Working Paper 309

THE USE OF MICROSIMULATION IN TRAVEL-ACTIVITY ANALYSIS

H.C.W.L. Williams

School of Geography,
University of Leeds,
Leeds LS2 9JT.

July, 1981

ABSTRACT

While the use of "microsimulation" and Monte-Carlo methods in the Social Sciences may be traced to the mid 1950's, and specifically to dynamic studies of household behaviour, their use in travel/activity analysis is largely confined to the last decade. The adoption of a simulation approach for model solution has resulted both from the requirement of more detailed information generated by models and from attempts to improve the specification and aggregation of micro-relations.

In recent years a number of different applications, drawn from many different travel demand modelling contexts, have been reported which employ these procedures. It is partly through their discussion that we wish to consider the characteristics and potential of the simulation approach. By employing a common notation and drawing on simple probability statements, some of the features of these models which necessitate the use of sampling and simulation procedures, are made explicit. Provision is also made for comparison with related constructs and possible alternative models.

We discuss the general context of simulation in the development of behavioural models, and offer some views on the potential and drawbacks of employing such techniques for forecasting, policy exploration, evaluation and design. Throughout we emphasize the need to distinguish carefully solution procedures from model statements, with which they become easily entwined , so as to prevent an artificial distinction between "simulation models" and more conventional analytic counterparts.

INTRODUCTION

Over the last decade the micro-approach to the study of travel and transport policy impacts has been embraced widely and for a variety of reasons. These range from specific considerations relating to data efficiency, aggregation bias and information loss in comparison with more traditional methods, through the aspirations towards an integrated *behavioural* theory of travel, to the provision of what some have seen as an appropriate 'social policy' perspective to planning and problem analysis. In this period demand model research and applications have been dominated by a class of *analytic* micro-models typified by members of the logit family which are underpinned by discrete choice (random utility) theory. These have served both for the description and explanation of statistical patterns of travel behaviour, and as a basis for forecasting demand response to policy measures. With the greater availability of appropriate computer programs the models have been widely and are increasingly adopted for examining different aspects of decision making in travel related contexts.

Although present attitudes towards "conventional" micro-models show a predictable spread between profound dissatisfaction on the one hand, to the feeling that they are of considerable practical value on the other, it is increasingly held that their behavioural foundation is too simplistic for examining issues both of traditonal concern, and for addressing problems and policies of current interest in transport analysis. That the variability, interactions, interdependencies and the processes of choice formation are inappropriately treated is an argument commonly used in promoting new approaches.

In travel analysis and urban research, as in economics and the social sciences more generally, microsimulation is emerging as an increasingly adopted tood for examining problems in which the detail and subtleties of human affairs suggest an abandonment of the search for relatively simple *analytic* models which permit a general relationship between variables of that problem in favour of the computational manipulation - according to model relations - of ('pseudo') individuals or samples drawn from probability distributions. In this paper we offer some views on the characteristics, development and potential of micro-simulation in travel-activity analysis. Before embarking on the applications of microsimulation to this field we would make two general points related firstly, to the terminology and traditions of simulation and, secondly, on its development for social systems analysis.

Although specific characteristics of and motivation for the technique will emerge later, it has been observed generally that *simulation* has stood as a 'catch all' term embracing the computational analysis of systems which receive a formal representation in terms of model relations. Simulation is, in fact, not infrequently taken as synonymous with the representation, computational implementation and solution of models, and *microsimulation* is in this sense the computational analysis of relations characterising the behaviour of individuals and/or groups of individuals, households or other 'micro-units'. In more formal discussions of simulation, however, more precise characteristics, applications and traditions of the approach are usually identified. Tocher (1963), for instance, refers to the origins and problem contexts encountered in Statistics, Applied Mathematics and Operational Research which led to the establishment of simulation as a distinct approach to *practical* problem analysis. Specifically, simulation became associated with:

* the formulation of *sampling experiments* as a means of solving problems in statistics. These involved the formation and manipulation (transformation) of samples drawn from proability distributions, and examination of the characteristics of the resultant frequency distributions, particularly in comparison with the predictions of theory;

* the solution of a class of complex mathematical problems arising from the *deterministic* representation of systems, through the construction of formally equivalent problems in probability which were then amenable to solution by sampling experiments. This idea, which was christened the 'Monte-Carlo' approach by Von Neumann and Ulam has a powerful application in the evaluation of multidimensional integrals. As such, it makes an important contribution to the practical study of micro-model formation and the *aggregation* of micro-behavioural relations as described in the next section;

* the solution of probabilistic models subject to complex restrictions and/or irregular 'boundary' conditions such as in the representation of realistic queueing problems. This context of simulation is widely encountered in the examination of traffic flow systems in which supply and/or demand have a stochastic representation, as for example in the analysis of: bus lanes, junction characteristics, traffic signals, and demand-reponsive transport systems.

In all of these approaches the theory of probability arises as a common element. In this paper we shall not place too fine a restriction on what is understood by the term *simulation*, although those 'Monte-Carlo' applications which involve the manipulation of samples of micro-units whose characteristics or behaviour are determined by reference to probability distributions will receive particular attention.

Although *micro-simulation*, as such, is a terminology of relatively recent use in the literature on models of travel and activity analysis, its adoption and exploitation in the study of social systems, as in the physical and management sciences, is longstanding with its origins closely associated with the development of 'large' computers in the middle of the 1950's. The work of Orcutt and his colleagues in the United States in the late 1950's (Orcutt *et al.*, 1961) was among the first, and certainly constitutes the most substantial research effort, in the application to social systems, and culminated in a model of household dynamics and income (DYNASIM) which became operational recently. Despite its long traditions, the field, measured by the number of applications of the approach, has been slow to develop (an interesting feature itself to which we shall later refer) and there are relatively few readily available publications on the subject. The recent books by Orcutt *et al.* (1976), Haveman and Hollenbeck (1980), and the papers by Harris (1978) and Clarke, Keys and Williams (1981) review some recent applications in the study of socio-economic systems. It should be remarked that as a consequence of the dearth of publications in this area, the general microsimulation strategy, together with some of its particular features, become 're-invented' from time to time in circumstances where the problems tackled have rather similar characteristics.

Because microsimulation is essentially a tool for practical problem analysis, the present paper will centre its discussion on applications and application contexts. This discussion is, however, preceded by some general remarks in the following section relating to micro-representations and the behavioural approach and, in particular, the examination of decision contexts, which underpin a number of the simulation applications. In this section we also include an overview of the philosophy and theoretical assumptions associated with the "conventional" micro-approach. This aspect is important for three related reasons: firstly, it established a link between current models and more general behavioural constructs; secondly, current models may employ simulation techniques in their implementation; and, thirdly, a number of the applications of simulation use current *analytical* approaches as points of departure in what are considered to be improved specifications.

In the section following this general discussion, several applications of microsimulation applied to travel/activity contexts and socio-economic dynamics are described. As we indicated above, these applications will reflect different interpretations, characteristics and motivation for the simulation approach which we attempt to draw out. In a further section some general application contexts and specific research issues are examined, with regard to a consideration of the potential and applicability of microsimulation. In a concluding section we summarise and comment on what are believed to be the main advantages and drawbacks of the simulation approach as described in the paper.

MICROANALYSIS: MODEL DEVELOPMENT AND SOLUTION PROCEDURES

The modelling process essentially involves the transformation of information which is referred to some representation of the state of a system. The outputs of this process will relate to the specific purpose for its development, and may involve aggregate quantities of the traditional kind such as link flows, modal demand, interzonal trips and system-wide performance measures, or/and detailed information on the *distributions* of individual and household characteristics, under different policy measures. In this paper we shall not be concerned explicitly with different strategies of model development - viewed as distinct forms of transformation - which are discussed, for example, by Williams and Ortuzar (1981), but shall accept a commitment to establish micro-relations to be aggregated as appropriate. In 'conventional' micro-procedures this is understood as the process of specification, estimation and forecasting with probabilistic choice models.

In order to provide a general context for simulation modelling with micro-data sets, we must discuss in turn: micro-state representations; decision contexts; the aggregation of micro-relations; and solution procedures.

State Representation and Microanalysis

Central to the formal analysis of individual and household behaviour, whether in travel or more general contexts, is the provision of an unambiguous description of the system in the form of a state representation. In the present context this involves the appropriate description of individuals and their associated households (micro-units) in terms of demographic, social and economic characteristics together with information on various activities performed and the associated travel. The state of the system as a whole may then be identified in terms of the states of individual micro-units.

For the purpose of state definition we shall associate with each individual i of a population Π_I of size N_I a set of M_I characteristics $x_1, \ldots x_\mu, \ldots x_M$, such as age, household status, sex, income, residential location, etc., and shall suppose that each characteristic x_μ, $\mu=1\ldots M$ may occupy one of n_μ categories or classes. If it is necessary to refer to the particular class of an individual we shall do so explcitly, otherwise an (arbitrary) state of an individual will be summarised in the vector

$$\underline{x}_I^i = (x_1^i, \ldots x_\mu^i, \ldots x_M^i)$$

the components of which contain the classes occupied for each of the M characteristics. This notational procedure is, of course, nothing other than is involved in the coding of a (travel) survey.

The state of the system of individuals may then be described by a vector X_I^*, the components of which are the N_I vectors associated with the micro-units

$$X_I^* = (\underline{x}_I^1, \ldots \underline{x}_I^i, \ldots, \underline{x}_I^{N_I}).$$

In a similar way we may refer to the state of a population of households with the vector

$$X_H^* = (\underline{x}_H^1 \ldots, \underline{x}_H^h \ldots \underline{x}_H^{N_H}).$$

A summary of the notation to be used in the paper is as follows:

Π_I, Π_H : Populations of N_I individuals and N_H households, respectively.

$\underline{x}_I^i$, $\underline{x}_H^h$: Vectors of attributes associated with individual i and household h, respectively.

$\underline{X}_I^*$, $\underline{X}_H^*$: System state vectors containing the components $\underline{x}_I^i$, $i=1\ldots N_I$ and $\underline{x}_H^h$, respectively.

S_h : The set of individuals belonging to household h of size n(h).

$S_I^{\nu}(\underline{x}_I)$, $S_H^{\nu'}(\underline{x}_H)$: Descriptions of samples drawn from Π_I and Π_H. ν, ν' will include information on size and sampling strategy.

$N_I(\underline{x}_I)$, $N_H(\underline{x}_H)$: Occupancy matrices, the elements of which are the numbers of individuals or households in states formed from given combinations of attributes. Thus $N_I(\chi_1,\chi_2,\ldots\chi_M)$ is the number of individuals in the state defined by class χ_1, of category x_1, χ_2 of x_2, etc.

$P_I(\underline{x}_I)$, $P_H(\underline{x}_H)$: Probability distributions expressing the variation of the vectors $\underline{x}^i$ and $\underline{x}^h$ over the populations Π_I and Π_H. These will be referred to as joint (multivariate) distributions of the characteristics.

The vector $\underline{x}_H^h$ will, in general, be of the form

$$\underline{x}_H^h = (\underline{x}_I^i,\ i=1\ldots n(h);\ \underline{y}^h)$$

containing the attributes of component members, together with an additional set $\underline{y}^h$, specifically relating to the household. $\underline{y}^h$ will contain information derived from individual attributes, such as total income, consuption vectors, etc. Both individual and household vectors will further contain identifiers which allow cross-referencing, preventing a loss of information if separate individual and household lists are formed.

