LOCAL GOVERNMENT OPERATIONAL RESEARCH UNIT

Royal Institute of Public Administration

COMMUTERS' VALUES OF TIME

Report No. T 55
January 1975

LOCAL GOVERNMENT OPERATIONAL RESEARCH UNIT

Royal Institute of Public Administration

COMMUTERS' VALUES OF TIME

A.J.Daly
S. Zachary

PREFACE

This report describes a study carried out by the Local Government Operational Research Unit under contract to the Department of the Environment. We are grateful to the Department for their support of this work, and in particular to Martin Dale and Alan Nichols for their helpful advice throughout the study.

We are also grateful to our former colleagues Tony Davies (now with City of Manchester District Council) and Ken Rogers (now with the United Nations Development Program) for their assistance and to Anne Dixson for her help with this report. Responsibility for any errors must, however, remain our own.

CONTENTS

Chapters 4 and 5 contain material of a highly technical nature and may be omitted by the non-specialist.

1. INTRODUCTION

This report presents the findings of the latest of a series of related studies undertaken by the Local Government Operational Research Unit (LGORU) for the Department of the Environment. These studies have been concerned with investigating the factors that influence individuals in making mode choice decisions, and with deducing from observations of individuals' choices the values of time implied by their decisions.

This particular study was concerned with the development of methodology to follow up an insight reported in the previous study (reference 5). This insight was that from a single observation of mode choice we could derive no more than a single limit on that individual's value of time. The objectives of this subsequent study were to develop the methodology further so that times spent in different activities could be valued differently, and to derive methods of statistical measurement of the validity and accuracy of the values of time derived. In the event, we found it desirable to make explicit the assumptions made concerning the different behaviour of different individuals in similar choice situations. In so doing it became possible to investigate the variation of time values within the population.

Although the major aim of the study has been the investigation of methodology for value of time estimates, the models of modal choice from which these estimates are derived are of interest in their own right. In particular, the variation of time values over the population suggested by some of the results obtained has important consequences for modal choice modelling. Some of these are discussed in Appendix 7.

The remainder of this chapter presents first a discussion of the concept of 'value of travel time'. We then go on to consider the data available for estimation purposes and the limitations that its form and reliability put on those estimates. Finally, we summarize the findings of our previous work, which derived a technique called Limiting Time Values, as a starting point for the investigations of the current study.

The remaining chapters of the report deal first, in Chapter 2, with the forms in which models of mode choice may be formulated to permit derivation of values of time. Chapters 3 and 4 specify the model we chose to investigate and the statistical techniques used for calibration. Chapters 5 and 6 respectively report the computer methodology, and give examples of some of the tests we performed with it. A final chapter summarizes our findings and gives recommendations for future work in this area.

1.1 The Concept of the Value of Time

It is common usage to talk about 'spending' time or 'saving' time, and this usage of itself renders plausible the idea that the speaker has some notion of time as having a value and being worth preserving. Nevertheless, as Lisco has pointed out (see reference 15), time is not a commodity in the economic sense; that is, its use cannot be transferred from one individual to another on payment of a price. Thus any determination of value, or of limits on value as are given by prices, must necessarily be indirect.

The urgency of deriving values of time comes from the need to evaluate proposals for new transportation schemes. Typically, such proposals will offer savings in journey times for individual travellers on the investment of a sum of money by the community. The determination of whether such an investment is worthwhile requires a means of reducing the money investment by the community on the one hand and the time savings to individuals on the other to a common basis.

It is not within the scope of this study to consider the questions of equity that arise when time or money savings of one individual are compared with time or money savings of another. Our considerations have been restricted to exchanges of time and money by individual travellers. The values derived are, therefore, to some extent arbitrary as to their units. It would be equally possible, for example, to talk of the value of money in time units, although we in fact follow the conventional practice of expressing time values in money units.

Apart from the familiarity of expressing time in money units, a further reason for retaining this mode of expressing results is that individuals cannot actually gain or lose time. What we are discussing when we talk of the value of travel time is the value individuals place on a change of activity during the time period considered. That is, we seek to estimate how much people would be prepared to pay to be able to divert time spent in travelling to other activities, to be selected by the individual concerned. We cannot measure the absolute value of time, but merely the marginal value of diverting it to other activities. This fact has two important consequences for our analysis.

The first is that different individuals, having different alternative activities in mind, may value time spent travelling very differently even when the travelling activity is the same (e.g. walking). Economists have often argued that the marginal utility of money to an individual will vary with his disposable income, and their argument would already lead us to suspect variations in the money value of time in a population with unequal incomes. Now we see that variations may be expected even between individuals of the same income.

The other consequence of the fact that we measure only the value of transferring time from one activity to another is that we can reasonably expect that time spent in different travel activities will be differently valued. It has been suggested (e.g. by Stopher[1]) that these differences can be approached by psychological measurements of stress, comfort, etc., experienced while travelling. The data available to us, however, did not include measurements of such factors, so that we have taken, for example, 'walking' to be a complete definition of an activity. The stress, etc., experienced by an individual while walking, and the value he puts on that stress, will be subsumed in what we give as 'the value of walking time'.

1.2 Limitations of the Data

In developing the methodology reported here, we have of course kept in mind the data available for statistical analysis. The data used in previous studies has commonly been binary modal choice data, giving details of the mode selected and an alternative for a number of individuals making their journeys to work. Although this is perhaps the most useful form of travel data, it does place certain restrictions on the analysis.

First, it is necessary to screen the data to remove from it individuals whose choice is determined by overriding considerations outside the scope of the model. For example, people who use their cars to get to work because they need to use the car during the day, or who give lifts to other members of their families, are best excluded because their choices are not influenced by factors within the model. That is, we define individuals for whom the model applies to be those with a free choice of modes.

Second, it is necessary to assume that the option of not making the journey is not open to the individuals about whom we have information.[2] This assumption is not unreasonable

1 Personal discussion

2 Nor do we admit of the possibility that a journey may be made to other destinations or by modes other than those for which we have data

for journeys to work.

A further limitation is imposed by the manner in which the data has been collected. The times and costs of the alternative modes in the data are those stated by individuals in answer to a questionnaire. They may therefore differ not only from actual times experienced – because of misperception – but also from the times perceived by the individuals. An element of justification of the choice made has been suggested by other workers.[1]

It is indeed difficult to determine which times – actual, perceived or reported – are most appropriate. The mode choice decision will not necessarily be directly related to actual times, which often vary unpredictably from day to day and may, in any case, be misperceived by travellers. Yet for a future situation it is difficult to predict anything other than estimates of the actual times. We shall therefore proceed on the basis that the reported times are the best estimates available of the times actually experienced.

It is with these reservations and assumptions that we proceed to the analysis of binary modal choice data. Some of the difficulties are reflected in the wide confidence limits found for the time values, but others remain. The true limits are, therefore, still wider particularly in practical applications of values of time.

1.3 Limiting Time Values

The previous LGORU study of the value of time, reported in reference 5, recommended a specific view of the information content of binary mode choice data. This interpretation is also fundamental to the present study, and distinguishes LGORU work from that of most other workers in this field.

In that previous study, we argued that the fact that a traveller chooses one mode in preference to another gives us a limit on the value he attaches to his time. For example, if he prefers mode 1 (time t_1, cost c_1) to mode 2 (time t_2, cost c_2), we can say that

$$vt_1 + c_1 \leqslant vt_2 + c_2 \qquad (1)$$

where v is the individual's marginal value of time. From this inequality it is straightforward to derive the limit:

$$v \leqslant \frac{c_2 - c_1}{t_1 - t_2} \qquad t_1 > t_2$$

$$\text{or} \quad v \geqslant \frac{c_1 - c_2}{t_2 - t_1} \qquad t_1 < t_2$$

Thus time 'spenders' give us an upper limit on their value of time, time 'savers' a lower limit. By comparison of the two sets of upper and lower limits we can obtain information on the way the value of time is distributed among the population.

The limited resources available for the previous study precluded a full exploitation of this insight. In particular, the study team did not examine fully two aspects of the problem that were suggested in the conceptual discussion of section 1.1

---

1 e.g. Heggie in reference 16.

The first of these aspects is that time spent in different travel activities is expected to be valued differently. That is, in the inequality (1) we should consider a series of components of time, each with its distinct value. Also, as investigated in Appendix 2 to the previous study, we should include the possibility of a 'bias', or preference for one mode over the other independent of the travel times or costs. This preference is due to convenience of timing, reliability, status or perhaps other factors not included in the data. With these components, the inequality becomes:

$$b + v_1 t_{11} + v_2 t_{12} \ldots + v_n t_{1n} + c_1 \leqslant v_1 t_{21} + v_2 t_{22} \ldots + v_n t_{2n} + c_2$$

where b is the preference for mode 2 over mode 1 in money units. It is the calibration of this generalized model that was the major aim of this study.

The second of these aspects is that different individuals may have different values of times or modal bias. As we shall see, this possibility colours the whole process of statistical calibration of the models, and constitutes the main departure of this study from the techniques of its predecessors[1].

* * * *

In the following chapter we discuss the possible formulations of models for obtaining estimates of values of time. In this way we can set the models considered in this study in the context of other work in this field.

---

1 These techniques are further discussed in Appendix 2.

2. METHODS OF ANALYSIS

The derivation of values of time from observations of mode choice requires us to set up a model of mode choice in which the value of time figures explicitly. In this chapter we discuss the possible ways in which such models may be and have been set up, and argue that only models that give an individual's mode choice as a determinate function of times and costs permit a direct inference of his value of time.

2.1 Empirical Methods and Statistical Models

Given a set of data on individuals' choices between two modes, let us suppose that there exists a linear function z of measures of time and cost differences with which the choices are correlated; i.e.

$$z = \alpha_1 + \alpha_2 x_2 + \cdots + \alpha_r x_r$$

where $x_2, \ldots\ldots, x_r$ are measures of time and cost differences between the two modes and $\alpha_1, \alpha_2, \ldots\ldots, \alpha_r$ are constants.

Each set of coefficients ($\alpha_1, \ldots, \alpha_r$) specifies a hyperplane ($z = 0$) that fairly well separates the individuals choosing one mode from those choosing the other. By selecting these coefficients to maximize in some way the separation we can find a generalized cost function associated with mode choice in this data set.

An empirical method is one that simply specifies some reasonable procedure for finding a hyperplane separating one set of points from the other. For example, in two dimensions the points may be plotted on a graph and a line drawn by eye. Alternatively, some mathematical criterion of optimal separation may be used, as is done in discriminant analysis.[1] None of the criteria of optimality of such methods is derived from an assumption that the times and costs influence the mode chosen; all are simply criteria of 'good' separation. Although it might be argued on an empirical basis that any set of coefficients giving a reasonable separation must be near the true set, we are not able to go beyond this argument for these techniques.

An entirely different and more satisfactory approach is to start with an explicit model of modal choice:

$$\text{mode} = f\,(\text{times, costs, other parameters})$$

in which all the elements of uncertainty are given explicit probabilistic expression. We can then use the most appropriate method of statistical analysis to estimate the parameters of the model. In this way the criteria of optimality are constrained to be those appropriate to the problem of modal choice, and thus to estimates of values of time.[2]

Explicit models of this type are of two kinds. The one that has been more commonly used is for the mode function to be a stochastic function of determinate variables. It is also possible, however, for the function to be determinate, while some of the variables over which it is defined are subject to stochastic variation. The distinction between these two types of model, which we have called probability and explanatory models respectively, is discussed in the following section.

---

1 Discriminant analysis as a technique for modal split modelling is further discussed in Appendix 1.

2 Once the probability distributions in the model are specified, the precise criterion of statistical analysis is not very important. All reasonable methods will give estimates of the parameters well within the confidence limits for those estimates.

2.2 Probability Models and Explanatory Models

It is clearly not possible to construct a model that will always predict correctly the mode chosen by every individual, and consequently any statistical model of mode choice must contain some stochastic element. We consider two different approaches to the specification of such a model.[1]

1. **Probability model**

This is a model expressed as:

$$\text{Pr (mode 1 chosen)} = \text{p (costs, times, other parameters)}$$

The model simply specifies the probability of choosing each mode as a function of the variables believed to influence that choice. Such a model may be very simple or very complex and could well represent very closely the real situation. Its weakness is that it incorporates no explicit explanation of how the probability function was arrived at. Apart from the general shape of the function p, the model has no *a priori* justification, and therefore the only way a model of this form can be justified is by comparison of observed and predicted results for very large sets of data.

2. **Explanatory model**

This is a model in which, for each individual, the mode chosen is specified *determinately* by a function incorporating time, cost and any other relevant information, but in which some of the parameters are assumed to vary from individual to individual. The exact values of these parameters are not known for any individual, only that they are drawn from some specified form of probability distribution over the population.

Such a model is expressed in the form:

$$\text{mode} = f(\underline{\alpha}, \underline{x})$$

where f is a function that, for binary modal choice, may take only two values; $\underline{x}$ is the vector of independent variables for each individual, including time and cost data; and $\underline{\alpha}$ is the vector of parameters which is distributed over the population. Typically the mode is determined by the sign of a function representing the difference in the generalized costs of the two modes.

The advantage of the explanatory type of model over the probability model is in the direct connection between the individual's valuation of the choices offered to him by the transport system and his decision. Consequently such models may be examined for their economic realism, and more sophisticated hypotheses concerning individual decision-making may readily be introduced. These points will be illustrated below.

However, in analysing an explanatory model – that is, in deriving estimates of the parameters of the various distributions – we no longer require the original form. All that is necessary is to derive from the original form a statement of the probability of choosing one mode or the other. The probability arises, of course, not because the choice for the individual is random, but because the analyst is unable to observe directly the values that the individual has of the stochastic variables in the mode choice function.

---

1 These are described by Stopher (reference 14) as 'Exact Utility' and 'Random Utility' respectively.

Thus there is a close connection between explanatory models and probability models, in that explanatory models must be rewritten for analytic purposes as probability statements. The conversion in that direction is always possible. The conversion from probability models to explanatory models is also usually possible (by somewhat artificial processes), so that the two types of statement are mathematically indistinguishable. We shall, however, continue to use the explanatory form as the basic statement of a model, for three reasons.

First, the existence of the explanatory form enables unequivocal deductions of the meaning of parameters. For example, consider an explanatory model of mode choice:

$$\text{mode} = \text{sign}\,(\alpha_{1i} + c + \alpha_{3i}\, t)$$

where α_{1i} is the preference for one mode by individual i;
c is the cost difference between the two modes for his journey;
t is the time difference;
α_{3i} is a parameter for individual i whose meaning we wish to interpret.

The parameters α_{1i} and α_{3i} are distributed over the population. The parameter α_{3i} may be seen to be the value of time for individual i. For, if his choice is such that $\alpha_{1i} + c + \alpha_{3i} t = 0$, then he is indifferent between the modes. If the time difference changes by one unit, then the cost difference must change by α_{3i} units to make him once more indifferent. That is, he is prepared to pay precisely α_{3i} units of money to save one unit of time – by definition his value of time. On the other hand, the equivalent probability model:

$$\Pr\{\text{mode} = 1\} = \emptyset\,\{\lambda\,(a_1 + c + a_3 t)\}$$

where $\emptyset$ is some s-shaped function;
λ is a non-zero constant;

gives no such ready explanation. An analogous argument to that given above merely shows that an exchange of a_3 units of money for one unit of time will retain the same probability of choosing mode 1. Any interpretation in terms of value of time requires arguments involving distributions over the population – implicitly returning to the explanatory form.

Second, it is much easier to introduce more sophisticated assumptions concerning individual behaviour to the explanatory form. For example, suppose we have an explanatory model:

$$\text{mode} = \text{sign}\;(\alpha_{1i} x_{1i} + \ldots + \alpha_{ni}\, x_{ni})$$

where the x_{1i} are the time and cost measures facing individual i;
α_{1i} are his valuations of those time and cost measures.

If we wish to investigate the hypothesis that the valuations α_{ji} vary multivariate normally with means a_j and covariance matrix Σ, it is simple to derive the probability statement:

$$\Pr\{\text{mode}_i = 1\} = N(\underline{a}'\underline{x}_i(\underline{x}_i'\Sigma\underline{x}_i)^{-\frac{1}{2}})$$

giving the probability of individual i choosing mode 1. It would, we suggest, be difficult to derive this probability statement by other means and still obtain a clear understanding of the meaning, for example, of Σ.

Finally, we use explanatory models because of the arguments of Harris and Tanner (reference 8), which show that the curve-fitting approach of probability models can (in some cases) lead to inconsistencies when using demand models for consumer surplus evaluation.

For the above reasons we decided to use, in the 2-mode situations under consideration, models of the form

mode = sign (f(times, costs, other parameters)).

We now turn to consideration of the possible forms of the deterministic function f.

2.3 Functional Forms of Explanatory Models

In the previous sections we argued for the use of a model of the form:

mode = sign (f(times, costs, other parameters)).

In choosing the forms of f to be investigated there is a large amount of freedom, limited only by the amount of information contained in the data. Some restrictions, however, can be valuable.

First, we are aiming principally to derive values of time for use in economic analysis. It would be totally inconsistent with such analysis to permit behaviour to be predicted on the basis of non-linear valuation of money by an individual. Consequently we may take f as being linear in costs, although the constant of linearity may vary for different individuals.

Second, as the only observed information we have concerning f is its sign for each individual, the units in which it is expressed are arbitrary. Consequently we may divide by the constant money value discussed in the previous paragraph and obtain:

mode = sign (cost difference + f(times, other parameters)).

Third, as the data available to us contains no information concerning journey parameters other than times and cost, it is inevitable that we take all these other parameters together as one overall preference, other things being equal, for one mode over the other. This preference, compounded of status, attitude, convenience of timing, reliability and many other factors difficult to measure and not related to times is often summarily termed a bias. The values of all time-dependent preferences are subsumed in the time values. Thus, we reduce the model to:

mode = sign (bias + cost differences + f(times)).

Finally, we shall consider in detail only formulations of f linear in times. Other formulations may easily be evaluated with the methodology we have developed. For example, an investigation of whether the value of time savings is considered proportional to the time saved would require a non-linear formulation. (This possibility is discussed in Appendix 6.) More complicated functions, however, would require more data for their analysis, and as we shall see, the simple models actually considered stretch the data available well beyond its limits of reliability.

3. LINEAR BIMODAL MODELS

In the previous chapter we presented arguments for the specification of models of mode choice in a specifically explanatory form. In this chapter we go on to specify more precisely the models we propose to investigate and to discuss the range of alternative hypotheses that can be evaluated by the calibration of these models.

3.1 Specification of Bimodal Model

We define a general class of linear explanatory models for bimodal choice as follows. Each individual chooses his mode according to the sign of a linear function:

$$\alpha_1 + \alpha_2 x_2 + \ldots\ldots + \alpha_r x_r$$

where $x_2, \ldots\ldots, x_r$ are functions of the times and costs, and possibly other determinants, influencing the modal choice. The coefficients α are weights, in general distributed over the population. In particular the coefficient α_1 may be regarded as a modal bias, or handicap. (It might, of course, be specified to be zero.) For ease of expression, we introduce a dummy variable x_1, to be identically one, and write

$$\text{mode} = \text{sign}\,(\underline{\alpha}'\underline{x})$$

Typically x_2 may be the money cost difference between the two modes, and $x_3, \ldots\ldots, x_r$ may be differences between the modes of each of the possible components of journey time (for example, waiting time, walking time, 'in-bus' time and 'in-car' time). But $x_2, \ldots\ldots, x_r$ are in no way restricted by the structure of the model to be such linear functions of the times and costs. We might, for instance, put

$$x_3 = t^2_{(1)} - t^2_{(2)}$$

where $t_{(1)}$ and $t_{(2)}$ are the waiting times for the two modes. Thus we would be hypothesizing that waiting time was valued according to its square. Alternatively we might put

$$x_3 = (t_{(1)} - t_{(2)})^2$$

where $t_{(1)}$ and $t_{(2)}$ are as before. Such a hypothesis might arise if we believed that the *saving* in any component of time was the critical factor, and that such a saving was valued non-linearly.

