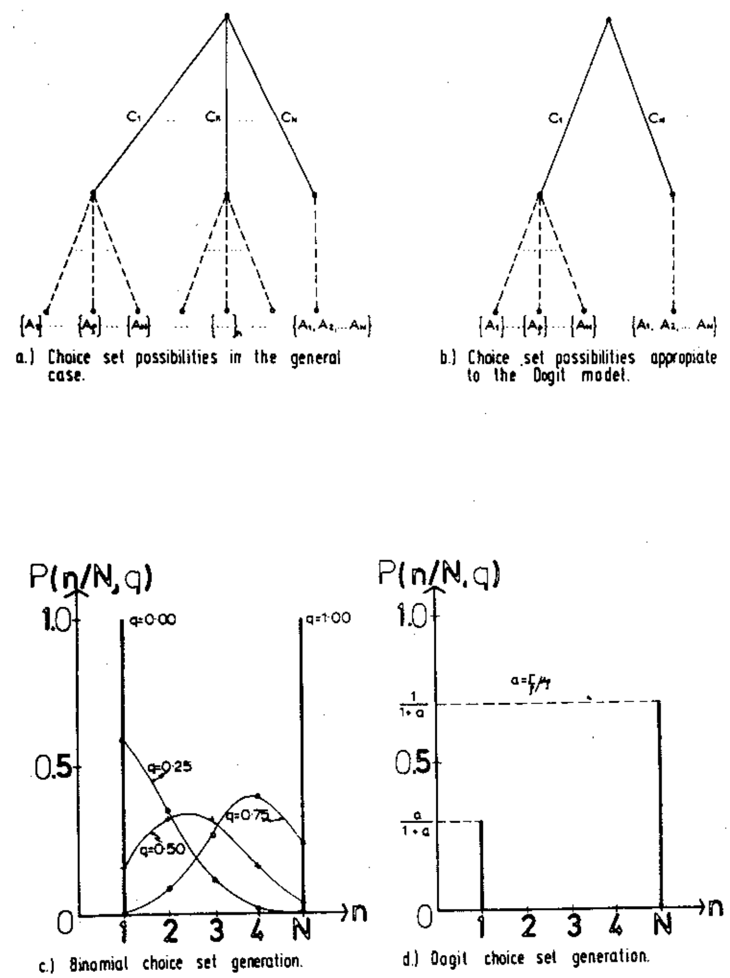


Fig. 10. Choice set generation partitioned by size of choice sets.

Λ^*:	Random generation of choice sets	D^*:	Utility maximisation from $W(o, \sigma^*)$

Λ:	All members search the full choice set available	D:	Utility maximisation from $W(o, \sigma)$

In the series of simulation tests, the data $\mathbf{P}^*$ will be generated from parametric probability distributions, and the probability of a choice set of size n being selected will again be taken as a truncated binomial function with parameter q. That is $P(C_n|\mathscr{C})$, which we shall write $P(n|N, q)$), will be taken as

$$P(n|N, q) = \frac{{}^N C_n q^n (1-q)^{N-n}}{1-(1-q)^N} \tag{84}$$

for which the mean set size $\bar{n}_q$ is given by

$$\bar{n}_q = \frac{Nq}{1-(1-q)^N}. \tag{85}$$

The forms of the discrete probability function $P(n|N, q)$ for different q values are shown in Fig. 10(c). (Continuous curves are used to distinguish the various functions.) The two special cases corresponding to $q = 1$ and $q = 0$ are again worthy of note. In the former

$$\begin{aligned} P(n|N, 1) &= 1 \quad \text{if } n = N \\ &= 0 \quad \text{otherwise,} \end{aligned} \tag{86}$$

and corresponds to the homogeneous and complete choice set assumption of conventional theory. For $q = 0$, on the other hand

$$\begin{aligned} P(n|N, 0) &= 1 \quad \text{if } n = 1 \\ &= 0 \text{ otherwise,} \end{aligned} \tag{87}$$

all individuals select a choice set containing a single alternative from the set of choice sets.

Having specified the probability of selecting a particular set C_n it is necessary to examine the probability of selecting a *given* set of alternatives $\{\ldots\}_n$, containing n members. In the first series of simultation test (A), it is once more assumed that the ${}^N C_n$ member sets in C_n occur with equal probability—that is, the probability of selecting a set of alternatives is a function of its size alone. In the second series of simulations (B), the probability of selecting any member C from C_n will be a function of the mean utilities of its component choice alternatives.

7.3 *Simulation experiments with DCS models*

As we have found no simple solution (generator functions) for the series expression (82) we have resorted to simulation in order to generate the data $\mathbf{P}^*(q)$ by sampling choice sets randomly from $P(n|N,q)$. In each set of tests the probabilities $P(A_\rho|C)$ defined in eqn (22) are determined according to the usual principles of utility maximisation with random components of utility. For numerical tests five alternatives ($N = 5$) were considered, and the Weibull curves adopted in Λ^* were taken to have a standard deviation σ^* of 5 units (the difference between the largest and smallest mean utilities was 1 standard deviation).

In tests (A) choice sets $C \epsilon C_n$ were drawn from a uniform distribution, and the choice model $P(A_\rho|C)$ was generated from Weibull distributions,† with mean values $(\bar{U}_1, \ldots \bar{U}_\rho, \ldots \bar{U}_{N=5})$. The maximum utility option is recorded for each sampled "individual" and the process repeated for the members of π.

For all values of $0 \leq q \leq 1$ it was found that the logit function

$$P_\rho(q) = \frac{e^{\theta(q)\bar{U}_\rho}}{\sum_\rho e^{\theta(q)\bar{U}_\rho}} \tag{88}$$

†Other distributions (e.g. normal functions) were also adopted in an alternative series of tests.

provided an excellent fit to the base data $P^*(q)$. The variation of the estimated dispersion parameter $\theta(q)$ with q is shown in Fig. 11a. For $q = 1$ the complete set $\{A_1, \ldots, A_\rho, \ldots, A_n\}$ is selected by all members of π and the estimated parameter $\theta(q = 1)$ may be compared with its exact value $\theta = (\pi/\sqrt{(6\sigma^*)}) = 0.257$. A reduction of q which has the effect of introducing smaller choice sets (see eqn 85), results in an increase in the standard deviation $\sigma(q)$ imputed by the observer in accordance with

$$\sigma(q) = \frac{\pi}{\sqrt{(6)}\theta(q)}. \tag{89}$$

The agreement between the model predictions **P** and the simulated results **P*** in the base system is shown for different q values in Fig. 12(a).

We can now compare the simulated and predicted response, $\boldsymbol{\delta}\mathbf{P}^*$ and $\boldsymbol{\delta}\mathbf{P}$ respectively, when the utility values $\{\bar{U}_1, \ldots, \bar{U}_\rho, \ldots, \bar{U}_N\}$ are modified. In fact a single value $\bar{U}_3$ was altered to allow the direct and cross elasticities to be measured. In Fig. 12(b) we plot the predicted share modifications from the base system against their simulated counterparts for the five alternatives A_ρ: $\rho = 1, \ldots, 5$. Although the agreement is slightly erratic, the important point to note is the lack of a systematic over- or under-prediction of response.

