Working Paper No. 269.

MICRO-ANALYSIS AND SIMULATION OF SOCIO-ECONOMIC SYSTEMS: PROGRESS AND PROSPECTS*

M. Clarke, P. Keys and H.C.W.L. Williams.

* To be published in:

Bennett, R.J. and Wrigley, N. (Eds)

<u>Quantitative Geography in Britain: Retrospect and Prospect</u>.

London: Routledge and Kegan Paul. 1980.

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

January 1980.

<u>Micro-analysis and Simulation of Socio-economic Systems: Progress and Prospects.</u>

INTRODUCTION

One of the more significant developments in analytic planning methods over the last decade has been the refinement and application of models specified and estimated at the level of the individual decision making unit (indivudual, household, etc.). This is well illustrated in the case of travel demand analysis in which there has been a significant shift, particularly in the United States, away from the conventional aggregate approach to policy analysis, involving models specified in terms of trip groups, to one embracing probabilistic consumer choice concepts in which theoretical statements are married to a representation of behaviour at the micro-level. This particular transition was prompted not only by the inefficiencies, expense and insensitivity of the modelling frameworks developed through the 1960s, but also by the new planning contexts and policy measures which have emerged in the 1970s.

While the changes in modelling style were initiated to a large extent by technical considerations - notably the exploitation of the full variability between units in a given sample size - it was apparent to some of the early exponents of micro-analysis that when the framework was endowed with an appropriate theoretical base the models were more likely to accord with a behavioural explanation of statistical patterns, which some would argue is a necessary pre-condition for the development of stable forecasting relations.

While the behavioural paradigm was already fashionable in human geography in the late 1960s, much of the early work was of an empirical nature and this style tended to dominate work in the field. It was notably Hagerstrand (1970) and Thomas (1969) who urged the reconciliation of aggregate "macro-level" descriptions of socio-economic phenomena and their spatial manifestations with disaggregate micro-level concepts. A discussion of the development of these ideas is given by Leigh and North (1978) and Thrift (1980, this volume).

The formal aspects of the aggregation process have, of course, long been of interest to economists (see, for example, Green 1964) but it was not until the advent of high speed electronic computers that the representational and computational aspects involved in aggregation have been studied in depth. Orcutt et.al. (1961, 1976) recognised that for many socio-economic phenomena and policy related issues of interest, several inter-related attributes of the relevant behavioural units had to be specified if a significant amount of variability between these units was to be realized. That is, a greater amount of *information* was required to discriminate realistically between the actions and behaviour of individuals in particular problem contexts. The focus of these developments was, as Guthrie (1972) has discussed, on the high degree of heterogeneity which exists in a given population.

Orcutt and his colleagues (1961) noted the possibility of the effecient manipulation of this information by storing and computationally processing samples of the relevant population multiply classified by the various attributes of interest. This approach was then used in conjunction with Monte Carlo techniques for the solution of models. Although simulation models based on Monte Carlo methods have been used very widely over the past thirty years in disciplines associated with Operational Research, to the authors' knowledge, Orcutt's work, dating back to the mid 1950s, was the first attempt to build specific operational models of socio-economic systems, and in particular one of household dynamics, based on micro-simulation.

In this paper we review the basic principles of micro-analytic simulation, together with some recent applications of the approach. Many of the important issues in model construction relate to the transformation of information and to the aggregation process itself. These aspects are discussed in Section Two as a basis for a discussion of several examples presented in Section Three. It is not our aim here to provide a comprehensive survey of model applications but to illustrate the variety and potential of the micro-simulation approach. The applications reveal several important representational and analytical issues which are treated more formally in Section Four. In a final section we offer some views on the prospects and limitations of the method.

It is worth remarking that, in a number of the applications the advantages and some of the principles of the micro-simulation approach involving list processing appear to have been recognised independently. One could speculate that this is partly due to the specialist nature of the applications dispersed over several disciplines, and the fact that much of the literature is of the form of 'in-house' documentation (eg. publications of the Urban Institute in Washington, United States Government reports) which has perhaps failed to reach a wide audience. Moreover, for much of the last two decades during which micro-simulation methods have been developed, there appeared to be no great incentive to jettison conventional, and easily accessible planning techniques.

It will become clear that the process of aggregation over micro-relations in both static and dynamic formats is a central theme of the paper. It is then also of significance to note that the development of travel choice models specified at the level of the individual is also traceable to the late 1950s. Because until fairly recently, little emphasis has been given in studies of travel and housing demand to the detailed process of aggregation over micro-relations, it is perhaps inevitable that these lines of model development attributable to Orcutt (1961) and Warner (1962) have not been generically related. The recent and increasing use of simulation in travel demand forecasting indicates a unified model building strategy.

REPRESENTATIONS, MODELS, AND THE AGGREGATION PROBLEM

Whether models are meant to provide a causal representation of behaviour or are simply intended as a device for the formal description of variability in a data set, the modelling process, in essence, consists of a series of *transformations* on *information*. We express this process simply as

$$\{\text{information out}\} \leftarrow T\{\text{information in}\}$$

in which the transformation operator T{...} may be in the form of simple data manipulation procedures; may express an analytic dependence between variables; or may assume the form of an algorithm or heuristic process.