The linear bimodal model may be interpreted as hypothesizing that each individual forms a generalized cost difference between the modes by taking a weighted sum of the various components which have the nature of a cost or penalty. He then chooses the mode for which the generalized cost relative to the other mode is negative. The values of the weights used, including the bias, may vary from individual to individual, so that individuals with identical explanatory data may nevertheless choose different modes. Any specific model will always incorporate hypotheses about the distributions of those weights over the population, for otherwise inference would be impossible.

Usually $\underline{x}$ will be of the form $\underline{y}_{(1)} - \underline{y}_{(2)}$ where $\underline{y}_{(1)}$ and $\underline{y}_{(2)}$ are vectors of data for each mode. In this case the model can be expressed as

$$\text{mode} = \text{sign}\,(\underline{\alpha}'\underline{y}_{(1)} - \underline{\alpha}'\underline{y}_{(2)})$$

Thus it is hypothesized that each individual forms (normally subjectively) a generalized cost for each mode, and then simply chooses the mode with the lesser generalized cost. (Such a model formulation has the considerable attraction that it generalizes immediately to the multi-modal choice situation, without the usual logical contradictions that are implicit in most multi-choice models, which become evident when two of the choices are very similar to one another. Such generalizations are discussed in Appendix 7.)

Whenever a component of time is represented in the generalized cost simply by

$$x_j = t_{(1)} - t_{(2)}$$

where $t_{(1)}$ and $t_{(2)}$ are the durations of that component in the two modes, then the coefficient α_j may always be interpreted directly as the generalized cost weight for time spent in that activity. The fact that an individual has been observed to choose a particular mode enables us to derive a limit on the collective values that his weights $\underline{\alpha}$ may take. Thus our model is a generalization of the Limiting Time Value model proposed in the earlier LGORU study.

For more formal statistical analysis, we need to derive from the model a statement of the probability of each individual choosing the mode observed. This requires an explicit assumption about the distribution of $\underline{\alpha}$ over the population; we shall turn to this consideration later in this chapter.

The model specified contains an excess degree of freedom: if any vector of coefficients $\underline{\alpha}$ explains an individual's modal choice for all vectors of costs and times $\underline{x}$, then so does the vector $\lambda\underline{\alpha}$ for any $\lambda>0$. In other words, we have not specified the *units* of generalized cost for that individual. Further, we may choose a different positive multiplier λ_i for each individual i without changing the model. Thus we are not able to compare one person's generalized cost with another's, unless we actually specify some normalization. We could, for instance, agree that the generalized cost valuation of money would be unity for all individuals, so that the units of generalized cost would be pence or pounds. Alternatively we could apply a scale factor λ_i to each individual so that his value of waiting time, say, became unity. The value of his generalized cost would then be the number of 'waiting minutes' to which he would attach the same disutility.

It is important to note that any inequality existing between two individuals' generalized costs on one such scaling may well be reversed when a different scaling is adopted.

Because of this excess degree of freedom, any individual i, with a specified vector $\underline{x}_i$ of costs and times, has a probability of choosing either mode determined by the distribution of the *ratio*

$$\alpha_1 : \alpha_2 : \cdots\cdots : \alpha_r$$

and independent of all but the sign of any overall multiplier λ_i that determines the magnitude of the components of $\underline{\alpha}$ for that individual. Thus any particular model is given entirely — both conceptually and statistically — by the specification of the distribution of this ratio.

In order to calibrate the model a family F of such distributions is specified, and the individual member that gives the best fit to the observed data is then chosen. The specification of the magnitude of each $\underline{\alpha}$ then corresponds to defining the units of generalized cost for the corresponding individual. It is natural to do this by defining one of the components of $\underline{\alpha}$ to be identically one; for example, to hold α_2 to be one is to define the units of generalized cost to be those of money.

The specification of the required family F is achieved in practice by specifying a family F* of joint distributions of the entire vector $(\alpha_1, \ldots\ldots, \alpha_r)$ such that F* generates F. For any F there are many such families F* and the choice of any one of them implicitly defines units of generalized cost for each individual. Consequently it is plausible to choose the family F* so that the implied generalized cost units are those in which we would finally wish to present our results. However, this choice is not necessary, and a different family F* may have a much simpler analytical representation. Alternatively, careful consideration of the choice of F* may enable us to specify a more general family F of ratio distributions than might at first seem possible.

As an example, consider the family F* of multivariate-normal distributions in the coefficient $\underline{\alpha}$ in which *all* the coefficients are allowed to have a non-zero variance. The generated family F of distributions in the ratios $(\alpha_1 / \alpha_2, \alpha_3 / \alpha_2, \ldots\ldots, \alpha_r / \alpha_2)$ – representing values of bias and time in money units – includes a large class of skewed distributions, and has more degrees of freedom than the simple family of normal distributions that would have resulted if the variance of the money coefficient α_2 had been held to be zero. Thus from a simple family F* of joint distributions of $\underline{\alpha}$ we are able by this device to specify in particular a wide class of possible distributions in the values of time $\alpha_3 / \alpha_2, \ldots\ldots, \alpha_r / \alpha_2$ from which the estimation procedure may choose. Although the same task could in theory have been achieved (more naturally) with α_2 constant this would have been analytically very much more complex.

3.2 Choice of Distributions for Analysis

It was stated in section 3.1 that the linear bimodal model specified there can, by suitable definition of the independent vector variable $\underline{x}$, be used to test a wide range of hypotheses involving non-linear valuations of time and money.

We now restrict consideration to strictly linear hypotheses and consider the model:

$$\text{mode} = \text{sign}\,(\underline{\alpha}'\,\underline{x})$$

where x_1 is identically one,
x_2 is the cost difference between modes,
x_j is the time difference between the modes for the (j-2)th activity involved, e.g. waiting, walking,
$\underline{\alpha} = (\alpha_1, \ldots\ldots, \alpha_r)$ is a vector of weights with a distribution over the population.

In order to analyse this model it is necessary to make explicit assumptions about the distribution of the parameter $\underline{\alpha}$. Clearly a wide range of possibilities is open, and here we indicate a few of the hypotheses that may be examined by appropriate choices of distribution.

If we assume that $\alpha_2 \ldots . \alpha_r$ have constant values $a_2 \ldots . a_r$ over the population, then the variations in individual choices are determined by the distribution taken for the bias α_1. Three cases are worth noting:

1. The bias α_1 is distributed about its mean a_1 according to the logit distribution:

$$\Pr\left\{\alpha_1 < x\right\} = \Phi\,(x-a_1)$$

where Φ is the logit function given by:

$$\phi(t) = \frac{1}{1 + \exp(-t)}$$

In this case we can easily derive the probability function of the model as:

$$\begin{aligned} \Pr\{\text{mode} = 1\} &= \Pr\{\underline{\alpha}'\underline{x} \geqslant 0\} \\ &= \Pr\{\alpha_1 \geqslant -(\alpha_2 x_2 + \ldots + \alpha_r x_r)\} \\ &= 1-\phi(-(a_1 + a_2 x_2 + \ldots + a_r x_r)) \\ &= \phi(\underline{a}'\,\underline{x}) \end{aligned}$$

using the symmetry of the logit function. That is, the simple logit probability model is a special case of the model we have defined.

2. Similarly, if the bias is distributed with the normal distribution:

$$\Pr\{\alpha_1 < x\} = N(x-a_1)$$

where N is the cumulative normal distribution:

$$N(t) = \frac{1}{\sqrt{2\pi}} \int_{-\alpha}^{t} \exp(-\tfrac{1}{2}u^2)\, du.$$

then we should similarly find that the probability function

$$\Pr\{\text{mode} = 1\} = N(\underline{\alpha}'\,\underline{x})$$

showed the simple probit model also to be a special case of our model.

3. Finally, if we choose a rather artificial distribution in the bias (see Appendix 2) we can also show that the Limiting Time Values model proposed by the previous LGORU study is a third special case of this model.

These special cases give a base line against which our current more general investigations may be evaluated.

In this study, however, we wish to investigate the more general assumption that the weights $\alpha_2 \ldots \alpha_r$ do vary between different individuals. We shall assume the variation to be multivariate-normal (subject to the approximation discussed in the next chapter). This is not because we believe the actual distribution of the weights to have precisely this form, but because many actual distributions can be approximated in this way (as explained in section 3.1).

More complicated distributions may, of course, be specified, but, as we shall see, the data sets currently available are not large enough to define accurately two parameters of the distribution of a particular weight. Investigation of distributions requiring more parameters would require much more copious data. Finally, the multivariate-normal distribution enables us to derive simply the probability statement for the model. It can be shown that no other distribution has this property.

*　　*　　*　　*

We have in this chapter specified the class of linear bimodal models that we propose to investigate in the remainder of this report. The next chapter describes the methodology of the maximum likelihood procedure that we developed for calibrating the model and the statistics that are generated for assessing the calibration. Chapter 5 gives a broad outline of the computer programs that were written to carry out these analyses. These chapters are of a technical nature, and the non-specialist may prefer to turn to Chapter 6, where the results of tests of the model are described.

4. CALIBRATION OF THE MODEL

This chapter and Chapter 5 deal with statistical matter that may prove difficult for the non-specialist. These chapters are not essential to an understanding of the remainder of the report.

In this chapter we turn our attention to appropriate methods of calibration of the linear bimodal model presented in Chapter 3. This model represents mode choice by:

$$\text{mode} = \text{sign}\,(\underline{\alpha}'\underline{x}) \qquad (1)$$

where $\underline{\alpha}$ is distributed over the population with the multivariate-normal distribution such that

$$E\,(\underline{\alpha}) = \underline{a} \qquad (2)$$

$$E\left\{ (\underline{\alpha} - \underline{a})(\underline{\alpha} - \underline{a})' \right\} = \Sigma \qquad (3)$$

and our problem is to estimate $(\underline{a}, \Sigma)$.

As remarked in section 2.2, the statistical analysis of any mode choice model requires only a statement specifying the probability of an individual choosing a given mode under any circumstances. (In an explanatory model of the type we are considering the probability arises because we do not postulate the exact values of each individual's parameters.) With the above distribution of $\underline{\alpha}$ the generalized cost difference $\underline{\alpha}'\underline{x}$ has mean $\underline{a}'\underline{x}$ and variance $\underline{x}'\Sigma\underline{x}$. Thus if we label the two modes 0 and 1 such that $\underline{\alpha}'\underline{x}$ is the generalized cost of mode 0 less that of mode 1, we have

$$\Pr\left\{ \text{mode 1 chosen} \right\} = N\left(\frac{\underline{a}'\underline{x}}{(\underline{x}'\Sigma\,\underline{x})^{1/2}}\right) \qquad (4)$$

where N is the integrated normal function

$$N(t) = \frac{1}{\sqrt{2\pi}} \int_{-\alpha}^{t} \exp\,(-\tfrac{1}{2}u^2)\,du. \qquad (5)$$

If Σ were known, the transformation

$$\underline{x}^* = \frac{\underline{x}}{(\underline{x}'\Sigma\underline{x})^{1/2}} \qquad (6)$$

would leave us with no greater a problem than that of simple multivariate probit analysis. Since Σ is not known, a possible approach to its estimation is to find the value $\hat{\Sigma}$ of Σ that, when the transformation (6) is made and the resulting probit model analysed, gives the best fit, by some previously defined criterion, to the observed results.

It is, however, arguable that, for any value of Σ, the estimate of the parameter $\underline{a}$ in the multivariate-normal model is excessively sensitive to the presence of outliers in the data set – individuals with very low-probability behaviour. For this reason it is desirable to modify the tail of the normal distribution represented by equation (5) so as to reduce this sensitivity. This is almost ideally achieved by replacing the probit function (5) by an appropriate logit function. Appendix 3 discusses this problem in detail.

In this chapter we discuss this logit approximation to the likelihood function, and then go on to present the methodology developed for estimating the parameters of the model and for assessing the accuracy of these estimates.

4.1 Logit Modification to the Likelihood Function

If we substitute a logistic distribution of the same mean and variance for the normal used in equation (4) the probability statement then becomes

$$\Pr\{\text{mode 1 chosen}\} = \phi\left(\frac{\pi}{\sqrt{3}} \frac{\underline{a}'\underline{x}}{(\underline{x}'\Sigma\underline{x})^{1/2}}\right) \qquad (7)$$

where ϕ is the logistic distribution function:

$$\phi(z) = \frac{1}{1 + \exp(-z)} \qquad (8)$$

(The factor $\pi/\sqrt{3}$ reflects the fact that ϕ is the cumulative distribution function of a random variable with variance $\pi^2/3$.)

Provided there are no, or very few, outliers in the data set the closeness of the cumulative logistic and cumulative normal distributions implies that, for any given value of Σ, the values of $\underline{a}$ estimated by using (4) and by using (7) should be very close to one another, and that, even more importantly, the ratios of the coefficients of the two estimates should be almost identical. The latter result follows from noting that the ratios of any estimate made by using the logistic approximation are independent of the exact factor used to normalize the logit function to the probit. We have chosen the scaling factor $\pi/\sqrt{3}$ because (7) then implies that the variance of the generalized cost difference $\underline{\alpha}'\underline{x}$ is $\underline{x}'\Sigma\underline{x}$. When outliers are present in the data set we believe that the logit modification will lead to better estimates of $\underline{a}$. (See Appendix 3.) However, there is no underlying distribution of the vector $\underline{\alpha}$ which is independent of $\underline{x}$ that leads to an expression of the form (7) for all $\underline{x}$, except in the circumstance where, effectively, only one component of $\underline{\alpha}$ varies over the population. In other words, there is no simple explanatory model in the class of linear bimodal models defined in Chapter 3 that leads to a probability statement of the form (7), except in cases of very simple hypotheses about the form of Σ.[1] As we are specifically interested in an explanatory model of modal choice in which the explanation is provided by distributions in the values of time and bias, we must, in general, continue to regard (7) as a close approximation to (4).

---

1 A particular instance in which (7) does correspond to an exact explanatory model of the type under consideration occurs when we assume the generalized cost weights, and hence the values of time to be all constant except for the bias, which is to have a logistic distribution of variance σ^2. Then we have exactly

$$\Pr\{\text{mode 1 chosen}\} = \phi\left\{\frac{\pi}{\sigma\sqrt{3}} \underline{a}'\underline{x}\right\} \qquad (9)$$

which is a simple instance of (7) capable of full analysis by any simple logit routine. Another hypothesis for which (7) would be exact is that only the component of time weights $\alpha_3, \ldots \alpha_r$ vary, all perfectly correlated, with a logistic distribution. We have no wish, however, to make such specific hypotheses about the distribution of the generalized cost weights, and prefer instead to consider our multivariate normal model which, as we have seen, incorporates a wide range of possible distributions of values of time and bias.

Before estimating the parameters $(\underline{a}, \Sigma)$ in either (4) or (7) it would be useful to have a statement of the dimension of the parameter space. If the dimension of $\underline{a}$ is r, then ½r(r+1) is the dimension of the space of possible covariance matrices Σ, so that at first sight there appear to be ½r(r+3) degrees of freedom in our parameters. We have noted, though, that for all positive λ, $(\lambda \underline{a}, \lambda^2 \Sigma)$ specifies exactly the same model, so that there are in fact at most ½r(r+3) - 1 degrees of freedom. Further it can be shown that if, for any model

$$\Pr\left\{ \text{mode 1 chosen} \right\} = f\left(\frac{\underline{a}'\underline{x}}{(\underline{x}'\Sigma\underline{x})^{1/2}}\right) \tag{10}$$

where f is a one-one function and $(\underline{a},\Sigma)$ and $(\underline{a}^*,\Sigma^*)$ give the same values of f for all $\underline{x}$, then

$$(\underline{a}^*,\Sigma^*) = (\lambda\,\underline{a}, \lambda^2\,\Sigma) \tag{11}$$

for some $\lambda > 0$. This result remains true even when one component of the vector $\underline{x}$ is a dummy variable always held constant – in our model we have x_1 identically equal to one. Hence the number of degrees of freedom in the parameters of our statistical model is exactly ½r(r+3) - 1.

For the purposes of all statistical inference about the parameters $(\underline{a},\Sigma)$ we now take (7) to be exact, and in the next section look at methods of estimating these parameters.

4.2 Choice of Method of Estimation

Our problem is now the purely statistical one of estimating, for any given data set, the parameters $(\underline{a},\Sigma)$ in the relationship

$$\Pr\left\{ \text{mode 1 chosen} \right\} = \Phi\left(\frac{\pi}{\sqrt{3}}\,\frac{\underline{a}'\underline{x}}{(\underline{x}'\Sigma\underline{x})^{1/2}}\right) \tag{7}$$

We solve this problem in two stages. First, we find the best estimate $\hat{\underline{a}}(\Sigma)$ of $\underline{a}$ for a fixed known covariance matrix Σ. Second, we find the value $\hat{\Sigma}$ of Σ such that $(\hat{\underline{a}}(\hat{\Sigma}),\hat{\Sigma})$ gives the best explanation of the data.

For a fixed Σ the problem of determining $\hat{\underline{a}}(\Sigma)$ may be reduced to logit analysis by the transformation

$$\frac{\pi}{\sqrt{3}}\,\frac{\underline{x}}{(\underline{x}'\Sigma\underline{x})^{1/2}} \longrightarrow \underline{x} \tag{12}$$

which reduces (7) to

$$\Pr\left\{ \text{mode 1 chosen} \right\} = \Phi(\underline{a}'\underline{x}) \tag{13}$$

We therefore consider the best estimation technique for logit analysis.

There are two main approaches: maximum likelihood estimation; and least squares estimation. The maximum likelihood estimate $\hat{\underline{a}}$ of $\underline{a}$ is given by solving the vector equation

$$\sum_{i=1}^{n} \underline{x}_i\, y_i = \sum_{i=1}^{n} \underline{x}_i\, \Phi(\hat{\underline{a}}'\underline{x}_i) \tag{14}$$

where $\underline{x}_i$ is the independent vector variable for the ith individual
y_i, zero or one, denotes the mode chosen.

On the other hand, the least squares estimate $\underline{\bar{a}}$ is obtained by minimizing the expression

$$\sum_{i=1}^{n} w_i (y_i - \bar{\phi}_i)^2 \tag{15}$$

where

$$\bar{\phi}_i = \phi(\underline{\bar{a}}' \underline{x}_i) \tag{16}$$

and where the fixed weights w_i must finally satisfy the relation

$$w_i = \frac{1}{\bar{\phi}_i (1 - \bar{\phi}_i)} \tag{17}$$

Both the estimators $\underline{\hat{a}}$ and $\underline{\bar{a}}$ are asymptotically multivariate-normally distributed about the true value $\underline{a}$, in both cases with the same asymptotic covariance matrix V defined by

$$V = I^{-1} \tag{18}$$

where

$$i_{jk} = E \left\{ - \frac{\partial^2 (\log L)}{\partial a_j \, \partial a_k} \right\} \tag{19}$$

and where L is the likelihood function. Commonly, I is referred to as the information matrix, and V is the minimum-variance bound, so that both estimators are also asymptotically efficient. In other words, as the number of observations increase, the variation of either estimator about the true value $\underline{a}$ tends to zero as fast as is possible consistent with the fact that any stochastically generated sample contains a maximum amount of information about the parameters of the underlying distribution.