In the above set of tests (A), the probability of drawing any one of the ${}^N C_n$ sets of alternatives$\{\ldots\}_n$ with n members was given by

$$P(C|C_n) = \frac{1}{{}^N C_n}. \tag{90}$$

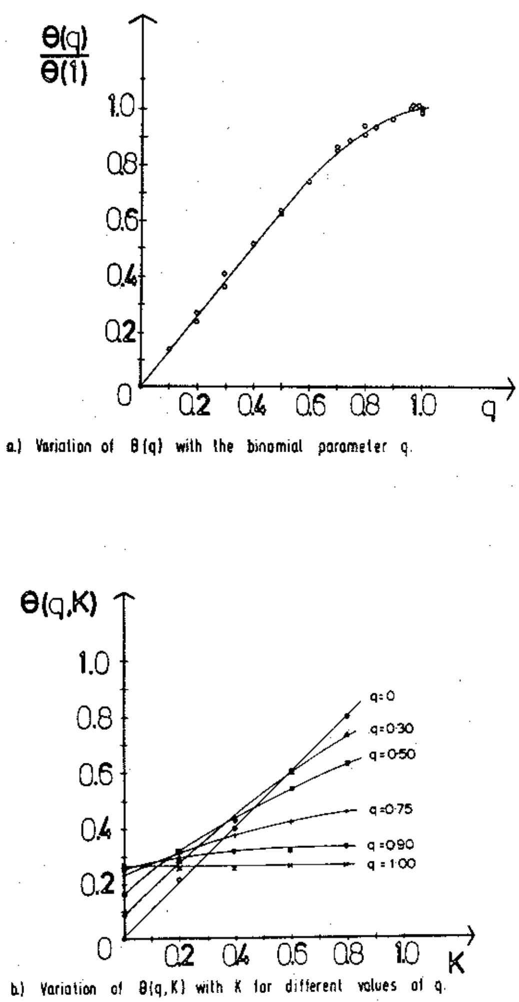


Fig. 11. Variation of the dispersion parameter θ with the choice set generation parameters.

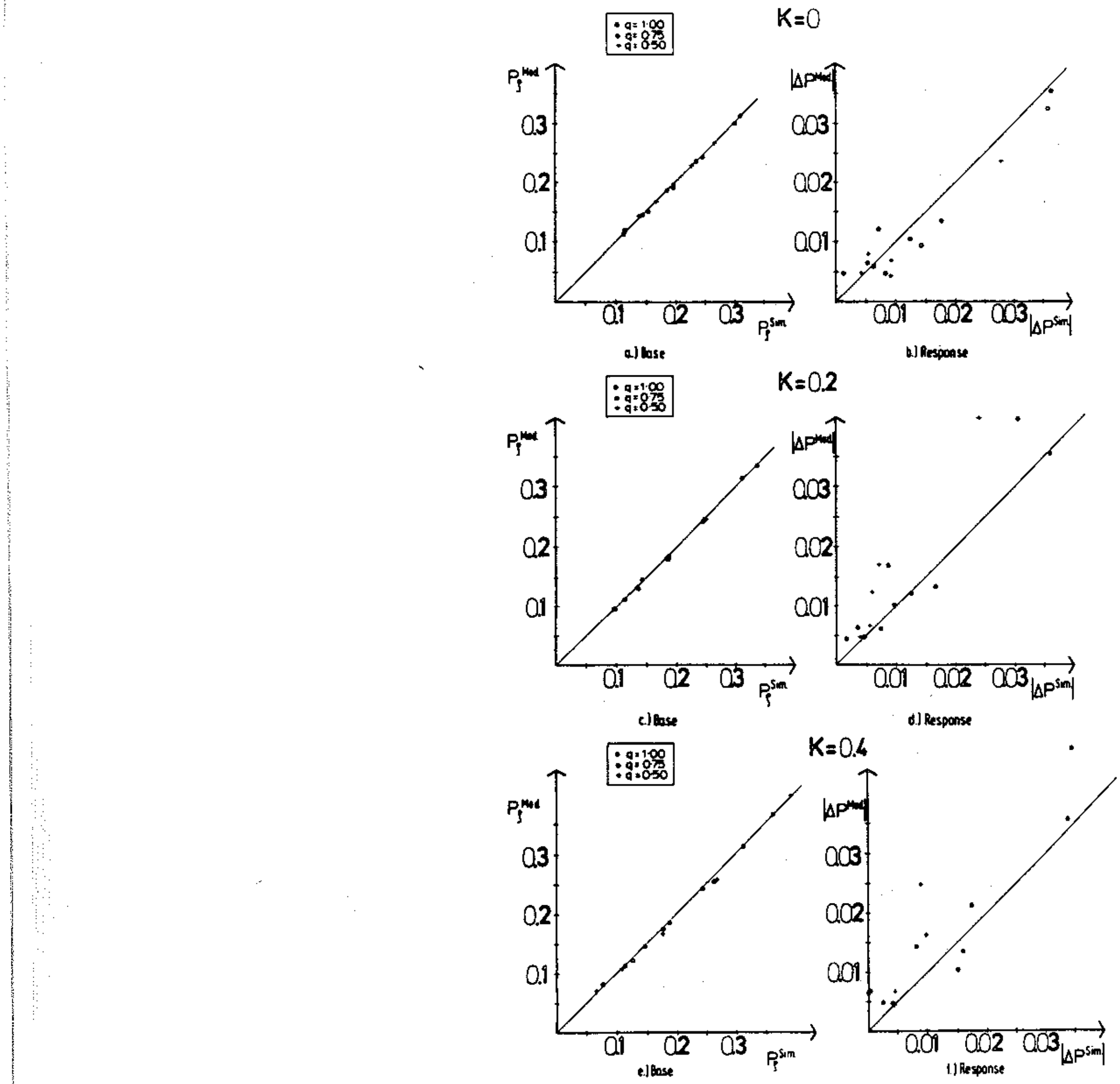


Fig. 12. A comparison of simulated and modelled results in the base and response contexts for various combinations of K and q.

It might be more realistic to assume that the probability of selecting a particular set $\{\ldots\}_n$ is related to the characteristics of the alternatives in the set—perhaps the utility values, or the spatial configuration. In the second series of tests (B), $P(C|C_n)$ involves a discounting factor which reduces the probability of an alternative A_ρ being selected for membership of any set in accordance with the difference between its mean utility $\bar{U}_\rho$ and that of the maximum, max $\{\bar{U}_1, \ldots, \bar{U}_N\} \equiv \bar{U}^+$. That is, the *relative* odds of membership of $A_1, \ldots, A_\rho, \ldots, A_N$ in a set of any given size are taken as

$$e^{-\kappa(\bar{U}^+ - \bar{U}_1)}, \ldots, e^{-\kappa(\bar{U}^+ - \bar{U}_\rho)}, \ldots, e^{-\kappa(\bar{U}^+ - \bar{U}_N)}$$

respectively. When $\kappa = 0$, all members are equally likely to be chosen and the results of the tests in series A are retrieved. As κ increases there is an increasing likelihood that utility options with high mean values will be sampled. (In the actual construction of the tests the options were considered to be spatially arranged and κ identified with a *spatial* discounting factor.) There are now two parameters characterising information heterogeneity which are identified with the two levels of the tree structure in Figs. 10. The binomial parameter q

governs the choice set size distribution, while κ governs the probability of selecting a given set $\{\ldots\}_n$ from C_n. Note that when $q = 1$ the single set $\{A_1, \ldots A_N\}$ is chosen and the discounting factor has no effect.