It should be noted that the individual identifiers $\underline{x}^i$ may contain information relating not only to the current state, but also (longitudinal) information derived from previous time periods. In dynamic models we shall be concerned explicitly with the possible transitions between individual states. The following quantities will therefore be defined:

$N(\underline{x},t:\underline{x}',t')$: the number of individuals (or households) making transitions between states $\underline{x}$ and $\underline{x}'$ in the time interval (t,t');

$P(\underline{x},t:\underline{x}',t')$: the probability that a transition between the states $\underline{x}$ and $\underline{x}'$ will be made by an individual in the time interval.

The transitions discussed here refer to different classes and combination of classes of the various characteristics.

It is appropriate at this stage to comment on the issue of *computational efficiency* of forms of state representation, a feature which has been mentioned by such authors as Orcutt *et al.* (1961), Wilson and Pownall (1976), and Kain *et al.* (1976). Recall that in 'conventional aggregate modelling procedures' information was stored, manipulated and output essentially in the form of occupancy matrices. If referred to a set of individual attributes of size M^*, $\prod_{\mu=1}^{M^*} n_\mu$ cells are required to store this information. When it is sought to increase M^*, either in response to the generational of new forms of distributional information, or to refine model specification, the storage requirement of this 'number representation' increases alarmingly. Furthermore, the vast majority of these cells will be filled with zero entries, corresponding to the very low probability of observing an individual with a *given* combination of several attributes (see the proof that 'every girl is one in a billion' Moroney, 1951, p.8). In contrast, the M^*N cell storage requirement for identifying the state of a system by listing the attributes of individuals *directly* (which essentially contains the same information) will typically be very modest, and this is particularly the case if samples are taken from the population Π. This argument, of course, holds with even greater force in relation to the storage of information relating to transitions between states.

Thus it is argued that by manipulating samples of information pertaining to individuals in the modelling process, such dimensionality problems may be eliminated or at least reduced. While this is valid for the storing of *given information*, such as a survey, ultimately the storage requirements of a *modelling process*, and the extent to which the variability and interdependency among numerous attributes has implications for computational storage and simulation involving samples drawn from distribution, must be related to, *among other things*, the assumptions involved in decomposing and simplyfying the joint probability distribution $P(\underline{x})$. For example, if $P(\underline{x})$ is written in the form

$$P(\underline{x}) = P(x_1)P(x_2|x_1) \; \ldots\ldots \; P(x_M|x_1,x_2\ldots\ldots x_{M-1}) \qquad (1)$$

the storage requirements of the information contained in this string of distributions will clearly depend on what interdependencies are assumed in the models developed for the conditional probabilities. A similar argument may be used in a consideration of the interaction between the processes in modelling the transition matrices. These general comments on the relationship between variability, conditional dependency and representational efficiency are also relevant to a consideration of data storage and information loss as discussed by Clarke, Keys and Williams (1981).

Before considering the procedure of model development in greater detail let us summarise the broad contexts in which sampling procedures and simulation applications may be found, characterised by the information contained in, and required from, the probability distributions $P(\underline{x})$:

* the generation and examination of micro-data sets, considered as samples drawn from $P(\underline{x})$

* the construction and solution of *forecasting* models for examination of the interaction between demographic, social and economic processes : $P(\underline{x},t)$

* an examination of the *response* of a population $\delta P^{*}_{G}(\underline{x},t)$ to policy measures G at some time t

* the determination of a policy G^{*} which meets the objectives of a *design process* for which the response of apopulation is $\delta P_{G^*}(\underline{x},t)$.

While the use of micro-data sets does not necessarily imply either a *behavioural* (or indeed *simulation*) approach to the construction and solution of models for the above contexts, it has become a common practice to seek an *explanation* for the present state of individuals and of the system as a whole - that is, the *variability* and *conditional dependency* reflected in $P(\underline{x})$- together with information on the dynamics and response characteristics, in terms of the *decision contexts* experienced by individuals and households. It is to a consideration of these that we now turn.

Decision Contexts

It is outside the scope of the present paper to discuss in detail the conceptualisation of decision *processes* and the means by which response patterns are established. These have been considered in particular contexts by Heggie (1978), Chapin (1978), Richardson (1980), Williams and Ortuzar (1981) among many others. Such discussions tend to vary in their formal content, and range between those which construct general scenarios of how information is acquired, processed and acted on by homo-psychologicus, those which condense out of these considerations a specific representation (model) of decision making in algebraic terms.

A fairly common conceptualisation of the process, which serves as a basis for constructing decision models involves, for each relevant decision maker i, the identification of

* all possible choices, $\underline{C}^i$;
* the set of constraints $\underline{K}^i$ to eliminate infeasible choices;
* the objectives $\underline{Q}^i$ which motivate the choice process;
* a set of perceived attributes related to the set of alternatives; which characterise the solution process.

A decision rule D^i is now constructed to allow the transformation of this information into a specific choice(s) in a deterministic or probabilistic model.

It is not necessary to elaborate on the heroic nature of the quest to generate an *objective* (and typically simple) decision rule from an essentially subjective process to serve as a basis for response prediction. Two general comments are, however, appropriate at this stage. Firstly, while recent work in travel/activity analysis has tended to emphasize the variety of *constraints* in the decision making process, seemingly in contrast to the traditional emphases on choices, the decision 'paradigm' not only provides a unifying context for the *symmetric* treatment of choices and constraints, but more generally allows a classification of different hypotheses of behaviour. It should be added that an emphasis on constraints is not infrequently associated with both a suspicion of 'revealed preference' methods, but also with a reluctance to accept that individual preferences should figure dominantly or even prominently in a planning context, a view advanced particularly eloquently by members of the "Swedish School" and discussed, for example, by Hägerstrand (1970), Lundqvist (1978) and Lenntorp (1978).

Secondly, there appears to be a growing reluctance to take the necessary leap required to reduce models of individual decision making to analytic form. To avoid a "corruption by mathematics" resort is increasingly being made to stated intention procedures within gaming constructs (see Jones, 1979 and Burnett, 1978), and/or through the adoption of 'interactive computing' procedures. To what extent the use of these devices will serve as a prelude to the construction of more formal modelling procedures, allowing abstractions of the traditional kind, awaits further developments.

It is hardly necessary to remark that the choices for model development implied by these comments relate to some of the most contentious issues in travel/activity analysis.

Within the behavioural approach many different (deterministic) rules have been suggested for resolving the decision making process and thus identifying distinct choices with an individual, while both 'revealed preference' and 'stated intention' methods have been used for quantifying these relations. Williams and Ortuzar (1981) have discussed the formation of decision rules within the framework of a multicriterion programming problem, which provides a unifying context for different hypothesized forms of behaviour, including 'compensatory', 'non-compensatory' and 'hybrid' schemes.

In what follows we shall be concerned at several stages with the formation of micro-models which employ specific decision making rules denoted as D^i. In effect, the range of all possible choices $\underline{C}^i$ will relate to the classes x_μ^ρ, $\rho=1\ldots n_\mu$; associated with one or more of the characteristics chosen to label an individual or household.

The Formation and Aggregation of Micro-relations

In developing models of travel behaviour it is necessary and customary to relate the differences in individual actions and response to the variation of *observable* and *unobservable* factors attributed to members of the population Π (or samples S drawn from it). To this end we shall use the vectors $\underline{w}^i$ and $\underline{\varepsilon}^i$ to denote the observable and unobservable characteristics of individual i which are considered relevant to the choice process. $\underline{w}$ will in general contain measured quantities relating to: the attributes Z_ρ of the alternative choices C_ρ in the possible choice set; the level of transport service $\underline{L}$; and socio-economic-demographic characteristics $\underline{S}$ which will in general be contained in the vector $\underline{x}$. $\underline{\varepsilon}$ on the other hand will refer to such factors as the individual valuation of attributes, or perhaps unmeasured quantities such as comfort, convenience, etc.

We shall now use

$$D^i = D^i(\underline{w}^i, \underline{\varepsilon}^i \mid \underline{C}, \underline{K}, \underline{Q}) \qquad (2)$$

to denote the decision model pertaining to individual i. This may be an extremal (optimisation) process as is employed in conventional utility maximising procedures, or some other model which is used to identify a unique choice within the possible set. This relation allows the determination of a quantity

$$R^i_\rho = R^i_\rho(\underline{w}^i, \underline{\varepsilon}^i) \qquad \text{for all } C_\rho \text{ belonging to the feasible set } C \qquad (3)$$

which specifies those combinations of $\underline{w}$ and $\underline{\varepsilon}$ which, when substituted into the decision rule (equation 2) results in the alternative C_ρ being chosen. R_ρ thus defines a region in the space of all possible values of $\underline{w}$ and $\underline{\varepsilon}$. The decision rule will be taken to be of *deterministic* form, and probability relations will be invoked to accommodate the uncertainty *of the modeller* on the exact choice which will result from a decision context at the individual level.

While the process of *behavioural specification* is concerned with the nature of the model (2), the *process of aggregation* essentially accumulates contributions to any particular choice (or aggregate quantity) from individuals in the system whose characteristics $\underline{w}$ and $\underline{\varepsilon}$ are distributed in a form $F(\underline{w},\underline{\varepsilon})$ over the population. In the aggregate procedure it is necessary to derive quantities such as $N(\underline{T})$ the occupancy matrix, defined by the classes associated with the vector of characteristics $\underline{T}$, from relations which are specified and estimated at the level of the individual. The size and classes of the vector $\underline{T}$ will not in general correspond to those in the vector $\underline{x}$. Symbolically, we shall write the process as follows:

$$N(\underline{T}) = \underset{R_T(x_\rho,\underline{w},\underline{\varepsilon})}{S} F(\underline{w},\underline{\varepsilon}) \qquad (4)$$

in which $\underset{R_T}{S}$ denotes the accumulation (integration over continuous variables and summation over discrete variables) over all combinations of $\underline{w}$ and $\underline{\varepsilon}$ which result in the class x_ρ being occupied, and which are associated with the individual classes in the vector $\underline{T}$. Rather than pursue the general case we shall assume, as is typically the case that the choices C_ρ correspond to the classes of one or more combinations of the characteristics common to both $\underline{x}$ and $\underline{T}$.

When expressed in terms of the specific variation of the components of the various vectors $\underline{\varepsilon}$ and $\underline{w}$, this conceptually straightforward process of accumulation appears as a forbidding multiple integral in which the region of integration required to relate micro-behaviour to the aggregate classes is *in general* highly irregular. An attempt may be made to solve this problem either analytically or numerically (including Monte Carlo simulation) and an example of where this has been done will be described in the following section. In the context of micro-model specification and estimation it is appropriate to consider the decomposition of this multidimensional simulation procedure, and represent the aggregation process symbolically as follows:

$$\begin{array}{c}\text{Aggregate}\\ \text{information}\end{array} = \underset{R(w)}{S}\ F(w)\ \underset{R_\rho(\varepsilon|w)}{S}\ F(\underline{\varepsilon}|\underline{w}) \tag{5}$$

in which we define

$F(w)$: the distribution of the vector of observable attributes

$F(\underline{\varepsilon}|\underline{w})$: the conditional distribution of the vector $\underline{\varepsilon}$ given the value $\underline{w}$.