All information received and transformed in the modelling process is specified with reference to a set of states relating to how a system can be described and observed. The input information sought may either be suggested by the theoretical framework within which the model has been developed, or and more usually, the form of the model is conditioned by the information available.

Of the many particular questions to which models are addressed a great number fall into three, often related, categories which we label: *dynamics*, *response* and *control*. We shall formally describe the first two categories by means of the equations for system output $\underline{y}(t)$, and its response $\delta\underline{y}(t)$ to a stimulus $\delta\underline{s}(t)$, as follows

$$\underline{y}(t) = \underline{f}(\underline{x}(t), \underline{u}(t), t, : \theta(t)) \qquad (2.1)$$

$$\delta\underline{y}(t) = \underline{g}(\underline{x}(t), \underline{u}(t), t, : \underline{\Psi}(t))\ \delta\underline{s}(t). \qquad (2.2)$$

In general the outputs will consist of several components and have been given vector designations. $\underline{x}(t)$ and $\underline{u}(t)$ are time dependent vectors of independent and controllable input variables respectively, while $\underline{\Theta}(t)$ and $\underline{\Psi}(t)$ are vectors of parameters to be estimated. In general, the stimulus $\delta\underline{s}(t)$ is expressed as a function of changes in $\underline{x}$ and $\underline{u}$. The functional dependencies, $\underline{f}$ and $\underline{g}$ may in turn be derived empirically or express an analytic relationship between variables consistent with theoretical statements embracing tested or untested hypotheses. The traditional method of solving the response problem (2.2) and obtaining the elasticities with respect to policy or other variable changes is essentially through differentiating a typically cross-sectionally estimated functional relationship, $\underline{f}$.

The third category into which many modelling applications fall, that of control, may be expressed in terms of an extremal problem in which it is sought to determine the vector, $\underline{u}$, which optimises a given objective function, subject to the system dynamics (2.1) and other constraints. It will become clear that this latter category is the least well developed in micro-simulation applications.

In order to introduce the micro-simulation approach and various aspects of the aggregation issue, we shall restrict our discussion to a study of households and the individuals comprising them. A micro-description of the household involves the provision of multiply-classified demographic, economic, social and activity/travel characteristics ($\underline{c}_1$, $\underline{c}_2$, $\underline{c}_3$, $\underline{c}_4$) for each of its members. At any time we may identify with a household H_j the following list

$$H_j\{\ldots, I_j^i(\underline{c}_1^i, \ldots, \underline{c}_4^i), \ldots\ i = 1, n_j : \underline{c}_1^j, \ldots\ \underline{c}_4^j\} \quad (2.3)$$

in which I_j^i refers to the i^{th} member of a household containing n_j members. We emphasise that the vectors $\underline{c}$ will, in general, refer to lists of attributes and in certain cases it may be appropriate to define aggregate summary quantities $\underline{c}^j$ pertaining to the household H_j. Traditional surveys (eg. Census, General Household Survey, Family Expenditure Survey, Transportation Surveys, etc.) collect overlapping subsets of individual and household attributes which for the purpose of confidentiality or presentation are often published in a form in which quantities are aggregated both over individuals *within* a household and those *between* households.

We may regard the household and/or individual as existing in a multiply classified state defined by the levels of the various sets of attributes which characterise them. Thus, if $c_1 \ldots c_m$ is a full list of relevant attributes and $\ell_1 \ldots \ell_m$ the corresponding levels associated with each, then the state of an individual i may be represented by a vector $\underline{\alpha}_i$ the components of which refer to the occupied levels of the various attributes. The state of the system as a whole may now be represented by the vectors $\{\underline{\alpha}_1, \underline{\alpha}_2, \ldots \underline{\alpha}_i, \ldots \underline{\alpha}_Q\}$ which correspond to the individual states of the Q members of a population or sample S.

It is the province of theory to understand how individuals come to be associated with the particular combinations of attributes which characterize a state. Behaviouralists argue that this association and individual transitions between states, which constitute events, are the outcome of *decision contexts* and attempt to formulate model relationships in terms of the choices, constraints and preferences of all relevant decision making units. A natural starting point for this analysis is a consideration of a decision context D

$$\left[D^{\underline{a}}\ \{\underline{A}(\underline{Z})\ :\ \underline{R}\}\mathfrak{P}\ \right]_i \qquad (2.4)$$

in which $\underline{a}$ represents the relevant attributes of a decision maker i confronted by a choice between a set of alternatives $\underline{A}^i$ each member of which A_μ is endowed by a set of attributes $\underline{Z}_\mu^i$ deemed to be of importance in the decision making process $\mathfrak{P}_i$. $\underline{R}$ represents a set of constraints which dictate the options which are feasible. Note that $\underline{a}$ will in general be a subset of $\underline{c}$. The choice set $\underline{A}$ may be composed of discrete options as in location decisions, or continuous as in the mortgage incurred in house purchase. Further, we may think of this choice context as endowed with very general characteristics. For example, each element of D may be dependent on the actions of one or more other decision units (through $\underline{A}$, $\underline{Z}$, $\underline{R}$ or $\mathfrak{P}$), whether this refers to the actions of fellow purchasers in a housing market, or to the constraints imposed by public sector housing management (Thrift and Williams, 1980).