We can now enquire, as before, if the logit function $\mathbf{P}(q, \kappa)$ with an estimated dispersion parameter $\theta(q, \kappa)$—now a function of q and κ—provides a good statistical fit to $\mathbf{P}^*(q, \kappa)$, in the base and response contexts. For the base system the fit is impressive, deteriorating only slightly for higher values of κ. The variation of $\theta(q, \kappa)$ with q and κ is shown in Fig. 11(b) and the base vectors $\mathbf{P}^*(q, \kappa)$ and $\mathbf{P}(q, \kappa)$ are compared in Figs. 12(c) and (e). Under conditions of change, however, there is a systematic *overprediction* of modelled response for values of κ greater than zero and q less than 1, as indicated in Figs. 12(d) and (f).

We can now summarise the results and implications of these tests as follows.

As all combinations of selected parameters $\{q, \kappa\}$ generate data sets $\mathbf{P}^*$ which are well fitted by the logit function we may say that the dispersion reflected in the estimated parameter $\theta(q, \kappa)$ is consistent with utility maximisation under *both* full and partial information assumptions. That is, the dispersion exhibited in the data $\mathbf{P}^*$ is statistically consistent with the following interpretations:

(i) Complete information and preference dispersion from utility distributions with standard deviation given by

$$\sigma(q, \kappa) = \frac{\pi}{\sqrt{(6)}\theta(q, \kappa)}. \tag{91}$$

(ii) Partial information characterised by the distribution parameters q and κ; in conjunction with random utility functions of standard deviation σ^*. For given σ^*, the (q, κ) combinations which generate the same dispersion as $\sigma(q, \kappa)$ defined by eqn (91) may be read off Fig. 11(b).

Further, if information about opportunities is incomplete ($q < 1$) but there are no biases in the selection of choice sets (i.e. $\kappa = 0$) no *systematic* response error will be involved with logit forecasts. On the other hand if the data is the outcome of a process characterised by preferential tendencies for particular choice sets to be selected ($\kappa > 0$) logit forecasts may well involve a systematic over-prediction of the response to policy stimuli. As the observer in Λ has no knowledge whatsoever what process did generate $\mathbf{P}^*$ it is not possible to discriminate between the interpretations (i) and (ii), and there remains an indeterminacy in the response forecasts.

Clearly the above simulations may be generalised in a number of ways, and it would be interesting to ascertain how the response error was dependent on further detailed aspects of choice set generation and decision models. (How for example the policy sensitivity of the parameters of the choice set generating function influence response results.) This is however, beyond the scope of this paper. Before leaving the theme of choice set generation however, we wish to relate the above considerations to two further models, namely: the DOGIT model introduced by Gaudry and Dagenais (1979) and a location model recently discussed by Kirby (1979).

7.4 *The Dogit and other DCS models*

The Dogit model is a generalisation of the logit model to accommodate varying degrees of interaction between alternatives. Its formation, originally achieved by means of transformation theory, has been given a behavioural interpretation and derivation by Ben-Akiva (1977a) who considered individuals to either be captive to a particular alternative or to have the full choice set available to them. Out of the set $\mathscr{C}$ the N sets associated with C_1 and the single set C_n are considered, as shown in Fig. 10(b). The spectrum $P(C_n|\mathscr{C})$, shown in Fig. 10(d) consists of the two "spikes" at $n = 1$ and $n = N$.

If the probability of selecting the individual choice sets $\{A_1\}\ldots\{A_p\}\ldots\{A_N\}, \{A_1\ldots A_N\}$ is taken to be

$$P(C|\mathscr{C}) = \frac{\mu_1}{1+\sum_{p=1}^{N}\mu_p}, \ldots, \frac{\mu_p}{1+\sum_{n=1}^{N}\mu_p}, \ldots, \frac{\mu_N}{1+\sum_{p=1}^{N}\mu_p}; \frac{1}{1+\sum_{p=1}^{N}\mu_p} \tag{92}$$

for which the odds of being captive and non-captive are

$$P(C_1|\mathscr{C}) = \frac{\sum_p \mu_p}{1+\sum_p \mu_p};\ P(C_N|\mathscr{C}) = \frac{1}{1+\sum_p \mu_p} \tag{93}$$

the DOGIT model is readily derived from eqn (81) and is given by

$$P_p = \frac{e^{\theta \bar{U}_p} + \mu_p \sum_p e^{\theta \bar{U}_p}}{(1+\sum_p \mu_p)\sum_p e^{\theta \bar{U}_p}} \tag{94}$$

The parameters $\boldsymbol{\mu}$ may be taken to be functions of the attributes associated with the alternatives.

For the particular simulations performed in this section "lumpy" distribution spectra for $P(C_n|\mathscr{C})$—which are appropriate to the Dogit model have not been adopted. We should not therefore be particularly surprised that the modifications to the logit function due to captivity, which inspired the formation of the Dogit, have failed to assert themselves in numerical tests on the model derived from the Binomial distribution. These modifications will inevitably become more important as the number of alternatives becomes small (as, e.g. in modal choice contexts).†

We turn finally and briefly to Kirby's model (Kirby, 1979) which involves search within locationally defined choice sets. In its simplest form the model incorporates a range function $\phi(r)$ which denotes the probability of an individual selecting a house up to range r from his or her place of work. (The housing market and choice making populations are actually considered stratified and r is considered to be expressed in terms of generalised cost.) Individuals are now considered to confine their search within their selected range and to choose a zone of residence within it with a probability proportional to the number of houses of the relevant type within the zone.

It is clear that the model is a special case of the choice set generating processes defined in eqn (82) above. For suitably defined choice sets the function $P(C|\mathscr{C})$, or $P(C|C_n)$ and $P(C_n|\mathscr{C})$, may be related in a straightforward way to the range function. The Kirby model appears in fact to be very similar to the intervening opportunities model, although the behavioural descriptions underpinning them are distinct.

7.5 *Discussion*

In this section we have examined the formulation of models which embody distributed choice sets and have investigated the mis-specification problems arising from relaxing the assumption of complete information in conventional micro-models. The parallels between the DAS model system developed in Section 6 and the DCS decomposition considered here are readily apparent (and could be further exploited). These characterizations of the choice model may be seen in terms of complementary aspects of information processing in the solution of the multi-criterion problem as described in Section 5. Indeed it is clear that the "attribute based" and "alternative based" decompositions could be integrated to produce a further and more general class of models characterized by a distribution of the number of alternatives *and* attributes searched in the choice process which could be underpinned by a variety of decision mechanisms. A consideration of these developments is however outside the scope of this paper.