The aggregation procedure now proceeds in two stages: firstly

$\underset{R_\rho(\varepsilon|\underline{w})}{S}$ accumulates the contribution from unobservables and forms a probabilistic choice model of the form:

$$P_\rho(\underline{w}) = \underset{R_\rho(\varepsilon|\underline{w})}{S} F(\varepsilon|\underline{w}) \tag{6}$$

while the accumulation over the distribution of observables then generates the aggregate information

$$\begin{array}{c}\text{Aggregation}\\ \text{information}\end{array} = \underset{R(\underline{w})}{S} F(\underline{w})P_\rho(\underline{w}) \tag{7}$$

in which $R(\underline{w})$ signifies an appropriate region of variation of the components of the $\underline{w}$ vector.

In view of the complexity of the general aggregation procedure it is not particularly surprising that modellers, often through practical necessity, have readily resorted to simplifying assumptions, and the typical combination which result in simple analytic forms of micro-model $P_\rho(\underline{w})$ will be described below. Given the adoption of simple closed form expressions, typically belonging to the logit family, the aggregation process is taken to refer to the expression (7), although here in a simulation context we shall generally make reference to the compound process (5) of aggregation over specified micro-decision rules.

There has been much study in travel demand analysis of the aggregation process expressed by equation (7), both in general terms and in specific contexts and we reference the articles by Koppleman (1976), McFadden and Reid (1975) and Watanatada and Ben-Akiva (1978) for further discussion. Here we shall merely summarise the main general approaches which involve:

* the adoption of *parametric distribution functions* for $F(\underline{w})$ and solution by *analytic* or numerical technique
* *classification* - involving the categorial representation of the distributions - with each class (consisting of a group of individuals or alternatives) represented by its average values. The aggregate quantity is then a weighted average of the choice probabilities for the class.
* *sample enumeration* in which aggregation is achieved by summing the choice probabilities for a sample of individuals and a sample of alternatives for each individual.

In the literature on aggregation, Monte Carlo integration appears as a general approach to the numerical solution of the expression (7). This process involves drawing a sample S from the (parametric) distribution $F(\underline{w})$ and for each element j in the sample the quantity $P(\underline{w}_j)$ is computed. For a sufficiently large sample the expression

$$\text{Aggregate Quantity} = \sum_{j\varepsilon S} P_\rho(\underline{w}_j) \qquad (8)$$

will approximate to the integral expression. Good examples of this use of Monte Carlo simulation may be found in the work of Watanatada and Ben-Akiva (1978) and Litinas (1980).

In practical problems combined methods, eg. sample enumeration and classification, may be adopted for different variables of integration. It should be noted that the sample enumeration approach is identical to Monte Carlo integration except that in the former case the samples are drawn from observed data while in the later they are drawn from specified distributions.

The process of Monte Carlo integration applied to the formulation of the micro-model through solution of the expression (6) is equally straightforward. Values of $\underline{\varepsilon}$ are sampled from the distribution $F(\underline{\varepsilon}|\underline{w})$ and are 'processed' by the decision model (2) to determine the selected choice. The proportion $\hat{P}_\rho(\underline{w}|\nu_N)$ in a sample of size ν_N will approximate to the required probability $P_\rho(\underline{w})$ for a sufficiently large N.

Let us now leave the general aspects of model formation and refer to those micro-models which have been fruitfully applied over the last decade.

The 'Conventional' Disaggregate Modelling Approach

Although still dominated by applications to modal choice, the scope of so called conventional disaggregate travel demand models based on random utility theory has broadened considerably in recent years and now include applications to: trip frequency, location and mode for a variety of purposes; home ownership and tenure selection; car ownership and car type choices; together with newer applications to trip chaining and activity and activity scheduling such as are described by Adler and Ben-Akiva (1979), Jacobsen (1978) and Damm (1979).

While the criticism that "such factors as life cycle, spatial and temporal constraints; and the interaction of decisions over the day and within the household, have not received sufficient *explicit* attention in these models" is valid, the extent to which they have received *implicit* recognition through problem organisation, market sequentation, choice set definition, and use of proxy variables in *some* applications has sometimes been overlooked.

The assumptions underpinning the *specification* of this class of discrete choice models is now familiar and will be outlined here, with a minimum of comment, for future reference.

By far the most popular discrete choice model, the linear in parameters multinomial logit (MNL) model is considered to be underpinned by the decision rule of utility maximisation in which parametric functions $U^i_\rho(\underline{Z}_p, \theta)$ are set up to record preferences. The alternative C_ρ is chosen if

$$U_\rho(\underline{Z}_p, \underline{\theta}) > U_{\rho'}(\underline{Z}_{p'}, \underline{\theta}) \quad \text{for all alternatives in the feasible choice set } \underline{C}, \tag{9}$$

a condition which defines the region R_ρ.

The derivation of MNL model requires the further assumption, of

(i) homogeneous choice sets for all individuals within a given market segment

(ii) independent and identical Weibull distributed utility functions

Decomposing the utility function into its traditional 'representative' and random components:

$$U_\rho(\underline{Z}_p, \underline{\theta}) = \overline{U}_\rho(\underline{Z}_p, \underline{\theta}) + \varepsilon_\rho \tag{10}$$

the MNL model

$$P_\rho = \frac{\exp(\sum_\lambda \theta_\lambda Z_\rho^\lambda)}{\sum_\rho (\exp \sum_\lambda \theta_\lambda Z_\rho^\lambda)} \tag{11}$$

is generated by the expression

$$P_\rho = \int_{R_\rho(\underline{\varepsilon})} d\varepsilon \; F(\underline{\varepsilon}) \tag{12}$$

when the representative utility component $\tilde{U}_\rho(\underline{Z}_\rho, \underline{\theta})$, involving the vector of parameters $\underline{\theta}$, has the form

$$\tilde{U}_\rho(\underline{Z}_\rho, \underline{\theta}) = \sum_\lambda \theta_\lambda Z_\rho^\lambda \tag{13}$$

For a highly restrictive class of distributions $F(\underline{w}, \varepsilon)$, integrals of the form (12) are amenable to analytic determination. Not surprisingly those like the MNL model are subject to highly restrictive conditions. Recently it has been shown that the nested logit (NL) model, which is a member of a class of General Extreme Value (GEV) models, may be generated from utility maximisation through this process. This model, which is less restrictive than the multinomial form, in that it permits the inclusion of 'simularity effects' between the alternatives in the choice set, is increasingly used in practical applications.

The Multinomial Probit (MNP) model, in which the distribution $F(\underline{\varepsilon}|\underline{w})$ is multivariate normally distributed, accommodates a more general structure of correlation between the residuals $(\varepsilon_1, \ldots, \varepsilon_\rho, \ldots, \varepsilon_n)$, and allows for interpersonal variation in the 'valuation' of travel attributes, and general 'similarity' effects. These bonuses must be paid for through analytic intractability, and resort must be made to numerical solution of the expression (12). One such approach as we described above is Monte Carlo integration, and an application of this procedure to probit analysis is discussed by Albright *et al.* (1977). We shall comment further on this application in the following section.

The solution of differential/difference equations representing dynamical behaviour

We have so far concentrated on the formation and solution of comparative static (cross-sectional) models. As indicated in the first section, important examples of microsimulation are drawn from the study of the dynamics of socio-economic systems in which the interaction of several demographic, economic and social processes may give rise to complex (and in general non-Markovian) model representations. By suitably enriching the individual state vector where necessary, Markov approximations are usually employed and typically appear in the following differential or difference equation forms

$$\frac{dN_p(t)}{dt} = \sum_{p'} A_{pp'}(t)\, N_{p'}(t) \tag{14}$$

$$N_p(t') = N_p(t) + \Big(\sum_{p'} A_{pp'}(t,t')\, N_{p'}(t)\Big)\, \Delta t \tag{15}$$

in which the elements of the matrix $\underline{\underline{A}}$ induce transitions between the classes of the components of the attribute vectors.

While these equations have a simple structure and are amenable to general solution forms, numerical procedures are invariably called for in practical cases, and this is particularly so where a complex web of inter-actions link individual and household characteristics to form a set of coupled differential or difference equations with time dependent transition matrices. Monte-Carlo simulation offers such a solution approach. Essentially, this involves the collection or synthesis of a sample S at an initial time t_0 and examining the pattern of development of *each* member under the influence of the various processes which may give rise to transitions.

Schematically we may represent this solution procedure in terms of successive operations through discrete time intervals

$$\hat{P}\, S(x,t) \rightarrow S(\underline{x},t') \tag{16}$$

in which the 'operator' $\hat{P}$ directs individual elements in the sample to the appropriate states in the interval (t,t'). The transitions are determined by the matrix $\hat{P}(\underline{x},t;\ \underline{x}',t')$ or more particularly, elements of it in the form

$$P(p,t:p',t') = F_{pp'} \tag{17}$$

Again the expression F may be in the form of empirically derived functions, analytic models or complex behavioural relations such as we have described above. The future state of an '*individual*' in S will be determined by random number draws and comparison with elements in the transition matrix. For example, if the probability that a woman with observable characteristics $\underline{w}$ will give birth in a particular year is $P(\underline{w})$ then a birth will be assumed to take place, if a random number r drawn from a flat distribution in the interval $0 \leqslant r \leqslant 1$, is such that

$$0 \leqslant r \leqslant P(\underline{w}) \tag{18}$$

Such a procedure must be repeated for all the relevant transitions for each member of the sample $S(\underline{x},t)$. This process is then repeated with the updated sample $S(\underline{x},t')$.

Technically, this process represents an integration of the dynamic equations by forward iteration through time, in which a probabilistic model has been set up whose solution properties approximate those of Equation (14) for a sufficiently large sample.

The use of simulation and Monte-Carlo integration: a summary of procedural contexts

Before considering specific applications we wish to draw attention to the different technical contexts of simulation and Monte-Carlo methods which will appear. These are summarised as follows:

* the generation of samples drawn from multivariate distributions p($\underline{x}$), which may be formed by direct synthesis from empirical data, or/and from specific model relations;

* the investigation of decision contexts and the formation of micro-models;

* the aggregation of micro-relations;

* the solution of dynamical equations.

In concluding this section we would emphasize the well known fact that if sampling procedures are used in conjunction with simulation techniques to solve either static or dynamical equations, these solutions are in the form of a *distribution* the properties of which will be determined by, among other things, the size of the sample.

SELECTED APPLICATIONS OF MICRO-DATA ANALYSIS AND SIMULATION

In this section we shall describe a number of examples which involve sampling and microsimulation techniques, and in their discussion bring out some important characteristics and motivations for the adopted approaches. These selected applications are derived from different disciplinary traditions including economics and geography, together with the more direct concerns with travel behaviour.

As we have indicated above, microsimulation may be considered as an *approach* to the solution of models specified at the level pertaining to individuals, families or households. Often there is in published work little attempt to summarise the work in the form of equations to be solved, authors tending to resort readily to flow diagrams and algorithms, with the result that different examples appear somewhat idiosyncratic. This only becomes a problem when the unfamiliarity of the solution methods cause an artificial distance to be created between a proposed model and more conventional micro-modelling approaches.