The issue of aggregation appears in many aspects of model development from the definition of alternatives, the measurement of attributes, the grouping of observations, to the summation over micro-relations. In a behavioural approach the aggregation process thus provides the essential link between the output of a model and the decision models pertaining to individuals within a population Q.

In an empirical study the modeller as an observer of a system to be represented, will note in an adopted frame of reference (eg. the way in which D is described) dispersion or variability between the actions of individuals who are endowed with apparently identical attributes, and may seek to explain this dispersion by the existence of unobserved attributes/constraints, dispersion of preferences, heterogeneity in choice sets etc. (Williams and Ortuzar, 1979). The probability that an individual with *observable* characteristics $\underline{a}^i$ and $\underline{R}^i$ will be associated with an alternative A_k will in general be expressed in terms of a parametric relation

$$P_k(\underline{a}^i, \underline{R}^i) = h(\underline{Z}^i : \underline{\phi}). \qquad (2.5)$$

This relation may either be stated simply to accord with the empirical regularities in a micro-data set or be consistent with a specific model of the decision process by which the alternative A_k was selected from $\underline{A}$. Such a model may for example be underpinned by extremal principles such as utility maximisation or by satisficing behaviour (Thrift and Williams, 1980). We shall comment on the form and application of analytic micro-models in Section Four below.

The total number of individuals or decision units N_k in Q associated with the alternative k may now be obtained by general summation (addition over discrete variables, integration over continuous variables) as follows

$$N_k = \underset{\underline{a},\underline{R}}{S} \rho(\underline{a}, \underline{R})\, P_k(\underline{a}, \underline{R}) \qquad (2.6)$$

in which $\rho(\underline{a}, \underline{R})$ represents the probability density of the set of characteristics over the population Q.

In practise it is usually necessary to perform this aggregation process S by numerical methods. One such approach is Monte Carlo simulation in which a representative sample of the population is formed according to the probability density $\rho(\ldots)$ and each unit assigned to an alternative $A_k \varepsilon \underline{A}$ according to a comparison between the value of a random number t $(0 \leqslant t \leqslant 1)$ and a cumulative probability index formed from P_k. If the sample is sufficiently large a statistically reliable estimate of N_k may be obtained.

The above example is a simple illustration of aggregation over a probabilistic relation estimated at the micro-level. The alternative courses of action $\underline{A}$ may be defined in a dynamical context corresponding to a wide range of processes, and the Monte Carlo method used to aggregate over individual actions whether or not they are provided with a behavioural basis.

Before describing some recent applications of micro-analytic simulation, we emphasise a number of distinctions in terminology. *Micro-analysis* is an approach to problem solving and model building based on a smallest unit representation, while the *Monte Carlo technique* is used for the experimental realisation or simulation of a process expressed by means of a set of equations or operations. It is also used, as described above, as a numerical method for the integration of functions, and as such forms an important approach to the aggregation process. In turn *List Processing* is simply a device for the computational storage and manipulation of information.

AN OVERVIEW OF MODEL APPLICATIONS

The applications of micro-simulation fall broadly into the two categories of dynamics and response described above. We include static considerations as a special case of the former. To our knowledge there have been no applications of the approach in the social sciences which involve aspects of control.

The use of micro-simulation has in each case been partly justified on the basis that it was necessary to include a large amount of information about individual decision units. This aspect will form a prominent part of the discussion in the next section. Here we discuss a number of recent examples which deal successively with demographic, economic and spatial issues.

It is appropriate to start with a brief discussion of the work of Guy Orcutt and his colleagues, because of its pioneering nature. In his original book (Orcutt, et.al. 1961) a framework was provided for the micro-simulation of the United States demographic/economic system. In subsequent work particular emphasis was paid to the dynamics of households and the derivation of income. This effort culminated in a flexible set of computer programs known as DYNASIM (Dynamic Simulation of Income, Orcutt et.al., 1976), designed as a policy testing instrument applied at the national level. Although the degree of detail is high, requiring a large number of sub-models, the essential elements of the approach involve the economic and demographic characteristics of individual members of a population sampled at a given time. The empirical complexity of the model is largely a consequence of the need to establish the dependence of a wide variety of processes (eg. household formation/dissolution, birth, death, marriage, entry into the labour market, etc.) on a multitude of individual and household attributes. For example, labour force participation is derived as an empirical function of an individual's race, sex, age, marital status, disability status, the previous year's labour force participation, the household's last year's transfer income, number of children under six, last year's unemployment rate. Estimation of this and other functions was achieved by resorting to a wide variety of data sources and special purpose studies.

