Monograph:
Bonsall, P.W. (1979) The Simulation of Organised Car Sharing (2) – The Simulation Models and their Calibration. Working Paper. Institute of Transport Studies, University of Leeds , Leeds, UK.

Working Paper 109

Working Paper 109

May 1979

THE SIMULATION OF ORGANISED CAR SHARING (2) - THE SIMULATION MODELS AND THEIR CALIBRATION

by

Peter Bonsall

Working Papers are intended to provide information and encourage discussion on a topic in advance of formal publication. The work reported here was sponsored by the Transport and Road Research Laboratory, but reports in this series represent only the views of the authors and do not necessarily reflect the view or approval of the sponsors.

CONTENTS

ABSTRACT

BONSALL, P.W. (1979) The simulation of organised car sharing II - The simulation models and their calibration. Leeds: University of Leeds, Inst. Transp. Stud., WP 109 (unpublished).

This paper is one of a series describing the methods and findings of a TRRL sponsored project to simulate organised car sharing. This volume describes the simulation model and its calibration. The model requires as input a description of each individual tripmaker within the system (there are 180,000 such individuals in the study area under investigation). Using these individuals as actors, the model simulates, within a calibrated choice model, the decision of each of them in turn whether or not to apply to join a hypothesised car sharing scheme. The matching of such applicants to form potential car pools is then effected in a special routine. A second calibrated choice model then simulates the decisions by each of the applicants whether to enter an arrangement with any of the potential partners with whom they have been provisionally matched. A mutual evaluation of utilities then determines which arrangements will actually come to fruition.

The microsimulation approach to transport modelling involves consideration of choice options by the fundamental actors within the system - the individual tripmakers. The approach is inherently attractive but only recently have decreasing computer costs made microsimulation a viable branch of travel demand modelling.

The choice models were calibrated on the basis of a series of field surveys which were designed to mirror exactly the simulation models - respondents were invited to make decisions and to evaluate proposals drawn from the simulation models themselves. Thus the respondents effectively became actors within the simulation and their reactions to the choices available to them were used to calibrate the models.

Previous investigations of car sharing have concentrated either on the behavioural and psychological aspects of the mode with little attempt to estimate the global consequences of these or have concentrated on the probable demand for the mode with little attempt to consider the interaction between potential matchees.

The microsimulation approach adopted here has successfully combined a calibrated model of demand for the mode with an accurate rendition of the supply side - the matching of individual trip makers with compatible requirements and, finally, a calibrated model of the decision whether or not to enter an arrangement with a specified individual.

This approach has been particularly appropriate to the modelling of organised car sharing but can obviously provide the basis for a whole range of behaviourally orientated planning models.

THE SIMULATION OF ORGANISED CAR SHARING. II THE SIMULATION MODELS AND THEIR CALIBRATION

1. INTRODUCTION

1.1 Aims of the project

The project was conceived during 1976 and submitted to the SRC for grant funding. The submission was then modified in the light of interest expressed by the TRRL who subsequently undertook to fund the work on a contractural basis.

The primary objective of the study was to provide guidance for policy makers contemplating the implementation or modification of car sharing schemes. This guidance to be in the form of the relationships established between performance of schemes, the policy environment in which they operate and the nature of the schemes themselves.

The performance of the schemes is expressed in terms of their effect on transport system indices including peak period car mileage, peak period public transport patronage and numbers of cars 'liberated' for off-peak use.

The project seeks to predict how the performance of schemes would reflect their operational characteristics such as size and location and how they would be affected by economic/financial deterrents and incentives or by traffic restraint policies. The project concentrates on car sharing schemes for peak period work trips.

Several studies have addressed themselves to the potential market for organised work journey car sharing (Tomlinson and Kellett 1977, Vincent and Wood 1979, Cambridge Systematics Inc 1976, Atherton et al 1976) but they have been concerned mainly with the potential and theoretical impact of car sharing given present journey-to-work patterns and characteristics. They have been able to contribute little to the estimation of likely impact because they could not estimate how many of the potential matches could or would be realised. Another line of research has been concerned with attitudes to car sharing in an attempt to understand the likely response at the micro level (Margolin et al 1976, Dobson and Tischer 1976, Levin et al 1978, Tomlinson and Kellett 1978, Hawker Siddeley Dynamics 1977). This attitudinal work has provided useful insight into the likely behaviour of individuals but it is, in itself, not readily adapted for predictive purposes because it is concerned with individuals rather than populations.

It was our aim in this project to bridge the gap between theoretical modelling and attitudinal investigation by developing a model which, while being based on the attitudes and consequential decisions of individuals, could take into account the availability and characteristics of potential partners and could thus predict the impact of a carsharing scheme at both the micro and macro level. The form of model best suited to this task is microsimulation. The resulting model seeks to represent the interactions between individual decisionmakers and the manner in which an organised car sharing scheme would operate.

1.2 Microsimulation

Microsimulation is a technique of computerised modelling within which the decision making process is replicated for individual decision makers within the system. These decision makers effectively become 'actors' within the modelled system. The model is driven by Monte Carlo type sampling.

Monte Carlo simulation has, of course, a long and respected pedigree particularly in the field of Operational Research. But it has not been much applied to travel demand modelling. Recently, however, Monte Carlo simulation has been used in the theoretical investigation of logit and probit models (Albright et al 1977, Ortuzar and Williams 1979) and as the basis of a model reported by Kreibitch (1978). In his model the population is divided into groups ('situation groups') deemed to share a common 'decision profile'. The decision profile is expressed as a table of probabilities of making a particular decision and is activated using a random number generator.

The main difference between Kreibitch's approach and microsimulation as presented in this paper, is that in its pure form, microsimulation makes explicit the mechanisms of decision making rather than relying on correlation.

1.3 Microsimulation compared with other model forms

The development of microsimulation techniques should be seen in the context of the current emphases (Manheim 1979, Williams 1979) within travel demand modelling. A changed emphasis in planning, away from the blueprint plans of the post war years and towards the incrementalism of the post oil crisis, has been matched by increasing disillusion with aggregate planning models and greater interest in disaggregate and behaviourally orientated models.

Much of the work on disaggregate modelling has been concerned with the development of the logit models and their derivatives. Advantages quoted (Atherton and Ben-Akiva 1977) for this type of model when compared with aggregate models include greater statistical efficiency, transferability, behavioural structure and policy sensitivity. Other authorities, however, dispute that the structure of logit based models is behaviourally valid, and insist that a radical change in model framework is required if the behavioural dimension is to be given a place. This thinking has developed the pioneering work of Hagerstrand (1976) into the activity based gaming model, 'HATS' (Heggie 1977, Heggie and Jones 1978). A problem with the HATS approach has, however, been its computational intractability.

The mathematical expression of conventional planning models (which for this purpose must include logit and probit models) tend to obscure any behavioural basis which they may have. This makes it difficult or impossible to represent the nuances of behavioural logic within them. Against this background it will never be possible to develop a truly causal model. Nor will it be possible to convince the layman-politician that planning models are anything more than black boxes. The development of HATS and of microsimulation promises to allow progress on both these fronts.

Microsimulation seems to allow for a combination of some of the philosophical advantages of the HATS approach with the computational advantages of more conventional model forms. The main advantages which the HATS approach and microsimulation have over more conventional model forms are their detailed representation of the decision making process and their essential simplicity. These twin advantages make them uniquely suitable vehicles for testing paradigms of behaviour adnd as aids to policy formulation. The computational intractability of HATS however, restricts its role in predictive planning and it is here that the value of microsimulation lies.

1.4 Revealed preference or stated intentions?

The dangers of basing predictive models on stated intentions are well known, they stem from the known divergence between what a respondent says he would do in a given, hypothetical, situation and what he in fact does if and when that situation arises. This divergence is due to the difficulty of replicating the environment in which the real decision would be made. It is important that the respondent should state his intentions in proper cognisance of the facts and under the same constraints which would affect his actual decision. Furthermore he must act as if his very statement of preference would entail a real commitment - if he thinks he can state intentions willy-nilly his decision is unlikely to be as cautious or realistic as it would be in reality. Arguments of this kind, although rarely articulated, have discouraged the use of stated intentions for the calibration of predictive models.

There are, however, a number of objections which can be raised to the conventional use of revealed preference data. Firstly data availability usually forces the use of cross-sectional rather than time-series data; this necessitates the heroic assumption that spatial and circumstantial variation in behaviour can be used to predict temporal changes. It means that models can at best be correlative - they can never be causal.

A second drawback of revealed preference data, even if it be time-series, is that it is not retrospective - it shows behaviour in the context of existing circumstances rather than in the context of the circumstances which prevailed when the behavioural decision was actually taken. Environmental factors may have caused the behaviour to be adopted but inertial effects will almost certainly ensure that the behavioural pattern outlives its causes. Correlation between co-existing behaviour and environment will rarely reveal causality and is thus a dubious basis for prediction.

A lengthy comparison of the relative merits of revealed preference and stated intention data will conclude that they share the same basic problem - an inability to construct the environment in which decisions are actually taken.

In the current project we seek to model a mode choice which does not yet exist in the field - organised car sharing. Clearly we have to choose between two options:

- a revealed preference model vased on observed behaviours which we may assume to be correlated with reaction to organised car sharing;
- or a stated intention model based on reactions to a hypothesised car sharing scheme.

Both of these options are compatible with a microsimulation framework. If we chose the revealed preference option then we would be working at two removes from the phenomenon we wish to model - we would be observing behaviour which we assume to be correlated with organised car sharing behaviour but which itself may have arisen in circumstances different from those which prevail at the time that the behaviour is observed. In such circumstances we would be unable to capitalise on all the advantages of microsimulation which were outlined in previous sections of this paper. The Monte Carlo simulation model described by Kreibitch was based on revealed preference data and, in the view of the present author, this must detract from its usefulness.

If the problems associated with using stated intention data can be overcome then its combination with microsimulation can prove an attractive basis for predictive modelling. It is this combination which the current model seeks to achieve.

2. STRUCTURE OF THE MODEL

2.1 Introduction

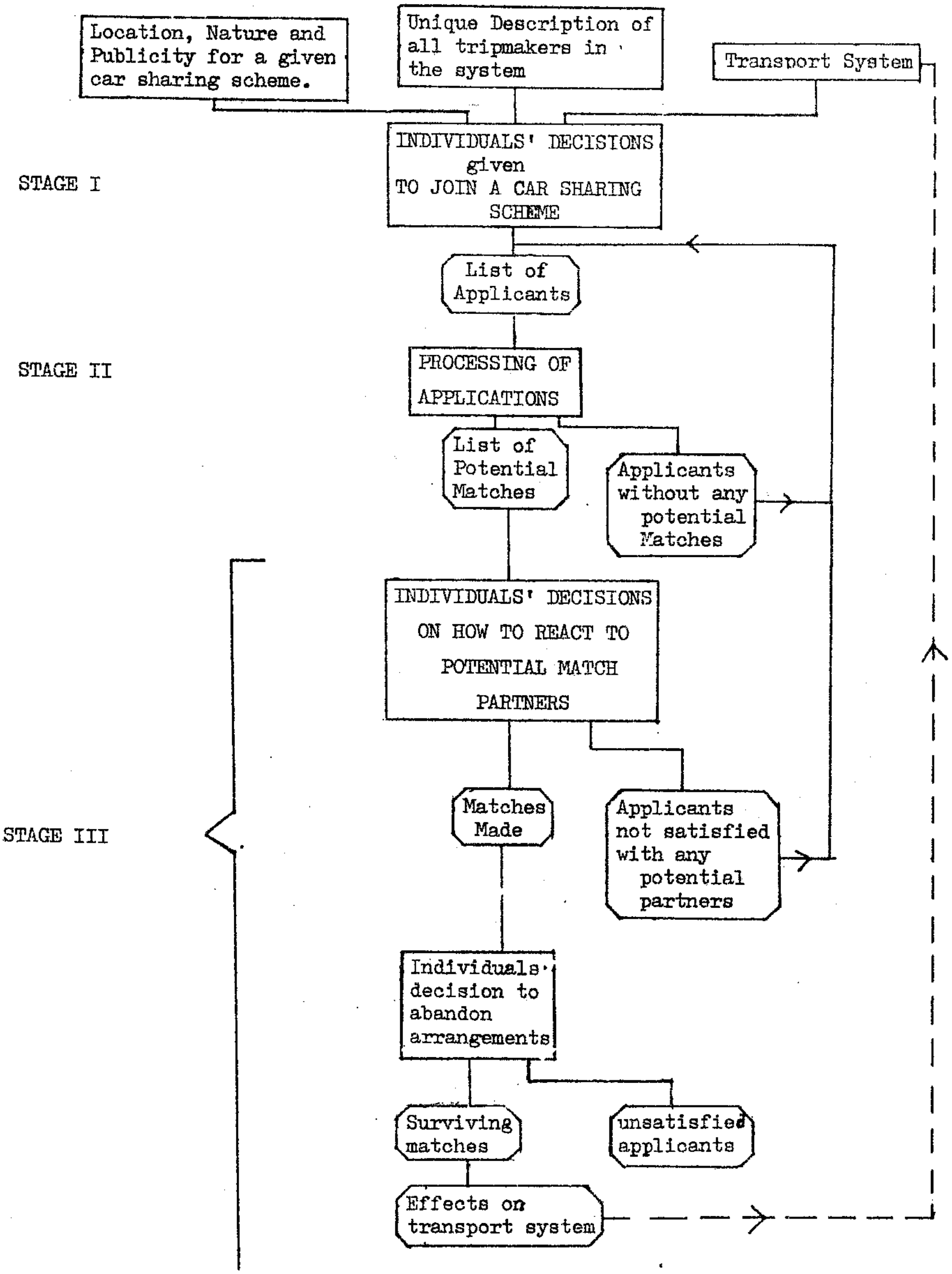
The simulation suite itself has three stages, each representing a distinct process in the stablishment of an organised car sharing scheme. These three stages are represented in Figure 2.1.

The first stage is concerned with the scope and intensity of the scheme being simulated and the decisions by members of the public to be associated with it. The second stage deals with the mechanics of attempting to match up potential partners, the third with the reactions of the participants in the scheme to their proposed partners. An appendix to the main model translates the performance of the scheme into its effect on some critical components of the transport system.

2.2 Synthesis of the population base

The microsimulation approach to modelling requires, as a fundamental input, a description of each of the actors in the system of interest. It is not possible to replace this list of unique individuals with the combined probability matrices which define them because of the problem of accounting - as each individual passes through the system records must be kept of his progress. This is particularly important in the present case because it is a fundamental feature of carsharing that there be absolute equality between supply and demand (each lift is given once and once only); clearly this equality can best be guaranteed if accounts are kept.