In the following examples we have tried to draw out their essential characteristics by reference to the comparative static and dynamic analyses of multivariate probability distributions, and where appropriate the decision contexts which are deemed to be the source of behavioural explanations. In this way some of the similarities of purpose may be recognised and comparisons made with possible alternative models. We discuss in turn: models of household dynamics; the examination of static interdependencies and the formation of synthetic samples; 'integrated' transport models; activity/travel frameworks; allocation models for demand and supply interactions; and the generation of discrete choice models from different decision contexts and rules.

Microsimulation studies of household dynamics and related models

Although not directly concerned with problems in the transport sector, it is not inappropriate to give precedence to the work of Orcutt *et al.* (1961,1976) on household dynamics for three related reasons: firstly, this work is of a pioneering nature in the application of microsimulation to the social sciences, and the motivation, strengths and weaknesses of the approach are common to a number of other studies; secondly, its emphasis on variability, interactions and interdependencies suggests a relevance both to the dynamic (or static) study of travel behaviour in which the importance of *life cycle* (*life-stage*) has been recognised only relatively recently, and in the economics of travel related consumption patterns; and thirdly, through the consumption and production functions of the individuals and households, dynamic links are established with other urban activites which will in general have direct implications for the transport sector in spatially defined systems.

In his books Orcutt (1961, 1976) has discussed a wide range of advantages for adopting a micro (-simulation) approach both for the *general* study of the interaction of social, demographic and economic processes, and for the detailed consideration of a wide range of policy tests. These are concerned both with *specific* statistical and aggregation issues and with a related and rather more philosophical position, which is well summarised by Haveman and Hollenbeck (1980) in their comments on the reasons for the emergence of microsimulation methods in economics:

> "First they reflect the basic tenet of micro-economics that a complex entity composed of many components can best be explained and predicted through an analysis of its constituent parts. Second, rational decision making in policy formation requires information about the benefits and costs of proposed policies and the gainers and losers experiencing these impacts. Micro-economic simulation provides policy makers with the capability of examining the *entire* distribution of effects...".

(Haveman and Hollenbeck, 1980, our emphasis).

Indeed, the DYNASIM (Dynamic Simulation of Income Model) system developed by Orcutt and his colleagues, which has been described (by them) as a 'unique social science tool' was designed to avoid many of the giant leaps of faith made by the widely accepted econometric models when they are used for macro-simulation. The reasons for constructing *microsimulation* models of economic systems and, more specifically, those relating to household behaviour are then, not surprisingly, closely related to the arguments which were expounded a decade later in support of disaggregate travel models in relation to the conventional transport study approach in the early 1970's.

While we indicated above that simplifying assumptions are generally invoked in dynamic models to render the future state of a system independent of its history, the interdependencies and time lags incorporated in the specification of DYNASIM made simulation an almost inevitable choice as a solution procedure. In spite of the complexity of the model - we shall not attempt to do justice to the years of research which were involved in the specification, estimation and verification of the micro-relations which are involved in successively updating the attributes $\underline{x}_I$ and $\underline{x}_H$ belonging to individual and family members in a population sample - the *conceptulisation* of that solution procedure is as we described above, simple enough. To give some idea of the detail entailed in both the updating of attributes according to Monte-Carlo testing for transitions (events), and in the conditional dependency between attributes, we include in Table 1 the determinants of the major events simulated by DYNASIM.

Table 1.

Determinants of Major Events Simulated By DYNASIM (From Harris, 1978).

Event or Characteristic	Determinants
Birth	Marital status, age, race, education, number of previous live births.
Death	Age, race, sex, education, marital status, parity of women, current simulation year.
First marriage	Age, race, sex, education, hours worked, wage rate, transfer income, current simulation year, year of birth.
Remarriage	Age, sex, marital status (widowed or divorced), current simulation year, year last marriage ended.
Divorce	Age, race, disability status, unemployment status of husband, earnings of wife, length of marriage, year of marriage, current simulation year.
Education (probability of advancing a grade)	Age, race, sex, education of head of family, number of grades completed.
Geographic location (region and size of Standard Metropolitan Statistical Area - SMSA)	Age, sex, education, and marital status of family head or single individual, duration of marriage, region and current SMSA size.
Disability	Age, race, sex, education, marital status, whether disabled in previous year.
Labour force participation	Age, race, sex, presence of disability; whether participated in previous year; other income; marital status and presence of child under six for women; aggregate unemployment rate.
Hours of labour supplied	Age, race, sex, education, marital status, presence of child under six, expected wage, labour force supply in previous year.
Hours of unemployment	Age, race, sex, education, marital status, presence of child under six, unemployment in previous year, aggregate unemployment rate.
Wage rate	Age, race, sex, education, marital status, region, disability status, wage in previous year, aggregate level of income.
Job change/tenure	Age, race, sex, education, tenure, sector of employment.
Sector of current employment (private industry, federal government, and state or local government)	Education, sector in previous year.
Income from major government transfers	Computed according to program rules.
Non-social security pension benefits	Sex, age, marital status (widowed), education, hours worked, race, whether family has wealth, whether receives social security.

The initial condition $S(\underline{x},t_0)$ which primes the forward iteration process consists of a sample drawn from the 1960 or 1970 decennial census, edited and augmented 'by imputation' to include attributes which are not present in the survey. The outputs of the dynamic model consist of socio/demographic/economic information of three kinds.

* *Aggregate* : consisting of total births, deaths, marriages, etc.

* *Disaggregated Cross-Sectional* : consisting of the complete range of distributional information cross-classified and aggregated as appropriate.

* *Longitudinal* : yielding historical information on the development patterns of individuals and families.

Because DYNASIM was concerned with the full multivariate richness and dynamic development of a population, it is not surprising that its main applications to date have been concerned with the assessment of government social and economic programs and especially public transfer policies (see Orcutt *et al.*, 1980).

Of all the applications of microsimulation the work on DYNASIM devotes most effort to the (experimental) study of output variability due to the two sources:

* the variance related to the specification of an initial population

* Monte-Carlo variability associated with the sampling approach which induces transitions among (particular) members of the population - a prediction produced by a single run of the model is regarded as a sample estimate of the true model prediction.

A number of tests were carried out with different sample sizes, and the following conclusions indicate the sort of concerns with the development of dynamic simulation models:

> "...the variance of predictions from the model are not large even for samples with as few as 1000 persons. Second, there is no substantial evidence that the variance in predicted outcomes grows rapidly after about the third or fourth year of simulation. Third, preliminary analysis suggests that the variance of the model's predictions is roughly inversely proportional to sample size... Finally, there is no indication of small sample bias in the predictions"

and with respect to Monte-Carlo variability

> "...the single most impressive finding from the Monte-Carlo experiments ... is how small the impact of a specific random number strain is on the results of the model. DYNASIM is very complex and involves a large number of stochastic decisions for each person each year. Yet with samples of only 1000 to 2000 persons the standard error of most aggregate statistics from the model are less than 5 per cent of the prediction. With samples of 4000 persons the errors are in many cases less than 1%".
>
> (Orcutt *et al.*, 1976: Chapter 11).

To minimise Monte-Carlo variability in policy analysis, in which the *difference* between two or more model predictions is often of key relevance, a facility is in fact included to use in each model run the same 'seed' which generates the same stream of random numbers.

It should be emphasized that the conclusions above pertain to *aggregate* information generated from the model. The sample size and Monte-Carlo errors for *distributional* and *longitudinal* information outputs are not however reported. This is a significant omission.

Interactions and Interdependencies in Urban Systems Analysis

The concern for information loss through classification and aggregation, and the treatment of interactions and interdependencies were also important aspects of Alan Wilson's work in the early 1970's on the development of urban systems theory (Wilson, 1974; Wilson and Pownall, 1976). This work was conducted at a time when considerable criticism was being directed against urban and regional planning models, which were deemed to exhibit many theoretical weaknesses not least an appropriate treatment of the interdependencies between the 'major subsystems' - demographic, housing, employment, retailing, public services, etc. An aspect of this criticism coincided with the concern over the classifications adopted in transportation models, specifically in the course definition of trip purpose classes and the assumption of zero elasticity of time and expenditure substitution between them. Wilson, like Orcutt, was concerned with a *strategy* of model development, seeking to relate aggregate information and *systemic effects* directly to micro-dependencies and interactions, and adopting sampling procedures and the processing of lists of individuals and households by probability relations to obviate difficulties of information storage.

A central feature of the work was the development of a model for the multivariate quantity $P_H(\underline{x}_H)$ and through a procedure of sampling from this distribution, the distributional effects of imposing particular policies was sought. We shall briefly describe the *synthesis* of a 1 per cent sample of the population of Leeds County Borough in 1971. Further details may be found in the papers by Wilson and Pownall (1976), and Pownall (1975).

The individual and household attributes which were considered are as follows:

$\underline{x}_I$ = ($\underline{b}$ = status in household; $\underline{a}$ = age; $\underline{d}$ = sex; $\underline{e}$ = ethnic group; $\underline{u}$ = educational attainment; $\underline{q}$ = employment status; $\underline{\nu}$ = socio-economic status; $\underline{w}$ = wage; $\underline{jw}$ = work location)

and

$\underline{x}_H$ = ($\underline{n}_h$ = household size; $\underline{x}_I^i$, $i=1\ldots n_h$; I = household income; C_1 = housing expenditure; C_2 = shopping expenditure; C_3 = capital car expenditure; $\underline{j}_R$ = residential location; C_4 = journey to work expenditure; js = shopping location; C_5 = journey to shop expenditure).

$P(\underline{x}_H)$, the probability that a household is endowed with the multivariate characteristics contained in the vector $\underline{x}_H$ is taken to be of the following form

$$P(\underline{x}_H) = P(n_H) \cdot \underbrace{\prod_{t=1}^{n_h} P(\underline{x}_I^i : i\varepsilon S(h))}_{\text{individual attributes}}$$

$$\times \underbrace{P(C_1|I,n_h,\nu)\, P(C_2|I-C_1,n_h,\nu)\, P(C_3|I-C_1-C_2,n_h,\nu)\, P(C_4|I-C_1-C_2-C_3)}_{\text{consumption functions}}$$

$$\times \underbrace{P(j_R|\underline{\hat{x}}_H^R,\underline{Z}^R)\, P(j_s|\underline{\hat{x}}_H^s\ ;\ \underline{Z}^s)}_{\text{residential and shopping locational determinants}} \tag{19}$$

which illustrates the structure of dependency between household attributes. The probability distribution $P(\underline{x}_I^i)$ pertaining to individual members is in turn given by

$$P(\underline{x}_I) = \underbrace{P(\underline{b})\, P(\underline{a}|\underline{b},n_h)\, P(\underline{d}|\underline{b})\, P(\underline{e})\, P(\underline{u}|\underline{a},\underline{d})\, P(\underline{q}|\underline{a},\underline{d},\underline{e})}_{\text{personal characteristics}}$$

$$\times \underbrace{P(\underline{j}_w|\underline{\hat{x}}^w,\ \underline{z}^w)}_{\text{workplace location determinant}} \tag{20}$$

The location models, P_{jR}, P_{js}, P_{jw} corresponding to residence, shopping and work, include the dependency on sets of attributes associated with individuals/households $\underline{\hat{x}}$ and a set of exogenous factors $\underline{Z}$ as shown in Table 2.