8. HABIT, HYSTERESIS AND TRAVEL RESPONSE

Many authors have remarked on the relevance and role of habit, learning and "triggers" in the decision process accompanying (re-) location (i.e. migration) and travel choice behaviour (see, e.g. Banister, 1978; Heggie, 1978; Hensher, 1975; Goodwin, 1977, 1979; and the references cited therein). In spite of the widespread recognition of the influence of habit, few empirical or

†Note that in the development of DCS models we defined market segments in order to generate homogeneous populations with respect to certain obvious constraints. The notion of captivity as discussed here is a residual characteristic pertaining to those individuals who are constrained to a particular alternative on the basis of *non-identified* factors.

theoretical results concerning this phenomenon have been forthcoming, although the papers by Blase (1979) in the former class, and those of Goodwin (1977, 1979) and Wilson (1976) in the latter are noteworthy.

The work of Goodwin is of particular interest and our intention is to offer some elaboration of the ideas presented by that author within the framework of perspectives developed in this paper. We shall adopt a random utility model of binary choice incorporating habit developed by Goodwin.

The existence of habit, or what might be considered as inertia accompanying the decision process of the individual is possibly the most insidious of behavioural aspects which represent divergencies from the traditional assumptions underpinning choice models, for its existence appears *directly* in the *response* context. In order to examine the effects and the implications of habit it is appropriate to return once more to the assumption underpinning the conventional cross-sectional approach.

In Fig. 13(a) the familiar S-shaped curve relevant to binary choice is reproduced. We can think of this in terms of a continuum of populations $\pi(\bar{U}_2-\bar{U}_1)$ characterised by an *imputed* utility difference distributed along the curve. For a given difference $\bar{U}_2-\bar{U}_1$ there exists a single population π with a *fixed* proportion of members associated with the alternatives, and identified by a point on the curve. Under conditions of change $(\bar{U}_2-\bar{U}_2\rightarrow\bar{U}'_2-\bar{U}'_1)$ the population $\pi(\bar{U}_2-\bar{U}_1)$ will simply acquire the characteristics of $\pi'(\bar{U}'_2-\bar{U}'_1)$ observed in the base system. The *response* is determined from the cross-sectional *dispersion.*

An implication of this assumption is that the response to a particular policy or change will be exactly reversed if the stimulus is removed. The stimulus-response relation is symmetric with respect to the sign and size of the stimulus.

These features are consistent with the "rationality" assumption attributed to an individual, prepared to *continually* monitor his present and alternative options. They must be modified in the presence of habit, although we should re-emphasise that the necessity for modification is not a refutation of *homo-economicus* but simply a re-interpretation by the observer of the actions and behaviour of the individual. To understand the required modifications, we consider a *single* population π subjected to a continuous modification in characteristics of the alternatives A_1 and A_2 which are reflected in $\bar{U}_2-\bar{U}_1$.

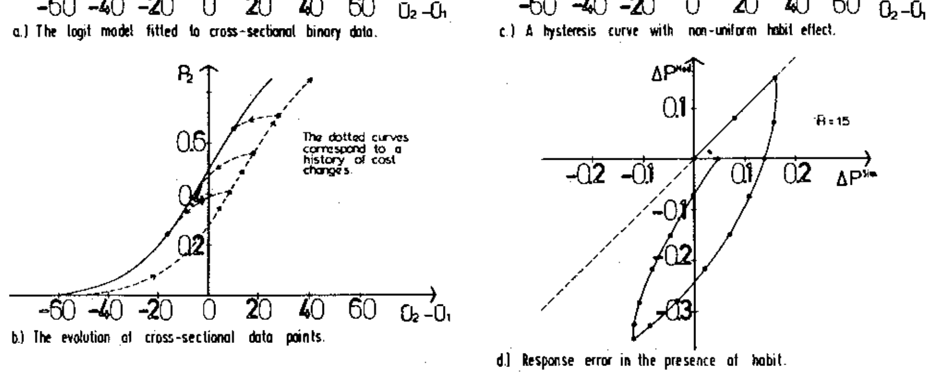


Fig. 13. The influence of habit on the interpretation and response of cross-sectional models.

In the conventional derivation of random utility models we identify with each individual $I_i \epsilon \pi$ utility functions U_1^i and U_2^i and decree that alternative A_1 will be selected if

$$U_1^i > U_2^i. \tag{95}$$

Let us assume that this inequality is satisfied and A_1 is selected. A stimulus of magnitude δU_{12}^i in favour of the rejected option A_2 must be applied before a response—that is, a change of option—will be observed. The minimum size of δU_{12}^i is $|U_1^i - U_2^i|$.

In the presence of habit, which will be interpreted as an equivalent utility value h^i, this minimum required stimulus is raised to

$$\delta U_{12}^i = U_1^i - U_2^i + h^i$$

and the total population response to change is given by

$$R = \sum_{I_i \epsilon \pi} \eta(\delta U_{12}^i) \tag{96}$$

in which

$$\eta(x) = 1 \text{ if } x > 0$$
$$= 0 \text{ otherwise.} \tag{97}$$

The existence of the habit term will clearly accompany those members of π who are *currently* associated with an alternative(s) experiencing a stimulus to the relative advantage of another option(s). This feature introduces a basic asymmetry into response behaviour and gives rise to the phenomenon of hysteresis (Goodwin, 1977; Wilson, 1976). Now, the *present* state of the population π, identified in terms of the proportions on each alternative, is dependent on not only the utility values $\bar{U}_1$ and $\bar{U}_2$, but on how these utility variables attained their current value. Formally, the state of the system P_ρ may be expressed as a path integral in the space of utility components $(\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_N)$. The value of this integral is path independent when habit is absent, but path dependent when it is present.

The features of these hysteresis curves have been reproduced by Monte Carlo simulation and the results shown in Fig. 13(b). As Goodwin (1977) has noted, when there exists a distribution of habit h^i over the population, the response behaviour of π, which is dependent on the prior state of the system, can be complicated. In the figure we have identified the system "history" by means of arrows. (Utilities were sampled from Weibull distributions and habit values from a negative exponential function with mean $\bar{H}$.) Now it can be seen that for a given utility difference $\bar{U}_2 - \bar{U}_1$, a multiplicity of P_2 values can exist, each of which is dependent on the path of utility values by which $\bar{U}_2 - \bar{U}_1$ was achieved. The response to change is likewise history dependent.