Figure 2.1. OUTLINE STRUCTURE OF SIMULATION SUITE



Thus we require descriptions of each of the actors in the system. In the current case this means a unique description of every peak period work tripmaker in the study area. These descriptions have to define the individuals in such detail that we can estimate their propensity to join a carsharing scheme, their reaction to proposed matches and their influence on the transport system. In many instances a sample of the actors would suffice but in the modelling of car sharing a complete population is necessary. This is because, for car sharing, successful arrangements are a function of the compatability between individual suppliers (lift givers or poolers) and individual demanders (lift receivers or poolers); one of the factors bearing on this compatibility is clearly the spatial relationship between the potential partners and this is a function of residential densities which can not be satisfactorily represented with a sample population.

Had it been possible, the use of a sample population would obviously have reduced the computational requirements of the model. In some circumstances the amount of computation required for modelling the total population by microsimulation would prove prohibitive. In the case of car sharing, however, the model can be arranged sequentially so that that part of the population (the majority) who express no interest in car sharing can be discarded in the early stages of the simulation. The more complex parts of the simulation (stages II and III of Figure 2.1) can then proceed with a manageable number of actors.

Ideally, of course, the population of actors would be taken directly from a 100% household census. Clearly such censuses are rarely available and so a second-best solution must be adopted. We did have available a sample survey of 9,500 households in our study area with files for individual trip makers. (WYTCONSULT 1976) It would not have been appropriate merely to multiply the sample data by the sampling function because to have done so would have produced a population of sets of identical people - whose mutual interactions could not be taken as representative of a true population! A more sophisticated method of synthesis was therefore necessary.

A full description of the method of synthesis is described in a companion volume of the present paper (Bonsall and Champernowne 1979). In summary, however, the method was based on the use of inter-characteristic probabilities revealed in the household survey to generate individuals within control totals derived from published census material (OPCS 1973a,b).

Table 2.1 lists the personal characteristics which were synthesised for each number of our population. They are characteristics which *a priori* can be expected to influence an individual's propensity to join and be accepted in a car sharing scheme. Other characteristics would no doubt be equally important but there would have been no point in our synthesising characteristics which could not be used in our simulation - this ruled out such things as race, education and income for which we could not expect to achieve reliable attitudinal data for calibration purposes. Other characteristics had to be ignored because of poor data on their distribution within the population (eg. smoking habits and political persuasion).

2.3 The definition of the scheme to be tested

2.3.1 The simulation suite accepts parameters which describe the location and intensity of the car-sharing scheme being simulated. These parameters comprise a list of residential areas and of work locations to be included in the scheme and a 'Threshold of interest' below which individuals are deemed not to participate in the scheme.

By manipulating the list of residential zones and work areas, it is possible to simulate anything from a county wide scheme to one which links a single city centre zone to a given suburb. By manipulating the threshold of interest it is possible to represent publicity campaigns of varying intensity, from one which results in the participation of all trip makers in the target area down to one which interests only a minute proportion of the population. Also by manipulation of the threshold of interest it is possible to order the list of applicants on the assumption (uncalibrated) that the keenest applicants apply quickest. By manipulating the list of zones in conjunction with the threshold of interest it is possible to simulate the complex effects of a publicity campaign whose intensity varies in time as well as space. This may turn out to be an important element in the organisational strategy of car sharing schemes.

2.4 The simulation of individuals' decisions to join a given car sharing scheme

2.4.1 This submodel, together with its inputs and outputs, is represented in Figure 2.2

TABLE 2.1 CHARACTERISTICS DEFINED FOR EACH MEMBER OF OUR POPULATION

1) Precise location of home (6 Figure grid-reference)

2) Precise location of workplace (6 Figure grid-reference)

3) Sex

4) Age (under 30, 30 to 50 or over 50)

5) Whether head of household.

6) Driving licence tenure.

7) Employment category (Manual/shop floor, tech-nical/clerical or professional/management)

8) Whether car needed at work (business use)

9) Current mode of travel to work (ie. prior to introduction of car sharing scheme)

   Possible modes = 1) Solo car driver
   2) Car driver with one passenger
   3) " " " two passengers
   4) " " " three or more passengers
   5) Car passengers
   6) Public transport
   7) Any other mode

   the evening mode is not constrained to equal the morning mode thus there are 49 possible modal combinations.

10) Normal time of arrival at work.

11) Normal time of departure from work.

12) Number of cars available in the household.

13) Number of licensed drivers in the household.

14) Total number of people in the household.

15) Houshold telephone?

In addition to these 15 characteristics each individual is allocated a reference number indicating which household he is a member of and his unique identity within that household. Each individual is also allocated a random number with which to seed the Montecarlo sampling.

FIGURE 2.2 STRUCTURE OF SIMULATION OF APPLICATIONS

unique description
of each tripmaker
within the system
(15 characteristics)

For each tripmaker:

Given his characteristics

1. How likely to apply to give lifts morning & evening?
2. " " " " " " morning only?
3. " " " " " " evening only?
4. " " " " " receive lifts morning & evening?
5. " " " " " " " morning only?
6. " " " " " " " evening only?
7. " " " " to pool (alternate driving)?
8. " " " allow more than one passenger?
9. " " " " " " two passengers?

(likelihoods expressed on a scale 0-15)

Calibrated
Coefficients
(logit model
parameters)

unique description of each tripmaker
with his likelihood of applying/
allowing each of the 9 modes

FOR EACH TRIPMAKER:

Given his Characteristics:

Is he within the target population?
if no, reject him

Given his likelihoods:

Would he apply to a scheme of this intensity?

if no, reject him.

What type(s) of application(s) would he make?
(the 9 modes)

DESCRIPTION OF
SCHEME (location
and intensity)

unique description of each tripmaker
deemed to make an application and
description of the type(s) of
application(s) he would make

The model allows applications for up to seven types of car sharing arrangement:

1. Car pooling
2. Giving lifts morning and evening
3. Giving lifts morning only
4. Giving lifts evening only
5. Receiving lifts morning and evening
6. Receiving lifts morning only
7. Receiving lifts evening only.

An individual may make an application to include any number of these types of arrangement subject only to the following restrictions:

- an individual without a full car driving licence cannot apply for arrangement types 1, 2, 3 or 4
- an individual without a car available to him cannot apply for arrangement types 1, 2, 3 or 4
- an individual needing his car at work for business purposes cannot apply for arrangement types 1, 5, 6 and 7
- an individual whose normal arrival time at work is earlier than 0638 or later than 1022 cannot apply for arrangement types 1, 2, 3, 5 or 6
- an individual whose normal departure time from work is earlier than 1523 or later than 1907 cannot apply for arrangement types 1, 2, 4, 5 or 7.

(These last two restrictions are introduced because of the high marginal cost of processing applications for times so far outside the main peak periods, the rather peculiar time bands are a result of our desire to include the 15 most popular quarter hour periods. 15 because of computational requirements). Individuals making applications of types 1, 2, 3 or 4 are required to indicate the maximum number of passengers that they would want in their car.

2.4.2 The likelihood of a tripmaker making any of the seven types of application mentioned above is deemed to be a function of certain of that tripmaker's characteristics. The relevant characteristics are listed in Table 2.2.

In order to establish the importance of these characteristics a series of binary logit models were calibrated. We recognise that this is a departure from our desire to make explicit the mechanism of all choices within the model suite but it was a compromise forced on us by constraints

TABLE 2.2 DETERMINANTS OF APPLICATION

0. A dummy (always set to 1)
1. The length of the individual's journey to work.
2. 1 if individual's normal mode of travel to work is solo driver, otherwise 0.
3. 1 if individual's normal mode of travel to work is accompanied driver, otherwise 0.
4. 1 if individual's normal mode of travel to work is private transport passenger, otherwise 0.
5. 1 if individual's normal mode of travel to work is public transport, otherwise 0.
6. 1 if individual's normal mode of travel from work is solo driver, otherwise 0.
7. 1 if individual's normal mode of travel from work is accompanied driver, otherwise 0.
8. 1 if individual's normal mode of travel from work is private transport passenger otherwise 0.
9. 1 if individual's normal mode of travel from work is public transport, otherwise 0.
10. 1 if individual is under 30 years of age, otherwise 0.
11. 1 if individual is over 50 years of age, otherwise 0.
12. The number of cars available to the household and not needed for business use.
13. 1 if individual has a full car driving licence, otherwise 0.
14. 1 if individual is a factory or manual worker, otherwise 0.
15. 1 if individual is a professional or managerial worker otherwise 0.
16. 1 if individual is female, otherwise 0.
17. Number of licensed persons in the individual's household.
18. Number of unlicensed persons in the individual's household.
19. 1 if individual's journey to work is between 0638 and 1022, otherwise 0.
20. 1 if individual's journey from work is between 1523 and 1907 otherwise 0.
21. 1 if individual's household has a telephone, otherwise 0.

of time and resources. The models were regression transformations of the standard logit model. They can be expressed as:

$$P = \frac{e^{\sum_{i=0}^{21} a_i x_i}}{1 + e^{\sum_{i=0}^{21} a_i x_i}} \qquad 2.1$$

where P is the probability of making an application

x_i is the value of the ith characteristic of the individual being considered

a_i is the calibrated coefficient.

The calibration process is described in section 3.2

Application of this logit model for each individual in the population produces, for him, a probability of applying to join a car sharing scheme. This probability is then compared with a random number drawn from a rectangular distribution between 0 and 1. - the ratio of the probability to the random number is then deemed to be the 'likelihood' of that individual making an application (it is this likelihood upon which the 'threshold of interest' described in section 2.3.1 operates).

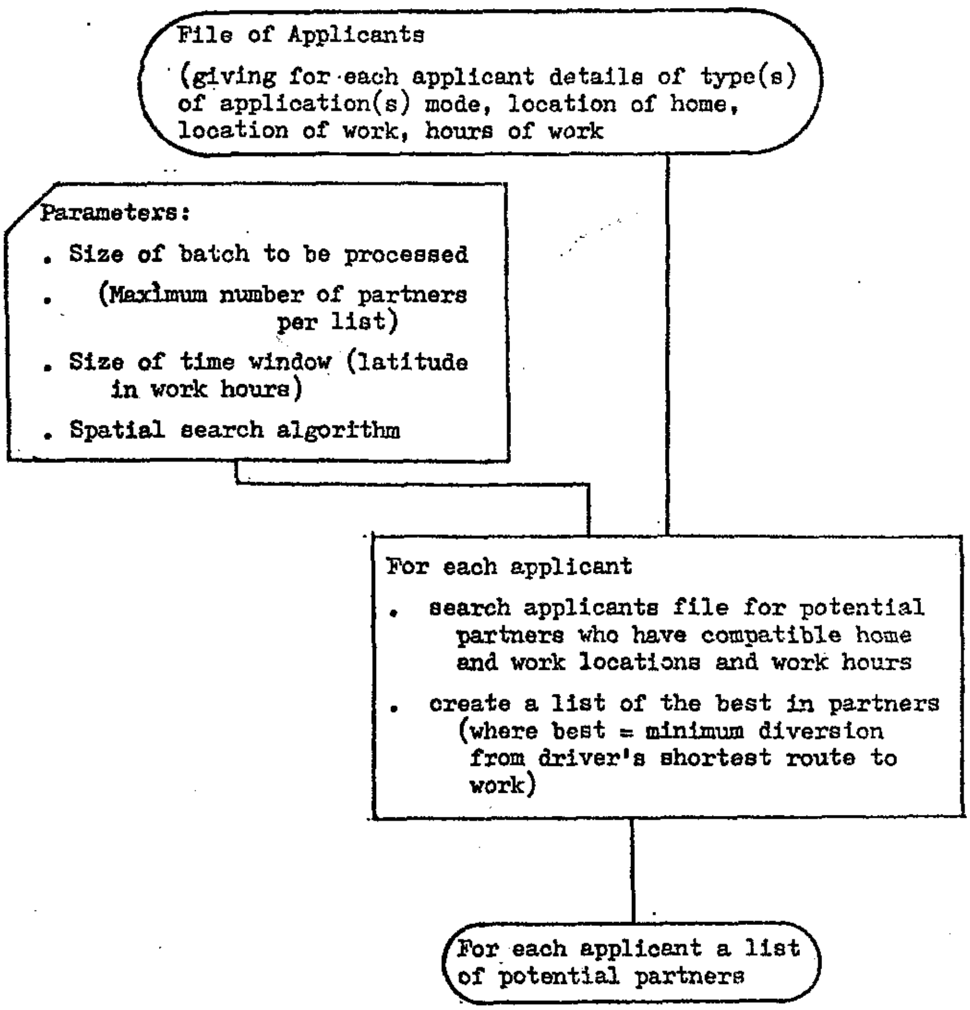
2.5 The processing of applications

2.5.1 This submodel is shown in Figure 2.3. It is a direct representation of the matching process which is fundamental to organised car sharing schemes. A matching system will accept a file of applicants, and will produce, for each applicant, a list of people whose journey to work characteristics and expressed interest in car sharing make them, prima facie, viable travelling companions. In practice this matching process may be manual, using pigeon holes or pins on a map, or computerised (several packages exist in the USA). (See for example USDOT 1974).

The simulation model perhaps bears a closer resemblance to computerised matching because its search routine is based on co-ordinate geometry rather than on a detailed road network.* The simulation suite was, however,

* The advantages of matching on the basis of a detailed road network are most pronounced on areas with a sparse population or with substantial barriers to movement. However, the computational costs of matching on the basis of a detailed road network are considerable and this has precluded such refinements in the current generation of matching algorithms.

Figure 2.3 THE PROCESSING OF APPLICATIONS



designed in such a way that the processing of applications could be done manually by bypassing the computerised matching routine. It will be interesting to examine how different the model results are using such manual intervention compared to the computerised method.

2.5.2 Within the simulation suite a matching algorithm has been programmed to process a batch of applicants (ie. individuals who made an application to join the scheme at stage 1 of the model) producing, for each active applicant, a list of potential partners whose work hours are comparable with his own, whose trip ends are close to his own and who have made applications for modes of car sharing which are compatable with those of the applicant. The algorithm accepts parameters to control the following aspects of the simulated matching process:

1. The size of batch of applicants to be processed at one time (a larger batch is more expensive and time consuming but will result in more successful matches).
2. The rigorousness of the constraints on compatability of work hours. (Varies from insistance that both parties' work hours be within the same 15 minute band to total relaxation of the time constraint).
3. The number of potential partners to be included on each applicant's list.
4. The nature of the search routine used to create the list of potential partners.

The fourth parameter is the most complex, it defines the spatial search algorithm to be used in the matching process. Two alternative spatial search algorithms have been programmed for inclusion in this model. The first is a highly efficient (and unique!) routine which assumes that one end of the trip is common to all applicants (as will be the case in employer-based schemes); it is based on an ordered search in concentric elipses. The second method is somewhat more expensive computationally but allows for variation at both ends of the trip (as would be the case in area-wide schemes). Details of these two algorithms are given in appendix A.