The conditional dependency between the household attributes is subject to two sets of constraints:

* the total household income I is equal to that derived from individual members;

* the consumption variables are confined within lower and upper bounds.

The attributes of individuals and households in a synthetic sample are now formed by successively creating - by Monte-Carlo methods - the component attributes, according to Equations (19) and (20), in the order: household size; individual characteristics (iterated for all members in a household of the selected size); consumption characteristics; and finally, residential and shopping locational characteristics. A set of checks were included to prevent violation of the above constraints - the first being trivially incorporated.

The specific model relations could be elaborated in many ways, particularly those pertaining to the economics of consumption and the mutual dependency between the attributes of individuals within a household - work which is now taking place at Leeds. The above 'prototype' model was prepared to examine a number of problems existing at an intra-urban level, including: the study of multiple disadvantages in connection with education priority areas; and of more direct relevance in a transport context, a determination of the response of expenditure patterns and of workplace and shopping location due to changes in urban housing - specifically the relocation of families within the public sector. In both these problems the interdependencies between the attributes are central factors in modelling distributional effects.

Table 2. The determinants (conditional dependencies) of location behaviour in the Wilson-Pownall model

Location model $P_j = P_j(\hat{\underline{x}}, \underline{Z})$	Individual/household determinants $\hat{\underline{x}}$	Exogenous factors $\underline{Z}$
Work	Status within household; age; sex; ethnic group; educational attainment; employment status	number of jobs of wage w in zone j
Residence	Income; housing expenditure; household size; socio-economic status; ethnic group	number of houses of type k in the zone; price of houses of type k in j; generalised cost of travel from residence to work
Shopping	Income; shopping expenditure; capital car expenditure	attractiveness of zone for shopping; generalised cost of travel between residence and shopping location

The consideration of *dynamic interdependencies* at an urban and regional level suggests a combination of various features of the work of Orcutt *et al.* (1976) and of Wilson and Pownall (1976), and such an approach is being developed at Leeds. Some applications and early results of this work are discussed by Clarke, Keys and Williams (1979, 1980). It should be added that in these models a more modest approach is adopted for the simulation of economic effects at the household level than is present in DYNASIM.

The work of Kain *et al.* (1977), developed independently, also appears to incorporate micro-dynamics of the household sector within the context of an urban housing model. The motivation for this study appears to be related to the problems of data storage (of partial matrices) and execution time, of an early version of the NBER model designed to consider the impact of housing allowances on a spatially disaggregated housing system. The authors comment

> "Our decision to use list storage (for households and dwellings) resulted from our evaluation of four factors: the growing inefficiency of the summary matrix technique as model sized increased; the greater accuracy entailed in the list storage method; the direct mapping between theory and computer model permitted by list storage; and perhaps most importantly, the increased availability of disc storage with greater capacity and lower access costs".
>
> (Kain and Apgar, 1977).

We would add a final comment here on the generation of synthetic samples. In the models developed by Wilson and Pownall (1976) and by Clarke *et al.* (1979) a number of secondary data sources are used to 'build up' the vectors of individual and household attributes - a full survey not being available. While an initial and hopefully reasonable structure of dependency may be imposed on the joint probability $P(x_1 \ldots x_M)$, as for example appears in Equations (19) and (20), it is often necessary to obtain sample information from the joint distribution of a subset of the attributes, say $x_1 \ldots x_\ell$, when the available data is only in the form of marginal tables. That is, the contingency table corresponding to the set $x_1 \ldots x_\ell$ is not obtainable but only contingency tables for subsets of these variables. An attempt must therefore be made to reconstruct as far as possible the quantity $P(x_1 \ldots x_\ell)$ from available marginals. A common approach to this problem of combining contingency table data from two or more sources is *iterative proportional fitting* which is discussed in depth by Bishop, Fienberg and Holland (1975). Its application in a transport related context is considered in depth by McFadden *et al.* (1977), in the formation of SYNSAM - a method for generating a synthetic representative sample of households for an urban area for any specified date.

'Integrated' Travel Models

The motivation for Kreibich's simulation study lies in his perception of the deficiencies of conventional transport study models which

> "...divide the inter-connected decision process of transportation behaviour into separate models of trip generation, modal split and trip distribution, relying on a rudimentary theoretical base. Socio-economic, demographic and spatial determinants are not sufficiently considered and the models are not able to depict complex structures and processes. They are not exhaustive of the available information...".
>
> V. Kreibich (1979)

sentiments which became familiar in the 1970's. Furthermore, with regard to disaggregate models it is suggested that

> "...their implementation, however, was hampered by methodological difficulties. They could not treat sufficiently the large amount of data, nor could they offer a reasonable logic to reproduce the decision processes of a transportation problem on an individual level".

Kreibich (1979) has sought to present an alternative model of an integrated decision process involving modal split and trip distribution, based on a system of constraints, in which Monte-Carlo methods are used for model solution. In contrast to current disaggregated choice models which tend to emphasize *preferences*, a model is proposed which is based on the notion of mobility *constraints*.

It is an important part of model development to identify market segments $\underline{s}$ (called 'situation groups') which 'represent significantly different decision situations with respect to socio-demographic position and territorial location'. The segments are formed from factors based on: family status; social and occupational position; sex; and general work place location. It is suggested that these variables

> "...reproduce T. Hagerstrand's system of major constraints: availability constraints resulting from an individual's social position, coupling constraints, which can be traced back to sex and family status, and authority constraints, which are caused by accessibility conditions"
>
> Kreibich (1979).

Using a clustering algorithm for the analysis of travel (commuting) behaviour, six market segments were finally identified which we shall identify as the classes of the (composite) characteristic $\underline{s}$.

The model is of 'interaction form' in that it determines the residential location i, workplace location j, and mode of travel k for employed persons. The dependent variable is the joint probability distribution $P(\underline{s},\underline{i},\underline{j},\underline{k})$, and the conditional dependency between the attributes appears to be taken in the following form:

$$P(\underline{s},\underline{i},\underline{j},\underline{k}) = p(\underline{s})\, p(\underline{i}|\underline{s})\, p(\underline{k}|\underline{s})\, p(\underline{j}|\underline{k},\underline{i},\underline{s}) \tag{21}$$

in which we define

$p(\underline{s})$: the probability that an employed person belongs to the market segment (situation group) $\underline{s}$

$p(\underline{i}|\underline{s})$: the probability than an employed person belonging to segment $\underline{s}$ resides in a source community (residential location) i

$p(\underline{k}|\underline{s})$: the probability that an individual belonging to $\underline{s}$ will select mode $\underline{k}$ (rail, car, pedestrian, bicycle)

$p(\underline{j}|\underline{k},\underline{i},\underline{s})$: the probability that workplace location j is chosen given $\underline{i},\underline{s}$ and $\underline{k}$.

Having identified the market segments, the mode and location are determined by reference to 'decision profiles' of each situation group which appear to be interpreted as the conditional (prior) probabilities $\tilde{p}(\underline{k}|\underline{s})$, $\tilde{p}(\underline{\emptyset}|\underline{s})$, $\tilde{p}(\underline{t}|\underline{s})$ that an individual with characteristics $\underline{s}$ will: select mode $\underline{k}$; select a workplace in a centre with an identified number of jobs $\underline{\emptyset}$ (< 100, 100-1000, $\geq$ 1000); and be associated with a particular travel time class t, respectively.

(Subject to the present authors interpretation of V. Kreibich's paper) the model appears to be of the following form

$$P(\underline{s},\underline{i},\underline{j},\underline{k}) = \tilde{p}(\underline{s})\ \tilde{p}(\underline{i}|\underline{s})\ \tilde{p}(\underline{k}|\underline{s})\ p(\underline{j}|\underline{k},\underline{i},\underline{s}) \qquad (22)$$

in which the quantities $\tilde{p}(\underline{s})$ $\tilde{p}(\underline{i}|\underline{s})$ $\tilde{p}(\underline{k}|\underline{s})$ are determined directly from observed data, $p(\underline{k}|\underline{s})$ being essentially a category analysis of modal split determined by demographic and socio-economic factors.

The distribution (workplace location) model is, in turn, taken to be

$$P(j|\underline{k},\underline{i},\underline{s}) = \{ \sum_{\emptyset} P(\underline{j}|\underline{\emptyset})\ \tilde{P}(\underline{\emptyset}|\underline{s}) \}\ \hat{P}(j|\underline{k},\underline{i},\underline{s}) \qquad (23)$$

in which $P(j|\underline{\emptyset})$ is the probability of selecting location $\underline{j}$ given the selection of a workplace with a selected number of jobs $\emptyset$.

The quantity $\hat{P}(j|\underline{k},\underline{i},\underline{s})$ is introduced essentially to impose a feasible choice set on the decision process, in which the constraints are of two kinds:

* a modal accessibility constraint $K_1(\underline{j}|\underline{k},\underline{i})$ which eliminates location $\underline{j}$ if they cannot be reached from $\underline{i}$ by mode $\underline{k}$
* a time constraint $K_2(t_{ij}|\underline{k})$ which eliminates any location j if the interzonal time t_{ij} is outside the available budget $t(\underline{s})$,

$$t_{ij} \leq t(\underline{s}) \quad : \quad K_2 \tag{24}$$

The maximum travel time available is determined with reference to the prior distribution $\hat{p}(\underline{t}|\underline{s})$. It appears that $\hat{P}$ is a uniform distribution within the feasible region defined by these two constraints (given that $t(\underline{s})$ is distributed) and zero outside it.

In the solution of the model, attributes of individuals are built up by Monte Carlo sampling in the following order: $\underline{i} \to \underline{s} \to \underline{k} \to j$, iterative loops being included to incorporate the constraints on the accessibility of the workplace with the available mode, and on the maximum travel time available. (The above model has apparently been extended to include a trip generation component by Mentz and Kutter).

Results from a sample (a 2.5% sample, consisting of 25,000 individuals, in the case of Nürnberg) could be aggregated as appropriate.

With regard to the use of the Monte Carlo approach Kreibich comments

> "The major methodological advantage of the procedure is its linear character. It avoids complex matrix operations and resulting problems of data storage and system time. It represents a real chance to produce alternatives in transportation planning models." (Kreibich, 1979).

The above model is reported to be 'extremely accurate and sensitive'.

While Professor Kreibich has clearly introduced a new and very simple transport model based on a relatively unfamiliar solution procedure, critics of the model would take issue with a number of its features and claims of the author. We wish to mention three aspects which are relevant to the discussion in this paper.

Firstly, while simplicity is certainly a virtue, when other things are equal, here, other things are manifestly not equal in the range and form of attribute dependencies and the policies which can be addressed by the model compared with either conventional or currently available disaggregate models. The ease of implementation, which should not *artificially* inflate the claims of the Monte Carlo approach, relates here very much to the simple specification, and particularly to the amount of prior information used in a category analysis form. It would in fact have been rather surprising if a well fitting model had not resulted, given its specification. It should also be added that due to the heavy reliance on *particular* constraining concepts at the expense of preference structures, there are no free parameters to be estimated - again resulting in simplification - which limits the available response properties of the model.