Thus there is little to choose between the two methods of estimation, at least for large samples, and in all cases the two methods would give answers much closer to each other than the width of the confidence limits involved. There are, however, some slight theoretical grounds for preferring the maximum likelihood estimate $\underline{\hat{a}}$. In particular it is a one-one function of the sufficient statistic

$$\sum_{i=1}^{n} \underline{x}_i \, y_i$$

for the parameters $\underline{a}$, and hence contains all the information about $\underline{a}$ inherent in the original sample. The maximum likelihood estimate is to be preferred on practical grounds also. The solution of equation (14) presents a simpler problem from a numerical analysis point of view than the minimization of the expression (15) in which the weights w_i are not known in advance.

Methods exist for finding approximate solutions for both the above estimates when the observations may be collected into groups such that all those in each group have the same independent vector variable $\underline{x}$. These are, respectively, the empirical logistic transform and the method of minimum logit chi-squared.[1] However, neither of these methods is available for the analysis of value of time data, as the approximation relies on a large number of individuals in each group, while our data is specifically individual with no possibility of creating more than an occasional group with even two individuals. The use of either method would produce totally unreliable results.

Our method of estimation is therefore the formal solution of the maximum likelihood equations (14).

1 Then, provided all the groups are sufficiently large, we may make the linear approximation

$$y_i - \phi_i \simeq \phi_i (1 - \phi_i)(\ell_i - \underline{a}'\underline{x}_i) \tag{20}$$

where

$$\ell_i = \log \left\{ \frac{y_i}{1 - y_i} \right\} \tag{21}$$

and where y_i is now the proportion of individuals in the group choosing mode 1. This approximation may be used to provide an empirical method either of solving the equation (14) or of minimizing the expression (15).

Our second problem is the estimation of Σ. Again our choice is basically between a maximum likelihood and a least squares approach. In either case we vary the parameter Σ attempting to find that value that, under the transformation (12), yields the logit model providing optimal explanation of our observations. It seems sensible to choose the same criterion of optimal explanation as that used within the logit model itself, and consequently we again choose maximum likelihood estimation. (Once more we would expect the results produced by either method to be much closer to each other than the width of the confidence limits involved.)

Our overall calibration procedure is therefore the determination of the maximum likelihood estimates $(\hat{\underline{a}},\hat{\Sigma})$ of $(\underline{a},\Sigma)$.[1]

4.3 Derivation of Confidence Limits

Having chosen the method of estimation we continue the consideration of the relation

$$\Pr\left\{\text{mode 1 chosen}\right\} = \Phi\left(\frac{\pi}{\sqrt{3}}\;\frac{\underline{a}'\underline{x}}{(\underline{x}'\Sigma\underline{x})^{1/2}}\right) \qquad (7)$$

by examining the problem of deriving confidence limits for the value of the vector $\underline{a}$.

First, however, we should distinguish clearly the two types of variation that have so far arisen in consideration of the generalized cost weights. Our basic multivariate-normal linear model postulates that each individual chooses his mode in accordance with the sign of a personal generalized cost difference $\underline{\alpha}'\underline{x}$. The value of $\underline{\alpha}$ varies from individual to individual and we write

$$E(\underline{\alpha}) = \underline{a} \qquad (2)$$

$$\text{cov}(\underline{\alpha}) = \Sigma \qquad (3)$$

This variation in $\underline{\alpha}$ is postulated actually to exist, and is in no way a reflection of the finiteness of the available data from which it may be determined. Indeed as we acquire more data we should be able to determine the exact degree of this variation, as measured by Σ, more accurately. The second type of variation arises in the estimation of the mean weight $\underline{a}$. Because we have only a finite amount of data we cannot determine $\underline{a}$ exactly, but estimate it first by deriving a 'best' estimate $\hat{\underline{a}}$, and second by deriving confidence limits for the true value of $\underline{a}$. As we acquire more data both the variation of $\hat{\underline{a}}$ about $\underline{a}$ and the width of the confidence limits decrease, and indeed become arbitrarily small for a sufficiently large quantity of data.

Confidence limits are concerned with this second type of variation – that inherent in estimation – and provide a basic measure of the amount of information contained in a data set.

When deriving confidence limits for the mean weights in our model it is necessary to make three basic assumptions . Where these do not hold, the true limits will be wider than those quoted. The assumptions are:

---

1 If instead of approximating (4) by (7) we were to carry out direct statistical analysis based on the probit function (4), the same basic approach would be appropriate, probit analysis being used instead of logit. The relative merits of maximum likelihood and least squares estimation would be very much as discussed. On large samples they would be equally good, both being asymptotically efficient, and would produce virtually identical answers; and we would again adopt the same criterion of estimation for Σ as was chosen for $\underline{a}$.

1. That the relation (7) is an exact statement of the truth – whereas it is in fact only a reasonable model;

2. That our observations are entirely free from experimental error;

3. That our maximum likelihood estimate $\hat{\Sigma}$ of Σ is the true value of Σ.

The first two assumptions are always made in the derivation of confidence limits for the parameters of any statistical model, though the inaccuracy arising from making the second assumption may sometimes be countered by including in the statistical model itself terms describing the distributions of the experimental errors. The third assumption is made so as to obtain an analytically manageable problem.

To express these assumptions analytically, we make the transformation:

$$\underline{x}^* = \frac{\pi}{\sqrt{3}} \frac{\underline{x}}{(\underline{x}'\hat{\Sigma}\underline{x})^{1/2}} \qquad (22)$$

for each independent variable $\underline{x}$ in our data set – which is simply the transformation (12) with Σ replaced by its best estimate $\hat{\Sigma}$ – then the relation (7) becomes

$$\Pr\{\text{mode 1 chosen}\} = \phi\{\underline{a}'\underline{x}^*\} \qquad (23)$$

where, as always, ϕ is the logit function

$$\phi(z) = \frac{1}{1 + \exp(-z)} \qquad (8)$$

The three assumptions are equivalent to saying that the equation (23) specifies the true probabilities of the recorded observations for some vector $\underline{a}$. On this basis we consider the derivation of confidence limits for that $\underline{a}$. The limits thus obtained will therefore be narrower, perhaps considerably narrower, than the true limits. It is difficult to assess the impact of the first two assumptions, though we might do something about the third by examining how fast the likelihood function falls away from its maximum as the value of Σ deviates from $\hat{\Sigma}$. However, this is a case of considering the errors in our estimates of errors, and as such is beyond the resources of this study.

The transformation (22) reduces our problem to that of deriving confidence limits for the parameter $\underline{a}$ of the simple logit model (23). In addition, the model hypothesizes the ratio α_j/α_k to be the value of time for the appropriate j and k. If we must pick a single value to represent this distribution, the ratio of the means a_j/a_k has the best claim, for reasons given in Appendix 4. We therefore wish to derive confidence limits for this ratio.

The first derivation – that for the parameter $\underline{a}$ – is relatively simple provided we have enough data available to be able to make use of the asymptotic distribution of the maximum likelihood estimate $\hat{\underline{a}}$ of $\underline{a}$ about the true value. This distribution is multivariate-normal with density function

$$f(\underline{a},\hat{\underline{a}}) = \frac{1}{(2\pi)^{n/2}\,|V|} \exp\{-\tfrac{1}{2}(\hat{\underline{a}} - \underline{a})'\,V^{-1}\,(\hat{\underline{a}} - \underline{a})\} \qquad (24)$$

where the covariance matrix V is the inverse of the information matrix I defined by

$$i_{jk} = E\left\{-\frac{\partial^2(\log L)}{\partial a_j\,\partial a_k}\right\}$$

$$= \sum_{i=1}^{n} x^*_{ij}\, x^*_{ik}\, \phi_i\,(1 - \phi_i) \qquad (25)$$

Using this asymptotic distribution, confidence limits for $\underline{a}$ may easily be derived in the manner usual when estimates have a multivariate-normal distribution, provided we replace ϕ_i in (25) by $\hat{\phi}_i$. The validity of making this substitution, and indeed of using the asymptotic distribution at all, must of course be verified for each data set, and we describe later in this section one practical way of performing both these validations simultaneously. For the present we simply remark that both approximations appear to be reasonably justified for those data sets that are sufficiently large to make practical inference about the value of time.

When we turn to the second problem, deriving confidence limits for a ratio a_j/a_k, we cannot assume that the ratio $\hat{a}_j/\hat{a}_k$ is normally distributed about the true value a_j/a_k -- this is in general decidedly not the case. Indeed the multivariate-normal distribution of $\underline{\hat{a}}$ about $\underline{a}$ implies that the distribution of $(\hat{a}_j/\hat{a}_k - a_j/a_k)$ is skewed and that the various parameters describing it are dependent on both a_j/a_k and a_k. This intrusion of the nuisance parameter a_k makes confidence limits, as defined in the traditional way, difficult in concept, technically virtually impossible to derive, and liable to be misleading in interpretation.

We therefore tackle the development of limits for a_j/a_k by the alternative Bayesian approach, which requires the specification of some prior distribution of the vector $\underline{a}$ before the analysis of the data. (Although the merit or otherwise of this alternative approach to almost the entire field of statistics is still a matter of debate among statisticians, it is generally agreed that neither approach, correctly applied, will lead to inaccurate results.) In this case we take as our prior distribution a uniform distribution of $\underline{a}$ throughout R^r, and modify this in the light of our experimental results to provide a posterior distribution of $\underline{a}$ and hence of a_j/a_k, from which confidence limits of any desired level may immediately be obtained. To derive the posterior distribution of $\underline{a}$ we require a likelihood function for our observations and we make use of the asymptotic result

$$L \propto \exp\{-\tfrac{1}{2}(\underline{a} - \underline{\hat{a}})' V^{-1} (\underline{a} - \underline{\hat{a}})\} \qquad (26)$$

where, as before, V is the inverse of the information matrix defined in (18). (The function L specifies how the probability of having obtained our *fixed* set of observations varies as the parameter $\underline{a}$ varies. Again V is strictly a function $V(\underline{a})$ of $\underline{a}$, but, as before, we approximate it by using the fixed value $V(\underline{\hat{a}})$.[1]

It is necessary to verify that the likelihood function obtained by using the asymptotic result (26) and approximating $V(\underline{a})$ by $V(\underline{\hat{a}})$ is sufficiently close to the true likelihood function for our particular data set. Because of the close relation between the results (24) and (26) this verification will simultaneously establish the validity of the similar approximation of the distribution of $\underline{\hat{a}}$ around $\underline{a}$ in the traditional approach.

One simple procedure is to use (26) and the approximation to V to compute a series of values of $\underline{a}$ for which log L should lie a certain specified amount, typically ½, 1 or 2, below the maximum given by $\underline{\hat{a}}$. The true drops in the log-likelihood are then calculated directly from the relation (23) and the observed data, enabling predicted and actual results to be compared. The execution of this procedure for two or three specified deviations in log L is normally sufficient to establish whether the likelihood function L may be approximated as described. If it can, and assuming our uniform prior distribution, the posterior distribution of $\underline{a}$ is multivariate-normal with covariance matrix V, and it is possible to derive an explicit analytical expression for the density function of the posterior

1 It is worth noting at this point that if we use the Bayesian approach to derive confidence limits for the parameters $\underline{a}$ itself, then using the same uniform prior distribution and the approximation to the likelihood function outlined above, we obtain exactly the same limits as those produced by the more traditional approach when the asymptotic result (24) is used with its corresponding approximation to V.

distribution of any ratio a_j/a_k. Tabulation of the distribution function then provides a method of directly reading off confidence limits at any desired level.

To obtain confidence limits, including say, 90 per cent of the posterior distribution, we remove the lower 5 per cent and the upper 5 per cent from the cumulative distribution function of a_j/a_k. The limits obtained are not, of course, symmetric about the ratio $\hat{a}_j/\hat{a}_k$ of the maximum likelihood estimates, but extend considerably further above it than below. Any other result would be self-contradictory, for confidence limits for a_j/a_k when inverted, must also be confidence limits for a_k/a_j.

4.4 Use of Significance Tests

In this section we derive significance tests for the parameters of the model and show how the results of such tests may also be interpreted as providing measures of goodness-of-fit of the model itself.

Suppose that a statistical model has a parameter space $\Theta = \{\underline{\theta}\}$ of dimension p, and that Θ^* is a subspace of Θ of dimension $q < p$. For a given data set $\underline{x}$, let the maximum likelihood estimates within each space be $\hat{\underline{\theta}}$ and $\hat{\underline{\theta}}^*$ respectively. If $L(\underline{x},\underline{\theta})$ is the likelihood function define the likelihood ratio

$$\ell(\underline{x}) = \frac{L(\underline{x},\hat{\underline{\theta}}^*)}{L(\underline{x},\hat{\underline{\theta}})} \qquad (27)$$

Then, provided certain regularity conditions concerning the asymptotic properties of maximum likelihood estimates are satisfied, if the *true* value of the parameter $\underline{\theta}$ lies within the subspace Θ^*, the statistic

$$-2 \log \ell$$

is asymptotically distributed as chi-squared with $p - q$ degrees of freedom.

This result may be used to test the significance of improvements obtained in the value of the likelihood function by increasing the dimension (number of degrees of freedom) in the space of parameters $(\underline{a},\Sigma)$ of the relationship (7)[1].

In addition the results of these significance tests may be seen as measures of improvement in the goodness-of-fit of the relationship (7), and therefore as measures of goodness-of-fit for our basic model.

Normal goodness-of-fit measures are difficult to devise for any binary model in which we have only one observation for each value of the independent variable. (This is for purely analytical reasons, as the amount of information available in the data set is determined by the *overall* number of observations.) The normal chi-squared test, for example, is inappropriate. Further, even if an appropriate test were devised, there could be difficulties

---

1 It should be remembered that while the achievement of a significant result on increasing the dimension of the parameter space strongly suggests the necessity of the more complex model, we may *not* infer from a non-significant result that the more complex model is *not* necessary – it may simply be that we do not have the data available to distinguish the two models. In this case we must either suspend judgement to await more data, or else turn our attention to a detailed examination of the confidence limits involved.

in the interpretation of the results, for we do not wish to test whether the model exactly fits the observations (apart from anything else these are subject to experimental error), but only whether it represents an adequate statement of the truth for the purposes of inference about modal choice or the value of time. What we really want to know, therefore, is how much better the explanation produced by the model is than that produced by a simpler model, or indeed by a purely random 'explanation'.

In the specific case of logit modal choice models of the form (7), we may regard the purely random explanation as being the degenerate case where all the components of Σ are zero except for σ_{11} (which is identically one) and where all the components of $\underline{a}$ are also zero. In this case the dimension of the parameter space is zero.

Introduction of the model corresponds to increasing the dimension of the parameter space from zero; and making the model more sophisticated (e.g. by allowing Σ to vary so that we no longer have the simple logit model) corresponds to widening the parameter space still further. We wish to know whether, as the parameters $\underline{a}$ or Σ are allowed to vary to their optimal values for a given data set, we achieve a genuinely better explanation of this data. The introduction and optimal estimation of any new parameter will of course always give a better likelihood, but we may use significance tests in the manner already outlined to see if the new model gives a truly improved fit.

5. DESIGN OF A COMPUTER PROGRAM

This chapter deals with topics in statistics and numerical analysis that may prove difficult for the non-specialist. It is not essential to an understanding of the remainder of the report.

In this chapter we outline the methods that were used in this study to apply the calibration techniques described in Chapter 4 to actual modal split data. This was done by writing a FORTRAN IV program for an IBM 370/145 machine.

5.1 General Approach

The multivariate-normal linear explanatory model introduced in section 3.3 can be entirely represented, for the purposes of statistical analysis, by the probability statement

$$\Pr\{\text{mode 1 chosen}\} = N\left(\frac{\underline{a}'\underline{x}}{(\underline{x}'\Sigma\underline{x})^{1/2}}\right) \qquad (1)$$

where N is the integrated normal function defined by

$$N(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-\tfrac{1}{2}u^2)\,du \qquad (2)$$

In Chapter 4 we argued that the calibration of the model, i.e. the estimation of $\underline{a}$ and Σ, could be carried out with minimal loss of accuracy by replacing (1) with the alternative probability statement

$$\Pr\{\text{mode 1 chosen}\} = \phi\left(\frac{\pi}{\sqrt{3}} \quad \frac{\underline{a}'\underline{x}}{(\underline{x}'\Sigma\underline{x})^{1/2}}\right) \qquad (3)$$

where ϕ is the integrated logistic function

$$\phi(z) = \frac{1}{1 + \exp(-z)} \qquad (4)$$

In particular any differences in the estimates of $\underline{a}$ and Σ resulting from the use of (3) instead of (1), would be very much less than the width of the confidence limits inherent in any data set ever likely to be available. We also argued that there were theoretical and practical grounds for choosing maximum likelihood estimates of the parameters $(\underline{a},\Sigma)$ of the relation (3).

A computer program was written to derive these estimates from actual data sets and, further, to obtain the confidence limits and measures of significance as specified in Chapter 4. The program finds maximum likelihood estimates $(\hat{\underline{a}},\hat{\Sigma})$ of $(\underline{a},\Sigma)$ where the components of time in the independent variable $\underline{x}$ (defined in section 3.3) may be either:

1. total time;
2. excess time and in-vehicle time;
3. walking time, waiting time and in-vehicle time;
4. walking time, waiting time, mode 0 in-vehicle time, and mode 1 in-vehicle time, (e.g. in-bus time and in-car time).

The maximum likelihood estimate $\hat{\Sigma}$ is chosen from within a specified space S of possible covariance matrices. A large variety of different spaces may be specified –

corresponding to hypotheses of differing degrees of generality – from the degenerate space in which S contains the one matrix:

$$\Sigma_{11} = 1, \Sigma_{ij} = 0 \text{ otherwise} \tag{5}$$

to the space S of all possible covariance matrices for the multivariate-normal distribution.