It is important to remember that these curves correspond to a single population π responding to changes in the utility differences. We now return to the interpretation of the S-shaped curve in Fig. 13(a) which is assumed to provide a good statistical fit to underlying observations on dichotomous choice. Admitting the possibility of habit, we must consider the history of cost changes which accompanies each population $\pi(\bar{U}_2 - \bar{U}_1)$ at the cross-section. One such history might be a continually increasing difference of $\bar{U}_2 - \bar{U}_1$ so that all populations simply move up a curve identified at the cross-section. Another history is that shown in Fig. 13c which accompanies an initial rise and subsequent fall of utility differences experienced by all populations, π. We cannot be sure of the response of each population identified on an observed S-curve unless we know something of its history. The traditional assumption of zero habit, and the movement of all populations π along a single curve, will in general introduce a response error—but of what size? To examine this numerically, we introduce the perspectives Λ^* and Λ characterised as follows:

Λ^*:	Distribution of habit;	D^*:	utility maximisation from $W(o, \sigma^*)$

Λ:	No habit;	D:	utility maximisation from $W(o, \sigma)$

and look for the response error under a given "path" of utility changes. We have assumed that a given population $\pi(\bar{U}_2 - \bar{U}_1 = 10)$ is identified on the right-hand curve in Fig. 13(d) which is appropriate to a continually increasing utility difference. The stimulus is now identified with three stages: an initial increase of $\bar{U}_2 - \bar{U}_1$ of 10 units, a subsequent reduction of 40 units and a final increase of 30 units. (This kind of situation might correspond to an initial increase in the relative advantage of car travel (1: bus; 2: car) which is in line with previous cost movements, followed by a sudden large rise in car costs, accompanying say a series of petrol rises. Finally, the cost difference is eroded when public transport operators increase charges to further increase revenue.)

The response error corresponding to this situation is mapped in Fig. 13(d). The simulated response to cost changes with $\bar{H} = 15$ units are plotted against the modelled response derived from the single logit curve under the assumption of zero habit. On the first stage no error is involved as the past trend is reinforced. A significant over-estimate of the modelled response takes place on the second stage, and this decreases on the third.

We shall discuss the more general implications of habit for response forecasting, together with the other implications of the paper in the next section.

9. DISCUSSION AND CONCLUSIONS

The topics considered in this paper may all be accommodated under the umbrella of model specification issues. This broad field embraces a variety of problems ranging from the process of embedding theoretical statements within a model, through the statistical analyses of competing functional forms, to the important and controversial aspects of model transferability. At the outset of the work we were conscious of the different emphases placed by commentators in the discussion of these issues, and this is particularly true of the problems associated with *mis-specification*. There are the clear influences and priorities of practise on the one hand, and the fastidiousness of formal theoretical research on the other. It is equally the case that individual perspectives are derived from the various disciplines whose interests intersect in the analysis of travel behaviour.

In the absence of firm evidence to validate travel *response* forecasting models in the wide range of applications contexts, it is not surprising then that statements on the validity issue span the confident expression of faith in the state-of-the-art through considered agnosticism to downright disbelief. There is the danger in these circumstances that any suspected deficiencies in the theoretical base are treated either with complacency or as having a profound significance for the forecasting process. In this paper we have adopted a broad experimental base for the discussion and assessment of *potential* mis-specification issues, and have attempted to unify a number of critical themes within an extended choice theoretic framework. We have placed emphasis on the rather obvious point that the rational choice paradigm is far less restrictive, and many distinctions less significant, than a number of commentators have tended to imply in their discussions on the current generation of travel choice models.

Before offering some general comments and revealing our own prejudices of the validity issue we shall briefly review what we feel are the implications of the tests themselves. We would emphasise that out of the whole family of possible mis-specification experiments the four selected are not of course necessarily the most important in practise, their discussion has been motivated by a number of criticisms raised against analytic models of probabilistic choice. What are regarded as the most important sources of mis-specification in any particular circumstance is perhaps the major contentious issue in the validity debate (see Ben-Akiva, 1979; Horowitz, 1979b, 1979c; and Louviere, 1979). The rationale for the experimental approach adopted here is essentially that a demonstration that one or other of the assumptions was suspect would serve to direct further theoretical and empirical investigations.

We feel that the tests themselves give some cause for optimism and some cause for concern. Reconsider them from a purely numerical viewpoint. An attempt has been made to fit a function with p parameters and w variables to data generated by a function characterised by p' $(> p)$ parameters and w' variables. Now unless special conditions are present it is expected that there would be circumstances in which the simulated and estimated responses diverge. We have however constrained the experiments by the requirement that the fitted model pass some test of statistical scrutiny, and have sought to discover the extent to which such special conditions are present for plausible behavioural postulates, and in addition, how the parameters of a fitted model attempt to "compensate" for its restricted degrees of freedom. We have aimed to design the experiments in such a way the choice model under test (typically the linear multinomial logit model) may be regarded as a special (extreme) case of a family of models characterised by one or more additional parameters. This has usually been possible when one of the assumptions in the formation of a model is relaxed.

The question then was: in the range for which the test model—say the LPLA multinomial logit model—was deemed acceptable on statistical grounds, would the response error be *significant* or not? This raises the issue of what are the criteria for "acceptability" and "significance" and leads us to a further point. Because of the construction of the tests the size of the response error will tend to increase with the size of the extra parameter(s) governing an additional source of variability and this will *tend* to be accompanied by a deterioration of the statistical fit. Unless the response errors are dramatic or negligible we are again faced with the problems associated with the modeller's own standards for accepting or rejecting a model on *statistical* grounds. This issue will be influenced by the experimental design itself and in particular whether we are judging the acceptability of a single model or *comparing* the performance of two models estimated with synthetic data, as in the first test on structural mis-specification. We must of course also be aware that in these experiments on *theoretical* misrepresentation we have been able to carefully control the sources and size of variability in the (synthetic) data. In practise the problems of functional specification compete with data problems, and theoretical deficiencies must be explored within a "fuzzy" environment, and in the context of the inaccuracies of the model as a whole.

Because of some of the well known restrictive properties of the linear multinomial logit, we were preconditioned to believe that that model could be rather easily confounded—that is, a range of conditions could readily be found which would lead to serious mis-specification. Although these conditions can certainly be found, and not surprisingly with respect to the inclusion of similarity effects, we found that the model was rather more robust than we had initially thought, and that when the nested logit model is added to the "modelling kit" the logit family becomes rather powerful (notwithstanding the problems of taste variation examined by Hausman and Wise, 1978; Horowitz, 1979a, 1979b; by Cardell and Reddy, 1977, and others).

The problems of similarity, or correlation, once exposed are now appearing considerably less formidable in theory and practice. Circumstances can occur in which the nested logit model of a particular design becomes suspect, particularly when a complex web of similarities exist between the attributes and alternatives. Inappropriate nested models may readily be diagnosed through the violation of the condition for consistency with the theoretical base.

The recognition that models based on alternative decision processes may provide mutual numerical *approximations* may revive the long debated issue of the appropriate rationale for alternative model structures (Brand, 1973; Ben-Akiva, 1973; Williams, 1977). Within the framework in which individuals are concerned to *optimise* their choices on the basis of particular utility functions, the question as to whether *simultaneous* or *sequential* decision making processes provides an appropriate rationale for alternative model structures is, we believe, largely irrelevant. The question is essentially one of determining the appropriate utility function and from that deducing the structure. If, on the other hand, distinctions *are* to be made on the basis of *optimising* and *satisficing* behaviour (the latter interpreted in terms of, for example, priority ranking and information content of the decision process) there are two immediate concerns: firstly, how do we discriminate empirically between such expressions, and secondly, if we are indifferent from a *numerical* point of view to the response properties of such models, are we also indifferent to the implications for measuring user benefits? These questions must await further research.