Secondly, much of the computational complexity arising in the application of 'conventional' (and in certain cases disaggregated) models is due to the introduction of demand-supply interactions, both in the transport system and in the housing/land market. To nest the present Monte Carlo procedure within loops to incorporate supply side interactions would greatly increase the computational burden.

Thirdly, in the solution of models by Monte Carlo sampling in which the attributes of 'pseudo-individuals' are successively updated, there is some danger of imputing forms of decision process at the micro-level which are not consistent with the model specification. This issue is further discussed, in relation to the rationale for different model structures, by Williams (1981). In terms of the structure of dependencies and forms of (22) the model can be recognised as a good old-fashioned 'pre-distribution modal choice model' of tabular form, with a 'singly constrained' journey to work structure (albeit generated by constraints).

In short, it would be argued by critics that the model falls rather short of the standard implied by the earlier comments put forward against conventional or disaggregate models.

In the discussion of integrated transport models and attribute interdependencies it is appropriate here to make (brief) reference to the work on disaggregate model systems, particularly that developed through the 1970's at MIT (Ben-Akiva, et al., 1977). It is now well known that the specification of this system is based on probabilistic choice models, underpinned by random utility theory, in which liberal use is made of multinomial and nested logit forms. This system may also be considered to relate to the investigation of the dependencies in a multivariate joint distribution $P(\underline{x})$ pertaining to individuals/households. The demand model is based on a structure of dependency in which travel choices are modelled conditional on mobility choices, as follows:

Mobility choices

Employment location $\underline{j}_W$
Residential location and housing $\underline{i}_R$
Automobile ownership $\underline{CO}$
Mode to work $\underline{k}_W$

Travel Choices (Shopping)		Travel Choices (Social/Recreational)		Travel Choices (Other)	
Frequency	$\underline{\nu}_S$	Frequency	$\underline{\nu}_{SR}$	Frequency	$\underline{\nu}_O$
Destination	$\underline{j}_S$	Destination	$\underline{j}_{SR}$	Destination	$\underline{j}_O$
Mode	$\underline{k}_S$	Mode	$\underline{k}_{SR}$	Mode	$\underline{k}_O$

If the vector $\underline{x}$ is denoted

$$\underline{x} = \{ \underbrace{\underline{j},\underline{i},\underline{CO},\underline{k}}_{\underline{x}_M} ; \underbrace{\underline{\nu},\underline{j},\underline{k}}_{\underline{x}_S} ; \underbrace{\underline{\nu},\underline{j},\underline{k}}_{\underline{x}_{S/R}} ; \underbrace{\underline{\nu},\underline{j},\underline{k}}_{\underline{x}_O} \}$$

then $P(\underline{x})$ is written in the following *general* form

$$P(\underline{x}) = P(\underline{x}_M)\, P(\underline{x}_S|\underline{x}_M)\, P(\underline{x}_{S/R}|\underline{x}_M)\, P(\underline{x}_O|\underline{x}_M) \qquad (25)$$

The partial models defined are taken to be of *analytic* form (members of the logit family), their specification and aggregation being considered at length by Ben-Akiva et al. (1977).

The common *general* philosophy underpinning the different models discussed here - the design of policy sensitive structures based on an appropriate decision model - is clear enough. A framework for comparing *individual* models must be made with reference to the appropriate policy testing contexts and is discussed further by Williams and Ortuzar (1981B).

Activity-Travel Models and Space-Time Constraints[*]

Although the existence of interdependencies between individuals within the household and the influence of 'space-time' constraints has been recognised for a long time (considerably more than a decade), the problems of time allocation and formal studies of interactions have asserted themselves only relatively recently. Allaman (1980) has indicated that while the so-called activity approach is not a single and easily definable set of concepts and methods, among the recurrent themes which occur in this body of literature are the study of: *activity patterns*; the *scheduling* of activities in space and time; *interactions* in decisions over the day and within the household. An indication of recent developments in activity based approaches in Europe and the United States is given by Jones (1980). Here we shall merely point to some applications of simulation concepts and the examination of choices and constraints at the individual and household level.

A broad framework for discussing many of these studies may be made by appealing once again to the formation, through synthesis from *observed* distributions or causal modelling, of the joint multivariate distributions $P(\underline{x}_I)$ and $P(\underline{x}_H)$ of individual and household attributes $\underline{x}_I$, and $\underline{x}_H$, in which the vector $\underline{x}_I$ is of the form

$$\underline{x}_I = (\underline{\hat{x}}, \underline{A}, \underline{I}_A) \tag{26}$$

with $\underline{\hat{x}}$ denoting a set of personal socio-demographic-economic characteristics,

---

[*] Full justice could not be given to a discussion of these models in the time available. The author is grateful to Peter Jones for a discussion on recent work in the area.

$\underline{A}$ is a vector of activities performed by an individual, and $\underline{\tau}_A$ contains temporal information on the start and/or duration of these activities. As before $\underline{x}_H$ includes the individual vectors $\underline{x}_I$ of members belonging to a particular household, and their mutual dependence (corresponding to interactions) is included in the form of $P(\underline{x}_H)$.

Studies differ in their emphases and styles of model, some concentrating on *individual* activity patterns associated with particular family members or classes of people (eg. students). Some seek to synthesize a population from observed conditional probability distributions, others such as those of Dawn (1979) and Allaman (1980) are concerned with developing econometric models for the activity-travel pattern. In the former case a probabilistic choice model of the traditional kind was nested within a constraint procedure for time allocation, while as the latter multiple regression were developed directly for the time allocations τ_A in terms of socio-economic-demographic characteristics

$$\underline{\tau}_A = \underline{\tau}_A(\hat{\underline{x}}) \qquad (27)$$

and subject to the constraint on total time available.

Other studies typified by Zharvi's work attempt to formulate hypotheses of constancy (over time *and* under policy measures) of travel time budgets for different *groups* of individuals distinguished by their attributes $\hat{\underline{x}}$.

Not surprisingly, many of the same features characterise the synthesis of the quantity $P(\underline{x},\underline{A},\underline{\tau})$ as appeared in the money budget allocation problem in the model of Wilson and Pownall (1976). In the present case, constraints must be imposed to allocate individual time components $\underline{\tau}_A$ subject to the total time available for performing activities over the whole day and/or between certain fixed points (pegs). In any synthesis of $P_H(\underline{x}_H)$ constraints reflecting the logic of interaction between individual members, must be imposed.

Various studies on the synthesis of individual distributions $P(\underline{x}_I)$ have been reported. The work of Brail (1969) has involved the synthesis of a population by sampling from a joint probability distribution in which the times specified in various activities are generated in a sequential manner. The activity pattern model of Stevens (1976) adopts a similar simulation approach but includes the notion of pegs around which a day is structured. Using conditional probabilities for activity sequences the model records the type of activity that is selected and its timing, duration, and location, and has a constraint imposed to ensure that a proposed activity does not encroach on pegged time (Jones, private communication).

An alternative approach to the generation of $P(\underline{x})$ is by means of the entropy maximising approach. In the work of Tomlinson *et al.* (1973) the most likely distribution of the allocation of individuals to activity/time/location combinations is sought subject to a set of constraints relating to timing restrictions, and average time devoted to each activity groups, etc.

While the above studies focus on the generation of the probability distributions, an alternative set of studies has been concerned with the generation and examination of the feasible choice sets available to individuals and/or households in travel/activity contexts. These sets are determined either for a *choice situation* directly, or for *response contexts* given the characteristics of a policy to be examined.

One of the best known examples of feasible choices investigations in a computer simulation is that due to Lenntorp (1978) who, in the PESAPS programme, has developed a computational procedure for assessing whether a set of activities (including travel) is feasible in terms of the available time constraints and the 'temporal organisation of society'. The model 'maps' the set of alternative ways in which a particular activity programme can possibly be carried out, without trying to identify and predict which pattern of behaviour *will* be carried out. One application of the model was to determine the effects of modifications to a bus timetable.

As Lenntorp himself observes, a present limitation of the model is that it deals with an individual rather than with groups or populations

> "This limitation can partially be overcome, however, by testing complementary programmes for individuals belonging to the same household; for example where there are certain points of co-ordination in time-space between household members.... Individual programmes can also be designed which correspond to membership in certain household categories...".
>
> (Lenntorp, 1978; p.179).

Such extensions appear to be essential elements of the CARLA (combinatorial algorithm for rescheduling the lists of activities) programme described by Clarke (1980) and Clarke and Dix (1980). The essential purpose of CARLA is the generation of possible responses to change (feasible choice (response) sets) under modifications of the constraints which influence the allocation of time between activities. In their papers the authors have suggested that microsimulation may be an appropriate method to combine the choice set generation of the activity scheduling model with an activity choice mechanism and necessary aggregation techniques to generate macro information. This procedure would involve the synthesis and sampling from a joint probability distribution for the variation of socio-economic-demographic characteristics of the household along the lines described above.

Decision Model Formation and the Investigation of Approximations

From the description of sampling procedures and Monte-Carlo simulation for the examination of interactions and interdependencies, we return to consider their use in the resolution of decision models, in two contexts. Firstly, in the solution of models which entail the relaxation of some of the restrictive assumptions which underpin simple choice models; and secondly, in the study of specific approximations invoked in implementing models.

The motivation for the simulation studies of Williams and Ortuzar (1981) was the resolution of complex choice models in the experimental investigation of *mis-specification*, that is a determination of the consequences of employing a model in circumstances in which the data is assumed to be generated by an alternative set of theoretical assumptions. A specific field of enquiry was to determine under what conditions, and to what extent, the simplifying assumptions underpinning the logit model imply serious errors in response prediction.

Among the particular features examined were

* the existence within a given market segment, of a distribution of feasible choice sets in the formation of a probabilistic choice model;

* the existence of a distribution of attribute sets in the formation of choice models;

* alternative rules for resolving the multicriterion decision process, including elimination-by-aspects.

One such example will be taken to illustrate the procedure, further details being available in the reference cited above.

When the assumption of homogeneous choice sets is relaxed the appropriate model generator is

$$P_\rho = \sum_{\underline{C}}' P(C_\rho|\underline{C}) \;\; P(\underline{C}|\underline{C}^*) \tag{28}$$

in which are defined the following

$\underline{C}^*$: the set of all available choice sets

$P(\underline{C}|\underline{C}^*)$: the probability of selecting choice set $\underline{C}$ from the set $\underline{C}^*$.

$P(C_\rho|\underline{C})$: the probability of selecting alternative C_ρ from the choice set $\underline{C}$.

$\sum_{\underline{C}}$ denotes simulation over the choice sets $\underline{C}$ belonging to $\underline{C}^*$.

Models of choice set determination may be related directly to search rules (see, for example, Richardson 1980B) or other mechanisms by which information is processed. Alternatively, implicit behavioural mechanisms may be assumed to underpin parametric distributions for $P(\underline{C}|\underline{C}^*)$, which may then be used in conjunction with a choice model to determine the quantity P_ρ, according to Equation (28). In the work of Williams and Ortuzar (1981), such an approach was adopted and a further decomposition of (28) was introduced through the partitioning of choice sets by size. The choice set size distribution was then parameterized to allow a continuous variation between the complete captivity of an individual to a selected alternative, to the availability of the complete set of available options $C_1 \ldots C_M$. The variability in behaviour, reflected in the variable, P_ρ, may thus be traced to three sources: the heterogeneity in choice set sizes; the propensity to select particular choice sets of a given size; and the dispersion in preferences in the choice model $P(C_\rho|\underline{C})$ which was taken of random utility form.