As suggested in Chapter 4, we adopt a 2-stage approach to the overall maximum likelihood estimation procedure. In the first stage, for any fixed Σ, we determine the maximum likelihood estimate $\hat{\underline{a}}(\Sigma)$ of $\underline{a}$, and the corresponding log-likelihood $l(\Sigma)$ of the observations. The transformation

$$\underline{x}^* = \frac{\pi}{\sqrt{3}} \frac{\underline{x}}{(\underline{x}'\Sigma\underline{x})^{1/2}} \tag{6}$$

reduces this problem to one of traditional logit analysis by maximum likelihood estimation, and in section 5.2 we describe a logit analysis routine. In the second stage, the value $\hat{\Sigma}$ of Σ maximizing $l(\Sigma)$ is determined by the use of a general non-linear optimization routine in which the logit analysis routine is used to evaluate the objective function.[1]

An essential consideration in the structure of the program is the excess degree of freedom contained in the specification of the statement (3), i.e. the fact that, for any real $\lambda > 0$,

$$(\lambda \underline{a}, \lambda^2 \Sigma)$$

always describes the same model. This excess degree of freedom is removed in the estimation of Σ, rather than that of $\underline{a}$. Specifically, we ensure that the space S of possible Σ matrices never contains Σ_1, Σ_2 such that

$$\Sigma_2 = \lambda^2 \Sigma_1 \tag{7}$$

Two possible complications could arise for a particular data set. One would be the existence of a vector of coefficients $\underline{a}$ such that the sign of the function $\underline{a}'\underline{x}$ agreed with the chosen mode *for every individual.* Then in our statistical model the maximum likelihood estimate of Σ would be $\hat{\Sigma} = 0$. However, the method of analysis involves the continual reduction of the problem to logit analysis by means of the transformation (6), which always requires a non-zero matrix Σ. For any such Σ, in this circumstance of 'perfect discrimination', the maximum likelihood estimate $\hat{\underline{a}}(\Sigma)$ will always tend to infinity. Since such a data set is in no sense a bad one, but rather has some claim to be an unusually good one, it is necessary to be able to detect this situation automatically. Fortunately this can be done in the logit analysis routine itself since the phenomenon of perfect discrimination for a particular data set is retained under the transformation (6).

A much more serious complication could occur when there exist constants $b_1, \ldots, b_r$, not all zero, such that for every independent variable $\underline{x}$, we have

$$\underline{b}'\underline{x} = 0 \tag{8}$$

---

1 For efficiency it is necessary to have some degree of coupling between these two stages of the estimation process. If, for example, $l(\Sigma)$ is evaluated for two matrices Σ_1 and Σ_2 close to one another – as will happen in the optimization routine as the optimum value is approached – then the two maximum likelihood estimates $\hat{\underline{a}}(\Sigma_1)$ and $\hat{\underline{a}}(\Sigma_2)$ (determining $l(\Sigma_1)$ and $l(\Sigma_2)$) should also be close to one another, and it is inefficient not to make use of this information. Arrangements are therefore made for the passing of such information between successive passes of the logit analysis routine.

(This is equivalent to saying that when we view each observation $\underline{x}_i$ as a row of a matrix (x_{ij}), there exists a linear dependence between the columns of this matrix.) If the data set has this property, which is again retained under the transformation (6), then, for any Σ, all vectors of the form

$$\hat{\underline{a}}(\Sigma) + \underline{b} \tag{9}$$

are equally good maximum likelihood estimates of $\underline{a}$. Unique estimation is impossible, and the program is designed simply to report this circumstance if it occurs, and then stop. Fortunately, for a large data set, the possibility of this happening is remote, but it may occasionally be nearly true, in which case the confidence limits for $\underline{a}$ will be unusually wide.

5.2 The Logit Analysis Routine

The logit analysis routine is designed to find the maximum likelihood estimate $\hat{\underline{a}}$ of $\underline{a}$ for the simple statistical relationship

$$\Pr \{ y_i = 1 \} = \phi \{ \underline{a}' \underline{x}_i \} \tag{10}$$

holding for a set of observations $\{(\underline{x}_i, y_i)\}$ where each y_i may take the value zero or one. Then $\hat{\underline{a}}$ is the solution of the vector equation:

$$\frac{\partial (\log L)}{\partial \underline{a}} = \underline{0} \tag{11}$$

where L is the likelihood function.

The program uses the vector form of the Newton-Raphson method, in which successive solutions are obtained iteratively by:

$$\underline{a}^* = \underline{a} + I^{-1} \frac{\partial (\log L)}{\partial \underline{a}} \tag{12}$$

where I is the information matrix defined by:

$$i_{jk} = - \frac{\partial^2 (\log L)}{\partial a_j \, \partial a_k} \tag{13}$$

and $\underline{a}$ and $\underline{a}^*$ are successive approximations.

Both

$$\frac{\partial (\log L)}{\partial a_j} = \sum_{i=1}^{n} x_{ij}(y_i - \phi_i) \tag{14}$$

and

$$\frac{\partial^2 (\log L)}{\partial a_j \, \partial a_k} = \sum_{i=1}^{n} x_{ij} \, x_{ik} \, \phi_i (1 - \phi_i) \tag{15}$$

are readily calculated, and a subroutine is included for the inversion of I.

An initial solution $\underline{a}$ is supplied (frequently the final solution of the previous pass of the logit routine), and iteration continues until

$$\sum_{j=1}^{r} \frac{(a^*_j - a_j)^2}{a^{*2}_j} \leqslant h \tag{16}$$

for a specified small h. Then we set $\hat{\underline{a}}$ equal to $\underline{a}^*$. Should the iterative process start to diverge, the initial $\underline{a}$ is set equal to zero and the process restarted. (In many hundreds of runs this procedure has not yet failed.)

It is not in fact necessary to invert I at each iteration, and a new value of I^{-1} is only calculated when the rate of convergence falls below a specified level. Because the routine is usually called a large number of times by the non-linear optimization routine, it is normally possible to enter it each time with a good first approximation to $\hat{\underline{a}}$, and achieve the final value after only one iteration; the matrix I need not be inverted more than once in every few passes of the routine.

In addition to calculating the maximum likelihood estimate $\hat{\underline{a}}$, the program also calculates the final probability associated with each observation, the log-likelihood for all the observations taken together, and, on the final pass of the routine, a final value of I and of I^{-1} so that confidence limits may then be derived.

The routine will detect the occurrence of 'perfect discrimination' as defined in the previous section, or the linear dependence phenomenon also discussed there, which exists in the data if and only if the matrix I will not invert.

The routine will fail (with suitable diagnostics) if the probability associated with any individual falls below a user-specified level – indicating that that item of data is extremely suspect – or if convergence is not achieved within the specified number of iterations.

5.3 Estimation of the Covariance Matrix

To determine the maximum likelihood estimate $\hat{\Sigma}$ of Σ, a non-linear optimization routine is used to search a specified space S of possible covariance matrices. All the matrices in the space S must be positive semi-definite symmetric. (This is a necessary and sufficient condition for a matrix to be a possible covariance matrix of a multivariate-normal distribution.)

Most optimization routines search in a p-dimensional space R^p of independent variables (R representing the real numbers). It is therefore necessary to establish a mapping from such a space onto the space S. This is done in two stages as follows:

1. The function

$$\Sigma = \Upsilon^2 \tag{17}$$

maps the space U of all symmetric (r x r) matrices Υ onto the space S of *all* positive semi-definite symmetric matrices Σ. Both spaces have the same dimension, and there are only a finite number of matrices Υ – not more than 2^r – that yield the same matrix Σ.

If we restrict S to be the space of all Σ matrices of the form

$$\Sigma = \begin{pmatrix} \Sigma_1 & & O \\ & \Sigma_2 & \\ O & & \Sigma_3 \end{pmatrix} \tag{18}$$

where Σ_1 may be *any* $(r_1 \times r_1)$ positive semi-definite symmetric matrix, Σ_2 may be any $(r_2 \times r_2)$ diagonal matrix with non-negative elements, Σ_3 is a fixed $(r_3 \times r_3)$ diagonal matrix; and correspondingly restrict U to be the space of all matrices of the form

$$\Upsilon = \begin{pmatrix} \Upsilon_1 & & O \\ & \Upsilon_2 & \\ O & & \Upsilon_3 \end{pmatrix} \tag{19}$$

where the restriction on each Υ_i is the same as on the corresponding Σ_i, except that we no longer insist on positive semi-definiteness, then the function (17) maps the restricted U onto the restricted S.

2. Let p be the number of variable components in the restricted space U above (identifying υ_{jk} with υ_{kj} because of symmetry). If we identify each such component with a component in the space R^p, then we establish a one-one correspondence between R^p and U.

Hence we establish a mapping from R^p onto any space S of the form defined in (18) above.

In the program, the user may specify any such space S of possible covariance matrices for the multivariate-normal distribution. For any r_1 of the generalized cost weights it is hypothesized that the covariance matrix may be of any possible form, i.e. any variances and any possible correlations may exist in the weights. For a further r_2 of the weights, it is hypothesized that the distribution of each is independent of everything else but may have any variance. For each of the remaining r_3 weights $(r_1 + r_2 + r_3 = r)$, an independent normal distribution of given variance is hypothesized. Under the above mapping the searching of the space R^p – the natural domain of the optimization routine – corresponds to the searching of the space S.

One restriction is necessary. As remarked in section 5.1 we exclude the possibility of including Σ_1 and Σ_2 with

$$\Sigma_2 = \lambda^2 \Sigma_1 \qquad (7)$$

This restriction is achieved by always holding υ_{11} fixed at its input value (which is normally one).

The user specifies S by means of a vector locating each generalized cost weight in one of the three classes described. He also specifies an initial matrix Υ such that Υ^2 is in S. The program implements the required transformations and uses the optimization routine to find the maximum likelihood estimate $\hat{\Sigma}$ of Σ in S, which is returned together with a statement of the associated variances and correlations.

Several methods of non-linear optimization in R^p exist. Among these Powell's method has long been popular, as it is reasonably efficient. Although more recent methods may be slightly more efficient, we are familiar with the behaviour of Powell's method, and have therefore incorporated it in the program. It is written as a FORTRAN subroutine called VA04A available from AERE Harwell. The capability exists to substitute easily another method, should this be required. Any non-linear optimization routine will in fact only find local optima, but the log-likelihood function appears to be reasonably well behaved and no problems have so far been encountered.

5.4 Special Facilities

The program contains two special facilities to deal with:

1. **Confidence limits**
The confidence limits derived by the program, both for the vector parameter $\underline{a}$ in the relationship (3), and for the ratio of any two of its components, are subject to the considerations and limitations discussed in section 4.3, which also sets out in detail the theory used in their derivation.

Special routines are provided for the derivation of limits for a ratio of components a_j/a_k. A comparison is also obtained of the true likelihood function with that predicted by the use of the asymptotic result

$$L \propto \exp\{-\tfrac{1}{2}(\underline{a} - \underline{\hat{a}})' V^{-1} (\underline{a} - \underline{\hat{a}})\} \tag{20}$$

where $V = I^{-1}$ is approximated by using $\underline{\hat{a}}$ for $\underline{a}$. (This predicted likelihood function is used in the derivation of the limits.) This comparison is achieved as suggested in section 4.3. The set of points of the form

$$\underline{a} = \underline{\hat{a}} \pm (0, \ldots\ldots, 0, 1/\sqrt{i_{jj}}, 0, \ldots\ldots, 0) \tag{21}$$

(where in each case we use $1/\sqrt{i_{jj}}$ in the jth place) lie on the contour of points for which the asymptotic prediction of log L is exactly 0.5 below the optimum. Similarly if we replace $1/\sqrt{i_{jj}}$ by $2/\sqrt{i_{jj}}$ we obtain a set of points on the contour for which the asymptotic prediction is 2 below the optimum. The program evaluates the true log L at each of these points, subtracting it from the maximum log-likelihood so that we may easily compare the sets of results with the predicted values of 0.5 and 2.0.

The program calculates the density function of the posterior Bayesian distribution of the ratio a_j/a_k, and a subroutine, QA03A – again supplied by AERE Harwell, is used to tabulate the integral of this function. That is, we tabulate the cumulative distribution function, expressing our beliefs, in the light of the data, about the possible true values of a_j/a_k. Confidence limits may be read off directly from this tabulation.

2. **Distributions in the values of time**

The underlying multivariate-normal explanatory model (section 3.3) postulates a distribution in the vector $\underline{\alpha}$ of generalized cost weights. The basic program is designed to evaluate the maximum likelihood estimates $(\underline{\hat{a}},\underline{\hat{\Sigma}})$ of $(\underline{a}, \underline{\Sigma})$ where

$$E(\underline{\alpha}) = \underline{a} \tag{22}$$

$$\text{cov}(\underline{\alpha}) = \Sigma \tag{23}$$

If we take these estimates as representing the true values then we can derive the distribution of the ratio of any two of the weights α_j/α_k. In particular, if α_k is the generalized cost weight attached to money, and α_j that attached to a particular component of time, then α_j/α_k is the value, in money units, attached to that component of time. The program is designed to evaluate any such distribution of α_j/α_k.

Analytically, the task is very similar to that involved in the derivation of confidence limits for the true value of the ratio a_j/a_k of mean weights. (Conceptually, however, the two ideas are very different, in that confidence limits are a measure of the lack of complete information about a single value and shrink as the available data increases, whereas the distributions in the ratios α_j/α_k are postulated actually to exist.) The AERE Harwell subroutine QA03A is again used.

6. TESTING THE PROGRAM

Chapters 4 and 5 described respectively the techniques we proposed to use for calibrating our model and the computer program written to apply these techniques. This chapter describes the tests we carried out with the program. The main objective of these tests was to ensure that the program was operating satisfactorily and that the techniques suggested were in fact applicable to real data. Nevertheless, certain of the results are of interest in themselves, and are presented here for that reason.

Because our main objective was to test the techniques, we used for most of the runs the data collected by Quarmby on bus-car choice in Leeds in 1964/65. This data has already been the subject of several studies (e.g. references 5, 12 and 6), and its general properties are well known. Thus it makes a good subject for program testing. The data itself, however, is far from perfect for purposes of modal choice or value of time modelling. The inaccuracies revealed by the investigations of Appendix 5 and the age of the data make it inappropriate for application in the late 1970's, even if the confidence limits associated with the estimates were thought to be acceptable.

Thus the results reported below should be treated as proof that the techniques are working, and in some instances giving insight into choice mechanisms. They should not be taken as calibrations for application in current work.

6.1 Number of Components of Time

The first runs of the program aimed to investigate the number of components of time necessary to explain mode choice. The results of these runs are shown in Table 1.

In this table are shown the explanatory variables, the weights attached to each and the log-likelihood for each run. The first two results are shown for comparative purposes only, to give the log-likelihood for the most simple choice 'models': the first assuming that each individual has a probability ½, and the second that each individual has a (calibrated) constant probability, independent of times and costs, of choosing mode 1. The subsequent runs show the improvements obtained by introducing first cost, then a successively increasing number of components of time. By comparing the log-likelihood improvement with the additional number of degrees of freedom at each stage, we can determine the significance of each additional component, as explained in Chapter 4.

The results in Table 1 indicate that time and cost are indeed important explanatory variables for mode choice. Further, the splitting of total travelling time into excess time and in-vehicle time is also very important. The worse significance level recorded for the split of excess time into walking and waiting does not mean that this split does not improve the model, merely that the necessity for the split is not proved by this test. The necessity is in fact established by later tests. Finally, the separate valuation of in-bus time and in-car time is also shown to be very important by the last run.

This final result, which has not been found by previous researchers, is nevertheless intuitively plausible. It might be thought that the result was due to too high a valuation being put on car running costs (obviously highly correlated with in-car time), but in fact a later result showed that this was not the case.

The weights given in this table are scaled so that the distribution of the generalized cost over the population has unit variance.

6.2 Variations in the Generalized Cost Weights

The probability statement of the model we investigated is

$$\Pr \left\{ \text{mode 1 chosen} \right\} = \Phi\left(\frac{\pi}{\sqrt{3}} \frac{\underline{a}'\underline{x}}{(\underline{x}'\Sigma\underline{x})^{1/2}} \right)$$

where Φ is the logit function. The runs described in section 6.1 were those for which $\sigma_{11} = 1$ and all other elements of Σ were zero. We now go on to consider non-zero values of these other elements, retaining $\sigma_{11} = 1$ throughout.

As explained in Chapter 5, control of these non-zero elements is achieved by the specification of a control vector. Elements of this vector may take the values 0, 1 or 2, as follows:

Control = 0 means that the corresponding weight in the generalized cost function is assumed to have zero variance over the population.

Control = 1 means that the corresponding weight is assumed to vary over the population (with a variance to be estimated), but independently of all other weights.

Control = 2 permits the mean and variance of that weight to be estimated as well as the correlation of that weight with all others for which the control is 2.

For all runs presented on this report, the bias is assumed to vary independently of all other weights. The controls, therefore, refer only to cost and time weights. The last run in Table 1, for example, is identified by the control 00000. A run allowing the weights of cost and total time to vary, but without the possibility of correlation, would be identified by the control 11.

Table 2 gives a summary of all the runs of this model that we performed on the Leeds data. From a comparison of the degrees of freedom and the log-likelihoods, we can see that the runs with three and four components of time gave significantly better explanations than the runs with only one or two components.

Detailed results of some of the three component runs are shown in Table 3. Two results in this table worthy of particular note are the third (0110) and fourth (0111). The former shows that allowing variation in walking and waiting time weights gives a significantly better explanation of the traveller's behaviour. In both runs, however, we note that the best estimate of the standard deviation of the walking time weight is very small, so that similar log-likelihoods could have been achieved by holding this weight fixed. In this case, the numbers of degrees of freedom would be reduced and the significance levels therefore improved. We would conclude specifically that allowing variation of the waiting time weight gives a significantly better explanation than holding the weight constant.[1]

For the models with four components of time we were unable for budgetary reasons to carry out a full investigation. Three of the most interesting results are shown in Table 4. Here again we see from the second and third results that the waiting time weight has a considerable variation, but that walking time is weighted constantly over the population. Comparing the in-bus time weight with the in-car time weight, we see that the former is constant, the latter varying. The significance level recorded for the second result indicates that the variations in waiting time and in-car time weights are

---

1 It is also possible that these variations are an attempt by the model to compensate for rounding by the travellers in reporting their waiting times. See Appendix 5.

jointly significant, but we were unable to investigate their significance separately. [1]

6.3 Confidence Limits and the Value of Time

As we noted in Chapter 4, the ratio of mean values of normally distributed variables has reasonable claim to be seen as the best single representative of the distribution of the ratio. The reasons for this are discussed in Appendix 4. Accordingly, in presenting the value of time – the ratio, for an individual, of his weights for time and money – we propose to use the ratio of the mean weights.

These ratios have been given in Tables 3 and 4. Table 5 gives, for two typical results from those tables, the 90 per cent confidence limits associated with our estimation of the ratios. Note that the ranges given are for the ratio of the *mean* values of the weights – i.e. the range represents the error in estimation, *not* the distribution over the population.

The wide variations found, typical of all the runs, indicate that great caution should be exercised in using these results. Clearly, where results of an evaluation are at all dependent on the value of time input, sensitivity analysis is essential.

6.4 General Results

In this section we draw together a number of useful but unrelated minor results that have been derived from the analysis:

1. The results summarized above indicate that all the features of the program are working satisfactorily. The most complicated run of the program – run 2222 reported in Table 3 – took 19 minutes to run on the IBM 370/145. This run involved 275 calls of the logit routine, estimating five mean weights, and ten degrees of freedom in Σ. Thus the cost of computer analysis of data, while by no means negligible, is very small by comparison with the other costs incurred in studying the value of time.

2. It will be seen from the tables that the analytic device of permitting variations in the weight attached to the money cost has not given any significant improvements in log-likelihoods. This in effect means that we have not found definite evidence that distributions of the value of time – i.e. the distributions of ratios of coefficients – are significantly skewed away from the normal distribution. Some such evidence has been found, however, and it remains more likely that the distributions are not normal than that they are. More data will be required to determine the truth in this question.

3. In section 6.1, we remarked that a further analysis had shown that the lower weight associated with in-car time compared with in-bus time was not due to the assignment of too high a cost to car running. This analysis is shown in Table 6.