This work, undertaken at the Urban Institute in Washington, has spawned a number of interesting applications both in academic and government institutions. A typical use of DYNASIM and its predecessor TRIM (Transfer of Income Model) was the identification of particular classes of the population eligible for certain services and welfare provision. An overview of the use of static and dynamic micro-simulation models for public welfare and demographic analyses and their joint interaction is given by Harris (1978).

Many of these applicaitons have been performed at the national level, though a few regional applications do exist, including a model of household energy consumption developed by Caldwell et.al. (1979) for New York State, and a model of household production and consumption for the Yorkshire and Humberside region developed by Clarke, Keys and Williams (1979, 1980). In the latter application the authors were concerned with the projection up to 1982 of changes in the economic circumstances of households under alternative regional macro-economic assumptions. In this example, explicit models of matching are developed in multi-sector labour and housing systems.

The model developed by Kain and his colleagues at the United States National Bureau of Economic Research is one of the most detailed representations of a metropolitan housing system. The explicit spatial representation of household residential and workplace location combined with the socio-economic identifiers produced very large data storage and retrieval problems in the earlier version of the model (as in the Detroit prototype, see Ingram et.al., 1972). List processing was seen as a means of alleviating these problems and was incorporated into the most recent version of the model (Kain et.al., 1976). While the model may address many aspects of the private housing system, a prominent use has been in the analysis of housing allowances.

There are a number of alternative, and less ambitious, activity location/travel demand models that employ micro-simulation concepts. Of considerable interest are the applications of Wilson and Pownall (1976) and Kreibich (1979) who explicitly use Monte Carlo methods for examining micro-level interdependencies. In the former the interdependency between workplace, residential and shopping location was examined, under conditions of change (relocation of households in the public sector) while the latter investigated the relationship between location behaviour, car availability and modal choice using a heuristic process in which time budget constraints played a prominent role. An alternative and related treatment of constraints in activity/trip analysis is provided by Lenntorp (1976) in the time-geographic tradition of Hagerstrand (1970).

Recently, Monte Carlo simulation has been used for the purpose of aggregating micro-analytic travel demand relations (see, for example, Watanatada and Ben-Akiva, 1979), and for the purpose of examining travel decision processes. The car-pooling model developed by Bonsall (1979) is a good example of this type of work. The decision to share a ride (car pool) was found to be dependent on a large number of detailed characteristics of the driver, passenger and trip. Simulation proved to be a convenient tool for the aggregation over the decision processes of a highly heterogeneous population group.

THE SOLUTION OF MODELS BY MICRO-ANALYTIC SIMULATION: SOME FORMAL CONSIDERATIONS

The applications discussed in Section Three combine a number of methodological characteristics which can be subject to individual scrutiny. We now examine several of these issues in the context of model development and solution. Although sometimes of a rather technical nature they are of prime importance in the implementation of models developed within the micro-simulation framework.

List Processing and the Storage of Information.

A number of authors, including Orcutt et.al. (1961), Wilson and Pownall (1976) and Kain et.al. (1976), have commented on the advantages of list processing for the efficient storage and retention of information. Some comments on this important aspect are therefore necessary. A central feature here is the computational representation of the state of a system. With reference to the categorical definitions used in Section Two it is easily seen that each individual or unit may exist in any one of $\prod_{i=1}^{m} \ell_i$ possible states. The system as a whole may than be described in terms of the traditional occupancy matrix, $\underline{N}$, the elements of which, $\underline{N}(\underline{\alpha})$, record the number of units in each alternative state, $\underline{\alpha}$. Typically, for large m, $\prod_i \ell_i$ will be an exceedingly large number and, moreover, the matrix $\underline{N}$ will be very sparse - a large proportion of states (cells) will not be occupied by *any* individual.

An alternative means of recording the state of the whole system and storing equivalent information is to list the individual states of each component member $\underline{\alpha}_1, \ldots \underline{\alpha}_i, \ldots \underline{\alpha}_Q$. For a population of size Q, this will involve recording QM quantities. Which of these methods is the more efficient in its capacity for storage of information appears to depend therefore on the relative sizes of $\prod_i \ell_i$ and QM. For units which are described by a large number of categories the former may well be greater, and often considerably greater, than the latter. The efficiency of the list representation is further enhanced when it is noted that for many purposes in which simulation is adopted for the numerical solution of models, a satisfactory computation of the *required outputs* may be achieved with a sample, q, which will be a small fraction of Q. In summary, typically

$$\prod_i \ell_i \gg qM. \qquad (4.1)$$

One method of overcoming this dimensionality problem with the occupancy matrix is to form summary matrices from $\underline{N}$, in which summation is performed over certain categories, and this is often done in government and other survey publications (eg. Census, F.E.S., G.H.S., etc.) in which it is required to portray large amounts of information with due regard for confidentiality. It is commonly remarked that considerable information is lost through this process of aggregation.