2.6 The simulation of the decision to match

2.6.1 This part of the simulation suite is the most ambitious and is closest to the ideal of microsimulation. It represents the consideration, by each applicant, of the list of potential travelling companions sent to him by the car sharing scheme organisers. This consideration is assumed to involve an evaluation by the applicant of the net expected utility associated with each possible arrangement presented by his list of potential partners. This evaluation is made on the basis of the known and expected characteristics of the arrangement postulated. If an arrangement has a positive net expected utility to all participants within it and has a higher utility than any other arrangement to at least one of them then it is deemed a successful car sharing arrangement.

The model is thus based on utility maximisation with a satisficing constraint. The utility to a given person P of a given arrangement A is a function of personal characteristics of the person P, of the personal characteristics of his partners in arrangement A and of the operational consequences of the arrangement (delays, diversions etc) on the participants. These utilities can be represented as

$$U_{AP} = \sum_{n=1}^{N} \sum_{m=1}^{M} a_n \, p_m \, x_{nm} + e_{nP} + \text{feepaid} \qquad 2.2$$

where U_{AP} is the utility of the arrangement A to person P

$a_1 \ldots a_n$ are characteristics of the arrangement A (see Table 2.3a)

$p_1 \ldots p_m$ are characteristics of the person P (see Table 2.3b)

$x_{11} \ldots x_{nm}$ are components of utility associated with any person with characteristic m engaging in an arrangement with characteristic n

$e_i \ldots e_{nP}$ are stochastic elements associated with the utility to person P of an arrangement with characteristic n.

feepaid is the net sum of money, if any, passing to this person in respect of his participation in the scheme.

The calibration of the components x was on the basis of a series of regression equations using data from a special field survey (see Bonsall 1979a). It is described in section 3.3 of this paper. The calibration procedure leaves a residual error term which we take to be normally distributed.

TABLE 2.3 CHARACTERISTICS CONTRIBUTING TO THE UTILITY OF AN ARRANGEMENT TO A GIVEN INDIVIDUAL

(a) Characteristics of the arrangement (s_n in equation 2.2)

From a passenger's point of view:

1. standard arrangement
2. whether the driver is female
3. number of minutes earlier than previously that the arrangement will require passengers to set out
4. number of minutes later than previously that the arrangement will require passengers to arrive home
5. whether the driver has a telephone at home
6. number of miles between the driver's workplace and that of the passenger
7. number of miles between the driver's home and that of the passenger
8. whether the driver is over 50 years of age.

From a driver's point of view:

9. standard arrangement
10. whether the passenger is female
11. number of minutes earlier than previously that the arrangement will require drivers to set out
12. number of minutes later than previously that the arrangement will require drivers to arrive home
13. whether the passenger has a telephone at home
14. number of miles between the passenger's workplace and that of the driver
15. number of miles between the passenger's home and that of the driver
16. whether the passenger is over 50 years of age
17. extra mileage incurred due to diversions
18. whether this is not the driver's first passenger.

From a pooler's point of view:

19. standard arrangement
20. whether the partner is female
21. number of minutes earlier that the arrangement will require participants to set out (when they are passengers)
22. number of minutes later that the arrangement will require participants to arrive hime (when they are passengers)
23. number of minutes earlier that the arrangement will require participants to set out (when they are drivers)
24. number of minutes later that the arrangement will require participants to arrive home (when they are drivers)
25. whether the partner has a telephone at home
26. number of miles between workplaces of pooler and partner
27. number of miles between the homes of pooler and partner
28. whether the partner is over 50 years of age
29. extra mileage incurred due to diversions
30. whether this is not the pooler's first partner.

(b) Characteristics of the individual (P_m in equation 2.2)

1. standard person
2. whether female
3. whether has a home telephone
4. whether under 30 years of age
5. whether over 50 years of age
6. whether a manual worker
7. whether a professional worker
8. distance to work
9. whether previously gave someone a lift to work
10. whether previously a non driver to work
11. whether more drivers than cars in this household
12. whether the arrangement is for only one journey per day.

We use this error term to recreate the stochastic element of individual decisions. This is done by random sampling from a normal distribution with mean zero and standard deviation equal to the standard error of the residual.

2.6.2 Within the simulation model each individual will consider the utility to himself of a car sharing arrangement with each of the persons on his match list (which was prepared for him from among his fellow applicants during stage two of the model).

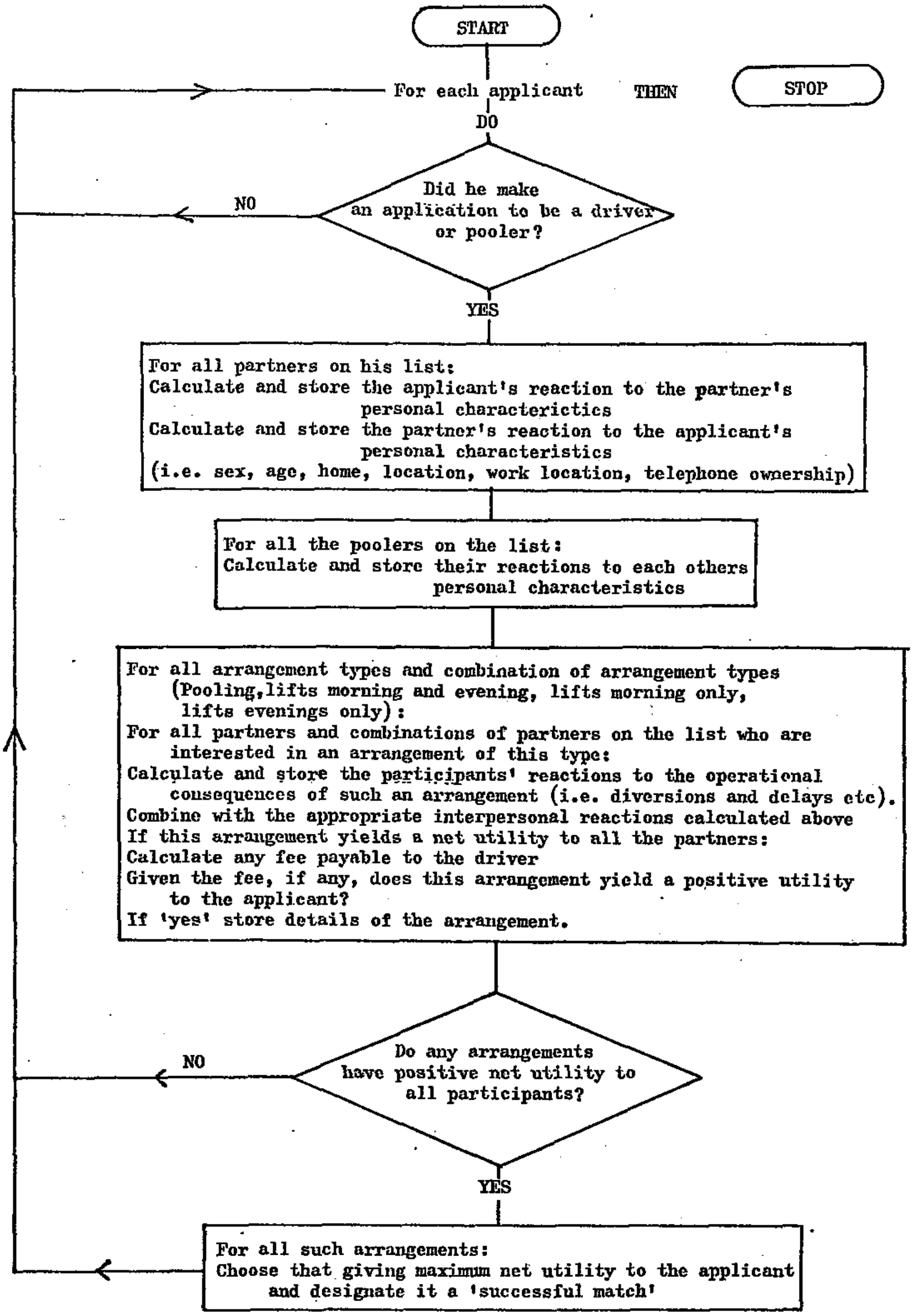
The model maximises utility for individuals rather than for the system as a whole (this is fundamental to microsimulation). A system optimum could exist only if all decisions could be made simultaneously and in perfect knowledge of all other decisions. Such a circumstance is as impossible computationally as it is in reality. An optimum might be approached by means of an expensive iterative routine but such a routine is not justified given that we seek to mirror reality. Since we do not seek this unrealistic system optimum, the end state reached in the model (as in reality) will be a function of the order in which bargains are struck. (Once person A has firmly contracted to travel with person B they are both of them effectively out of the market). In the absence of data on the complex question of the order in which bargains are likely to be struck we must assume that the order will be random. The sensitivity of model results to this assumption will be tested (see Bonsall and Kirby 1979).

2.6.3 It is not known precisely how an individual will make his decision on how to react to a list of possible car sharing partners nor is it at all likely that everyone would make their decision in the same way. For the purposes of this simulation, however, a single decision making algorithm had to be adopted. This algorithm was constrained to be computationally possible but was designed to be intuitively reasonable. Several algorithms have been programmed in the project to replicate the decision to match; the preferred version is presented here as Figure 2.4. The important points to note about this algorithm are:

a. We assume that applicants will consider all possible partners and combinations of partners within all possible types of arrangement or combinations of arrangement. This consideration may however be an almost immediate rejection on the grounds that (say) he is not interested in carrying that many passengers or he does not want to set off that early in the morning.

b. If an arrangement has a net utility to any of its participants of $\leqslant 0$ then that arrangement will not come to fruition (this is equivalent to the assumption that the utility of the status quo is zero).

FIGURE 2.4 THE DECISION TO MATCH ALGORITHM



c. We assume that optimisation will be from the point of view of the driver rather than the passenger (ie. the driver chooses the most attractive passenger(s) on his list rather than passengers choosing the most attractive driver on their lists). This is done for computational reasons but can be defended on the grounds that car-sharing is a sellers' (drivers') market.

d. The algorithm will calculate any fees payable by passengers to their drivers on the basis either of a fixed fee per mile or of an offer by the passenger on the basis of his utility (this may be subject to a maximum rate per mile in line with insurance company regulations).

2.7 The failure of matches to survive

The decision to match is based on the expected utilities of the arrangements in question. In pratice, however, these utilities may be revised after the arrangement has been in operation for a week or so. The revised utility may be smaller than the original utility; where it is so much smaller as to be negative, we may assume that the arrangement would fail to survive. The process by which these revised utilities are calculated would be similar to that for the decision to match but would include more accurate estimates of the operational consequences of the arrangement and more intricate evaluations of the personal characteristics of the partners.

Data was not available for the calibration of this model and it has consequently not been implemented in the car sharing microsimulation suite. Without it we are effectively simulating the establishment rather than the survival of car sharing arrangements. As an uncalibrated proxy for the survival calculation it is proposed to use a random number generator in conjunction with the utilities to all parties that were calculated at the time of the decision to match. This program would then accept 'thresholds' to determine the level at which arrangements are deemed to survive, to be modified or to be terminated. In such a model the 'death rate' of arrangements would be determined exogenously.

2.8 System performance indicators

2.8.1 It will be recalled from the introduction to this paper that the model was to help planners and policymakers considering the implementation of organised car sharing schemes. In order to do this it is necessary to indicate how a given scheme would perform and, in particular, what effect it would have on the transport system as a whole.

TABLE 2.4 IMPORTANT MODEL OUTPUTS

Type	Indicator
PROFILE OF APPLICANTS AND PARTICIPANTS IN EACH TYPE OF ARRANGEMENT	Location of homes and workplaces* Length of journey to work Previous mode of travel to work Sex, age and employment status Household background (including cars owned, number of drivers, number of members and telephone ownership) Perceived utility of arrangements Fees changing hands Diversions and delays accepted
OPERATIONAL PERFORMANCE OF THE SCHEME	Number of applicants for each type of arrangement Number of applicants given a match list Number of arrangements initiated Computational cost of matching program
SYSTEM EFFECTS	Work journey public transport patronage numbers of passengers lost passenger kilometres lost Private vehicle usage: kilometres saved kilometres driven within car sharing arrangements* net saving in kilometres driven change in car occupancies vehicles 'liberated' for possible off-peak usage

* also displayed on a map.

This information is provided via an analysis package into which model predictions are fed. The package provides for a range of performance indicators including graphical display. The main indicators are listed in Table 2.4. The analysis package produces values and 90% confidence intervals for each indicator.

The profiles of applicants and participants will be of use to policy makers wishing to consider the distributional effects of a car sharing policy. A description of operational performance of the scheme will obviously be of particular interest to the organisers and managers of schemes. The most important indicators, however, and the ones of greatest general interest, will be those which describe the effects that the scheme would have on the transport system as a whole.

2.8.2 From figure 2.1 it will be apparent that there is provision within the model package for the effects of the car sharing scheme to be fed back into the transport system description file. This allows the car sharing scheme to become iterative. The possible effect of a scheme on congestion levels and (via patronage levels) on public transport service provision and performance, can thus be allowed to influence modal choices and decisions to apply to the car sharing scheme in the next time period. This feed-back loop is presently designed only to include mode choice decisions (destination/origin decisions being regarded as longer term phenomena). It must be stressed that these feed-back effects are purely speculative and should not be seen as an integral part of the model design.

3. CALIBRATION OF THE DECISION MODELS

3.1 Introduction

The discussion in section 1.4 explained our decision to calibrate on stated intentions rather than on revealed preferences. The method by which we gathered the stated intentions data is quite unique and was developed in order the the respondents might give as 'accurate' replies as possible and that the data be in a form readily input to the microsimulation models.

The calibration process involved special surveys within our study area (Bonsall 1979a). These surveys effectively treated the respondents as actors within a 'field simulation' running parallel to the microsimulation model itself. The respondents were invited to react to precisely the same range of options and were given exactly the same information as were our actors in the microsimulation model.

A sample population were invited to react to the proposition that they should join an organised car sharing scheme and those respondents who reacted positively were then asked to express their reaction to a series of potential partners; this reaction to be in terms of the amount of money that they would require in compensation, or would be prepared to pay as a price, for participation in the arrangement as proposed.

3.2 Calibration of the decision to join a given car sharing scheme

3.2.1 The first element of the survey involved the distribution to 10,000 randomly selected households of publicity material similar to that which would accompany the establishment of an actual car sharing scheme.

This publicity material invited the public to indicate whether they would like to make use of a car sharing information system and, if so, what type(s) of car sharing arrangement would they be interested in. Questions were also asked of the individual in respect of his home and work locations and work hours - this information being required in the matching process. We took a positive reaction to this publicity material to be indicative of likelihood to make an application to an actual car sharing scheme.

Using the results of this survey we were able to create two data sets; the first containing descriptions* of the respondents deemed to have made applications and details of the type(s) of applications that they made (data set A). The second data set contained a synthesised sample of individuals from 10,000 households on the basis of the known characteristics of the original sample (data set B). (For method of synthesis see Bonsall and Champernowne 1979).