   We carried out a series of runs of the model 0000 that gave a significant difference in the in-car time and in-bus time weights. For these runs, we multiplied the car cost given in the data by the factors in the first column of Table 6. From the table it can be seen that the best explanation is given by a car cost actually *higher* than the original figure. The

---

1 As in the previous footnote, these apparently significant variations may be the result of attempts to compensate for rounding by the respondents.

log-likelihood improvement is not, however, significant, and we continued to use the original figures. However, the results do indicate that the car cost is not too high. For example, the second result in the table, with multiplier 0.6, gives a significantly worse explanation of the observations, but already shows in-bus time weighted more highly than in-car time. We conclude that the difference in these weights is established by this data.

*   *   *   *

The final chapter draws together the most important points arising from our work, and makes recommendations for future studies of the value of travel time.

Table 1: Leeds data Simple logit models

Explanatory variables	Generalized cost weights	Log-likelihood	Degrees of freedom	Significance level
None (random mode choice)		-374.30	0	
Bias only	0.451	-333.92	1	< 0.1%
Bias Cost	0.626 0.0201	-323.42	2	< 0.1%
Bias Cost Total time	-0.167 0.0255 0.0336	-268.79	3	< 0.1%
Bias Cost Excess time In-vehicle time	-0.397 0.0253 0.0519 0.0239	-259.11	4	< 0.1%
Bias Cost Waiting time Walking time In-vehicle time	-0.468 0.0249 0.0637 0.0496 0.0240	-258.63	5	> 10.0%
Bias Cost Waiting time Walking time In-bus time In-car time	-0.950 0.0321 0.0624 0.0451 0.0279 0.0126	-252.82	6	< 0.1%

The significance level for each experiment is that corresponding to the improvement on the preceeding experiment.

Units of cost are old pence, and units of time are minutes.

Table 2: Degrees of freedom and log-likelihoods of all experiments

Bias, cost and total time			Bias, cost, excess and in-vehicle time			Bias, cost, waiting, walking and in-vehicle time			Bias, cost, waiting, walking, in-bus and in-car time		
Exp. identifier	D.F.	Log-likelihood	Exp. identifier	D.F.	Log-likelihood	Exp. identifier	D.F.	Log-likelihood	Exp. identifier	D.F.	Log-likelihood
00	3	-268.79	000	4	-259.11	0000	5	-258.63	00000	6	-252.82
10	4	-267.62	100	5	-258.74	1000	6	-258.35			
						0110	7	-254.89	01001	8	-249.29
01	4	-268.79	011	6	-258.89	0111	8	-254.85	01111	10	-249.29
11	5	-267.43	111	7	-258.19	1111	9	-254.81			
			122	8	-256.77	1222	12	-252.74			
22	6	-267.33	222	10	-253.20	2222	15	-250.15			

Table 3: Leeds data Explanation by bias, cost, waiting, walking and in-vehicle time

Experiment Identifier	Variables	Distribution of generalized cost weights		Correlation matrix					Log-likeli-hood	D.F.	Significance level
		Mean	Stan. dev.	Bias	Cost	Wait time	Walk time	In-veh. time			
0000	Bias	-18.8	40.1	1.0	-	-	-	-	-258.63	5	Base experiment
	Cost	1.00	-	-	1.0	-	-	-			
	Waiting time	2.55	-	-	-	1.0	-	-			
	Walking time	1.99	-	-	-	-	1.0	-			
	In-vehicle time	0.96	-	-	-	-	-	1.0			
1000	Bias	-18.9	37.5	1.0	-	-	-	-	-258.35	6	> 10%
	Cost	1.00	1.05	-	1.0	-	-	-			
	Waiting time	2.46	-	-	-	1.0	-	-			
	Walking time	1.97	-	-	-	-	1.0	-			
	In-vehicle time	0.97	-	-	-	-	-	1.0			
0110	Bias	-25.6	34.9	1.0	-	-	-	-	-254.89	7	2.4%
	Cost	1.00	-	-	1.0	-	-	-			
	Waiting time	3.70	2.60	-	-	1.0	-	-			
	Walking time	2.00	0.01	-	-	-	1.0	-			
	In-vehicle time	1.00	-	-	-	-	-	1.0			
0111	Bias	-25.5	34.2	1.0	-	-	-	-	-254.85	8	5.6%
	Cost	1.00	-	-	1.0	-	-	-			
	Waiting time	3.68	2.59	-	-	1.0	-	-			
	Walking time	1.99	0.00	-	-	-	1.0	-			
	In-vehicle time	1.03	0.35	-	-	-	-	1.0			
1111	Bias	-26.0	32.60	1.0	-	-	-	-	-254.81	9	> 10%
	Cost	1.00	0.95	-	1.0	-	-	-			
	Waiting time	3.57	2.43	-	-	1.0	-	-			
	Walking time	2.01	0.00	-	-	-	1.0	-			
	In-vehicle time	1.10	0.55	-	-	-	-	1.0			
1222	Bias	-27.0	31.6	1.00	-	-	-	-	-252.74	12	> 10%
	Cost	1.00	1.11	-	1.00	-	-	-			
	Waiting time	3.55	3.59	-	-	1.00	-0.95	-0.66			
	Walking time	2.06	1.18	-	-	-0.95	1.00	0.38			
	In-vehicle time	1.14	0.73	-	-	0.66	0.38	1.00			
2222	Bias	-23.1	23.5	1.00	-	-	-	-	-250.15	15	7.5%
	Cost	1.00	1.61	-	1.00	-0.32	-0.05	0.99			
	Waiting time	2.96	2.52	-	-0.32	1.00	-0.93	-0.20			
	Walking time	1.78	0.90	-	-0.05	-0.93	1.00	-0.17			
	In-vehicle time	1.13	1.34	-	0.99	-0.20	-0.17	1.00			

Units
Money - old pence
Time - minutes

Means and standard deviations have been scaled so that the mean weight attached to money is one.

Significance levels. For each experiment we give the significance of the improvement on the simple logit model i.e., the first result in the table.

The standard deviations and correlations refer to the hypothesized joint distribution of the generalized cost weights. They are not measures of the reliability of the estimates.

Table 4: Leeds data. Explanation by bias, cost, waiting, walking, in-bus and in-car time

Experiment identifier	Variables	Distribution of generalized cost weights: Mean	Distribution of generalized cost weights: Standard deviation	Correlation matrix: Bias	Cost	Wait time	Walk time	In-bus time	In-car time	Log-likelihood	D. F.	Significance level
00000	Bias	-29.6	31.1	1.00	-	-	-	-	-	-252.82	6	Base experiment
	Cost	1.00	-	-	1.00	-	-	-	-			
	Waiting time	1.94	-	-	-	1.00	-		-			
	Walking time	1.40	-	-	-	-	1.00	-	-			
	In-bus time	0.87	-	-	-	-	-	1.00	-			
	In-car time	0.39	-	-	-	-	-	-	1.00			
01001	Bias	-35.6	26.5	1.00	-	-	-	-	-	-249.29	8	2.9%
	Cost	1.00	-	-	1.00	-	-	-	-			
	Waiting time	2.84	2.01	-	-	1.00	-	-	-			
	Walking time	1.44	-	-	-	-	1.00	-	-			
	In-bus time	0.90	-	-	-	-	-	1.00	-			
	In-car time	0.41	0.18	-	-	-	-	-	1.00			
01111	Bias	-35.6	26.5	1.00	-	-	-	-	-	-249.29	10	>10%
	Cost	1.00	-	-	1.00	-	-	-	-			
	Waiting time	2.84	2.01	-	-	1.00	-	-	-			
	Walking time	1.44	0.01	-	-	-	1.00	-	-			
	In-bus time	0.90	0.00	-	-	-	-	1.00	-			
	In-car time	0.41	0.18	-	-	-	-	-	1.00			

Units Money - old pence
Time - minutes

Means and standard deviations have been scaled so that the mean weight attached to money is one.
Significance levels are by comparison with the first experiment

The standard deviation and correlations refer to the hypothesized joint distribution of the generalized cost weights. They are not measures of the reliability of the estimates.

Table 5: 90% confidence limits on values of time (ratios of weights) for the two most-significant results

Experiment identifier		Ratio of estimated weights	90% confidence limits
0110	Value of waiting time	3.70	2.44 - 5.86
	Value of walking time	2.00	1.42 - 3.01
	Value of in-vehicle time	1.00	0.67 - 1.55
01001	Value of waiting time	2.84	1.88 - 4.35
	Value of walking time	1.44	1.00 - 2.15
	Value of in-bus time	0.90	0.64 - 1.31
	Value of in-car time	0.41	0.12 - 0.82

Table 6: Leeds data Car cost sensitivity analysis (Simple logit models)

Car cost multiplier	Bias weight	Cost weight	Waiting time weight	Walking time weight	In-bus time weight	In-car time weight	Log-likelihood
0.0	-0.845	0.0143	0.0658	0.0431	0.0195	0.0196	-267.23
0.6	-0.915	0.0305	0.0640	0.0437	0.0228	0.0167	-257.55
1.0	-0.950	0.0321	0.0624	0.0451	0.0279	0.0126	-252.82
1.2	-0.997	0.0307	0.0622	0.0459	0.0301	0.0106	-251.75
1.4	-1.010	0.0283	0.0619	0.0465	0.0318	0.0093	-251.57
1.6	-1.026	0.0257	0.0619	0.0471	0.0330	0.0083	-251.71
2.0	-1.020	0.0210	0.0617	0.0477	0.0346	0.0072	-252.70

7. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

In this final chapter we draw together the results of the study, and from them derive recommendations for the conduct of future studies.

7.1 Summary and Conclusions

The argument in this report has been developed in five main stages, which are summarized below.

7.1.1 Explanation of the Behaviour of Each Individual

In Chapter 2 we argued that methods of analysis relying on empirical separation techniques cannot be regarded as deriving (except by a general association) parameters for use in modal choice modelling or estimates of values of time.

We further argued that the curve-fitting approach of constructing probability models of the form

$$\Pr\left\{\text{mode 1 chosen}\right\} = p(\text{times, costs, } \ldots\ldots)$$

is not an attempt to explain the behaviour of individuals, but merely to fit the existing situation. For this reason it is not possible with models of this type to investigate more detailed hypotheses of choice mechanisms, nor test the economic soundness of the modelling. In particular, values of time cannot be derived.

We recommend instead the initial formulation of modal choice models of the explanatory form:

$$\text{mode} = f\,(\text{times, costs, } \ldots\ldots)$$

where f is a deterministic function, but where some of its parameters are distributed over the population. Such explanatory models do not have the defects of the curve-fitting approach described above, but may be as readily analysed statistically by deriving from them a statement of the probability of choosing a given mode.

7.1.2 Explicit Assumption of Parameter Distributions in the Population

In this study we have investigated models of the form:

$$\text{mode} = \text{sign}\,(\underline{\alpha}'\,\underline{x})$$

where $\underline{x}$ is a vector of the time and cost differences between the modes and $\underline{\alpha}$ is a vector of weights distributed multivariate-normally over the population.

In Chapters 2 and 3 we argued that it is not possible to draw inferences about the value of $\underline{\alpha}$ without making an assumption about its distribution, even if this is only that it is 'constant apart from bias'. The assumption we have made seems particularly attractive: first, because it is a simple step forward from the stronger assumption implicit in previous studies that $\underline{\alpha}$ is constant for all individuals, apart from the bias; second, because it includes a wide range of possible distributions in the ratios α_i/α_j, i.e. in the values of time for individuals; and, finally, because it permits an analytically simple derivation of the probability statement.

7.1.3 Development of Statistical Techniques for Calibrating the Model

In Chapter 4, we described a calibration method for finding maximum likelihood estimates of the parameters of the model. The calibration is achieved by a 2-stage process using a logit approximation to the probit model. We argued that the approximation was very close, and indeed suggested that the logit has distinct advantages over the probit, as explained in detail in Appendix 3.

We went on to discuss measures of significance derived from the log-likelihood of our observations, and to develop techniques for obtaining confidence limits on the parameters. These enable us to assess objectively the significance and accuracy of the results obtained.

7.1.4 Specification and Testing of a Computer Program

Chapter 5 described the specification of the program that was written for this study. This program enables a wide range of possible models to be calibrated and assessed using the 2-stage procedure. Many useful statistics are also generated by the program.

Various devices have been incorporated to speed the operation of the program, which is now capable of fairly rapid analysis of quite complicated models. All aspects of the program have been tested satisfactorily.

7.1.5 Implications Drawn from Tests of the Model

The trials of the model carried out in this study had as their main objective the testing of the program and techniques we had developed. Nevertheless two interesting positive results were obtained from the Leeds data:

1. That it is necessary to value time in at least four separate components: walking time, waiting time, in-bus time and in-car time;

2. That at least two of these components (waiting time and in-car time) have significant variations over the population.

On the basis of the second result we conclude that analogous variations should be investigated in future studies on other data. Further variations were suggested by the results, but we were unable to establish unequivocally their significance.

Indeed, the major conclusion from the tests was that the data used was inadequate for deriving estimates of values of time without very wide confidence limits being given. The wide limits quoted in Table 5 of Chapter 6 were typical for the Leeds data set. Similarly wide limits were found in a trial using London bus-tube mode choice data.[1] Moreover, these limits are 'internal' – that is, they rely in their calculation on the accuracy of the data and the model formulation – and more realistic limits would be wider still.

7.2 Recommendations

Throughout this report we have stressed the limitations of the data available for model calibration. Consequently, our main recommendation for future work on the value of time is that these limitations should be attacked to bring the data more nearly in line

---

1 Data collected in 1970 and subsequently analysed by LGORU. See Reference 17.

with the techniques, which we believe to have run rather ahead of the data available. The attack should be on three fronts.

First, data sets should be larger. The data sets analysed in this and previous studies have contained fewer than 1,000 individuals. The 542 individuals in the Leeds data, as we have seen, give 90 per cent confidence limits on values of time of around ± 70 per cent, and to reduce these errors it is essential to have more data. Data sets of at least 2,000 reliable observations would be necessary to achieve internal limits of ± 35 per cent, at the same level of confidence. Such limits would be more acceptable and easier to handle in evaluation of proposed transport schemes.

Second, evaluation procedures requiring the input of values of time, and the mode choice models from which these values are derived, should be made compatible. One of the problems that has been made explicit by this study is that the demand model implied by standard evaluation procedures, which assume constant values of time, is incompatible with the more realistic demand models suggested by this study.[1] To overcome this problem requires reformulation of evaluation procedures so as to bring them into line with the demand models on which they are based. A further incompatibility is that evaluation procedures are necessarily based on observed and predicted actual times, whereas much of the data from which values of time are derived is based on reported times. Very little work has been done to relate actual travel times and travellers' reporting of them, and until some reliable results are available we can only speculate what the distortions of using reported times are likely to be. To use actual times for modelling would be greatly preferable, but the standard procedures of transportation planning are not appropriate, as they are designed to generate average values for a group of travellers.

Finally, we believe that some reductions in *real* confidence limits might be made by a theoretical treatment of errors in the data. Transportation data is notoriously error-prone. Some 'errors' will be those introduced by the respondent in reporting travel times, through misperception, rounding of his estimates, etc; but others will be simple errors of copying or illegibility. Some mode choices we have observed are so unlikely as to make a transposition of the selected and alternative modes the most reasonable explanation. Clearly, treatment of this latter type of error would be very difficult, but it might be possible to carry out an analysis of the biases introduced by reporting errors or by estimation from network analysis.

These recommendations for further work would, we believe, bring the data available for analysis closer to the level at which reliable results could be obtained. Of the three, the need for more and better data is the most urgent.

---

1 This incompatibility may give rise to biases caused, for example, by the fact that the people who change their behaviour as the result of a change in a network (e.g. to reduce walking time) will be those who are particularly sensitive to such changes – i.e. have a high value of walking time. Thus the introduction of the possibility of variation in time values in the demand model requires the introduction of corresponding sophistications in the evaluation to avoid bias.

REFERENCES

1. M.G.Kendall & A.Stuart: *The Advanced Theory of Statistics* (Volume 2). Griffin, 1967.
2. D.R.Cox: *Analysis of Binary Data.* Methuen, 1970.
3. W.D.Ashton: *The Logit Transformation.* Griffin, 1972.
4. D.J.Finney: *Probit Analysis.* Cambridge, 1971.
5. A.L.Davies & K.G.Rogers: *Modal Choice and the Value of Time.* LGORU report No.C143, March 1973.
6. K.G.Rogers, G.M.Townsend & A.E.Metcalf: *Planning for the Work Journey.* LGORU report No. C67, April 1970.
7. P.R.Stopher: *Goodness-of-Fit Measures for Probabilistic Travel Demand Models.* Northwestern University, Evanston, Illinois, July 1974.
8. A.J.Harris & J.C.Tanner: *Transport Demand Models Based on Personal Characteristics.* TRRL supplementary report 65UL, 1974.
9. A.D.Grigg: *A Program for Multiple Logit Analysis.* TRRL supplementary report 16UC, 1973.
10. J.Barber & G.Searle: *The Value of Travel Time Savings: Recent UK Studies.* Department of the Environment, 1973.
11. N.W.Mansfield (editor): *Research into the Value of Time* (Conference Proceedings). Department of the Environment, Time Research Note 16, July 1970.
12. D.A.Quarmby: 'Choice of Travel Mode for the Journey to Work: Some Findings', *Journal of Transport Economics and Policy,* September 1967.
13. M.J.D.Powell: 'An Efficient Method for Finding a Minimum of a Function of Several Variables without Calculating Derivatives.' *The Computer Journal,* Volume 7, pp 155 – 162.
14. P.R.Stopher & A.H.Meyburg: 'Travel Demand Estimation: A New Prescription.' *Traffic Engineering & Control,* November 1974 pp 879 – 884.
15. T.E.Lisco: 'The Common Economics of Value of Time'. Presented to Conference on *Issues in Behavioural Travel Demand Modelling and the Valuation of Travel Time.* July 1973.
16. I.G.Heggie: 'Value of Time, Modal Choice, and the Justification for Road Improvement.' Presented to *PTRC Symposium,* July 1974.
17. W.A.Steele & K.G.Rogers: *Predicting Multi-Modal Choice.* LGORU Report No. C139. December 1973.

DISCRIMINANT ANALYSIS

Suppose there exist two populations, which we label A and B. Associated with each individual in these populations is a vector $\underline{x}$ such that the distribution of $\underline{x}$ over each population is multivariate-normal, the two vector means being different but the co-variance matrices being identical. We wish to divide the space X of $\underline{x}$ vectors into two regions that can be used to assign any new individual, whose $\underline{x}$ vector is known but whose population is not, to his correct population with the minimum probability of error. If we choose axes in our space X such that the covariance matrix is diagonal, then the optimal division of X is given by a hyperplane

$$a_0 + a_1 x_1 + \ldots\ldots\ldots + a_r x_r = 0 \qquad (1)$$

perpendicular to the line joining the two means. The value of a_0 is determined by any prior beliefs we may have about the relative likelihood of the individual belonging to one population or the other, and by the relative importance of the two types of mis-classification error.

Strictly the assumption of multivariate-normality in each distribution is not required, but we do require some fairly strong similarity and symmetry conditions between the two distributions.

We now relate this to the problem of bimodal choice by defining the two populations to be the users of the two modes and each vector $\underline{x}$ to consist of the time, cost, and possibly any other relevant data for the corresponding individual. Suppose we manage to verify that the vectors $\underline{x}$ do indeed satisfy the required distributional assumptions throughout each population – at least to a sufficiently high degree of approximation. (This would probably be achieved through the use of a goodness-of-fit test on a sufficiently large sample drawn from each population.) Then for any individual whose vector $\underline{x}$ (time and cost data) we know, but whose chosen mode we do not know, we may use discriminant analysis to assign him to his most probable mode.