The experiments on alternative decision processes and information do emphasise again the potential hazards of the "correlation = causation" syndrome. If biases exist in the selection of choice sets there seems to be some danger that serious response errors may occur. This is not as surprising as might appear at first sight, when the actual choice process is decomposed into a model for choice (or attribute) set formation, and one for the decision process, it is evident that dispersion arises from at least two sources. If these models are endowed with different sets of explanatory variables, some of which may exhibit high correlation (e.g. distance, time and cost in a locational context) then an attempt to explain the full variability in terms of a restricted set of such variables may well have serious implications.

The results on the influence of habit are not particularly surprising, and we consider these the least satisfactory of the experiments. Because the influence of habit manifests itself directly in the response context our results, and those of Goodwin (1977) have simply shown that if it exists then it can be significant, and that the absolute size of the effect will depend on the size and distribution of the habit effect over the population, the present position on the "*S*-curve", and on the history of cost changes.

While the existence of habit will influence the response of a population and indeed the interpretation of data in the context of model development, the *detailed* implications for location and mode switching in the presence of, for example, petrol price increases are unclear. We must not forget also that changes of say mode or location may be triggered by a series of other stimuli (life cycle effects, etc.) and the extent to which the system is considered to be in "disequilibrium" because of inertia will be dependent on these. The significance of inertia will thus be dependent on context and in particular on the time scales over which cost *changes* are introduced and the response to them measured.

The recent evidence of Blase (1979) suggests that the effects of habit can be of practical significance and that should treat the phenomenon seriously. It would seem that two ways forward would be through the greater use of transfer pricing methods and in particular the use, where possible, of time series data. We shall return to the issue of "calibration" below.

Let us now proceed to a more general issue. At the start of this paper we suggested that a large number—the majority—of transportation study models made little explicit use of behavioural theory. The criterion for selecting an appropriate functional explanation was basically a statistical goodness-of-fit measure. In the cross sectional approach the prime task is that of identifying the variability which is associated with one or more policy variables in a larger set. We argued that one of the dangers of an indifference to behavioural postulates and the source of dispersion in the data, was that of obtaining elasticity parameters which were simply not sensible. This being the case we were therefore encouraged to look for the implications and restrictions which are introduced when a set of behavioural postulates is invoked. Of course, from a *numerical* viewpoint an observer will be totally indifferent between sets of postulates which result in the *same* functional expression for forecasting response, and there will be no particular incentive to discriminate between them. Furthermore, it *may* be that a specified model, through luck or judgement, may override a particular critique. Because a particular assumption in a particular formulation is not plausible, the functional expression—the model—is not necessarily damned for ever; we have reported on such cases. Conversely and typically the analyst may not have the evidence or the conviction to reject a model which is in fact seriously mis-specified. Theoretical fastidiousness tends to be sacrificed very readily to practical considerations.

In the same way as a "perfect" model of behaviour is not our goal, so too can we not expect to *explain* the totality of variability in a data set. It does however, seem ominous to us that we are not able to discern and discriminate with any conviction between the contributions to behavioural variability at the cross section from the multitude of sources: preference dispersion; the heterogeneous disposition of a wide variety of constraints; the role of information; the "inertia" or "disequilibrium" associated with habit effects; sub-optimal behaviour by individuals because "... they have not the wits to maximise" (Simon, 1955); or the measurement, aggregation and representational error on the part of an observer in the process of providing the framework within which dispersion is recorded!

We can of course, entertain the possibility of embellishing the theoretical framework of present models to include more complex features of the choice process. Indeed, is it not part of the demand analysts terms of reference to discover whether such features as information imperfections, etc. exist rather than use as a starting point a model which precludes them? Furthermore, were not the nested logit, probit, dogit and GEV models inspired by simple deficiencies in the traditional theoretical base? Unfortunately the relaxation of additional assumptions often leads to models which are not particularly amenable to empirical investigation. We can, of course, always resort to simulation as adopted in this paper but to estimate such models *in practise* will not be inexpensive unless approximations are found. Moreover the immediate consequence of such generalisations (even if relatively simple models could be found) is an increase in the number of parameters to be determined, raising problems associated with their estimation. Would we have any more confidence then that we had captured the essentials of human behaviour suitable for predicting response? Our concern is that with the present emphasis on cross-sectional models and revealed preference estimation methods, it would seem difficult, if not impossible, to discriminate in practise between hypotheses relating to behavioural *responses*.

Let us return to the problem of "calibration". It has been argued and we have some sympathy with the view, that

> "the model should not be calibrated on the same travel components that it was called to reproduce, for instance a truly behavioural model based on the constraints under which travel choices are made should produce such choices independently that would then be *compared* with the observed choices for its validity and not *calibrated* to them..." (Zahavi, 1978).

This represents a basic sentiment and optimism behind the transferability issue and ultimately the notion that some "universal law" may be found to explain behaviour. The search for transferability however seems to be an elusive one—indeed commentators are yet to agree on whether human behaviour is extraordinarily complex or amenable to a relatively simple explanation.

The "new option" policy will always prove somewhat of an embarrassment for current methods simply because the attributes of existing choices may not be satisfactorily matched to those apparent in the test system and, as more "unconventional" policies are introduced, we will have to rely increasingly on stated preference methods for determining elasticities. In spite of their limitations, transfer pricing methods should, we believe, be more widely explored (see the discussions by Daly, 1978; and Bonsall, 1979).

It is a characteristic of a number of recent approaches to travel forecasting, that different emphases are placed both on the *sources of variability* on the one hand, and the means for gleening information on *response*. This is particularly evident in some constraint-based activity-travel analyses, which are geared towards the response context. In the HATS framework, for example, (Jones, 1979), it is attempted through loose structuring of interviews to decrease the likelihood that an observer (in this case the interviewer) will impose an inappropriate framework for the analysis of behavioural response. At present such studies are in their infancy, and must be broadened when significant degrees of choice are found to be relevant to the decision context. Traditional problems may be met in representing in a formal analytical framework (if indeed this is indeed this is sought) and distinguishing between the influences of constraints *and* preferences.

In our opinion real progress in understanding and assessing the effectiveness of forecasting models will be made only when more information on behavioural *response* becomes available—and it would be useful if this exercise were co-ordinated. From a theoretical viewpoint, it is desirable that appropriate frameworks for the analysis of choice contexts be designed which allow the *means* for both the direct testing and *refutation* of hypotheses relating to response to be established. Until this is achieved, the problems of misrepresentation reflecting the uncertainty of the observer, will we fear continue to plague cross-sectional studies.