It appears that both the *storage efficiency* and *information loss* arguments have been used for the strong justification of the list processing approach, which retains full information in a compact form. There are, however, a number of crucial points to be made here relating both to the nature of a model and the information involved. If it is desired to record and process information on the attributes of each member of a real population of individuals and examine the full individual information as output (eg. for tax or health service purposes) the above arguments hold. For simulation models in which the aim is to *represent* and *replicate* real world processes we are not interested in any *particular* member of the real world but in information pertaining to samples, representative in terms of the joint distribution of characteristics over that population. The properties of this joint distribution and the nature of model operations rather than the characteristics of any *particular* sample itself are then the decisive determinants of both the storage efficiency and information retention issues. It is clear that the structure of dependence between the attributes is at the heart of the matter. If for example, in an extreme case, the attributes $c_1, \ldots, c_m$ were independent in the sense that the joint density funciton, $\rho(c_i, \ldots, c_m)$, could be factored

$$\rho(c_1, \ldots, c_m) = \prod_{i=1}^{m} \rho_i(c_i) \tag{4.2}$$

then the storage of the probability matrix, ρ, may be reduced to that of the right hand side of equation (4.2). The $\sum_{i=1}^{m} \ell_i$ storage requirement will typically be considerably less than that required to store the attributes of a list of individuals sampled from ρ. This argument will hold in suitably modified form when partial dependencies occur.

Equally, it is illusory to regard information to be lost in the formation of summary matrices or summary lists, if the attributes "separated" in the aggregation process are independent. In any given application involving many attributes the arguments of efficiency and information must therefore be related to the structure of the model itself. Often considerable interdependencies *will* exist between the attributes and list processing may well be a strong contender on the above grounds. It is possible to employ mixed representations in which some information is stored in an occupancy matrix form while other partial probability matrices are recreated from list processing after model operations have been employed. This method is useful in interfacing micro- and macro-models and has been adopted by Clarke, Keys and Williams (1979).

<u>The Generation of an Initial Population.</u>

The issue of attribute dependency is also an important ingredient in the provision of a sample population, which is adopted in all the above examples. In dynamical models this represents a starting point for the foward iterative solution of dynamical equations.

The information required for the construction of the multiply classified sample $q(\underline{c})$ may be sought directly in a survey, or indeed sampled from another survey, as in the applications of Orcutt et.al. (1976), Kain et.al. (1976), and Caldwell et.al. (1979). In the absence of a suitable survey an alternative strategy is to attempt to generate $q(\underline{c})$ by sampling from a *synthesised* joint probability matrix, $\rho(\underline{c})$, or corresponding contingency table. The theoretical basis for generating entries to a full contingency table consistent with available conditional and marginal probability distributions is long established and discussed in detail in the synthetic sampling procedure outlined by McFadden et.al. (1977). In general the more complex the structure of relationships between the variables the more available information in the form of marginal probabilities will be required to reproduce this structure in a synthesised joint distribution $\rho(\underline{c})$.

A strategy which has been used in sample generation is to express the joint probability as a product of conditional distributions:

$$\rho(\underline{c}) = \rho(c_1)\ \rho(c_2|c_1)\ \ldots.\ \rho(c_m|c_{m-1}\ \ldots\ c_1) \qquad (4.3)$$

and adopt approximations which impose a simplified structure on the conditional dependencies. That is, assumptions are made about important inter-relationships between the characteristics. Examples of this approach have been discussed by Wilson and Pownall (1976), McGill (1978) and by Bonsall (1979).

In the applications examined the sample size q adopted was found to vary considerably from about 4000 in the study of Orcutt et.al. (1976) to 78,000 in the housing market study of Kain et.al. (1976). It is difficult to give specific guidelines for the required size. This clearly depends on a number of factors which include: the nature of the model output; the number of attributes and the levels selected; and, importantly, on the structure of dependency between the attributes.

Models of Characteristic (Attribute) Dependence and Matching.

Models of dependence between individual characteristics may appear both in the formation and manipulation of the samples. This process of association usually involves a parametric function specified in terms of a set of attributes which include policy variables, as in the expression (2.5). In the application of Wilson and Pownall (1976), for example, spatial interaction models were employed to express an association of workplace, residence and shopping locations in terms of the generalized cost of interzonal travel.

As we noted in Section Two,models of association at the micro-level may or may not involve a behavioural rationale, but where a representation of the decision process *is* sought the resultant model may be of very general form and include a complex set of constraints and preference relations. Individual choice models, particularly those belonging to the logit family† underpinned by random utility theory (see, for example, Williams and Ortuzar, 1979) have now been widely used in travel demand and residential location models, and indeed have been employed in the housing study of Kain et.al. (1976) and car pooling model of Bonsall (1979).