The resolution of this choice model was achieved by Monte-Carlo integration. A choice set was sampled from the distribution $P(\underline{C}|\underline{C}^*)$ and the maximum utility option determined. The proportion $\hat{P}_\rho$ which selects alternative C_ρ is then taken as an approximation to the true model prediction P_ρ for sufficiently large sample sizes. The data generated by this procedure could then be used to estimate the parameters of a logit or more general model, and in this way the response error, which is a measure of the deviation between simulated and estimated behaviour under a policy stimulus, can then be assessed under various assumptions underpinning the hypothesized decision process. This experimental approach may be used to direct further theoretical and empirical attention to possible sources of serious mis-specification.

Closely related to the problems of specification and mis-specification is the study of approximations in the formation and solution of choice models, and we shall give brief mention to two issues: the use of the Clark approximation in the solution of the probit model; and the adoption of the multinomial logit model as an approximation to the Nested Logit and more general structures.

The Clark approximation concerns the solution of the multidimensional integral

$$P_\rho = \int_{R_\rho} d\underline{\varepsilon} \;\; F(\underline{\varepsilon})$$

in a utility maximising context, in which $F(\underline{\varepsilon})$ is multivariate normally distributed. Essentially it involves the successive application of a simple approximation in which the distribution of the maximum of two normally distributed variables is itself normally distributed. This allows the *multivariate* integral to be reduced to a particular *univariate* integral which may be simply evaluated by numerical methods. Albright, Lerman and Manski (1977) have employed Monte-Carlo simulation - successively sampling utility values from a multivariate normal distribution and determining the maximum utility option - in order to investigate how the deviation of the approximate Clark probabilities from the true choice probabilities varies with: choice set size; true probability magnitude; covariance structure of the utility distributions; etc. The authors were led to remark

> "While the Clark approximation seems to dominate the simulation approach...the approach provides some security to the empirical researcher wishing to estimate a multinomial probit mode".

An alternative area of approximation analysis, which has been closely associated with studies of mis-specification, is the investigation of 'similarity effects' in choice processes, and the errors involved in employing models which cannot accommodate them. It is well known that the recognition of the restrictive elasticity properties of substitution of the MNL model was a prime motivation for the development of more general structures. There are now several studies which have employed Monte-Carlo simulation and other numerical techniques to examine such issues. For example, Ortuzar (1978) has used Monte-Carlo integration to generate mixed-mode demand models in a comparative study with a multinomial logit alternatives, while Robertson and Kennedy (1979) use simulation in a joint destination, mode and route choice analysis. A Monte-Carlo study of mis-specification errors involved in approximating the MNL model to the Nested Logit and other forms is discussed by Williams and Ortuzar (1981), and Horowitz (1980) has conducted a similar type of investigation on the use of the MNL model as an approximation to the general probit form.

Matching Models and the Interaction of Demand and Supply

There are several applications of microsimulation in which it is sought to bring about an association of one item with another, as in the following examples:

Housing Context	:	{Household Characteristics : Dwelling Characteristics}
Labour Market	:	{Individual Characteristics : Job Characteristics}
Household Formation (Marriage Market)	:	{Bride Characteristics : Groom Characteristics}
Car Pooling	:	{Driver Characteristics : Passenger Characteristics}.

It is specifically desired to model the matching of two sets (lists) of items according to particular conditions of association between their attributes, whether these are reflected simply in prior probabilities or underpinned by a detailed micro-consideration of the mechanism responsible for the matching - possibly involving the presence of intermediaries. Examples of the above contexts may be found in Kain *et al.* (1976); Clarke *et al.* (1979); Orcutt *et al.* (1976) and Bonsall (1979). We shall briefly describe the transportation example (car pooling) before making some general points about the interaction of demand and supply.

The work of Bonsall (1979; 1980A,B) is concerned with the development and testing of models to predict the impact and performance of an organised car sharing scheme - a mode illegal and therefore unobservable in Britain at the time. The author was attracted to microsimulation for the following reasons:

> "The performance of a car sharing scheme is obviously a function of reconciling individual suppliers (drivers) and consumers (passengers) within a formalised market mechanism (the matching service). Furthermore, since the model results could not be directly tested against any observed phenomena, it was desirable to build a model with representational transparency so that the reasonableness of the mechanism which produced the predictions could provide a means of evaluating the reasonableness of the predictions themselves. Clearly microsimulation is the only means by which the performance of organised car sharing schemes could properly be predicted". (Bonsall, 1980B).

In essence the model may be summarised in terms of the following three stages (Bonsall, 1980b):

Stage 1: Eligible members of a sample $S(\underline{x})$, derived from a synthesized distribution $P(\underline{x})$ decide whether or not to join a scheme of a particular type k (seven car pooling arrangements were available).

Stage 2: For each applicant a search is made among other applicants for those whose home and work locations, work hours and stated interest in car sharing (either driving, riding or pooling) make them potential partners. This process determines match lists for each applicant.

Stage 3: Those applicants who have received match lists decide, on the basis of a mutual evaluation of the benefits and inconveniences of forming an arrangement, whether or not to initiate a car pooling arrangement with any potential partners on the match list.

These stages are considered to simulate the decision processes of individuals confronted by the choices implicit in the organised car pooling scheme.

At the kernel of the model a probabilistic choice model $P_{ij}(\underline{x}_i,\underline{x}_j|\underline{\Xi},\underline{\theta})$ is used to determine the probability that individuals i and j will form a satisfactory pooling arrangement, for given values of 'socio-economic' and transport system characteristics $\underline{\Xi}$. In terms of this micro-model, aggregate information is derived relating to demand and supply characteristics. For example the number of shared rides is given by

$$N(\text{share ride}) = \underset{i\epsilon\tau_1}{S} \; \underset{j\epsilon\tau_2}{S} \; P_{ij}(\underline{x}_i,\underline{x}_j|\underline{\Xi},\underline{\theta}) \tag{29}$$

in which the summation $\underset{i\epsilon\tau_1}{S} \; \underset{j\epsilon\tau_2}{S}$ accumulates all contributions to the pool made from applicants in the match lists. In fact in the matching model described, the expression (29) is given by the utility maximising form

$$N(\text{share ride}) = \underset{i\epsilon\tau_1}{S} \; \underset{j,j'\epsilon\tau_2^i}{S} \; \text{Prob}\,\{U_{ij} \geq U_{ij'}|U_{ij'} \geq 0\} \tag{30}$$

An applicant $i\epsilon\tau_1$ to a scheme contained in a (supply) list τ_1 selects the most satisfactory arrangement in his available match list τ_2^i of potential partners in terms of the highest non-negative joint-utility arrangement. To account for dispersion arising from unobserved attributes, the utility functions $U(\underline{Z},\underline{\Theta})$ are taken of the form

$$U(\underline{Z},\underline{\Theta}) = \sum_{\lambda} \Theta_{\lambda} Z^{\lambda} + \varepsilon \tag{31}$$

in which the random variate is assumed to be normally distributed. The parameters $\underline{\Theta}$ were determined from stated preference information.

Essentially the micro-model P_{ij} is of (truncated) multinomial probit form and is generated by Monte-Carlo sampling according to expressions (30) and (31), in which 1 and 0 values are accumulated according to the success or failure of a match. The operation $\underset{i}{S}\ \underset{j}{S}$ in Equation (28) aggregates contributions in a manner described by Equation (5).

As Bonsall notes, in the equilibration procedure involving the interaction of supply and demand, the end state of a model will be a function of the order in which matches are made, and some attention has been given to alternatives (Bonsall, 1979). It is important that bias is not introduced into a model due to the specific ordering of individuals in the match lists. A further discussion of this issue in a general context of matching models is contained in the paper by Clarke *et al.*, 1979.

CONCLUDING REMARKS

In the paper we have referred to several applications of microsimulation techniques, and without seeking a comprehensive coverage of the material, have attempted to draw out the essential motivations for the approach. These have often been associated with refinements to model specifications which have in turn had implications for model solution procedures. We have attempted to present microsimulation as a solution method within the context of the disaggregate modelling philosophy, and have, where appropriate, related applications involving simulation to alternatives which include 'conventional' analytic micro-model specifications. This exercise is important to qualify a view sometimes expressed or implied that microsimulation represents a 'new approach' to modelling, rather than a *solution* approach the merits of which are dependent upon, and must be subjugated to, the general issues of model specification, design and application context.

It would be desirable in a paper on the application of a relatively new technique to establish some necessary preconditions and guidelines for its use, but precisely for the reasons mentioned above, the "advantages" and "drawbacks" of simulation must be closely linked to possible alternative approaches for generating the required information. It is equally inappropriate to infer the general suitability and applicability of simulation from an individual application especially as the whole issue is complicated by different interpretations of alternative models and indeed solution procedures. In this context it is rather interesting to contrast the rather optimistic comments of Kreibich (1979) to the more guarded conclusions of Bonsall (1981), and this is not surprising in view of the great difference in the complexities of the models involved.

There have, however, over the years emerged what are considered to be certain desirable aspects of practical modelling procedures, namely: policy sensitivity; understandability; responsiveness; and computational efficiency, etc. In the following general remarks we will draw attention to the implications of using microsimulation, in terms of these features, and through their discussion point to some issues, which we feel merit further research.

One of the most positive aspects of microsimulation is that as a *technique* it is indifferent, in principle, to the complexity of the equations which may represent the interactions, interdependencies and variability in a population. Such is its appeal that there is an almost irresistable inducement to build the "*ultimate*" (behavioural) micro-model. Indeed, Orcutt confesses to the dream, very much apparent in his early work, Orcutt (1961), of simulating the American economy from the "bottom up" involving individuals, families, households, firms, government agencies, etc. interacting through markets. This was an intellectual quest of a high order - the establishment of a comprehensive micro-economic basis for aggregate economic behaviour. The developments of DYNASIM fell considerably short of this mark, essentially involving the embedding of a micro-model of household demographics, income and consumption within a macro-economic "environment". Even so, the work has encouraged the view that it is difficult to imagine in the social sciences a more ambitious research task, and its relatively slow development precisely relates to the specification and estimation of the host of static and dynamic micro-relations which characterise the model, and the resources required for this endeavour. Nevertheless, Orcutt (1976) and his colleagues are clearly of the opinion that the detailed specification of DYNASIM was necessary for understanding the time dependence of distributional and interdependency issues at the household level.

The seduction of simulation methods - and in this we refer to the availability of general model solution procedures - *may* however encourage a level of elaboration which would have important theoretical and practical consequences. As has been discussed elsewhere (Williams and Ortuzar, 1981B), the weaknesses of the theoretical basis for response models are often cruelly exposed at higher levels of problem resolution. Here we face fundamental, and indeed unresolved, problems in micro-model development, whose confrontation will clearly have implications for the approach to model solution.