In the paper we have addressed a number of themes concerned with the general debate on the validity of travel forecasting models. During the process of research we have focused on a number of related questions, specifically: what is the *nature* of the criticisms raised against the current generation of travel choice models? Under what conditions are they justified? Can a model—which is, after all, simply a functional relation—over-ride the criticism raised against its underpinning theory? Do alternative theoretical postulates produce empirically testable differences? To what extent can an observer (the modeller) be indifferent to the many facets of behaviour which may give rise to the measured dispersion in a data set? To what extent does the framework set up by an observer influence the interpretation of data? Does the choice of alternative models lead to major differences in the interpretation of evidence even when the models differ little in principle? To what level of realism are we committed if we assume the task of building a "behavioural" model; and, more centrally, can the elasticity parameters estimated from cross-sectional data give a reasonable indication of behavioural response?

Although we have touched upon all these topics we would not of course, claim to have provided comprehensive answers. Even less would we claim to have a monopoly of interest in them! We would, however, suggest that they are important, under-researched, and deserve a more explicit treatment in the literature.

Acknowledgements—We have benefited from discussions of our ideas—at various stages of their development—in seminar sessions in the Institute for Transport Studies, University of Leeds, and at the Annual Summer meetings of P.T.R.C., University of Warwick in July 1979. It is a pleasure to acknowledge, in particular, discussions with Hugh Gunn who also read and commented on an earlier draft of this report. As usual, we retain responsibility for opinions, errors and omissions.

We are also very grateful to Marguarita Green for her excellent work on the diagrams and to Mandy Walton for her patience and skill in typing the paper.

H.C.W.L. Williams and J. D. Ortuzar gratefully acknowledge financial support from the S.S.R.C. and S.R.C. respectively.

REFERENCES

Albright R. L., Lerman S. R. and Manski C. F. (1977) Report on the development of an estimation program for the multinomial probit model. Prepared for the Federal Highway Administration, U.S. Department of Transportation, by Cambridge Systematics Inc.

Banister D. J. (1978) The influence of habit formation on modal choice: a heuristic model. *Transportation*, 7, 5–18.

Ben-Akiva M. E. (1974) Structure of passenger travel demand models. *Transpn Res. Rec.* **526**, 26–42.

Ben-Akiva M. E. (1977a) Passenger travel demand forecasting: applications of disaggregate models and directions for research. *Proc. World Conf. on Transportation Research*, Rotterdam.

Ben-Akiva M. E. (1977b) Choice models with simple choice set generating processes. *Working Paper*, Department of Civil Engineering, Massachusetts Institute of Technology, U.S.A.

Ben-Akiva, M. E. (1979) Issues in transferring and updating travel behaviour models. *4th Int. Conf. on Behavioural Travel Modelling* 1979, Eibsee, July 1979.

Blase J. H. (1979) Hysteresis and catastrophe theory: a demonstration of habit and threshold effects in travel behaviour. *PTRC Proc. Summer Annual Meeting*, University of Warwick.

Bonsall P. W. (1979) Microsimulation of mode choice: a model of organised car sharing. *PTRC Proc. Summer Annual Meeting*, University of Warwick.

Brand D. (1973) Travel demand forecasting, some foundations and a review. In *Urban Travel Demand Forecasting*, Special Rep. 143, Highway Research Board, Washington D.C.

Burnett P. and Hanson S. (1979) A rationale for an alternative mathematical approach to movement as complex human behaviour. *PTRC Proc., Summer Annual Meeting*, University of Warwick.

Burnett K. P. and Thrift, N. J. (1979) New approaches to understanding traveller behaviour. In *Behavioural Travel Modelling* (Edited by Hensher D. A. and Stopher P. R.) Croom Helm, London.

Cardell N. S. and Reddy, B. J. (1977) A multinomial logit model which permits variation in tastes across individuals. Charles River Associates Inc. (Unpublished).

Cochrane R. A. (1975) A possible economic basis for the gravity model. *J. Transport Econ. Policy* **9**, 34–49.

Daganzo C. F., Bouthelier F. and Sheffi Y. (1977) Multinomial probit and qualitative choice: a computationally efficient algorithm. *Transpn Sci.* **11**, 338–358.

Daly A. J. (1978) Issues in the estimation of journey attribute values. In *Determinants of Travel Choice* (Edited by Hensher D. A. and Dalvi M. Q.). Saxon House, Sussex.

Daly A. J. and Zachary S. (1978) Improved multiple choice models. In *Determinants of Travel Choice* (Edited by Hensher D. A. and Dalvi M. Q.) Saxon House, Sussex.

Deaton A. (1975) *Models and Projections of Demand in Post-War Britain.* Chapman & Hall, London.

Domencich T. and McFadden, D. (1975) *Urban Travel Demand—A Behavioural Analysis.* North Holland, Amsterdam.

Eilon S. (1972) Goals and constraints in decision-making. *Oper. Res. Quart.* **23**(1), 3–15.

Foerster J. F. (1979) Mode choice decision process models: a comparison of compensatory and non-compensatory structures. *Transpn Res.* **13A** (1), 17–28.

Gaudry M. J. I. and Dagenais, M. G. (1979) The DOGIT model *Transpn* Res. **13B** (2), 105–112.

Gaudry M. J. I. and Wills, M. (1977) Estimating the functional form of travel demand models. *Publication* **63**, Centre de recherche sur les transports, Université de Montréal.

Gensch D. H. and Svestka, J. A. (1978) An exact hierarchical algorithm for determining aggregate statistics from individual choice data. *Working Paper*, University of Wisconsin at Milwaukee.

Golob T. F. and Richardson, A. J. (1979) Non-compensatory and discontinuous constructs in travel behaviour models. *Rep. Workshop* **6**, *4th Int. Conf. on Behavioural Travel Modelling* 1979, Eibsee, July

Goodwin P. B. (1977) Habit and hysteresis in mode choice. *Urban Studies* **14**, 95–98.

Goodwin P. B. (1979) Two problems of travellers' behaviour in mathematical models. *Bull. Instit. Math. Applic.* **15**, 300–303.

Gould P. and White, R. (1974) *Mental Maps.* Pelican Books, Harmonsworth, England.

Hausman J. A. and Wise, D. A. (1978) A conditional probit model for qualitative choice: discrete decisions recognising inter-dependence and heterogeneous preferences. *Econometrica* **46**, 403–426.

Heggie I. G. (1978) Putting behaviour into behavioural models of travel choice *J. Oper. Res. Soc.* **29**, 541–550.

Heggie I. G. and Jones P. M. (1978) Defining domains for models of travel demand. *Transpn* **7**, 110–125.

Hensher D. A. (1975) Perception and commuter modal choice: an hypothesis. *Urban Studies* **12**, 101–104.

Hensher D. A. (1979) Behavioural demand modelling: some recent issues and possible new direction, in particular choice set generation, endogenous attributes decomposition of utility functions and congestion modelling. *Regional Transport Research Workshop on Methods and Concepts in Transportation Modelling*, Melbourne, Jan.