3.2.2 For each type of appliation(shown in Figure 2.2)a binary logit model of the form shown in equation 3.1 was evaluated to give values of a_i that would, from the synthesised sample population on data set B imply the same expected number of applicants and the same expected average values of characteristics of the applicants as were observed in the survey (ie. in data set A).

---

* questions had been asked as to their previous mode of transport to work, employment type, age and sex.

This process may be written as:

$$O_i = \text{population} \sum_{j=1} \frac{S_{ij}\, e^{\sum_{i=0}^{21} a_i s_{ij}}}{1 + e^{\sum_{i=0}^{22} a_i s_{ij}}} E_j \qquad 3.1$$

where S_{ij} is the value of the ith characteristic of the jth synthesised person (relevant characteristics were listed in table 2.2)

E_j is 1 if that person is eligible to apply (otherwise 0)

O_i is the total of characteristic x_i over all applicants

a_i is the calibrated coefficient.

The equation actually solved is

$$O_i = \text{population} \sum_{j=1} \frac{S_{ij}\, e^{\sum_{i=0}^{21} a_i s_{ij}}}{1 + e^{\sum_{i=0}^{21} a_i s_{ij}}} E_j + a_i = P_i(\text{say}) \quad ....(3.2)$$

This modification (the addition of the a_i term) is made to avoid possible singularity in the matrix of derivatives $\frac{\partial P_i}{\partial a_i}$. In addition to the 21 characteristics listed in table 2.2 each individual has a dummy characteristic always set to unity which acts as a 'balancing factor' in the calibration.

The parameters a_i are calibrated by solving the equations $O_i = P_i$ using Newton Raphson iteration to a maximum likelihood solution.

3.2.3 As was indicated in section 2.3.1, applicants applying for one of the active modes (including some driving) are required to indicate the maximum number of passengers they would want in their car. This wish is represented in the simulation model on the basis of two coefficients calibrated on the survey data in much the same manner as was done for the seven types of application. The main difference being that for these two coefficients P_i was summed over eligible members of the observed population rather than of the synthesised population. Values of the coefficients a_i were chosen to make the expected total and mean values of x_i equal to the observed just as is described above. The two coefficients are:

- given that an application to carry at least one passenger has been made, is an application made to carry at least two passengers?

and

- given that an application to carry at least two passengers has been made, is an application made to carry at least three passengers?

3.2.4 In a few of the replies to the questionnaire some personal details were not divulged and some values of the characteristics x_i were therefore not known. These missing values were filled in according to the following procedure:

1. missing values initially assumed to be zero
2. a covariance matrix for the data was calculated on this assumption.
3. on the basis of this covariance matrix missing values were replaced by the most likely value given the known data for that individual

- steps 2 and 3 repeated until convergence.

This procedure was also used to determine the maximum number of passengers wanted by a given applicant if he had not indicated it on his questionnaire form.

3.3 Calibration of the decision to match

3.3.1 The second element of the survey was distributed only to those respondents who had indicated a desire to participate in organised car sharing. Respondents were invited to allocate a utility or disutility (*"how much would you be prepared to pay"/"how much would you have to be paid"*) to a postulated car sharing arrangement. Variations on that arrangement were then proposed and the effect on the utility or disutility sought. The results of this questionnaire were a set of utilities U_{AP} of arrangements A with characteristics a to persons P with characteristics p.

These utilities were processed, by simple addition or subtraction of utilities reported by a given person, to produce a distribution of 'value coefficients' for each of the 30 arrangement characteristics. (That is a distribution of valuations for each of the 30 arrangement characteristics listed on table 2.3a).

3.3.2 These distributions were fed into a series of linear regression equations (one equation for each coefficient). The regression equations were of the form:

$$SV_n = \sum_m p_m x_{Am} + C_a$$

where sV_n is the stated value of the coefficient n

P_m are the independent variables (the 12 personal characteristics of a potential car sharer listed in table 2.3b)

x_{nm} are the unknown components whose values are sought.

C_n is an error term.

The results of the regression equations provide us with linear components of utility for each of the 30 value coefficients (x_{nm}) and the residual (e_n) which can be input into the utility equation given in section 2.6.1. The assumption of linearity was forced upon us by data availability constraints.

The residual term is assumed to comprise a stochastic element together with error terms. The stochastic element is assumed to be normally distributed. Nothing is known about the distribution of the error terms but we may assume that they reflect the distribution of the original data. If the distribution of the original data approaches normality we can therefore assume that the residual term as a whole is normally distributed.

3.3.3 When the distributions of each of the value coefficients were plotted it was apparent that many of them had a pronounced skew. Since multivariate linear regression assumes normality and since we wished to be able to assume that the residuals of the regressions were normally distributed we attempted to eliminate any skew from the data prior to the regression. We also wished to approach kurtoses equal to unity. We therefore transformed the data using, according to the characteristics of the original distributions, either square roots, logs, or negations of these. The skews and kurtoses of the data with and without these transformations, are given in table 3.1. Values underlined in table 3.1 indicate which transformation maximises normality and which was therefore used in our transformed data set.

Note that although we prefer the transformed data we have run the model on the untransformed data in order to demonstrate the differences between them (see Bonsall and Kirby 1979).

Table 3.1 also indicates the number of observations contributing to each value coefficient - some are very much smaller than we would have wished. The mean values of each of the coefficients are intuitively reasonable (They are discussed elsewhere - Bonsall 1979a).

3.3.4 While the majority of questionnaire replies were able to be fed straight into the regression equations, some replies were inadequate or otherwise abnormal.

Some replies suggested that the respondent associated an infinite disutility with the adverse aspects of the postulated car sharing arrangement. These replies were withdrawn from the regression and were simply summed to give a percentage of replies for each aspect of an arrangement having infinitely negative values. These percentages were then stored for use in the simulation model where we assume that a corresponding proportion of applicants could not tolerate these aspects being adverse.

Other replies implied an infinitely positive utility for a given combination of coefficient and arrangement. These replies were modified to give realistic (though still large) values.

TABLE 3.1 DISTRIBUTIONS OF VALUE COEFFICIENTS (TRANSFORMED & UNTRANSFORMED)

code no. of value coefficient (see table 2.3a)	number of observations	untransformed distribution			transformed distributions*			
					y = log x		$y = \sqrt{x}$	
		mean	skew	kurtosis	skew	kurtosis	skew	kurtosis
1	68	52.4	0.8	0.4	-0.5	0.4	0.2	-0.2
2	68	-7.3	-5.1	26.4	-4.1	31.1	2.9	16.5
3	67	-1.5	-1.6	3.1	-1.7	9.8	0.5	3.0
4	55	-2.3	-1.5	0.9	0.8	-0.7	1.1	-0.1
5	67	-25.1	-3.4	12.4	-0.1	4.3	2.1	5.1
6	67	-91.2	-1.6	2.0	-1.7	5.8	0.8	0.1
7	67	-48.9	-1.9	3.2	-2.3	12.3	1.0	1.3
8	42	-3.8	-4.1	19.1	-4.2	22.2	1.1	10.5
9	51	-98.6	1.6	3.9	-4.2	21.3	-0.1	2.4
10	51	4.8	4.1	16.0	-3.0	20.5	2.5	11.2
11	41	-5.6	-1.9	2.6	0.5	-1.2	1.1	0.1
12	30	-6.6	-2.9	7.6	0.9	-0.7	1.7	2.1
13	46	-18.0	-1.7	4.8	-5.3	30.0	-0.7	7.4
14	33	-78.2	-1.8	3.4	-1.0	-0.3	0.5	-0.2
15	15	-91.0	-2.4	3.0	-0.3	-1.3	1.4	0.9
16	33	-4.5	-5.1	24.7	-4.2	20.9	2.0	15.1
17	46	-86.7	-3.9	16.5	-1.1	0.5	1.6	4.1
18	46	41.5	0.1	0.9	-4.1	15.4	-1.9	5.7
19	63	215.7	1.6	3.9	-5.2	31.2	-0.3	4.0
20	62	-0.4	-1.5	17.9	-6.7	45.6	-3.1	22.6
21	34	-10.9	1.0	6.4	-4.3	18.5	-1.6	6.3
22	23	-4.0	-0.8	3.9	-3.0	6.8	-1.5	3.7
23	63	-7.5	0.9	6.2	-4.9	28.7	-1.5	6.9
24	36	-36.6	-2.7	7.2	0.1	-1.4	1.3	1.1
25	61	-139.6	-1.7	3.0	-2.9	11.4	0.5	0.8
26	37	-138.8	-2.6	6.7	-3.4	15.2	1.1	3.0
27	20	-164.0	-1.6	1.3	-3.0	8.9	0.4	0.6
28	37	-22.6	0.9	8.4	-4.1	15.0	-2.1	6.6
29	58	-204.0	-1.4	1.1	-1.1	0.1	0.4	-0.6
30	55	-150.0	-0.3	1.8	-6.5	41.6	-2.1	10.7

* Clearly the transformation of a given value coefficient had to be preceded by the negation of the distribution when the skew of the untransformed distribution was negative.

4. PERFORMANCE OF THE MODELS

4.1 Introduction

This section reports on the results of the calibration of the two choice models and briefly presents results of a complete run of the model. More detailed results and a wider range of tests are described in a companion volume (Bonsall and Kirby 1979).

4.2 Results of the calibration of the decision to apply

4.2.1 The calibration of the decision to apply to an organised car sharing scheme was described in section 3.2. The calibrated coefficients are given in table 4.1. Inspection of these coefficients suggests the relative importance of the various determinants. We note the following as particularly significant:

- propensity to apply is directly correlated with journey length except in the case of evening only arrangements
- the longer the journey the fewer the desired passengers
- the coefficients based on each mode will normally be combined with the coefficient relating to that same mode in the evening. When this combination is effected it appears that solo drivers are particularly reluctant to become passengers or to give evening only lifts. Also it seems that public transport users are more unlikely to become lift givers than poolers.
- persons from high car availability households are most likely to become poolers
- manual workers, professional workers, females and persons from households with numerous non drivers are less likely to make applications than are others
- persons from households with no telephones are particularly unlikely to make applications - especially applications to pool.

TABLE 4.1 CALIBRATED COEFFICIENTS OF DECISION TO APPLY

no.	characteristic (range of values) see table 2.2	Application types								
		pooling	lift giving morning & evening	lift giving morning only	lift giving evening only	lift receiving morn.& evening	lift receiving morning only	lift receiving evening only	allow more than one passenger	allow more than two passengers
0	dummy (1)	-3.53	-2.96	-4.06	-0.77	-2.62	-3.24	-1.89	0.82	0.51
1	length of journey to work (kms)	0.16	0.13	0.11	-0.09	0.09	0.00	-0.14	-0.04	-0.02
2	normal morning mode solo driver?(0-1)	0.48	0.60	0.09	-0.57	-0.20	0.81	-0.60	0.25	0.35
3	normal morning mode accompanied driver?(0-1)	0.09	0.30	-0.27	0.20	-	-	-	-0.30	0.27
4	normal morning mode passenger (0-1)	-0.86	-0.18	-0.56	-0.07	-0.48	-0.77	0.14	0.16	0.26
5	normal morning mode public transport(0-1)	0.64	0.02	-0.31	-0.06	-0.06	-0.23	-0.30	0.07	0.23
6	normal evening mode solo driver?(0-1)	-0.36	-0.00	0.08	-0.69	-0.40	-1.04	-0.61	0.57	0.25
7	normal evening mode accompanied driver?(0-1)	1.03	0.94	-0.00	0.39	-	-	-	0.70	0.38
8	normal evening mode passenger?(0-1)	0.34	0.32	-0.36	-0.13	0.11	0.09	0.07	-0.43	-0.11
9	normal evening mode public transport?(0-1)	-0.43	-0.40	-0.60	-0.08	0.51	0.60	0.74	0.00	0.32
10	under 30 years old?(0-1)	-0.44	-0.18	-0.50	-0.37	0.29	-0.56	-0.17	-0.15	0.64
11	over 50 years old?(0-1)	-0.64	-0.24	-0.72	0.24	-0.08	-0.71	-0.96	0.20	-0.50
12	household cars available (0-4)	0.21	-0.43	-0.16	-1.26	-0.69	0.14	-0.09	-0.41	-0.02
13	full car driving licence?(0-1)	-	-	-	-	-0.19	-0.05	-0.53	-	-
14	factory or manual worker?(0-1)	-1.67	-1.39	-1.77	-0.61	-1.22	-0.15	-0.76	0.14	-0.24
15	professional or managerial worker?(0-1)	-0.74	-0.35	-0.13	-0.57	-0.19	-0.12	-1.28	0.63	0.12
16	female?(0-1)	-0.36	-0.32	-0.19	-0.44	-0.19	0.67	-0.45	0.05	-0.06
17	number of licenced drivers in household?(0-8)	-0.02	-0.09	0.19	-1.31	0.17	-0.48	-0.33	0.22	0.21
18	number of non-drivers in the household?(0-8)	0.44	-0.46	-0.49	-1.31	-0.31	-0.48	-0.89	0.37	0.02
19	morning journey off-peak?(0-1)	-	-	-	-0.63	-	-	-0.51	0.44	-0.10
20	evening journey off-peak?(0-1)	-	-	0.29	-	-	-0.58	-	0.04	0.09
21	household telephone?(0-1)	1.35	0.63	0.65	-0.05	0.31	0.23	0.26	0.17	-0.68

4.2.2 In order to test the success of the calibration procedure, the coefficients discussed above were input to the model which simulates the decision to join a car sharing scheme. The individuals synthesised to represent our survey sample* were then processed through this model. The results are presented in table 4.2. Clearly the simulation model has reproduced the observed applicants with a fair degree of accuracy. The discrepancies worthy of note are:

a - an overprediction of lift offerers (+0.2% of the eligible population)

b - an underprediction of lift requesters (-0.03% of the eligible population)

c - underprediction of the proportion of offerers who were previously solo drivers (-5% of applicants)

d - underprediction of the proportion of offerers who come from phone owning households (-5% of applicants)

e - underprediction of the proportion of offerers who are 'professional' workers (-4% of applicants)

f - underprediction of the proportion of requesters who wish to ride morning and evening (-3% of applicants)

g - overprediction of the proportion of requesters who were previously public transport users (+5% of applicants)

h - overprediction of the proportion of requesters who were previously solo drivers (+3% of applicants)

i - underprediction of the proportion of requesters who previously rode as car passengers (-10% of applicants)

j - overprediction of the proportion of requesters who have no driving licence (+3% of applicants)

k - underprediction of the proportion of requesters of age less than 30 (-3% of applicants)

The only discrepancy which should cause us much concern is (i). Even this underprediction is, however, less serious than it might be because in terms of the impact of car sharing schemes on the transport system (VMT and public transport patronage in particular) private car passengers moving from one driver to another will have little net effect. The importance of this and other discrepancies will lie in their complex effects on the supply and demand equations deep within the matching simulation. They will have to be borne in mind when the results of model predictions are analysed.