Many of the micro-simulation examples involve the matching of items of "demand" and "supply", each of which are endowed with lists of characteristics $\underline{c}_d$ and $\underline{c}_s$, respectively. The labour and housing systems provide obvious examples in terms of the respective matches {Individual $\leftrightarrow$ Job} and {Household $\leftrightarrow$ House}. Two more examples include the matching of individuals (passenger and driver) in car pooling schemes, and the merging of sexes in the marriage union, central to the process of family/household formation (see Orcutt et.al. 1976; Clarke, Keys and Williams, 1979).

† The logit family is comprised of particular funcitonal forms which may be adopted in the expression (2.5).

Certain rules of association which represent real world processes of choice or allocation are adopted by the modeller in the implementation of this matching process. From a computational viewpoint it is required to match items stored on two lists according to rules which entail a measure of association $B(\underline{\alpha}_d,\underline{\alpha}_s)$ between demand and supply micro-states. The elements of B may represent prior probabilities in information theoretic approaches, and utilities or costs of association in allocation models based on choice processes. Equally, each element of B could be expressed as a series of constraints in multicriterion satisficing approaches. The algorithm used to implement the process may thus embed probabilistic micro-models of the type discussed above and be suitably taylored to reflect "scarcity" on the demand or supply side. In the computational operation it is necessary to reflect competition between micro-units without introducing bias due to the position of each item in its respective list. In practise this bias may be rather expensive to completely eradicate - the successive double random sampling from the seperate lists and probabilistic (Monte Carlo) matching according to an index based on $B(\underline{\alpha}_d,\underline{\alpha}_s)$ can be very time consuming. It is however, often possible to employ less expensive procedures at the risk of introducing some inaccuracy. One such approach involves the solution of an aggregate allocation process at a lower level of resolution and sampling from its outcome. Such allocation procedures may be devised for interfacing individual circumstances with a macro-environment, and are of considerable theoretical interest (see, for example, the papers by Snickars and Weibull, 1977; Clarke, Keys and Williams, 1979; Anas, 1979; Los, 1979).

The matching process thus frequently embodies key elements of a model which determine, and ideally explain, the association of an individual or item with a particular state.

Dynamics and the Solution of Integrated Models.

While it is sometimes possible to provide a formal statement of the solution to a system of equations expressing socio-economic system dynamics, whether these are of a stochastic or deterministic form, it is usually necessary in applications to resort to some means of numerical integration of a set of differential (or difference) equations to achieve the required model outputs. Monte Carlo simulation is one approach to the solution of such equations and has found applicaiton in household models embracing demographic and economic processes. These include DYNASIM (Orcutt et.al. 1976), the NBER housing model (Kain et.al. 1976) and the model of household dynamics developed by Clarke, Keys and Williams (1979).

The household and its component individuals will in general partake in many dynamical processes, and indeed some induce others. This correlation between individual state transitions or events provides a dynamical interaction between individual (and household) characteristics. The complexity of dynamical behaviour derives precisely from this micro-level interdependence reflecting interaction between the various processes.

We can write a straightforward accounting identity for the change in state occupancy $\Delta N_{\underline{\alpha}}(t, t + \Delta t)$ in the discrete time interval $[t, t + \Delta t]$ in terms of the matrix $\underline{\underline{F}}(t, t + \Delta t)$ of flows or transitions between different states as follows:

$$\Delta N_{\underline{\alpha}}(t, t + \Delta t) = \sum_{\underline{\alpha}'} F(\underline{\alpha}', t : \underline{\alpha}, t + \Delta t) - F(\underline{\alpha}, t : \underline{\alpha}', t + \Delta t) \quad (4.4)$$

By a judicious choice of characteristics and associated levels (which may, for example, incorporate "duration of stay" effects in a particular state) the use of a Markovian assumption can often be justified in transition models

$$\Delta N_{\underline{\alpha}}(t, t + \Delta t) = \sum_{\underline{\alpha}'} r(\underline{\alpha}', \underline{\alpha} : t) N_{\underline{\alpha}'}(t) - r(\underline{\alpha}, \underline{\alpha}' : t) N_{\underline{\alpha}}(t) \;\; \Delta t, \quad (4.5)$$

and allows the onus of model building to be transfered to the matrix of rates $\underline{\underline{r}}(t)$. The number of characteristics which are "active" in any model operation will, of course, be dependent on the particular process involved.

The dynamics of a household may thus be formally reduced to a set of coupled differential equations in which the several processes and their respective time dependent rate matrices typically refer to overlapping sets of characteristics as discussed by Clarke, Keys and Williams (1979). This interdependency and the sparceness of the rate matrices comment the use of list processing in conjunction with the Monte Carlo method. In the sequential consideration of individual items in a list it is computationally expedient to arrange for an examination of the participation of each member in as many processes (giving birth, death, job/house change, etc.) as possible consistent with the causal restrictions imposed in the model, to avoid wasteful and repetitious examination of list information.