In the above situation the two populations are given *a priori* – there is no suggestion that the vectors $\underline{x}$, or rather the information they contain, in any way *cause* the individuals to belong to the populations they do. We must nevertheless postulate such a causal relationship if we are to make any inference about the effect of *changing* the $\underline{x}$ data on modal choice, or if we are to infer anything about the value of time. We must postulate that when an individual's $\underline{x}$ vector is changed there is a possibility that he may switch from one population to the other. Because the discriminating line (1) is not obtained from hypothesizing such a situation there is no direct justification for infering that the function

$$a_0 + a_1 x_1 + \ldots\ldots\ldots + a_r x_r \qquad (2)$$

provides such a causal determinant of modal choice – even in terms of the relative likelihoods of choosing one mode or the other.

The attempt to make such inference about the way in which the $\underline{x}$ data is responsible for the mode chosen *automatically* implies the belief in the existence of a relationship of the form

$$\text{mode} = f(\underline{b},\underline{x}) \qquad (3)$$

where the function f and the parameters b are to be determined – either or both of them may contain probabilistic elements. It should be noted that in this situation all individuals belong to the same population and the $\underline{x}$'s are independent variables. If we wish to use the function (2) obtained from discriminant analysis as representing, at least in part, the function $f(\underline{b},\underline{x})$ then we must show that it provides a satisfactory, and preferably an optimal, explanation of the data in terms of a relationship of the form (3). The only way we may do this is to *start* with a reasonable model of the form (3) and show that the use in this model of the coefficients obtained from discriminant analysis provides a satisfactory explanation of the data. A qualitative argument of this kind was given in section 2.1, but there was no suggestion that the discriminant analysis coefficients provided an *optimal* explanation for any such model (3). It could just be that for some very specific stochastic model (3), the mathematical expression for the coefficients $\underline{b}$ provided by, say, maximum likelihood or least squares estimation was identical with the mathematical expression for the coefficients of discriminant analysis, and if we believed that that model was the best obtainable, it could then be said that the discriminant analysis coefficients provided optimal explanation of the modal choice. However, the justification for using these coefficients would not be that they had been produced by discriminant analysis, but that they had been produced by the model (3).

If we can only justify the use of discriminant analysis for causal inference by showing that it produces sufficiently good explanation in terms of a satisfactory causal model of modal choice – and it remains doubtful whether even this can be done – then discriminant analysis has no standing in its own right for the purposes of such inference. We would do much better to throw it away and use the causal model itself. Then we would work with a correct understanding of what our basic model was, and of its status in relation to other models.

To sum up the above argument: if we are to make causal inference about modal choice we must first have a model of modal choice; discriminant analysis is not such a model.

LIMITING TIME VALUES

As discussed in section 1.3 the model specified in the previous LGORU study (reference 5) derived for each individual a limit on his value of time. This limit was inferred directly from his choice of mode. For this study we have generalized this interpretation to obtain the model:

$$\text{mode} = \text{sign}\ (\underline{\alpha}'\underline{x}) \qquad (1)$$

where $\underline{x}$ is an r-dimensional vector of times and costs for the individual, and $\underline{\alpha}$ an r-dimensional vector of weights attached to those times and costs. The previous study was limited by its budget to consideration of values of r up to 3 (bias, cost and time).

It is in the calibration of the model that this study departs from the previous one. Here we have assumed that $\underline{\alpha}$ has a distribution over the population and have calibrated maximum likelihood estimates for the parameters of that distribution. The previous study assumed $\underline{\alpha}$ to be constant and determined the value for which the number of individuals allocated to the correct mode was maximized. The behaviour of other individuals was considered to be unexplained.

We believe, however, that the original Limiting Time Values (LTV) model is not entirely satisfactory, in that it does not specify *prior to its analysis* the individuals to whom it is to apply. Further, the criterion of optimality is not amenable to the derivation of significance levels or confidence limits on the parameters. In short, the approach as stated is not statistically valid, and the difficulty of deriving significance measures is merely a symptom of this fact. What is required is the definition of a likelihood function, applicable to *all* individuals, which will permit statistical analysis of the model. The difficulties with significance and confidence measures will also, of course, be solved with the definition of a likelihood function by the subsequent application of standard statistical methodology.

In this appendix we specify a distribution for $\underline{\alpha}$ such that, when maximum likelihood estimates are found for the parameters of the distribution, the estimates are identical to those produced by the optimality criterion of the original LTV approach. By setting the model on this formal statistical basis, we are able to clarify the implicit distributional assumptions and examine the consequent strengths and weaknesses of this approach.

To set up the formal basis, suppose we have a model of mode choice (1), in which the distribution of $\underline{\alpha}$ is assumed to be:

$$E(\underline{\alpha}) = \underline{a}$$

$$\text{var}\ (\alpha_i) = 0\ , \quad \text{for } i > 1\ ,$$

and the distribution of α_1 (the bias) is

$$\alpha_1 = -\infty \text{ with probability } p$$

$$\alpha_1 = a_1 \text{ with probability } 1 - 2p$$

$$\alpha_1 = +\infty \text{ with probability } p \qquad (2)$$

and $0 \leqslant p < \tfrac{1}{2}$

Then the likelihood of any individual observation is given by

$$l_i = p \quad \text{(incorrectly classified)}$$

$$l_i = p + (1-2p) = 1 - p \quad \text{(correctly classified)}$$

The overall log-likelihood is thus given by

$$\log L = \Sigma \log l_i = n_1 \log p + n_2 \log (1-p) \qquad (3)$$

where n_1 is the number of incorrect classifications and n_2 the number correct. As $p < 1-p$, it is easy to see that L is maximized when n_2 is maximized – i.e. this model is equivalent to the original LTV model.[1]

It is apparent that the distributional assumption (2) necessary to carry out statistical analysis of LTV is somewhat implausible. It seems more reasonable to us to assume that individuals may have varying views of the modes available to them and that the information about $\underline{\alpha}$ contributed by an individual varies continuously with the value of his independent variable $\underline{x}$ of times and costs, rather than discretely as assumed by LTV. (The discussion in this paragraph does not include the possibility of variations in time and cost weights.)

In illustration of this criticism, we compared an LTV estimate on the Leeds data given in reference 5, with an equivalent run of the model analysed in this study. The maximum likelihood estimate of the parameter p in the LTV technique can be found by differentiating (3):

$$\frac{d}{dp} (\log L) = \frac{n_1}{p} - \frac{n_2}{1-p}$$

This is zero when $p = n_1 / n_1 + n_2$. It can easily be shown that this value *maximizes* the likelihood.

Taking the figures from reference 5, page 48, where the LTV model correctly classifies 417 individuals out of 542, we calculate the log-likelihood of this result from (3):

$$\log L = 125 \log \frac{125}{542} + 417 \log \frac{417}{542}$$
$$= -292.70$$

This result is achieved with three degrees of freedom: value of time, bias, and p. It compares with the third result in Table 1 of Chapter 6 of this report for which the log-likelihood is - 268.79, a very much better figure. Consequently the distribution assumed for the bias in this study appears to be very much more likely than that implied by the original LTV method.

It is interesting to assess the parameters found by the run of our model quoted above on the LTV basis. With these paramters, 406 individuals are 'correctly classified' – i.e. have a predicted probability of choosing the mode they actually chose of greater than 0.5. A further fifty have probabilities between 0.4 and 0.5, and thirteen have probabilities between 0.479 and 0.5, so that there are in this data a large number of individuals with close choice decisions.

---

1 Note that the maximum likelihood estimate of α is independent of p. The maximum likelihood estimate of p itself may be calculated afterwards.

We conclude, therefore, that the distributions assumed in this study give a much more realistic representation of behaviour than the original LTV work. One problem with our work, however, which was not present in the previous study, is the possibly excessive influence on the results of a few individuals with extremely low predicted probabilities – possibly the result of data errors or of behaviour influenced by factors not included in the model. The original LTV work is completely robust in the face of such individuals.

The sensitivity of our generalized logit model to outliers is discussed in Appendix 3, where we conclude that the logit distribution is much more stable than the probit. We conclude that the logit sensitivity is not excessive, and that the much better explanation given of the data compared with the original LTV makes it overall the best solution to the problem.

THE APPROXIMATION OF THE PROBIT CURVE BY THE LOGIT

Figure 1 is a graph of the integrated normal function

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\tfrac{1}{2}t^2)\,dt \qquad (1)$$

on a non-linear vertical scale such that the logit functions

$$\phi_k(x) = \phi(kx) = \frac{1}{1+\exp(-kx)} \qquad (2)$$

are represented by straight lines through the origin of slope k – using the linear vertical scale on the right of the figure, the graph is that of the function

$$z = \phi^{-1}\ (N(x)) \qquad (3)$$

The approximation of the probit function N by the logit function $\phi_k(x)$ therefore corresponds to replacing the curve by the straight line (through the origin) of slope k.

Suppose we have a probit model

$$\Pr\{y=1\} = N(\underline{a}'\underline{x}) \qquad (4)$$

relating binary observations y (taking values 0 and 1) to independent vector variables $\underline{x}$, and we estimate the parameter $\underline{a}$ by replacing (4) with the logit approximation

$$\Pr\{y=1\} = \phi_k(\underline{a}'\underline{x}) \qquad (5)$$

where k gives the best fit of the logit curve to the probit over the relevant range of values of N and then using any appropriate technique of logit analysis to obtain an estimate $\hat{\underline{a}}$. If, instead of k, we were to use a scale factor k'; the same method of estimation would produce the estimate

$$\frac{k\hat{\underline{a}}}{k'}$$

Thus the value of k chosen in the approximation – the straight line by which we approximate the curve in Figure 1 – is important for the determination of the absolute value of $\underline{a}$, but does not matter if we only wish to estimate $\underline{a}$ up to a constant multiple. For this purpose the estimation technique will effectively select the best fit of the two curves.

In the multivariate-normal linear model of bimodal choice that we have presented in detail in this report we approximate the associated probability statement

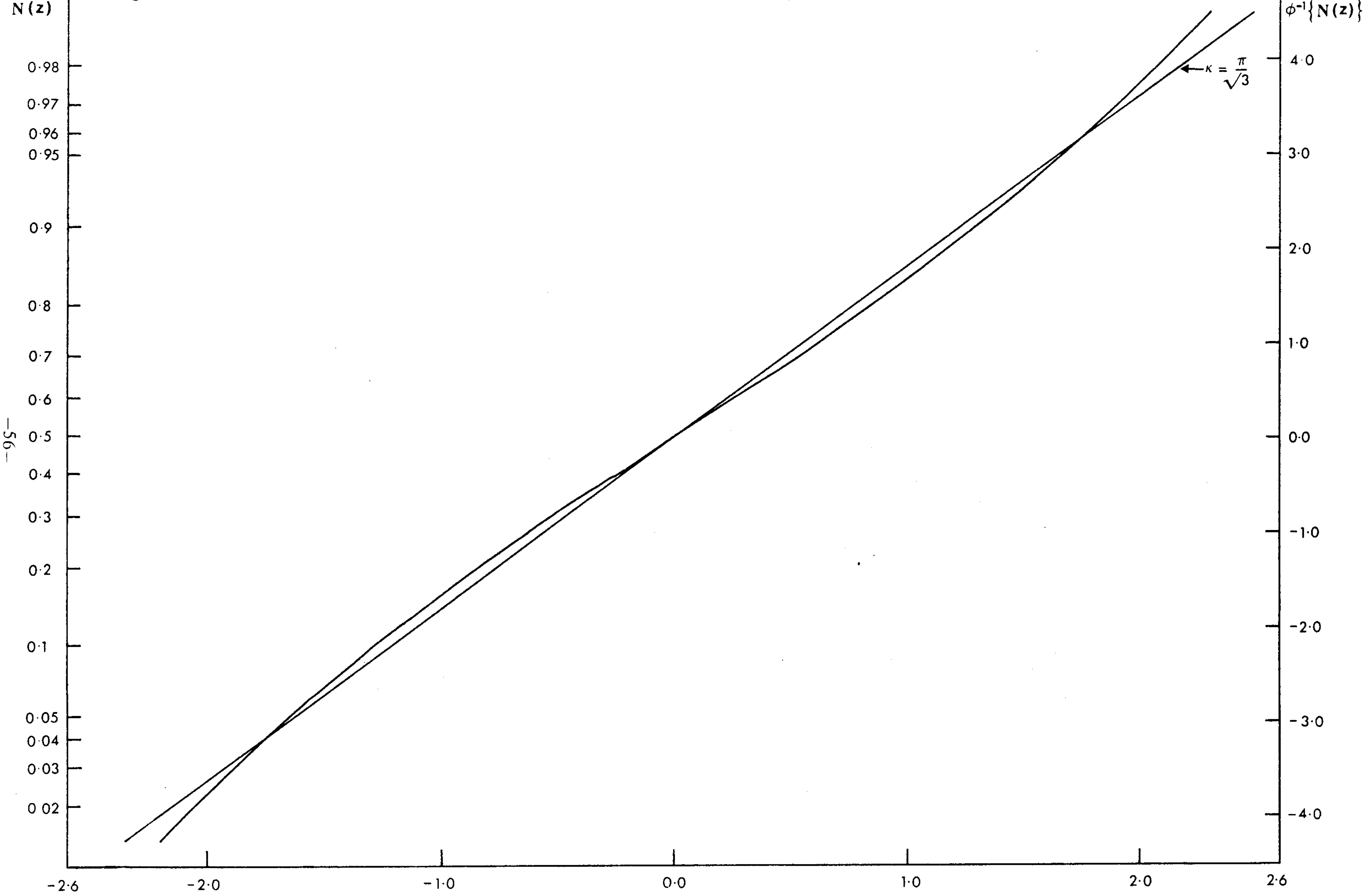
$$\Pr(y=1) = N\left(\frac{\underline{a}'\underline{x}}{(\underline{x}'\Sigma\underline{x})^{1/2}}\right) \qquad (6)$$

by

$$\Pr(y=1) = \phi\left(\frac{\pi}{\sqrt{3}}\ \frac{\underline{a}'\underline{x}}{(\underline{x}'\Sigma\underline{x})^{1/2}}\right) \qquad (7)$$

i.e. we choose the value of k to be $\pi/\sqrt{3}$ (numerically 1.814) as the need to relate $\underline{a}$ to Σ requires an estimation of the absolute value of $\underline{a}$, and there are good grounds for preferring this scale factor to any other (see section 4.1). Figure 1 also plots the logit function with $k = \pi/\sqrt{3}$.

Figure 1: The Integrated Normal (Probit) Function on a Logistic Scale



Our primary interest, however, is in the ratios of the means of the generalized cost weights, i.e. the ratios of the components of $\underline{a}$, and the estimation of these is independent of the value of k used.

We now discuss more formally the justification for the logit approximation, at least in the context of maximum likelihood estimation. First we introduce a measure of *sensitivity* for any model of the form

$$\Pr(y=1) = f(\underline{a}'\,\underline{x}) \tag{8}$$

which is symmetric between the modes, i.e. for which f satisfies the relation for all z

$$f(-z) + f(z) = 1 \tag{9}$$

We define the *sensitivity function*

$$f^*(z) = \frac{d}{dz}\log f(z) = \frac{f'(z)}{f(z)} \tag{10}$$

measuring the proportional rate of change of the probability function $f(z)$ with z.

Suppose we have a set of observations $\{(\underline{x}_i, y_i)\}$, where each $\underline{x}_i$ is an independent vector variable and each y_i takes the value 0 or 1 in such a way as to satisfy the relation (8) for some parameter $\underline{a}$. The relation (9) implies that if we replace any observation $(x_i, 0)$ by $(-\underline{x}_i, 1)$, then, for any value of the parameter $\underline{a}$, the probability associated with that observation (and therefore the overall likelihood of obtaining all the observations) will remain the same, and so the best estimate $\hat{\underline{a}}$ of $\underline{a}$ will be unaffected (no matter what the method of estimation used). Hence, without any loss of generality, we may assume that every $\underline{x}_i$ has been so defined that the corresponding value of y_i is 1. This mathematical reformulation of the problem, in which we discuss probabilities of obtaining the observations actually resulting (the modes actually chosen), and which is possible only when the model is symmetrical between the observations, gives a useful economy of presentation in the following discussion.

The maximum likelihood estimate of the parameter $\underline{a}$ is the solution of the vector equation

$$\frac{d(\log L)}{d\underline{a}} = \underline{0} \tag{11}$$

where the likelihood function L is given (in our reformulation) by

$$L(\underline{a}) = \prod_{i=1}^{m} f(\underline{a}'\,\underline{x}_i) \tag{12}$$

so that (11) becomes

$$\sum_{i=1}^{m} \underline{x}_i\, f^*(\underline{a}'\,\underline{x}_i) = \underline{0} \tag{13}$$

where f^* is the sensitivity function (10).

The maximum likelihood estimate of $\underline{a}$ is therefore that value of $\underline{a}$ for which the sum of the contributions

$$\underline{x}_i\, f^*(\underline{a}'\,\underline{x}_i) \tag{14}$$

from each of the observations exactly cancel one another. If we were to introduce a new observation i into a substantially sized data set then the change in the maximum likelihood estimate would be approximately proportional to the function (14) (this approximation tending to exactness as the size of the data set increased). The contribution of any observation to the maximum likelihood estimate is therefore given by the vector $\underline{x}_i$ multiplied by the sensitivity function f evaluated at the estimate.

Table 7 gives values of the sensitivity function $N^*(z)$ of the integrated normal curve and of the sensitivity function $\phi_k^*(z)$ of the logit function with scale parameter k, together with values of $zN^*(z)$ and $z\phi_k^*(z)$, the value of k for the tabulation being $\pi/\sqrt{3}$. These four functions all tend to zero as z tends to plus infinity.

Suppose that for a given data set $(\underline{x}_i)$, $\hat{\underline{a}}$ is the solution of (13) when f is the probit curve N. *Then provided none of the values of* $N(\hat{\underline{a}}'\underline{x}_i)$ *is small,* there is a value of k such that

$$\sum_{i=1}^{n} \underline{x}_i N^* (\hat{\underline{a}}'\underline{x}_i) \simeq \sum_{i=1}^{n} \underline{x}_i \phi_k^* (\hat{\underline{a}}' \underline{x}_i) \tag{15}$$

This is implied directly by the closeness of the functions N^* and ϕ_k^*, except possibly where $\underline{x}_i$ is large in absolute magnitude. In this latter case $N(\hat{\underline{a}}'\underline{x}_i)$ is almost certainly near one (we have excluded the possibility that it is near zero) and so $\underline{x}_i N^* (\hat{\underline{a}}'\underline{x}_i)$ and $\underline{x}_i \phi_k^* (\hat{\underline{a}}' \underline{x}_i)$ are both very small. *Under the above proviso,* $\hat{\underline{a}}$ is also an approximate solution to equation (13) when f is the logit function ϕ_k with scale parameter k – subject to the one additional condition that the rate of change with $\underline{a}$ of

$$\frac{d (\log L)}{d\underline{a}} = \sum_{i=1}^{m} \underline{x}_i f^* (\underline{a}' \underline{x}_i) \tag{16}$$

evaluated at $\hat{\underline{a}}$ for f equal to either ϕ_k or N, is not unduly small, i.e. that the information matrix I defined by

$$i_{jk} = \frac{\partial^2 (\log L)}{\partial a_j \, \partial a_k} \tag{17}$$

is not small.