We note that the standard deviations of the model predictions are generally very low and that the model is, overall, quite stable. Its stability is certainly within the margin of error which must be implicit in the model as a whole.

---

* ie. the data set 'B' mentioned in section 3.2.1.

Table 4.2 COMPARISON OF OBSERVED 'APPLICANTS' WITH SIMULATION MODEL PREDICTIONS FOR THE SAME STUDY AREA.

	Observed	Average* prediction	Standard deviation of prediction
			σ_n
Applicants for car pooling:			
total number	129	126.9	6.98
number as a % of theoretical total**	5.8	5.74	.32
length of journey to work (kms)	8.49	8.88	.50
% previously public transport users	6.2	7.00	2.61
% previously solo-driver	61.2	59.20	5.23
% previously accompanied drivers	27.9	25.33	3.00
% female	20.2	19.18	4.17
% having a home telephone	90.7	86.54	2.65
% professional workers	45.7	44.83	3.60
% under 30 years of age	28.7	28.70	3.93
Applicants to give lifts:			
total number	162	168.1	11.07
number as a % of theoretical total**	5.28	5.48	.36
% wanting morning and evening	69.1	68.70	4.21
% wanting morning only	30.9	33.57	3.89
mean length of journey to work (kms)	8.23	8.41	.20
% previously public transport users	1.9	2.02	.94
% previously solo drivers	70.4	65.23	2.69
% previously accompanied drivers	25.3	25.35	2.45
% female	22.2	23.20	2.75
% having a home telephone	84.6	79.33	2.98
% professional workers	54.9	50.49	4.01
% under 30 years of age	29.0	27.99	2.80
Applicants to receive lifts:			
total number	184	173.9	26.94
number as a % of theoretical total**	3.9	3.87	.85
% wanting morning and evening	77.2	73.82	2.79
% wanting morning only	20.1	21.83	3.53
mean length of journey to work (kms)	6.45	6.43	.45
% previously public transport users	60.3	65.07	3.64
% previously solo drivers	13.0	16.52	2.74
% previously car passengers	17.4	6.56	1.77
% female	53.8	53.03	1.46
% having a home telephone	72.3	69.72	2.70
% professional workers	26.6	27.33	2.60
% under 30 years of age	39.7	36.17	3.97
% having no household car	53.8	52.92	3.00
% having no driving licence	60.9	63.06	3.46

* Due to the stochastic element in the simulation model it was decided to run the model 10 times and to present here the mean value and its standard deviation.

** The theoretical total number of applications assumes one application from each eligible member of the population (ie. after taking account of licence tenure, car availability, work hours etc.). These theoretical totals are for pooling 2212, for lift giving 3067 for receiving lifts 4703.

4.2.3 The mechanism by which the microsimulation model uses the calibrated coefficients in table 4.1 can perhaps be appreciated by considering the case of one individual (number 109797) chosen from our synthesised population base. This individual has the following characteristics:

- locations of home: GR 847237
  work: GR 295299
  ($\therefore$ length of journey to work = 8.09 kilometers)
- normal mode of travel to work - solo driver
- normal mode of travel from work - solo driver
- age 30 to 50
- 1 car in household
- driving licence held
- professional worker
- male
- 2 licenced drivers in his household
- 2 non licenced members in his household
- work hours 0800 hrs to 1700 hrs.
- household telephone.

With these characteristics his likelihood of making each of the 7 types of application are achieved by utilizing in equation 2.1 the elements from the the following rows of table 4.1: 0, 1 x 8.09, 2, 6, 12, 13, 15, 17 x 2, 18 x 2, and 21.

Thus the probability (p) of applying to pool will be

$$p = \frac{e^x}{1+e^x}$$

$$\text{where } x = -3.53 + 0.16 \times 8.09 + 0.48 - 0.36 + 0.21 - 0.74 - 0.02 \times 2 - 0.44 \times 2 + 1.35$$

$$= -2.216$$

$$\therefore p = .098$$

for the other 6 types of application the p values will be .07, .037, .00005, .028, .005 and .0002 respectively.
Seven random numbers between 0 and 1 are then chosen for person number 109797, they are: .03, .84, .62, .85, .08, .46.
The seven p values are then divided by these seven random numbers to produce person 109797's likelihoods of applying, they are:

3.27, .08, .06, .00008, .35, .008 and .0004.

If we are deeming applications only where the likelihood is greater than 1 (which is the likelihood which was observed in the questionnaire survey) then person 109797 is deemed to apply for pooling but for nothing else.

4.3 Results of the calibration of match utilities

4.3.1 Information resulting from the regression calibration of match utilities is presented in tables 4.3 and 4.4. In these tables we present results from the regressions which were carried out after transformation of the value coefficients to maximise normality. The fact of this transformation makes between-row comparison of the regression coefficients rather difficult but it does not prevent other analysis. In table 4.3 results relate to value coefficients derived from the unedited* survey data. In this table we note that many of the regression equations have explained only a small part of the variance on the data (see R^2 values in column 4). We also note that the residual terms are frequently quite large in comparison to the regression coefficients.

These features result from our attempt to include at least 11 variables in each regression equation - more perhaps than the data might be thought able to support. (We did not however include any variables when the tolerance level fell below the SPSS default).

Another problem which is exacerbated by our small data set is that a small number of observations may dominate the values of certain independent variables and may lead to strong correlations between the values of those variables. Within a regression model this may result in the values of one independent variable being associated with another and although the model may give good results when both variables are taken together it may fail when they are applied separately. This, no doubt, is the reason for some of the counter-intuitive values.

Clearly we had to choose between having on the one hand, a small number of variables whose influences were all strong and intuitive but with a large amount of the variation left unexplained. (Low R^2s amd large residuals), or on the other hand a large number of variables in an attempt to extract the maximum information from the data (maximisation of R^2s and minimisation of residuals). After some considerable debate we chose the latter option.

---

* for significance of this fact see section 4.3.2.

TABLE 4.3 THE REGRESSION COEFFICIENTS - COMPONENTS OF MATCH UTILITIES DERIVED FROM THE TOTAL DATA SET ($U_{AP} = \sum_m \sum_n a_n p_m x_{nm} + e_n$)

value coefficient (a_n)*	number of observations	transformation used y =	R^2	Residual (e_n)	characteristics of individuals (p_m)**: constant term	female	no phone	<30	>50	manual	professional	DIJ (kms)	prev. as accomp. driver	prev. as non-driver	more licences than cars in h.h.	one way journey
standard driver	68	$\sqrt{x}$	.34	2.46	11.68	-1.56	0.04	-0.73	-0.81	-0.89	-0.97	0.3	n.a.	0.47	0.20	-0.72
female driver	68	$\sqrt{-x+51}$	.27	1.70	8.10	-0.57	-0.41	-0.07	0.41	-0.27	-0.25	0.1	n.a.	-0.81	-0.75	-0.25
1 min. early morning	67	$\sqrt{-x+7}$	.13	0.54	2.53	0.01	0.13	0.10	0.13	-0.26	-0.08	0.0	n.a.	0.14	0.20	0.09
1 min. late evening	55	$\log(-x+2)$	.33	0.63	0.21	-0.04	-0.05	-0.07	0.39	-0.56	-0.09	0.0	n.a.	0.91	0.09	0.03
no phone	67	$\log(-x+21)$	.26	0.77	3.06	-0.05	-0.05	-0.29	0.29	-0.45	0.17	0.0	n.a.	0.49	-0.23	-0.41
work separation (miles)	55	$\sqrt{-x+51}$	.36	4.40	11.38	-1.97	-3.00	-0.25	1.00	-4.40	-3.33	0.2	n.a.	1.51	2.11	-0.93
home separation (miles)	67	$\sqrt{-x+68}$	.25	3.39	8.87	-1.05	-.134	-0.06	1.73	-2.33	-1.81	0.1	n.a.	1.90	0.66	-0.28
over 50	42	$\sqrt{-x+46}$	.21	1.68	7.46	-0.08	-0.29	0.68	-0.59	-0.41	-1.24	0.0	n.a.	-0.69	0.70	-0.18
standard passenger	51	$\sqrt{x-401}$	.21	4.82	19.33	1.76	-1.18	-1.01	0.03	3.74	0.67	-0.2	-1.77	2.47	-1.02	-0.57
female passenger	51	$\sqrt{x-26}$	.15	1.47	5.15	0.60	0.08	-0.13	-0.65	0.23	0.66	-0.0	-0.17	0.09	0.11	0.07
1 min. early morning	41	$\log(-x+1)$	.54	0.95	0.52	1.46	0.36	0.66	-0.26	0.72	0.53	-0.0	-0.78	1.28	0.22	0.54
1 min. late evening	30	$\log(-x+1)$	.62	1.04	1.15	1.49	-0.09	-0.52	0.72	0.84	0.58	-0.0	-1.09	2.43	-0.55	0.05
no phone	46	$\sqrt{-x+101}$	.32	2.16	9.46	2.19	0.96	0.91	0.63	-1.80	0.99	0.1	-0.90	-1.02	0.21	-0.49
work separation (miles)	33	$\sqrt{-x+1}$	.62	3.68	10.63	3.63	-1.83	4.80	-3.65	-3.56	-4.25	-0.2	0.94	-0.77	0.73	2.14
home separation (miles)	15	$\log(-x+1)$	.90	1.32	4.64	0.00	0.36	0.88	-1.49	2.69	1.65	-0.1	-2.72	-1.14	-2.09	-0.23
over 50	33	$\sqrt{-x+51}$	.54	1.57	5.66	3.07	-0.92	2.13	-0.38	0.74	-0.29	0.1	-0.52	0.38	0.33	0.16
diversion (miles)	46	$\log(-x+1)$	.40	1.52	3.91	1.19	-0.54	1.33	-0.63	0.54	0.50	-0.1	0.00	1.18	-0.82	0.51
not first passenger	46	x	.32	70.67	27.62	-16.94	0.70	48.96	-32.05	-5.92	3.31	4.4	-23.04	-118.86	-2.62	-24.37
standard pooler	63	$\sqrt{x-301}$	.21	5.33	20.20	1.18	1.40	-2.36	-0.95	-0.17	-0.25	0.3	0.89	-0.08	-2.35	n.a.
female pooler	62	x	.21	17.91	-10.06	2.70	18.37	5.66	2.02	-1.00	8.71	-0.2	2.93	6.95	1.29	n.a.
1 min. early morning as passenger	34	x	.50	29.46	-41.41	18.88	4.63	4.18	-25.25	16.94	-11.75	0.4	39.07	4.84	28.95	n.a.
1 min. late evening as passenger	23	x	.53	28.50	14.89	-5.5	-0.23	-10.96	-11.30	-7.22	-25.92	-0.0	-11.48	-53.94	31.45	n.a.
1 min. early morning as driver	63	x	.11	29.46	2.72	-1.71	-11.67	1.79	-2.85	-5.22	5.64	-0.9	-10.57	-2.24	6.04	n.a.
1 min. late evening as driver	36	$\log(-x+1)$	.40	1.77	3.32	0.75	0.66	0.48	1.55	-1.82	0.27	-0.1	-0.77	-0.87	-1.22	n.a.
no phone	61	$\sqrt{-x+101}$	.18	5.58	17.84	0.91	0.24	-1.35	1.36	-2.18	1.00	-0.2	-3.54	-4.59	-0.18	n.a.
work separation (miles)	37	$\sqrt{-x+101}$	.40	4.89	12.82	4.02	-4.40	-3.66	-3.23	6.95	2.23	0.0	0.38	-2.49	1.01	n.a.
home separation (miles)	20	$\sqrt{-x+134}$	.40	7.47	21.07	3.91	3.75	-5.16	3.35	-4.96	-0.45	-0.3	0.20	-1.13	-3.89	n.a.
over 50	37	x	.32	145.45	59.17	-149.22	7.85	2.94	131.94	-83.77	72.17	-8.3	-38.05	-12.63	-27.69	n.a.
diversion (miles)	58	$\sqrt{-x+1}$	.17	7.97	14.18	1.43	-1.96	0.30	1.43	-2.90	-1.08	-0.3	0.01	4.81	1.86	n.a.
subsequent partners	55	x	.34	150.09	155.83	56.14	-61.91	122.43	-84.61	13.29	6.19	-8.6	76.06	84.33	38.41	n.a.

* for meaning of a_n see table 2.3a

** for meaning of p_m see table 2.3b

4.3.2 Table 4.4 is equivalent to Table 4.3 except that it contains results derived from a calibration on an edited version of our total survey data. The editing involved removal from the dataset of any records which violated the constraint that each of the following value coefficients be non positive: earlier morning departure, later evening return, no telephone at partner's house, separation of workplaces, separation of homes and diversion incurred. A further constraint was that the net value to the driver of an additional passenger should not be so positive as to cancel out the negative value of all existing passengers.

It is not surprizing that such modification of the original data produces a more successful set of regressions (compare for example the R^2 values in tables 4.4 with those in table 4.3). However, although one can manufacture some quite cogent arguments for excluding the records which violate the constraints set above (eg. that the respondent obviously did not understand the question). It is not thought that they are convincing enough to warrant substitution of the regression coefficients in table 4.3 by those in table 4.4 (The sensitivity of model results to the difference between these two data sets is reported elsewhere - Bonsall and Kirby 1979).

4.3.3 When we came to apply the results of the regressions in the simulation model we found that the linear combination of the coefficients sometimes resulted in positive valuations of the following quantities:

- the value of having to set out extra early in the mornings
- the value of having to arrive home extra late in the evenings
- the value of having to divert from one's shortest route to/from work
- the value of having one's partner living at a considerable distance from one's own house
- the value of one's partner working at a considerable distance from one's own workplace.

The occurrence of positive valuations for these quantities is serious because their application within the model will produce extra-ordinary predictions. /In some early runs of the model certain individuals seemed to delight in getting up each morning at the crack of dawn and driving 50 miles out of their way in order to give someone else a lift to work even though they neither lived nor worked within 20 miles of one another - Clearly such behaviour is counter intuitive/*

---

* It is one of the advantages of microsimulation that logical errors in the model are thrown into sharp relief in this way rather than being hidden within a complex set of formulae.