It has been suggested that a positive feature of microsimulation is its 'representational transparency' in which the simple conception of the solution procedure in terms of the actions of sample members is contrasted with the 'incomprehension' which characterises many existing model systems. While flow diagrams which often accompany solution by simulation may prove a useful conceptual guide, the above argument should not be overdone. Again we must return to a comparison of alternative model specifications and their theoretical bases. It seems that the essence of the above sentiment relates more to the desirability of underpinning a model with a behavioural explanation, and as we have empahsized earlier, many 'conventional' transport models whether of the traditional transport study type or those associated with the micro-philosophy may be given a behavioural interpretation (not to everyone's satisfaction, of course!). There is, we believe, in the range of examples cited in the paper, evidence that heavy reliance on flow diagrams *alone* as a means of specifying and solving a model, without due consideration to the properties of probability distributions has led to some confusion to the *modeller* both in terms of the nature of the model and its relationship to possible alternatives.

The immediate consequencesof solving a problem by (Monte-Carlo) simulation are that, not only is the solution in the form of a numerical expression which must be re-run for different combinations of inputs, but that in addition, and as was mentioned in Section 2, the solution is in the form of a distribution. For a given combination on input variables different sets of random numbers will generate different output information. Of the many applications referred to above few have given detailed consideration to problems of appropriate sample size determination and Monte-Carlo variability. Even the study of Orcutt *et al.* (1976, Chapter 11) which did proceed to examine these issues did so for the variation in specific (aggregate) output information. Because model specifications are often rather intricate, the cost of investigating this issue experimentally (through several computer runs) may be prohibitively high. As in many computational contexts, and Monte-Carlo simulation is no exception, the adoption of 'machine specific' facilities to ameliorate these problems, tend to be paid for in terms of the lack of probability of the programmes and the expertise required for implementation. While there do exist standard tricks for minimising problems of variability (by, for example, using the same seed for generating strings of random numbers when it is desired to compare outputs achieved under different sets of model inputs) we must conclude that in relation to the statistical properties of model outputs and the appropriate sample sizes for different information sought from a model, more research is needed. On such crucial issues depend the efficiency of the technique and the level of sophistication of model specifications which can be afforded with available resources.

Let us finally recall that simulation as a solution technique has always been seen as a 'last resort' when mathematical analysis is confronted by intractibilities. As a general approach to the solution of models, we believe it will be increasingly used in transportation analyses where the required information and model specification are deemed to demand it. Its use will as ever be subject to practical compromises, and the availability of information and other resources.

ACKNOWLEDGEMENT

I wish to acknowledge useful discussions with my colleague Martin Clarke, and with Peter Jones. I am also grateful for the financial support of the S.S.R.C.

REFERENCES

Adler, T. and Ben-Akiva, M.E. (1979) A theoretical and empirical model of trip chaining behaviour, *Transportation Research, 13B,* pp. 243-57.

Albright, R.L., Lerman, S.R. and Manski, C.F. (1977) Report on the development of an estimation programme for the multinomial Probit Model, Prepared for the Federal Highway Administration by Cambridge Systematics, Inc.

Allaman, P.M. (1981) Activity analysis: an outline of concepts and methods, Presented at the Annual Meeting of the Transportation Research Board, Washington DC, January, 1981.

Ben-Akiva, M.E., Adler, T.J., Jacobson, J. and Manheim, M. (1977) *Experiments to clarify priorities in urban travel demand forecasting research and development,* Summary Report, US Department of Transportation.

Bishop, T., Fienberg, S. and Holland, P. (1975) *Discrete multivariate analysis,* MIT Press.

Bonsall, P.W. (1979) The simulation of organised car sharing, II. The simulation models and their calibration, Working Paper 109, Institute for Transport Studies, University of Leeds.

Bonsall, P.W. (1980A) Microsimulation_ its application to car sharing and activity travel choice models, presented at the International Conference on Research and Applications of Disaggregated Travel Demand Models, University of Leeds, July 1980.

Bonsall, P.W. (1980B) Microsimulation of organised car sharing: Description of the models and calibration, Supplementary Report No. 564, Transport and Road Research Laboratory, Crowthorne, Berks.

Bonsall, P.W. and Champernowne, A.F. (1979) The simulation of organised car sharing I: the synthesis of a population base, Working Paper 107, Institute for Transport Studies, University of Leeds.

Brail, R.K. (1969) Activity systems investigations: strategy for model design, Unpublished Ph.D. thesis, University of North Carolina, Chapel Hill, North Carolina.

Burnett, P. (1978) Choices and constraints - oriented modelling: alternative approaches to travel behaviour, Evanston, Illinois, The Transportation Centre, Northwestern University.

Chapin, F.S. (1978) Human time allocation in the city, Chapter 1, in Carlstein, T., Parkes, D. and Thrift, N. (Eds.) *Human activity and time geography,* Vol. 2, London: Edward Arnold, 13-26.

Clarke, M., Keys, P. and Williams, H.C.W.L. (1979) Household dynamics and economic forecasting, Paper presented at the European Congress of the Regional Science Association, London, also as Working Paper 257, School of Geography, University of Leeds.

Clarke, M., Keys, P. and Williams, H.C.W.L. (1980) Microsimulation in socio-economic and public policy analysis, in Voogd, H. (Ed.) *Strategic planning in a dynamic society,* Delft: Delft University Press.

Clarke, M., Keys, P. and Williams, H.C.W.L. (1981) Microanalysis and simulation of socio-economic systems: progress and prospect, in Bennett, R.J. and Wrigley, N. (Eds.) *Quantitative geography in Britain: retrospect and prospect,* London: Routledge and Kegan Paul.

Clarke, M.I. and Dix, M.C. (1980) An activity scheduling model for generating feasible travel choice sets, Presented at the International Conference on Research and Applications of Disaggregate Travel Demand Models, University of Leeds, July 1980.

Damm, D. (1979) *Towards a model of activity scheduling behaviour*, Interdepartmental Doctoral Thesis, Massachusetts Institute of Technology, Cambridge, Mass.

Hagerstrand, T. (1970) What about people in regional science?, *Papers of the Regional Science Association*, 24, pp. 7-21.

Harris, R. (1978) Microanalytic simulation models for analysis of public welfare policies, Paper 819-1, Washington DC: The Urban Institute.

Haveman, R.H. and Hollenbeck, K. (1980) *Micro-economic simulation models for public policy analysis: Vols. 1 and 2*, New York: Academic Press.

Heggie, I.G. (1978) Behavioural dimensions of travel choice, in Hensher, D.A. and Dalvi, M.Q. (Eds.) *Determinants of travel choice*, Saxon House Press.

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

Jacobson, J. (1978) *Models of non-work activity duration*, Ph.D. thesis, Department of Civil Engineering, MIT, Cambridge, Mass.

Jones, P. (1980) Recent developments in activity-based approaches in Europe and the United States, *Australian Road Research*, *10*, pp. 41-6.

Kain, J.F., Apgar, W.C. and Ginn, J.R. (1976) Simulation of the market effects of housing allowances. Vol. 1: Description of the NBER Urban Simulation Model, Mass: Department of City and Regional Planning, Harvard University.

Kain, J.F. and Apgar, W.C. (1977) Simulation of the market effects of housing allowances, Vol. 11: Baseline and policy simulations for Pittsburg and Chicago, Report R77-3, Department of City and Regional Planning, Harvard University, Cambridge, Mass.

Koppelman, F.S. (1976) Guidelines for aggregate travel prediction using disaggregate choice models, *Transportation Research Record*, 610, pp. 19-24.

Kreibich, V. (1979) Modelling car availability, modal split and trip distribution by Monte-Carlo simulation: a short way to integrated models, *Transportation*, 8(2), pp. 153-66.

Lenntorp, B. (1978) A time-geographic simulation model of individual activity programmes, Chapter 9, in Carlstein, T., Parkes, D. and Thrift, N. (Eds.) *Human activity and time geography, Vol. 2*, London: Edward Arnold, pp. 162-80.

Litinas, N. (1980) Simplified methods for transportation policy analysis: a continuous modelling approach, Centre for Transportation Studies, MIT, Cambridge, Mass.

Lundqvist, L. (1978) Planning for freedom of action, Chapter 13 in Karlqvist, A., Lundqvist, L., Snickars, F. and Weibull, J.W. (Eds.) *Spatial interaction theory and planning models*, Amsterdam: North Holland.

McFadden, D., Cosslett, S., Duguay, G. and Jung, W. (1977) Demographic data for policy analysis, The Institute of Transportation Studies, University of California, Berkeley.

McFadden, D. and Reid, F. (1975) Aggregate travel demand forecasting from disaggregate behavioural models, *Transportation Research Record*, 534, pp. 24-37.

Moroney, M.J. (1951) *Facts from figures*, London: Cox and Wyman.

Orcutt, G.H., Greenberger, M., Rivlin, A. and Korbell, J. (1961) *Microanalysis of socio-economic systems: a simulation study*, New York: Harper and Row.

Orcutt, G.H., Caldwell, S. and Wertheimer, R. (1976) *Policy exploration through microanalytic simulation*, The Urban Institute, Washington DC.

Orcutt, G.H., Glazer, A., Harris, R. and Wertheimer, R. (1980) Micro-analytic modelling and the analysis of public transfer policies, Chapter 3 in Haveman, R.H. and Hollenbeck, K. (Eds.) *Micro-economic simulation models for public policy analysis, Vol. 1*, New York: Academic Press.

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

Pownall, C.E. (1975) The development, calibration and testing of a model to show micro-level interdependence in an urban system, Working Paper 117, School of Geography, University of Leeds.

Richardson, A.J. (1980A) Behavioural considerations in Transport modelling, Department of Civil Engineering, Monash University.

Richardson, A.J. (1980B) Search models and choice set generation, Paper presented to the International Conference on Research and Applications of Disaggregate Travel Demand Models, University of Leeds, July 1980.

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

Stephens, J.D. (1976) Daily activity sequences and space-time constraints, in Holly, B.P. (Ed.) *Time space budgets and urban research: a symposium*, Discussion Paper 1, Department of Geography, Kent State University.

Tocher, K.D. (1963) *The art of simulation*, London: English Universities Press.

Tomlinson, J., Bullock, N., Dickens, P., Steadman, P. and Taylor, E. (1973) A model of students' daily activity patterns, *Environment and Planning A*, 5, pp. 231-66.

Watanatada, T. and Ben-Akiva, M.E. (1979) Forecasting urban travel demand for quick policy analysis with disaggregate choice models: a Monte-Carlo simulation approach, *Transportation Research*, 13A, pp. 241-8.

Williams, H.C.W.L. (1981) Travel demand forecasting: an overview of theoretical developments, in Banister, D. and Hall, P. (Eds.) *Transport and public policy planning*, Mansell.

Williams, H.C.W.L. and Ortuzar, J.D. (1981A) Behavioural theories of dispersion and the mis-specification of travel demand models, *Transportation Research B*, (in press).

Williams, H.C.W.L. and Ortuzar, J.D. (1981B) Travel Demand and Response Models: some integrating themes. *Transportation Research* (to be published).

Wilson, A.G. (1974) *Urban and regional models in geography and planning*, London: John Wiley.

Wilson, A.G. and Pownall, C.E. (1976) A new representation of the urban system for modelling and for the study of micro-level interdependencies, *Area*, 8, pp. 246-254.