Therefore provided the data set contains no, or only a very few, individuals with an excessively low associated probability, the maximum likelihood estimates obtained by using probit and logit models are very close to one another. The exception occurs when the information matrix I is excessively small, but this exception is only to be expected as it is in precisely this circumstance that the confidence limits (under either model) for the true value of $\underline{a}$ are excessively wide.

The important feature of this result is that the presence of any observations with high associated probabilities does not affect the goodness of the approximation. With neither logit nor probit analysis do such observations make any significant contribution to the determination of the maximum likelihood estimate.

The approximation *is* affected, however, by those observations which may be considered the real outliers in the data set – those with very low associated probabilities. To examine the effect of such observations we may introduce an additional 'observation' $\underline{x}_i$ into the data set and examine the effect on the maximum likelihood estimate under each model as we progressively reduce its associated probability.

Under the probit model, as $N(\hat{\underline{a}}' \underline{x}_i)$ is reduced, the vector function

$$\underline{x}_i N^* (\hat{\underline{a}}' \underline{x}_i) \tag{18}$$

(measuring the impact of the observation $\underline{x}_i$ on $\hat{\underline{a}}$) increases in absolute magnitude, not only because $|\underline{x}_i|$ is almost certainly increasing, but also because $N^* (\hat{\underline{a}}' \underline{x}_i)$ is increasing – in fact

$$\frac{N^*(z)}{z} \longrightarrow -1, \text{ as } z \longrightarrow -\infty \tag{19}$$

Table 7: VALUES OF THE FUNCTIONS $N^*(z)$, $\phi_k^*(z)$, $zN^*(z)$ and $z\phi_k^*(z)$ when $k = \pi/\sqrt{3}$ (= 1.8138)

z	$N^*(z)$	$\phi_k^*(z)$	$zN^*(z)$	$z\phi_k^*(z)$
-4.0	4.22	1.81	-16.9	-7.25
-3.5	3.75	1.81	-13.1	-6.34
-3.0	3.28	1.81	- 9.85	-5.42
-2.5	2.82	1.79	- 7.06	-4.49
-2.0	2.37	1.77	- 4.75	-3.53
-1.8	2.20	1.75	- 3.96	-3.14
-1.6	2.02	1.72	- 3.24	-2.75
-1.4	1.85	1.68	- 2.60	-2.35
-1.2	1.69	1.63	- 2.03	-1.95
-1.0	1.53	1.56	- 1.53	-1.56
-0.8	1.37	1.47	- 1.09	-1.18
-0.6	1.22	1.36	- 0.729	-0.814
-0.4	1.07	1.22	- 0.428	-0.489
-0.2	0.929	1.07	- 0.186	-0.214
0.0	0.798	0.907	0.000	0.000
0.2	0.675	0.744	0.135	0.149
0.4	0.562	0.592	0.225	0.237
0.6	0.459	0.457	0.275	0.274
0.8	0.368	0.344	0.294	0.275
1.0	0.288	0.254	0.288	0.254
1.2	0.219	0.185	0.263	0.222
1.4	0.163	0.133	0.228	0.186
1.6	0.117	0.0944	0.188	0.151
1.8	0.0819	0.0667	0.147	0.120
2.0	0.0552	0.0470	0.110	0.0939
2.5	0.0176	0.0193	0.0441	0.0481
3.0	0.00444	0.00783	0.0133	0.0235
3.5	0.000873	0.00317	0.00306	0.0111
4.0	0.000134	0.00128	0.000535	0.00512

Under the logit model, as $\phi_k(\underline{\hat{a}}'\,\underline{x}_i)$ is reduced, the vector function

$$\underline{x}_i\,\phi_k^*\,(\underline{\hat{a}}'\,\underline{x}_i) \tag{20}$$

increases only insofar as $|\,\underline{x}_i\,|$ increases, the function ϕ_k^* tending to a constant. Specifically

$$\phi_k^*(z) = \frac{k}{1 - \exp(kz)} \longrightarrow k, \text{ as } z \longrightarrow -\infty \tag{21}$$

The maximum likelihood estimate under the probit model is thus very much more affected by the presence of outliers in the data set than the corresponding estimate under the logit model. For either model, the degree of sensitivity to such an outlier would be arguably correct if we could be sure that the observation genuinely fitted that model. However, an observation with such a low probability associated with it is quite likely either to be erroneous or else, although correct, not in conformity with the model for some other reason – for example, in the modal choice situation factors other than time and cost may be influential. Even if we could be sure that the observation fitted the model in a general sense, there is still the possibility that the exact curve, logit or probit, might not be perfectly correct – this probability is more of a certainty in modal choice models. In such a circumstance what is really happening is that the probit model is distorting its parameters violently in an attempt to explain an observation not really compatible with the model, whereas the logit model can more easily incorporate such an outlier with its longer-tailed distribution.

The (slightly paradoxical) conclusion is that in the one region where the logit and probit models seriously diverge the logit model is much to be preferred, so that even if we have what is in theory a probit model, as for example our multivariate-normal linear model of modal choice (6) presented in this report, its approximation by a logit model may well lead to better calibration. However, with anything so nebulous as a modal choice model, outliers in a data set are perhaps best regarded with enough suspicion to merit their removal from the data altogether.

RATIOS

We draw attention to two distinct problems:

1. *The representation, in the multivariate-normal linear model of modal choice, of the distribution of a ratio of generalized cost weights by the ratio of the respective means.*

Suppose that we *know* the true parameters $(\underline{a}, \Sigma)$ of the model. Then the distribution of the ratio α_j/α_k of any two of the generalized cost coefficients, representing say the value of a particular component of time in money terms, is completely specified. If we wish to make any practical use of that value of time, we should strictly work with the entire distribution α_j/α_k. This being excessively complex, however, the question arises as to what is the best single representative of the distribution.

Assume, for the sake of clarity, that a_j and a_k are both positive. Then, *provided* Σ_{kk}/a_k^2 *is sufficiently small for a negligible proportion of the distribution of* α_k *to be negative,* the distribution of α_j/α_k is positively skewed and

$$\text{mode } (\alpha_j/\alpha_k) < a_j/a_k \quad (1)$$

$$\text{median } (\alpha_j/\alpha_k) = a_j/a_k \quad (2)$$

$$E\ (\alpha_j/\alpha_k) > a_j/a_k \quad (3)$$

Analogous to (3) we also have, provided Σ_{jj}/a_j^2 is small

$$E\ (\alpha_k/\alpha_j) > a_k/a_j \quad (4)$$

i.e.

$$\frac{1}{E\ (\alpha_k/\alpha_j)} < a_j/a_k \quad (5)$$

Consequently, in isolation, the mean is not a stable representative of the distribution – the reciprocal of the mean is not the mean of the (equally important) reciprocal distribution. For a similar reason neither is the mode stable. However, the ratio a_j/a_k of the means a_j and a_k of α_j and α_k *does* have this stability property, and because it is also generally the median of the distribution of α_j/α_k, probably has the best claim to be the single representative of this distribution.

2. *The best estimate of* a_j/a_k *in the multivariate-normal model.*

Technically the problem is similar to that above. While, to a very good approximation, if $\hat{a}_j$ and $\hat{a}_k$ are the respective maximum likelihood estimates of a_j *and* a_k,

$$E\ (\hat{a}_j) = a_j \quad (6)$$

$$E\ (\hat{a}_k) = a_k \quad (7)$$

for the ratios we have

$$E\ (\hat{a}_j/\hat{a}_k) > a_j/a_k \quad (8)$$

and also

$$E\ (\hat{a}_k/\hat{a}_j) > a_k/a_j \quad (9)$$

provided a_j and a_k are both positive and have relatively small standard errors. The results (8) and (9) taken together are symptomatic of the fact that the concept of unbiassedness is not relevant in this situation. The best estimate of a_j/a_k in most circumstances is $\hat{a}_j/\hat{a}_k$.

THE LEEDS DATA SET

Figure 2 is a histogram that plots numbers of individuals in the Leeds data set (whether bus users or car users) against total time difference between modes (bus mode total time less car mode total time). The histogram covers the range

$$-20 \text{ min} < \text{total time difference} \leqslant 80 \text{ min.},$$

531 of the 542 individuals falling within this range. Each bar of the histogram covers a five minute interval:

$$5n \text{ min} < \text{total time difference} \leqslant 5(n+1) \text{ min.}$$

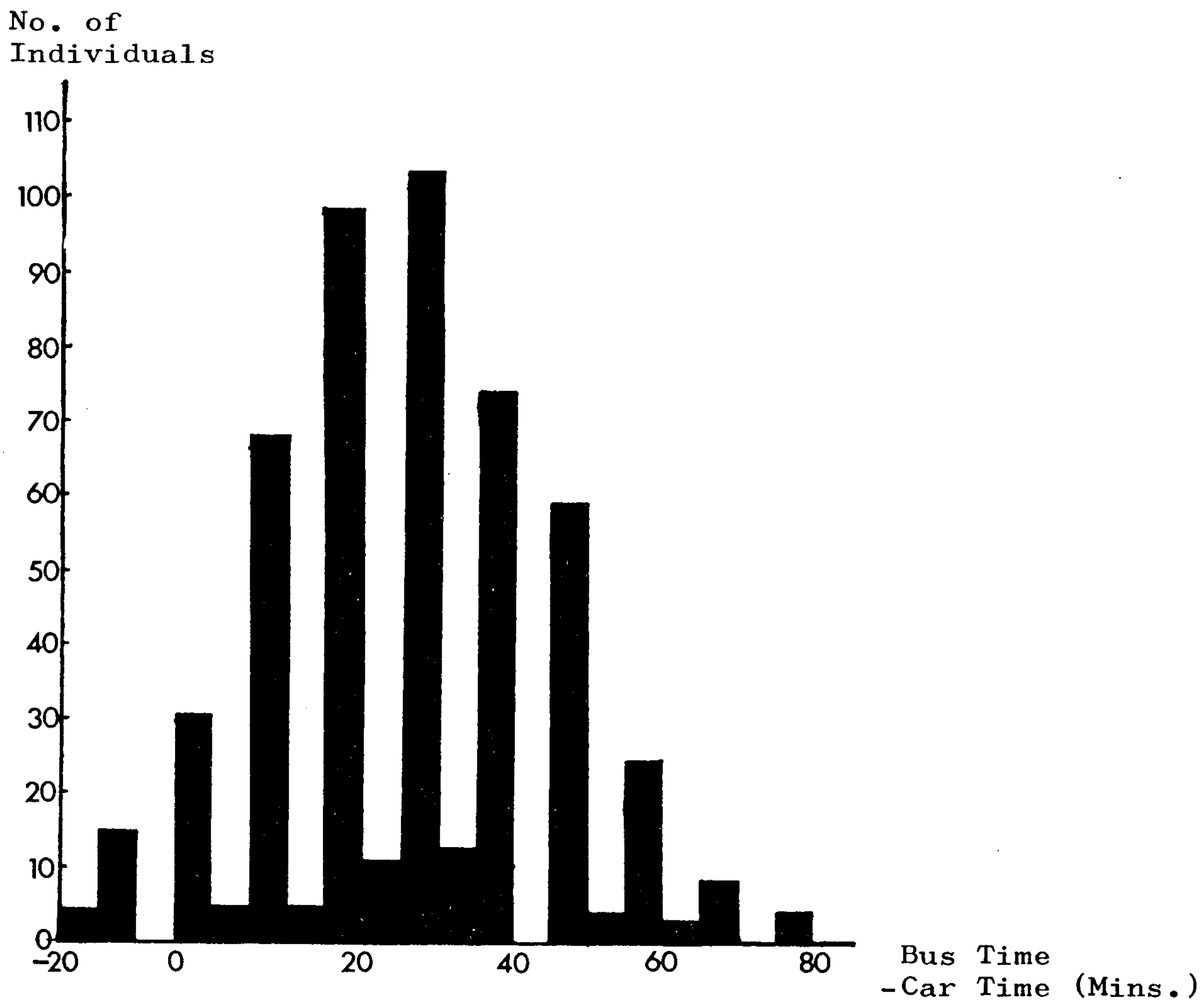
Despite this grouping it is obvious that the total times originally recorded were for the most part given to the nearest ten minutes (for the two-way journey), this being the only satisfactory explanation of the 'clumping' at alternate bars in the histogram.

Figure 3 presents a similar analysis of the waiting time difference between modes – effectively the bus mode waiting time. Separate histograms are given for those who actually choose the bus mode and for those who choose the car mode. These waiting times are all multiples of two minutes. The ten-minute rounding effect persists. Histograms are not presented for the remaining components of time, but the distributions of walking time differences and in-bus times are considerably smoother – indicative of less rounding in the reporting of these (independent) variables – while the distribution of in-car times suggests that a considerable degree of rounding has occurred in their reporting – again there is a grouping effect at ten-minute intervals.

These results support our suggestion in Chapter 6 that, when the data was analysed using the multivariate-normal model, the substantial variances obtained in the distributions of the generalized cost weights corresponding to waiting time and in-car time could be the result of an attempt by the analysis to compensate for these rounding errors – very low variances were obtained in the distributions of walking time and in-bus time. The question of whether such rounding errors introduce a systematic bias into the estimates of the *means* of the generalized cost weights has been beyond the scope of this study.

Finally, Figure 4 presents two histograms showing the distribution of the cost difference between modes for those who actually choose the bus and for those who choose the car.

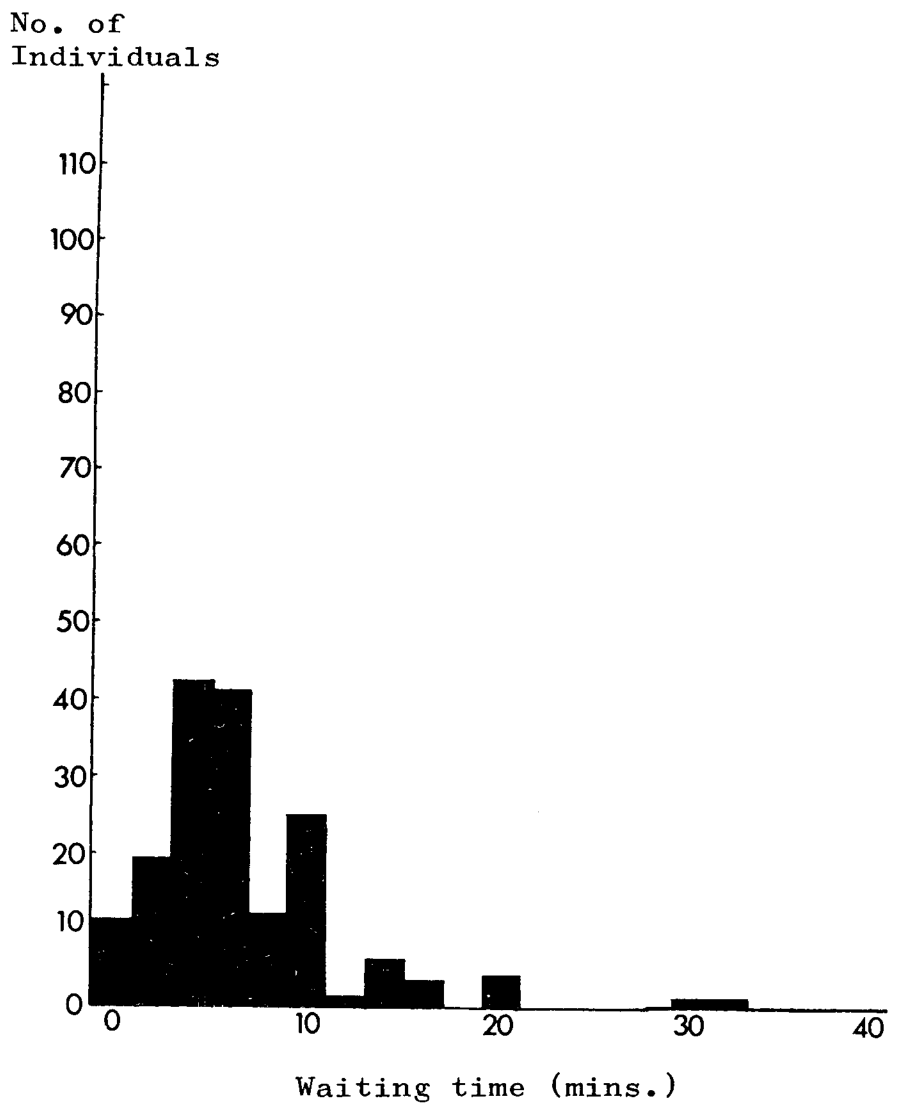
FIGURE 2 Leeds Data - Distribution Of Total Time Difference - Bus Users And Car Users



No. of individuals within range of histogram = 531 (from total of 542)