TABLE 4.4 THE REGRESSION COEFFICIENTS - COMPONENTS OF MATCH UTILITIES DERIVED FROM EDITED DATA SET ($U_{AP} = \sum_m \sum_n a_n p_m x_{nm} + e_n$)

value coefficient (a_n)*	number of observations	transformation used y =	R^2	Residual (e_n)	characteristics of individuals (p_m)** constant term	female	no phone	<30	>50	manual	professional	DIJ (kms)	prev.an accomp. driver	prev.a non-driver	more licences than cars in h.h.	one way journey
standard driver	55	$\sqrt{x}$	.38	2.50	11.30	-2.12	0.22	-0.75	-0.74	-1.21	-0.59	0.3	n.a.	1.23	0.07	-1.02
female driver	55	$\sqrt{-x+61}$	.31	1.85	8.17	-0.97	-0.74	-0.08	0.35	-0.51	-0.35	0.2	n.a.	-0.47	-0.90	-0.13
1 min. early morning	55	$\sqrt{-x+7}$	.11	0.55	2.75	0.07	0.15	0.12	0.13	-0.26	-0.18	0.0	n.a.	-0.03	0.18	0.01
1 min. late evening	46	$\log(-x+1)$	.39	0.83	-0.22	-0.20	0.04	-0.23	0.48	-1.17	-0.41	0.0	n.a.	1.12	0.14	-0.17
no phone	54	$\log(-x+1)$	.16	2.04	0.13	-0.14	-0.36	-0.39	0.75	-0.66	0.57	0.1	n.a.	1.55	-0.73	0.16
work separation (miles)	55	$\log(-x+1)$	.31	2.26	2.56	-1.06	-1.25	0.13	0.43	-2.40	-1.60	0.1	n.a.	0.91	1.03	0.03
home separation (miles)	55	$\log(-x+1)$	.32	2.09	0.94	-1.27	-1.07	-0.42	0.66	-1.48	-0.85	0.1	n.a.	1.81	0.61	1.07
over 50	35	$\sqrt{-x+46}$	.25	1.81	7.31	0.12	-0.01	0.89	-0.55	-0.30	-1.21	0.0	n.a.	-0.86	0.72	-0.30
standard passenger	42	$\sqrt{x-401}$	.25	4.79	17.52	2.97	-1.01	-1.79	-0.38	2.79	0.34	-0.0	-1.97	8.29	0.02	-1.42
female passenger	42	x	.29	7.26	1.60	0.78	1.03	0.38	-4.82	1.27	2.45	-0.2	-3.64	7.62	2.02	-0.48
1 min. early morning	33	$\log(-x+1)$	.70	0.80	0.05	1.57	0.37	0.71	-0.29	0.00	0.57	-0.0	-0.59	1.91	0.61	0.71
1 min. late evening	25	$\log(-x+1)$	.59	1.17	0.97	1.47	0.00	-0.49	0.70	0.52	0.46	-0.0	-1.02	2.98	-0.24	0.01
no phone	38	$\log(-x+1)$	.40	1.60	-0.32	1.21	-0.03	1.38	1.22	0.37	1.51	0.0	-1.57	-0.58	0.25	-0.65
work separation (miles)	28	$\sqrt{-x+1}$	.65	3.75	8.30	5.36	-3.90	4.26	-4.48	-2.67	-3.10	-0.1	1.02	3.42	2.51	1.08
home separation (miles)	12	x	.99	3.11	-92.68	41.81	-13.86	0.00	24.75	-8.62	-16.61	3.0	28.76	62.06	36.87	0.00
over 50	28	x	.76	23.87	0.31	-65.59	-10.02	-44.10	-10.63	-7.21	18.26	-0.7	19.96	1.02	18.27	-29.67
diversion (miles)	39	$\log(-x+1)$	.51	1.43	3.86	0.93	-0.94	1.59	-0.92	-0.04	0.60	-0.1	0.21	0.40	-0.26	0.73
not first passenger	40	x	.45	63.84	51.59	-80.26	19.37	63.44	-9.60	-45.44	-12.41	1.5	4.35	-155.19	-23.71	1.78
standard pooler	47	$\sqrt{x-71}$	.23	5.70	14.89	-0.72	0.42	0.35	1.43	-2.77	12.58	0.4	0.17	0.23	-1.70	n.a.
female pooler	46	x	.35	19.55	-11.43	5.11	37.36	9.29	2.59	-13.63	9.08	-0.4	4.98	6.11	3.10	n.a.
1 min. early morning as passenger	23	$\log(-x+1)$	.82	0.82	4.19	-0.22	2.93	0.12	0.38	-4.47	-1.83	-0.1	-0.25	-2.17	0.26	n.a.
1 min. late evening as passenger	18	$\log(-x+1)$	.73	1.09	-0.07	-0.17	-0.88	1.35	1.31	0.14	0.31	0.0	0.22	3.74	-1.10	n.a.
1 min. early morning as driver	47	$\log(-x+1)$	.28	1.52	0.49	1.22	1.28	0.93	0.53	-0.75	-0.07	0.0	0.90	0.05	-0.43	n.a.
1 min. late evening as driver	25	$\log(-x+1)$	.65	1.37	3.35	0.67	-2.17	2.09	1.93	-0.93	-1.05	-0.1	-0.95	-1.28	-0.12	n.a.
no phone	45	$\sqrt{-x+1}$	.17	6.58	11.21	5.42	2.06	-1.74	-2.53	-1.42	-2.00	-0.1	1.52	-2.51	-0.56	n.a.
work separation (miles)	26	$\sqrt{-x+1}$	.33	5.15	10.80	-1.35	-3.06	-0.36	-2.56	-3.03	-2.26	-0.1	1.82	-0.85	3.92	n.a.
home separation (miles)	16	$\sqrt{-x+1}$	.65	7.37	14.89	8.39	9.13	-10.39	-6.71	-4.65	-2.42	-0.2	2.59	-0.06	0.24	n.a.
over 50	28	$\log(-x+26)$	.43	0.89	3.39	0.05	0.37	0.70	0.48	;1.51	-0.29	0.0	-0.54	-0.28	0.18	n.a.
diversion (miles)	42	$\sqrt{-x+1}$	.14	7.89	13.96	0.05	3.96	0.04	0.82	-6.15	-1.34	-0.3	0.26	0.43	2.80	n.a.
subsequent partners	40	x	.51	130.78	-95.63	107.18	-89.34	131.38	-45.01	6.97	52.86	-14.2	28.93	7.75	37.12	n.a.

* for meaning of a_n see table 2.3a

** for meaning of p_m see table 2.3b

- 37 -

The problem of these counter-intuitive valuations required some attention. Investigation showed that, in the majority of cases, the valuation was made counter-intuitive by the addition of the stochastic element. In a minority of cases, however, the valuation was counter-intuitive even before additon of the stochastic element.

Occurrence of counter intuitive valuations is due in part to our failure to completely normalise the data prior to the regression and in part to the appearance in the simulation population of persons with a combination of characteristics each of which mitigated (in the observed data) against the intuitive value of the phenomenon in question. For example, if in our observed data women tended to dislike having to make diversions less than did men and if young people tended to dislike having to make diversions less than did older people then young women might actually be predicted to like making diversions. This would be most likely to occur if few young women occurred in the observed data. A solution to this problem would have been to invent a new independent variable in the regression to represent combinations of other variables all of which tended to have the same effect. An independent variable 'young woman' would, for example, have avoided the problem hypothesised above.

The problem with this approach howeer is that it is not feasible to include new variables for all the combinations of characteristics which result in a component of the same sign in the regression equation. If one were to have new variables for only the more important combinations then the anomaly would arise of one combination (for which no combinational variable was introduced) resulting in a valuation more counter-intuitive than that from a combination which might be expected to be more counter intuitive but for which a combinational variable was introduced. Moreover even if this approach were adopted it would not prevent the occurrence of counter-intuitive values due to the addition of the stochastic element. For these reasons we do not favour the introduction of these dummy-combination variables.

An alternative solution, which has the dual merits of simplicity and ability to deal with counter-intuitive values however they arrive, is to introduce a constraint that any counter-intuitive valuation be set to zero. Although it is regrettable to have to resort to such a constraint, it does in fact produce a distribution of valuations closer to that of the original observations than is achieved by the unmodified regression coefficients. (The original data was dominated by the modal value zero and this feature is recreated by using the constraint on counter-intuitive values described above).

The balance of argument persuaded us to adopt this constraint mechanism.

4.3.4 The mechanism by which the microsimulation model uses the regression coefficients in table 4.3 may be appreciated if we consider the case of person number 109797 whose application to join a car pooling scheme was considered in section 4.2.3 of this paper. 109797's personal characteristics mean that in considering the utility of a pooling arrangement we should make a linear combination of the values in the following columns of table 4.3: 5,12,13 x 8.09 and 16. These combinations are then to be supplemented by the stochastic element obtained by multiplying each residual (column 5) by a standard unit normal random number. This process is completed by retransformation of the resulting values (column 3). The values for rows 19-30 respectively then become:

	a_n	x_n (pence per week)
1	value of the standard arrangement	477
2	value of the partner being female	24
3	value of having to set off 1 minute early when a passenger	-16.81
4	value of having to arrive home 1 minute late when a passenger	10.43 → 0
5	value of having to set off 1 minute early when a driver	-25.86
6	value of having to arrive home 1 minute late when a driver	-13.1
7	value of partner not having a telephone	-138
8	value of having a distance of 1 mile between workplaces	-59
9	value of having a distance of 1 mile between homes	-219
10	value of having a partner over 50 years old	152
11	value of having to drive 1 mile out of one's way each day	-41
12	value of partner being not the first	-284

Note that we have applied to line 4 our constraint (see section 4.3.3 above) that the value of arriving home late be not positive. These components are then added to produce the utility to person 109797 of a given pooling arrangement. Thus for example:

If he is to pool with a female who causes him, when the passenger to set out 1 minute early and to arrive home 2 minutes late and, when he is the driver, to set out 5 minutes early and to arrive home 5 minutes late, who has a household telephone, who works at the same place as he does but lives 1 mile away from him, who is less than 50 years of age and who will cause him to drive 1 mile out of his way each day in order to pick her up.

The utility would be 477 + 24 - (16.81 x 1) + (0 x 2) - (25.86 x 5) - (13.1 x 5) - (219 x 1) - (41 x 1) = 29.39 pence per week

If the lady in question also puts a positive value on the arrangement and if this arrangement appears to person number 109797 to be the best on his list then the arrangement is deemed made. Note that if the lady had had no telephone at home then the utility to 109797 of the arrangement would have been reduced by 138 pence per week and since the net value of the arrangment would have been negative we assume that it would not come into operation.

If we had been considering the utility to 109797 of an arrangement to give lifts (as opposed to alternating driving) then any deficit in his utility might have been made up from a surplus utility accruing to his potential passenger. This transfer of utility might be by means of cash (a fare paid) or through some other medium (eg. periodic gifts). The model will calculate the magnitude of any such transfers of utility but does not have to consider how they would be effected.

4.4 Results of the microsimulation model

4.4.1 The framework for testing the model. Tests of the model are required to show whether the model predictions are intuitively reasonable, whether they are plausibly sensitive to the input parameters and what are the computational requirements of the suite. An attractive framework for these tests is the pivotal method of sensitivity analysis developed at Leeds in an earlier project (Bonsall et al 1977).

Within this framework 'most likely' values are chosen for each of the model parameters (thus reflecting a realistic policy environment) and best estimate values are derived for the coefficients (thus resulting in the optimal model). These values of parameters and coefficients are then used to produce a 'most likely future' (MLF) run of the model. The predicted values of important model outputs are noted. The values of the various parameters and coefficients are then systematically varied from their MLF values. The resultant changes in model outputs are then plotted to show how percentage changes from MLF values of a given parameter or coefficient, affect percentage changes from MLF values in the various model outputs. Examination of the curves thus produced can reveal a great deal about the model and provides a valuable test of its realism.

This method of model testing is clearly as appropriate to the present microsimulation model as it was for the macroscale transport demand model for which it was developed. The method requires that default parameter and coefficient values be derived for the MLF run which will serve as the pivot for all subsequent testing. Table 4.5 lists the 11 parameters and coefficients of the model and shows the default values assigned to them. A full test of the model would involve systematic variation of all 11 parameters and coefficients. It should be noted that coefficients 1, 5, 9 and 10 are groups of complex second order coefficients themselves derived from a number of primary coefficients. Table 2.4 in section 2.3 listed the important outputs of the model whose values will be closely monitored during the model testing.

TABLE 4.5 PARAMETERS AND COEFFICIENTS OF THE SIMULATION MODEL (AND DEFAULT VALUES THEREOF)

parameter/ coefficient Number	Name	Function	Refer to Section	default values
1	POPULATION	the population base who act on this the simulation model		'best' synthesised population as of March 1979
2	HOMESIN	defines which residence zones are valid for applicants	2.3	1 thru' 455 (entire study area)
3	WORKSIN	defines which employment zones are valid for applicants	2.3	1-13 (central Leeds)
4	THRESHOLD	threshold of interest	2.3	8 (level of publicity = that of survey)
5	APPLYCOEFS	coefficients of decision to make an application	2.4 and 3.2	'best' calibrated values as of March 1979
6	BATSIZE	number of applicants to be processed on current batch	2.5	1688 (all applicants)
7	NOONFORM	maximum number of potential partners to be included on each match list	2.5	10
8	TIMEWINDOW	extent, in time, of search for partners	2.5	± 15 minutes
9	SEARCH	extent and path of spatial search	2.5	spiral eliptical search routine as at March 1979
10	MATCHCOEFS	coefficients of utility of decision to match	2.6 and 3.3	'best calibrated values as of March 1979
11	MAXFEE	car running cost per 1/10 kilometer upon which driver can base the maximum fee that he may charge his passengers.	2.6	unlimited

4.4.2 In any model with a stochastic element the model results will vary depending on the random number string used. In order to test the stability of the model results the programs were run 12 times with different random number strings each time. The effect of each additional run on the mean values of important model outputs were then calculated in an attempt to demonstrate their relative stability and to define an acceptable minimum number of runs which shall be made each time the model is to test a new policy. Results of this investigation, which was carried out using the parameter values in table 4.5 are given in figure 4.1 which shows how the mean value after n runs approaches the mean after 12 runs. From this figure it is apparent that while the majority of the indicators reach some stability between 4 & 6 runs of the model, others (particularly the net saving in VKT) are still varying quite widely. Clearly some of the indicators cannot be regarded as stable even after 12 runs of the model. While we appreciate that the trends shown in figure 4.1 are to some degree of the function of the order in which the various runs were carried out (it was in fact a random order) we feel that there is a case for adopting 5 as the minimum acceptable number of runs. (5 observations being, by common rule-of-thumb, the minimum required for certain classes of statistical test).

The consequence which a decision to run the model five times rather than twelve times has for the confidence interval on the mean may be appreciated from table 4.6. In this table we show the mean and 90% confidence interval for model predictions using all 12 runs and for 3 different sub sets of 5 runs taken from the 12. The confidence interval was derived using the sample mean, the sample variance and the t distribution (t distribution rather than normal since we are estimating the population variance).

From the table we note that the means and confidence limits (and thus the upper and lower bounds) for each indicator vary somewhat depending which subset of 5 runs are used. We also note that the mean of the total sample is not always compassed in the 90% confidence range of a subset. Nevertheless the increase in accuracy consequent upon the larger sample size is not significant given all the other inaccuracies that must exist in any behavioural model. In the trade off between stability of prediction and computational cost of the runs it is not thought necessary to exceed 5 runs of the model.