Horowitz J. (1979a) The accuracy of the multinomial logit model as an approximation to the multinomial probit model of travel demand. U.S. Environmental Protection Agency, Washington D.C.

Horowitz J. (1979b) Identification and diagnosis of specification errors in the multinomial logit model. U.S. Environmental Protection Agency, Washington D.C.

Horowitz J. (1979c) Sources of error and uncertainty in behavioural travel demand models. Paper presented at the 4th Int. Conf. on Behavioural Travel Modelling, Grainau, Federal Republic of Germany.

Jones P. M. (1979) 'HATS': a technique for investigating household decisions. *Envnmt Plan.* **11**, 59–70.

Kent County Council (1976) Report on experimental park-and-ride services. *Maidstone Transportation Study*, Maidstone, England.

Kirby H. R. (1979) Is distance no object? Location behaviour and the journey to work. Presented to the *11th Annual Conf.* University Transport Study Group, University of Southampton. (Unpublished).

Lerman S. R. and Louviere, J. (1978) The use of functional measurement to identify the form of utility functions in travel demand models. *Transpn Res. Rec.* **673**, 78–85.

Louviere J. J. (1978) Modelling individual residental preferences: a totally disaggregate approach. *Tech. Rep.* **100**, Institute of Urban and Regional Research, University of Iowa.

Louviere J. J. (1979) Some comments on premature expectations regarding spatial, temporal, and cultural transferability of travel choice models. *4th Int. Conf. on Behavioural Travel Modelling*, Eibsee, July.

Makowsky G. G., Bakr M. M. and Sinha K. C. (1977) Estimating transportation mode utilities. *Transp. Engng J. ASCE* **103** (TE6), 807–823.

Manski C. F. (1977) The strucure of random utility models. *Theory and Decision* **8**, 229–254.

Manski C. F. and Lerman S. R. (1978) On the use of simulated frequencies to approximate choice probabilities. Cambridge Systematics Inc.

McFadden D. (1975a) Economic applications of psychological choice models. *Working Paper* 7519, Travel Demand Forecasting Project, Institute of Transportation and Traffic Engineering, University of California at Berkeley.

McFadden, D. (1975b) On independence, structure and simultaneity in transportation demand analysis. *Working Paper* 7511, Institute of Transportation Studies, University of California, Berkeley.

McFadden, D. (1976) The theory and practice of disaggregate demand forecasting for various modes of urban transportation. *Working paper* 7623, Institute of Transportation Studies, University of California at Berkeley.

McFadden, D. (1978a) On the use of probabilistic choice models in economics. *American Economic Association Meeting*, Chicago, Aug.

McFadden, D. (1978b) Modelling the choice of residential location. In *Spatial Interaction Theory and Planning Models* (Edited by Karlqvist A., Lundqvist L., Snickars F. and Weibull J. W.). North Holland, Amsterdam.

McFadden, D. (1979) Quantitative methods for analysing travel behaviour of individuals: some recent developments. In *Behavioural Travel Modelling* (Edited by Hensher D. A. and Stopher P. R.). Croom Helm, London.

McFadden D. (1980) Econometric models of pro .bilistic choice. In *Econometric Analysis of Discrete Data* (Edited by Manski C. F. and McFadden D.). M.I.T. Press Cambridge, Mass. (to appear).

McFadden, D., Tye, W. and Train K. (1976) Diagnostic tests for the independence from irrelevant alternatives property of the multi nomial logit model. *Working Paper* 7616, Institute of Transportation Studies, University of California, Berkeley.

Ortuzar J. D. (1978) Mixed mode demand forecasting techniques: an assessment of current practice. *PTRC Proc. Summer Annual Meeting*, University of Warwick.

Ortuzar J. D. (1979) Testing the theoretical accuracy of travel choice models using Monte Carlo simulation. *Working Paper* **125**, Institute for Transport Studies, University of Leeds.

Papaoulias D. B. and Heggie I. G. (1976) A comparative evaluation of forecast and use of park-and-ride in Oxford. *Traffic Engng Control*, **17**(4), 144–149.

Recker, W. W. and Golob T. F. (1979) A non-compensatory model of transportation behaviour based on sequential consideration of attributes. *Transpn Res.* **13B**, 269–280.

Richardson A. J. (1978) A comparative analysis of transport choice models. *Working Paper* 78/11, Civil Engineering, Monash University, Australia.

Robertson D. I. (1977) A deterministic method of assigning traffic to multiple routes of known cost. Department of the Environment, Department of Transport. *TRRL Rep.* 757, Crowthorne.

Robertson, D. I. and Kennedy J. V. (1979) The choice of route, mode, origin and destination by calculation and simulation. Department of the Environment, Department of Transport, *TRRL Rep.* 877, Crowthorne.

Senior, M. L. and Williams, H. C. W. L. (1977) Model-based transport policy assessment, I: the use of alternative forecasting models. *Traffic Engng Control* **18**, 402–6.

Simon H. A. (1955) A behavioural model of rational choice. *Quart. J. Econ.* **69**, 99–118.

Sobel K. L. (1980) Travel demand forecasting with the nested multinomial logit model. *50th Annual Meeting of the Transportation Research Board*, Washington, D.C., Jan.

Stopher, P. R. and Meybury A. H. (Eds.) (1976) *Behavioural Travel Demand Models.* Lexington Books, Heath & Co., Lexington, Mass.

Train, K. (1978) A validation test of a disaggregate mode choice model. *Transpn Res.* **12**(3), 167–174.

Tversky A. (1972) Elimination by aspects: a theory of choice. *Psychological Rev.* **79**(4), 281–299.

Weibull, J. W. (1978) A search model for microeconomic analysis with spatial applications. In *Spatial Interaction Theory and Planning Models.* (Edited by Karlqvist A., Lundqvist L., Snickars F. and Weibull J. W.) North Holland, Amsterdam.

Williams H. C. W. L. (1977) On the formation of travel demand models and economic evaluation measures of user benefit. *Envnmt Plan.* **A 9**, 285–344.

Williams H. C. W. L. (1980) Travel demand forecasting: an overview of theoretical developments. In *Transport and Public Policy Planning.* (Edited by Banister D. and Hall P.) Mansell Press.

Williams H. C. W. L. and Ortuzar J. D. (1979) Behavioural travel theories, model specification and the response error problem. *PTRC Summer Annual Meeting*, University of Warwick.

Williams, H. C. W. L. and Senior, M. L. (1977) Model based transport policy assessment, II: Removing fundamental inconsistencies from the models. *Traffic Engng Control* **18**, 464–469.

Thrift N. J. and Williams H. C. W. L. (1981) On the development of behavioural residential and industrial location model. *Econ. Geog.* (in press).

Wilson A. G. (1976) Catastrophe theory and urban modelling: an application to modal choice. *Envnmt Plan. A*, **8**, 351–356.

Young W. and Richardson A. J. (1978) Residential preference: a perceptual view of Melbourne. *Proc. 4th Australian Research Forum.*

Zahavi Y. (1978) Can transport policy decisions change urban structure? *PTRC Proc. Summer Annual Meeting*, University of Warwick.