FIGURE 3 Leeds Data - Distribution Of Bus Waiting Time

(i) Bus Users



(ii) Car Users

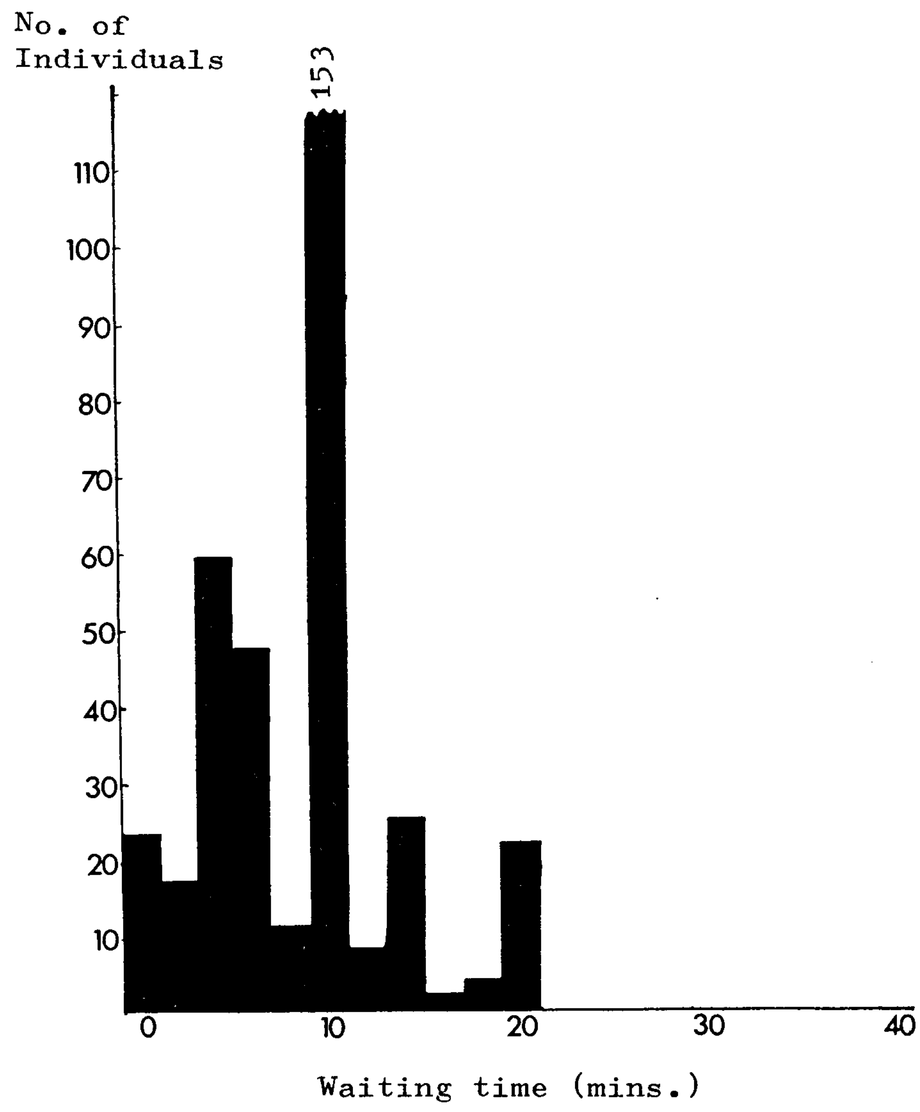
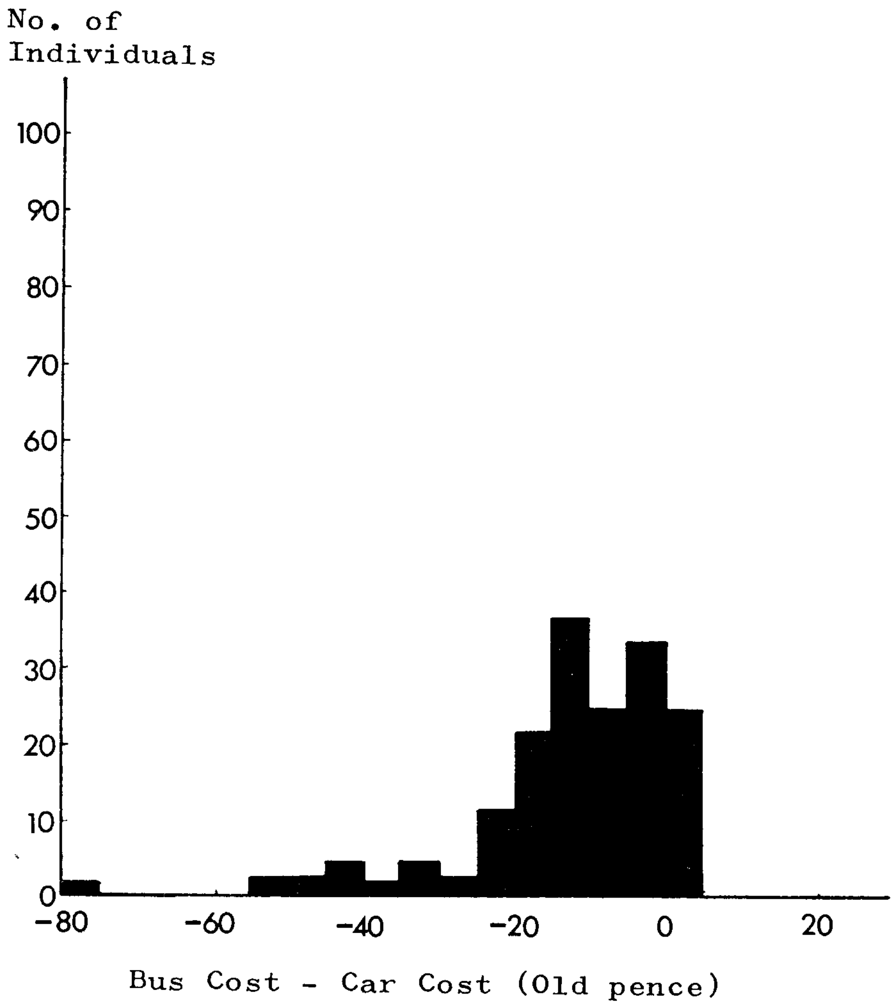
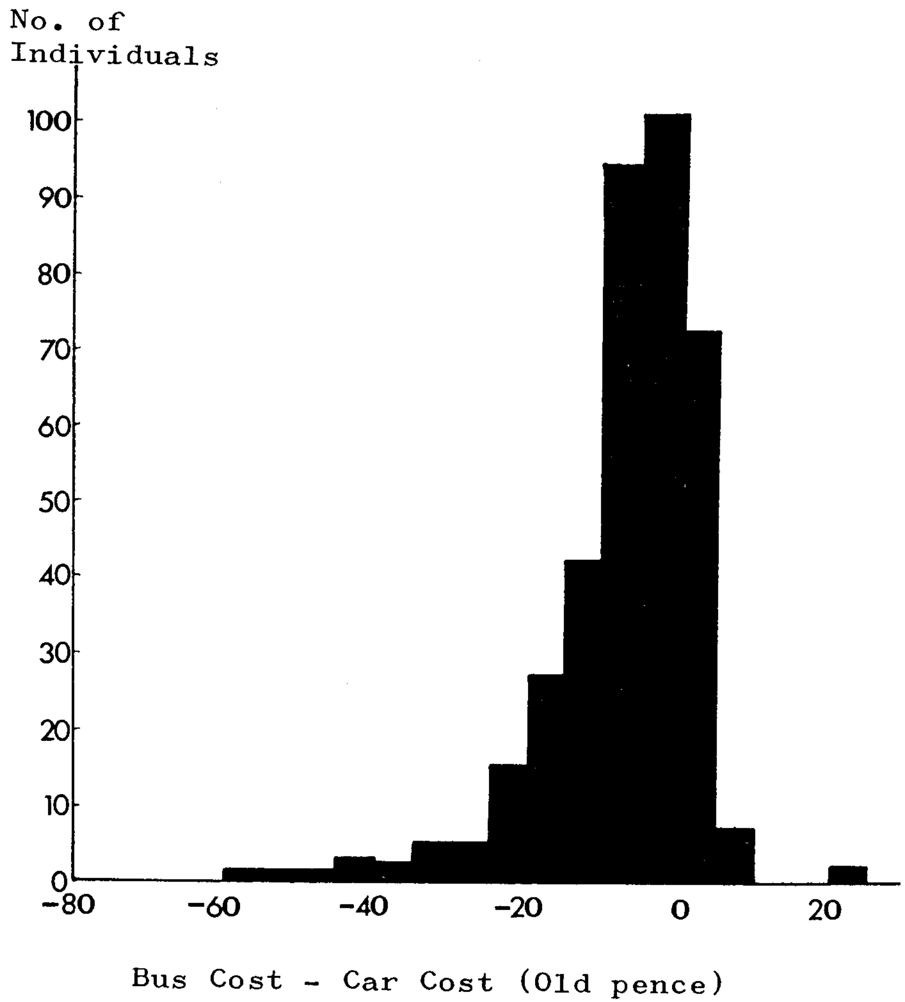


FIGURE 4 Leeds Data - Distribution Of (Money) Cost Difference

(i) Bus Users



(ii) Car Users



APPENDIX 6

SMALL TIME SAVINGS

It is frequently suggested that small savings in the journey time of one mode relative to another cannot be readily perceived, and therefore that the value placed on *unit* time in such circumstances is lower than is generally the case. This idea was originally envisaged for differences in the *total* time taken between modes. In view of our argument that the different components of journey time – walking, waiting, in-bus, in-car, etc., – are each separately and normally differently valued (and of the practical confirmation of this result from the Leeds data set), the idea that total time is what matters is no longer tenable. If, for example, an individual normally pays twice as much to save time travelling by bus as he pays to save time travelling by car, then in the situation where he is faced with a choice of a bus journey or a car journey of practically the same duration, he will not suddenly, because the differences between the two times are negligible, be equally happy to spend his time travelling by either mode. This is no more than the reiteration of the argument in section 1.2 that one pays not so much to save time as to save time in a particular activity, and therefore that the valuation of the difference in times spent in *different* activities is not a meaningful concept.

If, therefore, the idea that small time savings have a lower proportional valuation is to be seriously investigated we must look at small time differences between travel choices for each separate activity involved. This raises considerable difficulties in the situation where, for example, the only in-vehicle component of one choice is travelling by bus and the only in-vehicle component of the other is travelling by car. In such a situation we simply have no small time savings to consider for either in-bus or in-car travelling. All we may look at are small time savings for waiting time and walking time and even this can normally be done only when the in-vehicle components of both choices are public transport.

Assuming then that data is available in which the information necessary for the investigation of hypotheses about small time savings exists in theory there remains the practical problem of analysis. The only possible first approach is the stratification of the data by the size of the difference between modes in the relevant component of time. Separate analyses may then be carried out for each stratum, and some appropriate test devised to determine whether or not the resulting variations in the value of time (for that component) could have occurred merely by chance. (The construction of a formal test would be difficult, however, and an examination of the confidence limits in each stratum might well be sufficient.) If a significant variation were found more formal experiments might then be devised seeking the exact nature of the relation between the duration of the time and its valuation.

We have already indicated that the confidence limits for the values of time inherent in the data sets currently available are so wide that no accurate estimation is possible. Confidence limits derived from subsets of these sets would be even wider, and therefore inference about small time savings is currently quite impossible.

IMPLICATIONS FOR MODAL CHOICE MODELLING

The main objective of this study has been the derivation of methodology for estimating values of time. Nevertheless, this has been achieved by modelling travellers' mode choice, and the study has yielded some interesting insights into modal choice modelling itself. Two of these insights are presented in this appendix.

The models discussed in this report have represented the traveller as choosing the mode that minimizes his generalized cost. The generalized cost of a mode has been represented by:

$$\gamma_{im} = \underline{\alpha}_i' \underline{x}_{im} \qquad (1)$$

where γ_{im} is the generalized cost to individual i of mode m;
$\underline{x}_{im}$ is a vector of times and costs of mode m if used by individual i;
$\underline{\alpha}_i$ is a vector of weights attached to the times and costs by individual i.

Two specific problems arise when we seek to apply these models in a general transport planning context. These are first, that in this context data will usually be available only in aggregate form and, second, that we often wish to predict the behaviour of travellers confronted by more than two choices. We look at each of these problems in turn and explain how the methodology described in the main part of the report can be used to throw light on them.

1. Application to Aggregate Data

The models considered in this study of the form (1) have allowed that the variable $\underline{\alpha}$ should be distributed over different individuals. We have supposed, however, that the variables $\underline{x}_m$ are known and fixed for each individual. When we come to predict the behaviour of a group by aggregating the behaviour of individuals it is necessary to acknowledge that the $\underline{x}_m$ variables generated by transportation planning models are at best the mean values for that group.

The distribution of each $\underline{x}_m$ within a group presents considerable problems, as the component variables are of different kinds and subject to different influences. For example, bus waiting times have skewed distributions, and the distributions of walking times to bus stops will depend on the distribution of houses and workplaces relative to the routes.

To simplify the problem, however, it is not unreasonable to assume: first, that the distribution of each γ_m over individuals in a group is normal (appealing to the central limit theorem); and second, that its mean is $g_m = \underline{a}' \overline{\underline{x}}_m$ where $\underline{a}$ is the mean of $\underline{\alpha}$ and $\overline{\underline{x}}_m$ is the mean for the group of $\underline{x}_m$. Then for any individual in the group the probability of choosing one of two alternative modes is given by

$$\Pr\left\{\text{mode 1 chosen}\right\} = \Pr(\gamma_1 - \gamma_2 < 0) = N\left(\frac{g_2 - g_1}{\text{var}(\gamma_1 - \gamma_2)^{1/2}}\right) \qquad (2)$$

where N is the cumulative form of the normal distribution with mean 0 and variance 1.

As in the main study we can use the logit approximation to the normal distribution. This approximation gives us

$$\Pr\{\text{mode 1 chosen}\} = \emptyset\left(\frac{\pi}{\sqrt{3}} \quad \frac{g_2 - g_1}{\text{var}(\gamma_1 - \gamma_2)^{1/2}}\right) \tag{3}$$

This is, of course, a generalization of the model used in the main study, where we wrote $\underline{a}'\underline{x}$ for $(g_2 - g_1)$, and derived $\text{var}(\gamma_1 - \gamma_2)$ to be $\underline{x}'\Sigma\underline{x}$, where $(\underline{a},\Sigma)$ are the parameters of the multivariate-normal distribution of the generalized cost weights $\underline{\alpha}$.

The remaining analytical problem of aggregation, therefore, is that of estimating the variance of the generalized cost difference between the two modes. In the main study this variance arose solely because of different weights $\underline{\alpha}$ attached to travel times and costs by different travellers, but here, in considering aggregate data, we have the further variances of the times and costs themselves. The variance of the generalized cost difference may be computed from a full statement of all the distributions involved. Such analysis is outside the terms of reference of this study, and is likely to prove awkward.[1]

* * * *

The probability statement (3) covers only the situation of choice between two modes. We now turn to the problem of multiple choice between, for example, several modes. We show that the obvious extension of the logit function requires additional and implausible assumptions not consistent with the utility-maximizing approach, which we argued in Chapter 2 to be the only model that is both economically and behaviourally sound.

2. Extension to More Than Two Choices

In extending our modelling to cover situations where the traveller has a range of choices, we must be careful to distinguish between a 2-choice situation, although each choice may involve more than one mode, and a true multiple choice situation, where the traveller selects one of a number of exclusive choices. In a previous LGORU study (reference 17) it was shown that in a multi-mode situation 'trips' could be defined, each possibly made up of several modes, and that ordinary 2-mode choice models could reasonably be used to predict choices between two such trips. On this basis the models already proposed in this study could be applied in a multi-modal context. We now go on to consider choice between a number of possibilities, which for convenience we shall call 'modes', but which can be generalized to cover many combinations of modes, sub-modes, routes or even destinations.

Most researchers studying these multiple choice situations have used the n-dimensional logit model:

$$p_{m'} = \frac{\exp(-g_{m'})}{\sum_{m=1}^{n} \exp(-g_m)} \tag{4}$$

---

1 A possible simplifying assumption would be to take constant values of $\alpha_2, \ldots\ldots, \alpha_r$ over the population and the distribution of $\underline{x}_1$ to be independent of that of $\underline{x}_2$. Then, with the inclusion in γ_2 of an independently distributed bias α_1 $(x_{11} \equiv 0, x_{21} \equiv 1)$, we would have

$$\text{var}(\gamma_1 - \gamma_2) = \text{var}\,\alpha_1 + \underline{a}'(\Sigma_1 + \Sigma_2)\underline{a}$$

where $\underline{a}$ is the mean of $\underline{\alpha}$ and Σ_m is the covariance matrix of $\underline{x}_m$.

giving the probability $p_{m'}$ of choosing mode m', where g_m is a measure of the generalized cost of mode m. The model used in the main study is the particular case of (4) with n = 2, generalized to take account of the different variances in generalized cost for individuals in different situations.

The statement (4) is commonly derived by appealing to Luce's Axiom (the independence of irrelevant alternatives). This is taken as implying for modal choice that the ratio of the probabilities of an individual choosing a given pair of modes is independent of the cost distribution of every other mode. Predictions derived from (4) will automatically satisfy this condition.

Throughout this study we have insisted that the only valid models of modal choice are those that are consistent with every individual maximizing his perceived utility. Hence any statement such as (4) must be capable of being derived from a utility-maximizing formulation of choice of the form:

$$\text{mode} = m' : \gamma_{m'} \leqslant \gamma_m \quad \text{for all } m \qquad (5)$$

where γ_m is the generalized cost of mode m as perceived by the individual. At least one of the conditions that must be satisfied by the class of distributions of γ_m that lead to the probability statement (4) is so implausible as to cast doubts on the validity of (4) itself.[1]

It can be shown that for (4) to follow from a utility maximizing model, it is necessary that the variance of the generalized cost difference between two modes is constant for all pairs of modes. (This would follow if, as in the Weibull example (1), the generalized costs were independent with equal variance.) Stated analytically, the condition is:

$$\text{var}\,(\gamma_m - \gamma_{m'}) = k \qquad m \neq m' \qquad (7)$$

This requires that when a traveller considers two modes (or sub-modes, routes, etc.), the variance between their generalized costs introduced by his individual valuation is independent of the modes considered. This condition is intuitively implausible, and is in fact incompatible with some of the results suggested in the main part of this study.

For example, to take a case of three modes: bus, train and car. It would seem more reasonable to suppose that the bus and train, being public transport modes, would be viewed as being more directly comparable to each other than to the car. That is, we suggest that in general the distinction between the public modes is more exact than that between public and private, where the decision may be more strongly influenced by factors other than times and costs.

To be more specific, in the individual choice model investigated in the main part of the study, we have

$$\text{var}\,(g_1 - g_2) = (\underline{x}_1' - \underline{x}_2')\ \Sigma\ (\underline{x}_1 - \underline{x}_2) \qquad (8)$$

where $\underline{x}_1, \underline{x}_2$ are the vectors of times and costs for the modes for the individual, and Σ is the covariance matrix of the time and cost weights. From this equation (8) we can see that the more similar the time and cost vectors for a given pair of modes, the

---

1 We do not know at present how many distributions this class contains. There is at least one, viz: if we take the distributions of γ_m to be the independent Weibull distributions:

$$\Pr \left\{ \gamma < x \right\} = \exp\left(-e^{g-x}\right) \qquad (6)$$

then it can be shown that the probability of $\gamma_{m'}$ being the smallest is indeed given precisely by (4). We do not know whether any other distributions have this precise property. It is certainly true however, that many distributions of γ_m can lead approximately to (4).

less will be the variance of the generalized cost difference. Comparing bus and train, both involving walking and waiting, the variance will be less than comparing either mode with car, which usually involves little walking and no waiting. This argument reinforces the intuitive reasoning of the previous paragraph and leads us to regard condition (7) as implausible.

The extreme example of the failure of condition (7) is the notorious red bus/blue bus anomaly.[1] In this case we have the cost difference between two of the modes precisely zero, *with variance also zero,* and the application of the logit model (4) or other simple 'market shares' models will not work.

In the light of this reasoning, we reject the logit model as a *general* predictor of choice between more than two modes.

However, as we stated above, the logit model is usually derived by appealing to Luce's Axiom. This is commonly interpreted as meaning that the ratio of the probabilities of an individual choosing a given pair of modes is independent of the cost distribution of all other modes – i.e. the ratio of these probabilities is the same as would be given by a 2-mode choice model. With this interpretation, the arguments advanced in the main study for the 2-mode logit model would also lead us to the multiple logit model (4).

But, as we have seen, the use of the multiple logit for modal choice is implausible, and we are therefore led to question the application of Luce's Axiom. Implicit in our arguments against the multiple logit model was the *relevance* of other modes to the choice between a given pair. We now note that the Axiom merely requires independence of *irrelevant* alternatives, and that it therefore *does not apply to mode choice modelling,* where the alternatives are in general highly relevant.

To sum up, therefore, we are recommending the extension of the utility-maximizing approach, which we have applied to choice from two alternatives, to multiple choice. Although this approach leads naturally to the logit formulation in the case of choice from two alternatives, it does not, as we have seen, lead to the simple generalization (4) in the case of multiple choice. Instead, it is necessary to hypothesize realistic generalized cost distributions and derive from them the probability that each mode has the least cost. The distributional parameters may then be calibrated by appropriate statistical methods, analogous to those used in the 2-mode case.

---

1 The anomaly supposes that there exist a red bus mode and a car mode with equal generalized costs and 50:50 modal split.The ingenious bus operator paints half his buses blue, thus creating three modes with equal generalized costs, predicting 33:33:33 modal split, and 33 per cent rise in his revenues.