FIGURE 4.1 STABILITY OF MODEL RESULTS (% DIFFERENCE BETWEEN MEAN AFTER n RUNS AND MEAN AFTER 12 RUNS)

$$y = \frac{\left(\dfrac{\sum_{i=1}^{n} x_i}{n} - \dfrac{\sum_{i=1}^{12} x_i}{12} \right) \times 100}{\dfrac{\sum_{i=1}^{12} x_i}{12}}$$

A Number previously passengers
B Mean number of persons per arrangement
C Mean participant's journey to work
D Number of arrangements
E Number of drivers (morning and evening)
F Reduction in passenger kilometres of public transport patronage
G Number of participants
H Number previously public transport users
I Fees paid to participating drivers
J Mean utility per participant
K Number previously solo drivers
L Mean diversion per arrangement
M Net reduction in private vehicle kilometres travelled

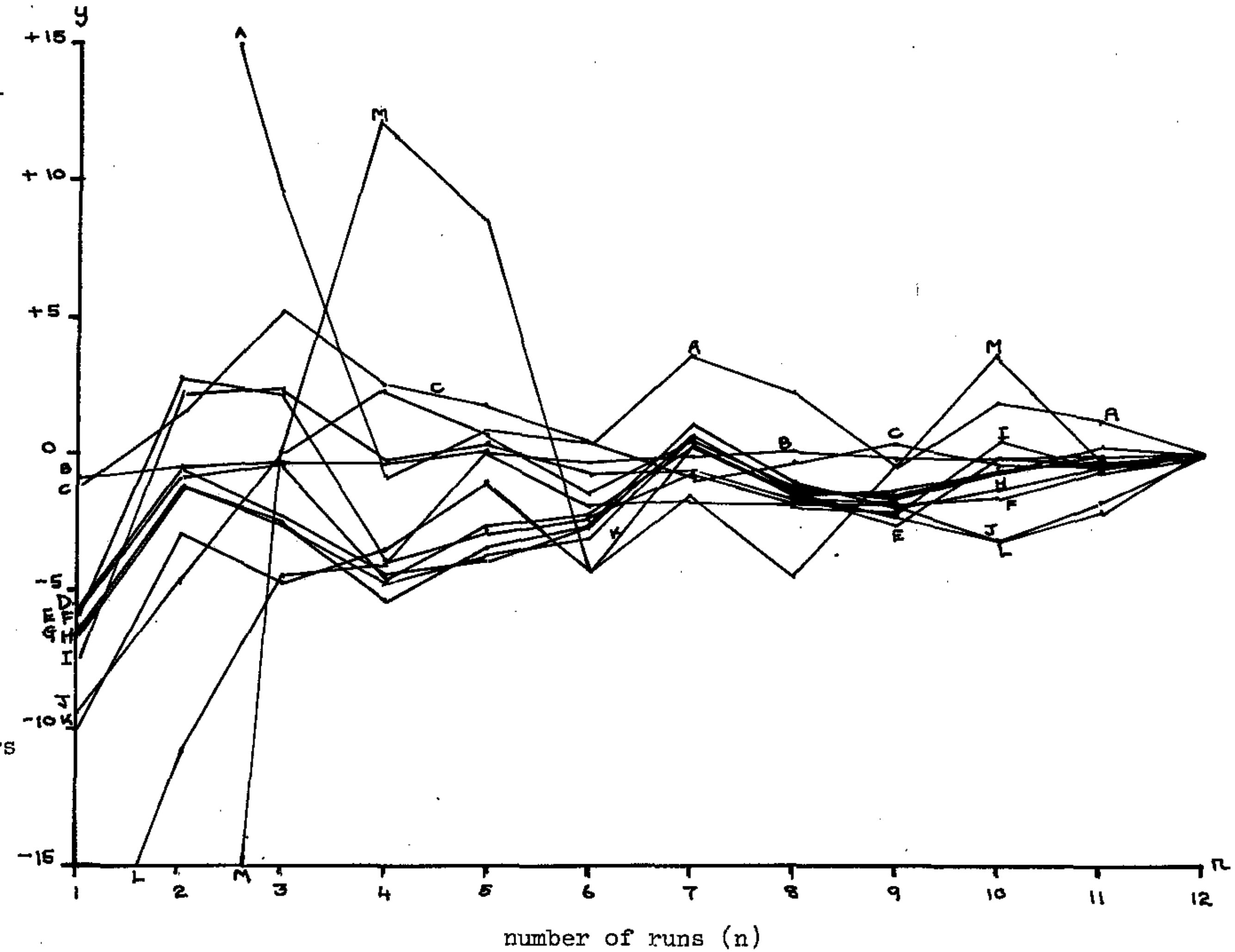


TABLE 4.6 CONFIDENCE INTERVALS FOR MODEL PREDICTIONS BASED ON DIFFERENT NUMBERS OF RUNS

Indicator	For all 12 runs		For first* subset of 5 runs		For second* subset of 5 runs		For third* subset of 5 runs	
	Mean	90% confidence interval (±)	Mean	90% confidence interval (±)	Mean	90% confidence interval (±)	Mean	90% confidence interval (±)
Total number of participants	326.8	20.79	314.6	25	323.6	37	338.0	40
Number of arrangements	133.4	8	128.2	9	132.0	16	137.4	18
Mean person/arrangement	2.45	0.03	2.45	0.04	2.45	.08	2.46	0.05
Number of poolers	49	10	42.8	8	50.0	23	51.8	21
Number of drivers and riders (morning and evening)	229	17	223.0	26	228.0	41	235.4	29
Number of drivers and riders (morning only)	47	6	48.0	15	43.0	11	50.8	13
Number of drivers and riders (evening only)	0.8	1.01	0.8	2.22	0.8	2.22	0.0	0.0
Number from public transport	134.8	9	130.0	14	132.8	17	139.2	22
Number of previously accompanied drivers	45.2	6	40.4	10	45.0	10	44.2	10
Number of previously solo drivers	120.0	10	118.8	12	119.2	12	127.8	20
Number of previous passengers	19.4	3	19.6	7	18.4	4	19.4	5
Number of other mode users	7.5	1.42	5.8	2.04	8.2	2.39	7.4	4
Number of vehicles liberated in households with more drivers than cars	17.33	2.63	16.0	2.64	18.8	2.84	19.0	5.05
Public transport patronage lost per week (passenger kms per week)	10707	630	10411	1649	10824	921	11159	799
Net reduction in vehicle kilometres travelled (VKT per week)	1422	412	1545	995	1450	615	1702	746
Total fees changing hands (1977 £ per week)	373	25	374	39	376	70	382	31
Mean utility per participant (1977 £ per week)	1.63	.13	1.65	0.17	1.62	0.32	1.68	0.17
Mean one way journey length of participants (kms)	9.26	.37	9.42	0.78	9.37	0.62	9.49	0.93
Mean diversion per driver (kms per week)	11.82	.94	11.82	2.07	12.10	2.01	12.28	1.20

* NB. subsets one and two are mutually exclusive

4.4.3 Comments on the model predictions themselves and their policy implications are discussed in a separate paper (Bonsall and Kirby 1979) which deals exclusively with the results of the various model runs.

4.5 Computational performance of the models

There are five main programs in the simulation suite; their computational requirements are set out in Table 4.7

Although these programs were not written with efficiency as the primary goal they are clearly not excessively expensive. Furthermore the most expensive programs (the calibration and simulation of the decision to match) need only to be run once. The large core requirement of some of the programs reflects the fact that the 1906A at Leeds has virtual storage and so no effort has been made to reduce core requirement by overlaying or the use of scratch files. If the programs were to be mounted on another machine it would be possible to reduce core requirement. Similarly, it would be possible to increase program efficiency if it were thought necessary.

At present the notional cost of a run of the simulation suite would be approximately £4. each time that a new scheme location or intensity is to be tested plus 1 penny per applicant, each time that new match lists are to be created and 6 pence per applicant each time that decisions to match are to be made. These costs are clearly very reasonable.

5. DISCUSSION AND CONCLUSIONS

5.1 Criticism of the model as formulated - possible enhancements

There are a number of respects in which the model and its calibration can be criticised. Firstly it must be admitted that the amount of data obtained from the calibration surveys is less than we would have wished and that it is perhaps overstretched in our calibration procedures. (For example 1800 responses have been used to determine the values of 340 coefficients in the match utility model). This deficiency, however, is put in its correct perspective when we consider that the survey cost less than £2400 to mount and that, in retrospect, the volume of data could have been increased substantially at very little extra cost.

TABLE 4.7 COMPUTATIONAL REQUIREMENTS OF PROGRAMS IN THE SIMULATION SUITE (On Leeds University ICL 1906A)

Program	Description	Maximum core requirement (words)	CPU Time (mins/secs)	Total Cost* (£)	Number of actors processed	Cost* per actor (pence)
CALAPALL	calibration of propensity to apply (9 logit models)	120K	42.15	253	-	-
CALAPALLSUP	supplementary programs to prepare the questionnaire data for input to CALAPALL	36K	0.38	4	-	-
PWBAPPLY	simulation of decisions to apply	146K	21.10	128	177836	.72
ADVERT	representation of scheme intensity and location	40K	0.42	4	177836	.002
SPSS	calibration of match utilities (30 regression equations)	109K	3.35	21	-	-
SPSSUP	supplementary programs to prepare questionnaire data for input to the SPSS package	186K	2.41	16	-	-
MATCH 1	creation of match lists	201K	2.31	15	1425	1.05
MATCH 2	simulation of decisions to match (to be run 5 times)	201K	2.04	12	1100	1.12

(* at notional commercial rates as set by University of Leeds 1979)

While the utility based model of the reaction of individuals to persons on their match lists can be said to have a behavioural basis it must be conceded that the binary logit models, which determine whether or not an application will have been made in the first place, are correlative rather than causal. Our only defence is that to have developed behavioural bases for both decision models within the microsimulation suite was beyond the resources available to us. Similarly, had we had more data we might have been able to improve upon the assumption of linearity in the utility equations.

In our attempt to model organised car sharing schemes we designed a model to simulate such schemes as they were then (1976) envisaged. Since that time it has become apparent (Wagner 1978, Bonsall 1979b) that for every 100 persons who start car sharing as a direct result of a car sharing matching system there may be another 100 who start because of the publicity and as a result of incentives other than the matching service itself. The microsimulation model presented in this paper was not designed to reflect this fact and nor in its present form can it do so.

The current model predicts the establishment of car sharing arrangements rather than their survival. Further attitudinal research would be required if a calibrated model of the survival of arrangements were deemed necessary.

The decision algorithm embedded in the match reaction model reflects a particular decision model; tests ought to be carried out of other decision models - is it, for example, the potential driver or the potential passenger who takes the active role in pool formation? In what order would a potential car sharer evaluate the matches on his match list? How would the payment of fees be negotiated? ....and so on.

The model as presented was severely restricted by the then availability of data. In retrospect we would wish to include a greater number of non-transport variables in descriptions of our actors - smoking habits and preferences would have been particularly useful. Other characteristics such as educational background, race, political stance and so on are of obvious importance but could not be included because of the difficulty which we would have had in obtaining honest reactions to them in the field surveys. In-depth interviewing revealed their importance but not in a way that is compatible with our model framework.

5.2 Conclusions

Even with the deficiencies noted above we believe that the microsimulation model presented in this paper is the best model yet developed for the prediction of the performance of organised car sharing schemes and that it also represents a contribution to the development of improved travel demand models.

The model predictions briefly presented in table 4.6 but discussed elsewhere (Bonsall and Kirby 1979) suggest that the model accords well with empirical evidence of the performance of organised car sharing schemes.

The unconventional calibration base ('field simulation') seems to have proved a very useful device.

The fact that the model deals with individual decision makers, rather than populations, has allowed the predictions to be closely scrutinised and verified in a manner quite impossible under conventional model frameworks.

In short, a microsimulation model calibrated on stated intention data has proved an attractive device that can be at once behaviourally based and yet computationally tractable.

ACKNOWLEDGEMENTS

This work was executed under contract to the Special Research Branch of the Transport and Road Research Laboratory. The author is pleased to have this opportunity to record the great contribution made by Arthur Champernowne to the construction of the models described in this paper. His particular contribution being in the design and implementation of the initial versions of the calibration procedures.

Thanks are also due to Drs. Webster and Bly of the TRRL for their support and advice. Similarly I wish to record my debt to Roger Kirby, Hugh Gunn, Juan Ortuzar and Christine Gill of the Institute for Transport Studies for their assistance and advice.

REFERENCES

ALBRIGHT, R.L., LERMAN, S.R., MANSKI, C.F. (1977) 'Report on the development of an estimation program for the multinomial probit model', prepared for FHWA, Oct. 1977

ATHERTON, T.J. and BEN-AKIVA, M.E. (1977) 'Transferability and updating of disaggregate travel demand models'. Transp. Research Record 610.

ATHERTON, T.J., SURHBIER, J.H. and JESSIMAN, W.A. (1976) 'The use of disaggregate travel demand models to analyse car pooling policy incentives'. Transp. Research Record 599.

BONSALL, P.W. (1979a) 'Attitudes to car sharing: The calibration surveys for a microsimulation model'. Institute for Transport Studies, WP 108, Leeds University.

BONSALL, P.W. (1979b) 'Car pooling in the USA: a British perspective' To be published as a TRRL Supplementary Report number 516.

BONSALL, P.W. and KIRBY, R.N. (1979) 'Microsimulation of organised car sharing - model predictions and policy implications.' Institute for Transport Studies, WP 114, Leeds University.

BONSALL, P.W. and CHAMPERNOWNE, A.F. (1979) 'The simulation of organised car sharing: I - the synthesis of a population base'. Institute for Transport Studies, WP.107, Univ. of Leeds.

BONSALL, P.W., CHAMPERNOWNE, A.F., MASON, A.C. and WILSON, A.G. (1977) 'Policy testing and sensitivity analysis'. Progress in Planning Vol.7. part 3.

CAMBRIDGE SYSTEMATICS INC (1976) 'Guidelines to travel demand analyses of program measures to promote car pools, van pools and public transportation'. Prepared for Federal Energy Administration. USA.

DOBSON, R. and TISCHER, M.L. (1976 draft) 'Beliefs about buses, car pools and single occupant autos: a market segmentation approach' Prepared for .976 Transportation Research Forum. USA.

HAGERSTRAND, T. (1970) 'What about people in regional science?' Papers of the Regional Science Association, vol. 24, 1970.

HAWKER SIDDELEY DYNAMICS (ADV.TRANSPOR. SYST.) (1977) 'Secondary public transp. in Herts.' - Appendix 2. Report of carsharing survey Produced for Transport Co-ordination Unit, Herts. C.C.

HEGGIE, I.G. (1977) 'Putting behaviour into behavioural models of travel choice'. Journal of the Operations Research Society vol. 29 1977 pp. 541-550.