The most challenging aspect of dynamic modelling involves an attainment and understanding of the dependence of the rate matrices

$$\underline{\underline{r}} = \underline{\underline{r}}(\underline{Z}(t), t) \tag{4.6}$$

on time and (time dependent) policy instruments $\underline{Z}(t)$. Time series information may be available, and again, the explicit derivation of models from decision contexts may be entertained. Although much progress has been made in deriving certain of these functional dependencies (see, for example, Orcutt et.al., 1976) the establishment of a comprehensive model of demographic-economic interactions at the micro-level remains a considerable challenge.

PROSPECTS AND PROBLEMS

We have described some aspects of the development and application of simulation models based on a smallest unit representation. It is clear from the above discussion that the method is interwoven by several themes, and the framework we have outlined is one in which data can be organised, transformation equations established and a model solution method devised.

In order to discuss the merits and limitations of the approach in any modelling context we must compare its advantages with those of its competitors. In addition we may ask whether subject areas which have not traditionally been associated with the approach would benefit from the use of microsimulation. In the applications described in Section Three, which involve the dynamics or statics of demographic and economic/activity systems, alternative approaches have involved some form of aggregation. As emphasised in Section Two the aggregation issue embraces many features. Some, for example the aggregation over functional relationships, concern the bias of parameter estimates, others relate rather more closely to model design, and the type of information required. One can always point to the worst excesses of the aggregate approach, and this form of criticism has adorned the literature in recent years (c.f. transportation models). However, we would point out that the arguments in favour of a highly disaggregate micro-approach are not as obvious as some of the proponents might suggest. For some purposes, as in the case of short-term population forecasts, it might be simply unnecessary to develop a micro-approach - trend forecasts developed within an appropriate accounting framework may be sufficient. This is equally true of the production sector where the firm, as a decision unit, bears many formal similarities with the household (Thrift and Williams, 1980). While the household sector is demonstrably complicated in its range of possible micro-behaviour it has considerably less variety both in its characteristics and its potential response and adaptive behaviour. Although some micro-models do exist at the firm level drawing on the initial impetus provided by Clarkson and Simon (1960), Shubik (1960) and Cyert and March (1963), and a mass of empirical work has been undertaken, there has been considerably less progress than in the household sector. Again for conventional forecasting issues econometric and input-output models may suffice, and indeed it was primarily the complexity of the economic behaviour of individual and interacting firms which inspired Leontief's aggregate approach.

In practice the problems of data acquisition and measurement can often frustrate the development of micro-models of the type discussed. It is not uncommon to have some aggregate information available in time-series form and some micro-information available at one cross-section. By generating aggregate forecasts from both macro- and micro-models one could hope to combine the distinct and complimentary features of both. Aggregate and disaggregate approaches should not necessarily be regarded as competitors but formal strategies for achieving complementarity await full development. This goal may be brought nearer by the wider availability of longitudinal data sets.

In any approach to problem solving or policy analysis computational issues and questions of efficiency should be clearly distinguished from representational and theoretical aspects. Many factors will determine the formulation of a model, including, the available inputs, the required outputs and the resources at hand. An assessment of a model must not be divorced from the context within which it is created or applied. The excessive promotion of a particular methodological perspective has in the past tended, ultimately, to be to its detriment. With regard to micro-simulation as an *approach* to problem solving we would, however, conclude on a note of guarded optimism. The smallest unit representation discussed here appears to be an extremely flexible one in the study of emerging and important themes, notably the inter-relationship between an individual and its environment, whether this be of a social, economic, institutional or spatial nature. Moreover, it does appear as Wilson and Pownall (1976) have suggested, highly suitable for an examination of extensive inter-dependencies between attributes which arise from a variety of processes and constraints. Our own work at Leeds is concerned with a number of applications concerning these inter-dependencies. Of particular interest is the mutual interaction both between individuals within a household, and between individuals and an environment created by the labour and housing systems, local and central government.

Many of our remarks have been implicitly concerned with traditional forecasting approaches. What might prove to be of particular value is the use of micro-simulation in a context of relatively new approaches to problem and policy analysis. The method is wholly complementary to gaming-simulation approaches in which the mutual interaction of decision makers is subject to experimental scrutiny, and may readily be embedded within control frameworks. The approach is equally suitable for application in a range of different planning frameworks as diverse as traditional cost-benefit analysis and planning for freedom of action in which the range of feasible alternatives available to individuals with widely different personal and environmental circumstances is the central focus of interest.

The method we have described can be viewed as a "first principles" approach to the representation of systems and as such provides the basis for the design of models and appropriate data sets, and the study of interdependencies and systemic effects originating at the micro-level.

REFERENCES

Anas, A. (1979) The Impact of Transit Investment on Housing Values. A Simulation Experiment. Environment and Planning, A 19, pp. 239-255.

Bonsall, P. (1979) Microsimulation of Mode Choice: A Model of Organised Car Sharing. P.T.R.C. proceedings, Summer Annual Meeting, University of Warwick, July 1979.

Caldwell, S., et.al. (1979) Forecasting Regional Energy Demand with Linked Macro/Micro Models. Working Papers in Planning, No.1, Dept. of City and Regional Planning, Cornell University, Ithaca, New York.

Clarke, M., Keys, P. and Williams, H.C.W.L. (1979) Household Dynamics and Socio-economic Forecasting: A Micro-Simulation Approach. Paper presented at the European Congress of the Regional Science Association. Working Paper 257, School of Geography, University of Leeds.

Clarke, M., Keys, P. and Williams, H.C.W.L. (1980) Models of Matching and Mobility in the Labour and Housing System. Working Paper, University of Leeds.

Clarkson, G.P.E. and Simon, H.A.S. (1960) Simulation of Individual and Group Behaviour. American Economic Review 50, pp. 920-932.

Cyert, R.M. and March, J.G. (1963) A Behavioural Theory of the Firm. Prentice Hall, Englewood Cliffs, New York.

Green, H.A.J. (1964) Aggregation in Economic Analysis. Princetown, University Press, Princetown, New Jersey.

Guthrie, H.W. (1972) Microanalytic Simulation of Household Behaviour. Annals of Economic and Social Measurement 1/2, pp. 141-169.

Hagerstrand, T. (1970) What about People in Regional Science? Papers and Proceedings of the Regional Science Association, 24, pp. 7-21.

Harris, R. (1978) Microanalytic Simulation Models for Analysis of Public Welfare Policies. The Urban Institute, Washington D.C.

Ingram, G.K., Kain, J.F. and Ginn, J.R. (1972) The Detroit Prototype of the N.B.E.R. Urban Simulation Model. National Bureau of Economic Research, New York.

Kain, J.F., Apgar, W.C. and Ginn, J.R. (1978) Simulation of the Market Effects of Housihg Allowances. Volume 1: Description of the N.B.E.R. Urban Simulation Model. Research Report, Dept. of City and Regional Planning, Harvard University.

Kreibich, V. (1979) Modelling Car Availability, Modal Split and Trip Distribution by Monte Carlo Simulation: a short way to integrated models. Transportation, 8, pp. 153-166.

Leigh, R. and North, D.J. (1978) The Potential of the Micro Behavioural Approach to Regional Analysis. In P.W.J. Batey (Ed.), London Papers in Regional Science, 8, Pion, London.

Lenntorp, B. (1976) Paths in Space-Time Environments. Land Studies in Geography, Series B, Human Geography, No.44.

Los, M. (1979) Discrete Choice Modelling and Disequilibrium in Land Use and Transportation Planning. Publication 137, Centre de Recherche sur Les Transports, University of Montreal.

McFadden, D., Cosslett, S., Duguay, G. and Jung, W. (1977) Demographic Data for Policy Analysis. The Urban Travel Demand Forecasting Project, Phase 1, Final Report Series, Volume 8. Institute of Transportation Studies, University of California, Berkeley.

Macgill, S.M. (1978) Multiple classifications in Socio-Economic Systems and Urban Modelling. Working Paper 212, School of Geography, University of Leeds.

Orcutt, G., Greenberger, M., Rivlin, A. and Korbel, J. (1961) Microanalysis of Socioeconomic Systems: A Simulation Study. Harper and Row, New York.

Orcutt, G., Caldwell, S. and Wertheimer, R. (1978) Policy Exploration through Micro-analytic Simulation. The Urban Institute, Washington, D.C.

Shubik, M. (1960) Simulation of the Industry and the Firm. American Economic Review, 50, pp. 908-919.

Snickars, F. and Weibull, J. (1977) A Minimum Information Principles - Theory and Practice. Regional Science and Urban Economy, 7, pp. 137-168.

Thomas, M.D. (1969) Regional Economic Growth: Some conceptual aspects. Land Economics, 45, pp. 43-51.

Thrift, N. (1980) Behavioural Geography: A Paradigm in Search of a Paradigm. In, Bennett, R.J. and Wrigley, N. (Eds.) Quantitative Geography in Britain: Retrospect and Prospect. Routledge and Kegan Paul. London.

Thrift, N. and Williams, H.C.W.L. (1980) On the Development of Behavioural Location Models. Economic Geography (in press).

Warner, S.L. (1962) Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice. Northwestern University Press, Evanston, Illinois.

Watanatada, T. and Ben-Akiva, M. (1978) Spatial Aggregation of Disaggregate Choice Models: An areawide urban travel demand sketch planning model. Presented at the Transportation Research Board Annual Meeting, January 1978.

Williams, H.C.W.L. and Ortuzar, J.D. (1979) Behavioural Travel Theories, Model Specification and the Response Error Problem. Paper presented at the P.T.R.C. Annual Summer Meeting, University of Warwick, U.K.

Williams, H.C.W.L. and Wilson, A.G. (1979) Some comments on the Theoretical and Analytic Structure of Urban and Regional Models. Working Paper, School of Geography, University of Leeds.

Wilson, A.G. and Pownall, C.E. (1976) A New Representation of the Urban System for Modelling and for the Study of Micro-Level Interdependence. <u>Area</u>, <u>8</u>, pp. 246-254.