HEGGIE, I.G. and JONES, P.M. (1978) 'Defining domains for models of travel demand'. Transporttaion, vol. 7 1978 pp.119-125.

KREIBITCH, V. (1978) 'Modelling car availability, modal split and trip distribution by Montecarlo simulation'. PTRC summer annual meeting 1978.

LEVIN, I.P., MOSELL, M.K., LARIKA, C.M., SAVAGE, B.E., GRAY, M.J. (1978) 'Measurement of psychological factors and their role in travel behaviour'. Transp. Research Record 6-49.

MANHEIM, M.L. (1979) 'What do we want from travel demand analyses?' Paper presented at research seminar on 'Transport and Public Policy Planning University of Reading March 1979.

MARGOLIN, J.B., MISCH, M.R., DOBSON, R. (1976) 'Incentives and disincentives to ridesharing behaviour: a progress report'. Presented at 55th Annual meeting of TRB, Jan. 1976.

O.P.C.S. (1973a) The 1971 population census. HMSO.

O.P.C.S. (1973b) The 1971 workplace analysis census. HMSO.

ORTUZAR, J.D. and WILLIAMS HCWLW.(1979) 'Testing the theoretical accuracy of travel choice models using Montecarlo simulation'. Presented at Institute of Mathematics and Appliations Conf. Univ. of Surrey, April 1979.

TOMLINSON, R.W. and KELLETT, J.S. (1977) 'The theoretical potential for organised carpooling in the UK'. TT 7710, Dept. of Transp. TEchnology, Loughborough University of Technology.

TOMLINSON, R.W. and KELLETT, J.S.(1978) 'Carpooling: two surveys investigating foundations for its application'. TT 7804, Dept. of Transp. Technology, Loughborough University of Technology.

USDOT (1974)'Review of matching software and procedures', FHWA Washington

VINCENT, R.A. and WOOD, L (1979) 'Carsharing and carpooling in Great Britain: the recent situation and potential'. DOE. DOT. TRRL Laboratory Report LR 893.

WAGNER, F.A. (1978) 'Evaluation of carpool demonstration projects - Phase 1. report, USDOT, FHWA Washington.

WILLIAMS, HCWLW (1979) 'Travel demand forecasting - an overview of theoretical developments', presented at research seminar on 'Transport and Public Policy Planning', University of Reading, March 1979.

WYTCONSULT (1976) 'Household interview survey report'. West Yorkshire Transportation Studies, document 511.

APPENDIX A - THE SPATIAL SEARCH ROUTINES

A.1 The eliptical search routine

This routine assumes that one (either) end of the trip is common to all applicants, for the purposes of this description we will assume that the destination is fixed and that origins may vary.

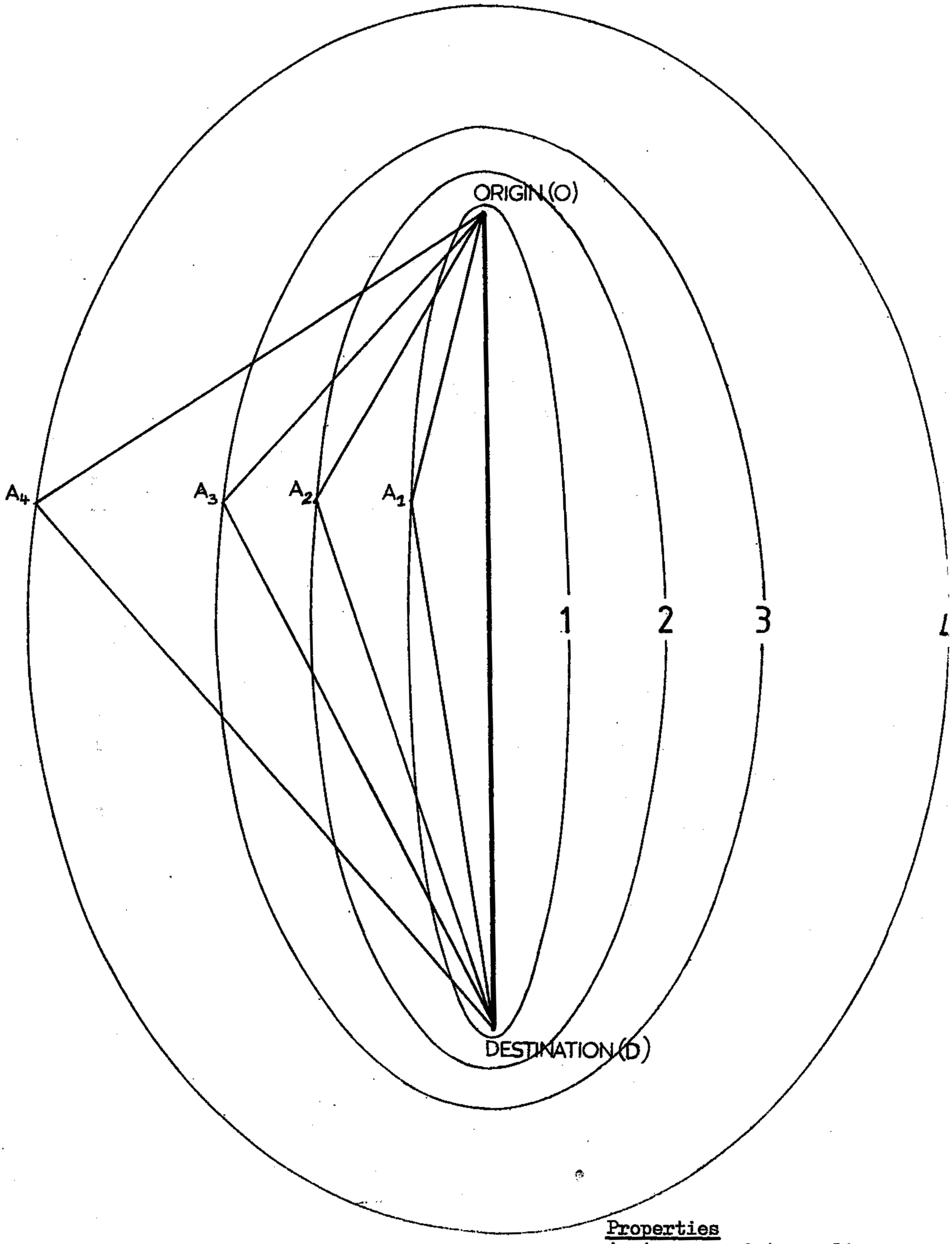
The efficiency of this search routine is due to the fact that it is so ordered that it considers locations involving progressively large diversions from the candidates minimum distance route. In this way the best locations are visited first and the search may be terminated as soon as sufficient partners have been found.

The order of locations to be visited is specified in a search table (so as to avoid the necessity for repetitive calculation). The search list contains a list of locations contained in progressively larger elipses whose major axis is the trip from which diversions are being considered. The relationship between increasing diversion and elipses will be appreciated from figure A.1:

Consider the journey from origin (O) to the destination (D) via A_1 which is located at any position on elipse number 1. All journeys $O \rightarrow A_1 \rightarrow D$ whatever the position of A_1 will have the same total diversion. Similarly all journeys via A_2, which is located at any position on elipse 2, will be of equal length to each other and longer than all journeys via A_1. Similarly for journeys via A_3 and A_4. Clearly if we wish to search locations at progressively large diversions from the straight line $O \rightarrow D$ then we should search on elipse 1, then on elipse 2, then on elipse 3 and so on.

The locations in the search table are expressed in terms of bearing and distance. The bearing being relative to the direction of the major axis of the elipse and the distance being relative to the length of that axis. The expression of locations in these terms allows them to be translated into locations in the study area so long as they too are expressed in terms of bearing and distance from the common destination. The presorting of all specified origins within the study areas into increasing distance within increasing bearing from the common destination then allows the search to proceed with great efficiency.

Figure A1 The relationship between elipses and journey length



Properties

- A_n is any point on elipse n.
- All journeys $O \rightarrow A_n \rightarrow D$ will be of equal distance
- Journey length increases with n.

This may be illustrated with reference to table A.1 which shows excerpts from the search table. Suppose we wish to find potential partners for a candidate whose origin is at a bearing of 100° and at a distance of 10 kilometers from the common destination. The search table tells us that our search must begin at (candidates' bearing $+0^{\circ}$) between distances (0.5 x candidates distance) and (1.01 x candidates distance). This is thus translated into bearing (100°) distances (5.0 kilometers) to (10.1 kilometers). The search then proceeds to (candidates' bearing $\pm 1^{\circ}$) between distances (5.0 x candidates distance) and (1.00 x candidates distance). This is translated into bearing 101°, distances 5.0 kilometers to 10.0 kilometers and bearing 99°, distances 5.0 kilometers to 10.0 kilometers, ...and so on.

It will be noted that the scale of the search is automatically adjusted to reflect the length of the candidates' journey.

TABLE A.1 THE ELIPTICAL SEARCH TABLE.

location number	bearing (relative to bearing of origin from destination) ±	distance from destination (in units, where one unit = the distance from origin to destination)	
		minimum	maximum
1	0	.5	1.01
2	1	.5	1.00
3	2	.5	.98
4	3	.5	.96
5	4	.5	.92
6	5	.5	.86
7	6	.5	.81
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
14	0	1.01	1.02
15	1	1.00	1.02
16	2	.98	1.01
17	3	.96	.99
.	.	.	.
.	.	.	.
.	.	.	.
32	0	1.02	1.04
33	1	1.02	1.03
34	2	1.01	1.03
.	.	.	.
.	.	.	.
718	46	.75	.78
719	47	.74	.77
720	48	.76	.76

The elements in the search table could have been determined computationaly but were in fact derived from a technical drawing which is reproduced as figure A2. Note the position of the line AB which represents our (arbitrary) decision to terminate the search when $\frac{1}{2}$ of the distance from origin to destination has been travelled. Note also that we terminate the search at that elipse which represents a diversion equal to 50% of the original distance from origin to destination.

This search routine is obviously at its best in a densely populated area where it is likely to find potential partners within the first or second elipse but its efficiency advantage over non-directed search algorithms is maintained even in sparsely populated areas.

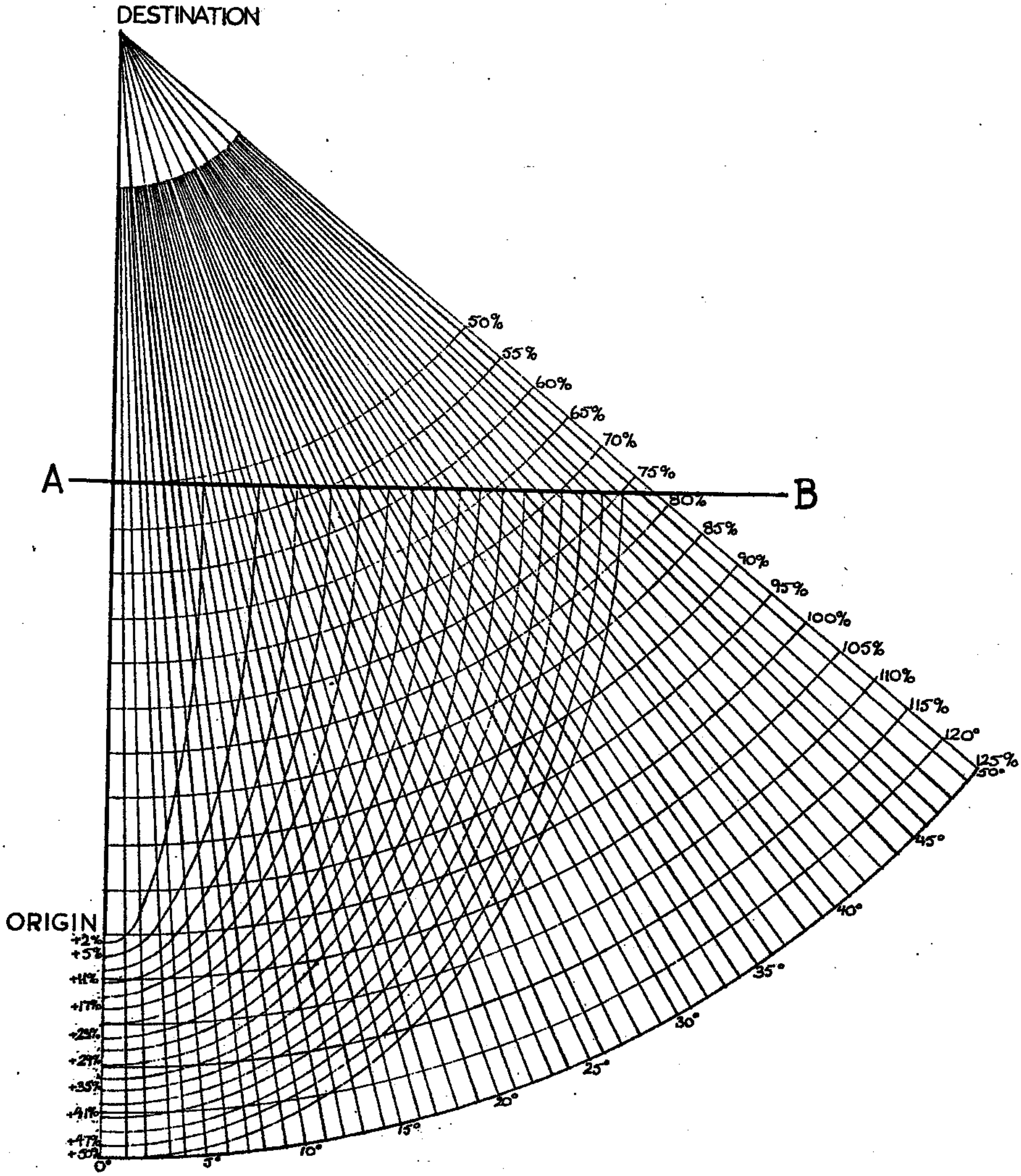
A.2 <u>Zone based search allowing for variable origin and destination</u>

The main advantage of this routine is that it can pair journeys even when neither end is shared but this flexibility is paid for in greatly increased computer usage.

The routine produces for each pair of zones (IJ) in the study area an ordered list of all pairs of zones ($\hat{I}\hat{J}$) which could be visited en route from I to J without excessive diversion from the shortest path from I to J. This list is then assessed by the program which simulates the matching process. Clearly with a 455 zone system we have over 200,000 pairs of zones for consideration. It is not feasible therefore to create a single check list for all possible zone pairs. The routine therefore creates a special list for the subset of zone pairs which are to be considered in a given run of the matching simulation. (The number of zone pairs to be considered at any given time therefore rarely exceeds 1000).

A flowchart of the routine is shown in figure A.3.

Figure A2 Eliptical search routine technical drawing



<u>FIGURE A.3</u> ZONE BASED VARIABLE TRIP END SEARCH ROUTINE