**Monograph:**
Gunn, F.H. (1987) Value of Time Estimation. Working Paper. Institute of Transport Studies, University of Leeds , Leeds, UK.

Working Paper 157

**Published paper**
Gunn, F.H. (1987) *Value of Time Estimation.* Institute of Transport Studies, University of Leeds, Working Paper 157

**UNIVERSITY OF LEEDS**
**Institute for Transport Studies**

*ITS Working Paper 157*

1987

# Value of Time Estimation

**F H GUNN**

# ABSTRACT

GUNN, Hugh F. (1981) Value of Time Estimation.
Leeds: Univ. Leeds, Inst. Transp. Stud., Work. Pap. 157

The statistical aspects of the procedures by which values are placed on savings in travel time, on the basis of stated or revealed preference data, are discussed and analysed. Conclusions are drawn for the design of such experiments.

This work was undertaken in the course of a larger project on value-of-time estimation commissioned by the Department of Transport.

CONTENTS

# 1. THE STATISTICAL PROBLEM.

## 1.1 Introduction

1.1.1 There are two principal approaches to the valuation of travel time savings:

a) by the analysis of the outcomes of choices made between numbers of options with differing travel time and cost characteristics, and

b) by the analysis of direct estimates of the difference in attractiveness of pairs of options which differ in respect of travel time and cost characteristics.

When the first approach is based on observed behaviour, i.e. real options and actual choices as revealed by subsequent actions, it is usually termed a 'revealed preference' method. When the choices are hypothetical in the sense of not committing the chooser to any action, it is usually termed a 'stated preference' approach. The second approach had been called 'transfer pricing', although in principle the measure of difference in attractiveness could be sought in terms of travel time.

We shall not discuss the way in which 'transfer price' estimates are obtained in any detail; put at its simplest, travellers are invited to consider changes in the cost and time attributes of options, and to indicate the amount of variation that would be needed to make the options equally attractive to them.

1.1.2. The analyses require that the link between behaviour and the chosen set of explanatory variables, or the link between the estimates of the difference in attractiveness and the set of explanatory variables, be made explicit in parameterised model. Theory suggests only general forms for such models; the final choice of a particular form and of a particular set of factors by which to characterise the options must be resolved by empirical means, which is to say by the data themselves.

The limitations of the accuracy of the model used in the analysis together with the amount of data available determine the accuracy with which the parameters in the models can be determined and thus the accuracy with which values can be ascribed to travel time savings.

1.1.3 The statistical aspects of the problem can be listed under five headings:

1. How should we draw our sample ?
2. How large must the sample be ?
3. How should we estimate the parameters in any model ?
4. How should we choose between rival models ?
5. How can we validate our preferred model ?

These five issues will arise in any given survey context, and indeed it will be shown that the accuracy with which we can estimate model parameters, for given sample size and survey method, varies from context to context, so that the choice of experimental context itself should be made with reference to the basic statistical problem.

## 1.2 Structure of the Working Paper

1.2.1 Throughout most of this paper, we shall assume ourselves in the position of considering the collection of disaggregate data sets for value-of-time estimation. The insights gained on the issue of sample size will then assist in the scrutiny of existing data sets, and of course the conclusions reached on model selection, estimation and validation apply equally to such data. In principle, the results also apply to the analysis of aggregate data sets such as conventional mode split or distribution data, where these can be interpreted as the outcome of a discrete choice process. Some approaches to value-of-time estimation are based on analyses of overall travel expenditure, or the variations in demand for a particular mode as a result of changes in journey times and changes in costs. The statistical problems associated with the analyses of such data are of a more conventional nature, and are also described briefly.

1.2.2. In contrasting the statistical properties of estimates of model co-efficients based on transfer-price measures of the size of the utility difference with those based only on the sign of the utility difference, we shall talk of 'maximal' accuracy under fairly strong hypotheses about the accuracy of the transfer price. In practice, the degree of success of any transfer price study must depend crucially on the skill with which the transfer price question is posed, as well as the suitability of the context for such an approach. These problems present issues which can only be tackled by empirical research.

1.2.3. Finally, we emphasize at the outset that this paper sets out a theoretical analysis of the statistical aspects of value of time estimation. The methods outlined below, and the formulae given for simple models, can be elaborated to address specific models and contexts once certain key facts are made available. Such information should be acquired during the course of a piloting exercise. In advance of this information, we provide rough guidelines wherever possible, based on past studies.

# 2. BASIC ASSUMPTIONS OF THE MODELS

## 2.1 Introduction

2.1.1 The concept of 'random utility' allows us to progress from the unfalsifiable and uninformative assertion that behaviour can be described in terms of utility maximisation to the stage of postulating concrete model forms to describe and predict behaviour and to establish a rate at which time saving can be substituted for cost saving to maintain the same level of satisfaction (the 'compensated marginal value of travel time saving' as defined by Bruzelius, 1979). The device of specifying the utility function only up to a random error term with unknown variance not only allows us to proceed with our (inevitably) approximate models of behaviour (as Daly, 1980, remarks) but also allows us to measure the relative importance of the factors omitted from the model specification in any particular context, by estimating that variance.

2.1.2 Specification of the 'representative' utility function and specification of the random error term, defines a complete model which can then be manipulated to yield both a probability density function for the difference between the utilities of any two options (and hence a distribution for the corresponding Transfer Price estimate, were we to equate that with the utility difference) and a corresponding expression for the probability that a particular one of the options has greater utility (and hence would be chosen by the 'rational decision maker') than any other. Both the p.d.f. for the transfer price and the probability that a particular option is chosen are defined by the 'complete model', the specified representative utility expression and the specified error term. Both are functions of the (initially unknown) parameters in both specifications. Standard statistical techniques can apply to either transfer price data or to observed outcomes of choices; in either case the analysis can be made consistent with the same underlying model.

## 2.2. Theory

2.2.1 Following the now classical account of the theory underlying discrete choice (see for example in Williams, 1980) we can describe our analysis of the preferences indicated by particular individuals over a fixed number of options, N say, as being based on the following postulates.

1) An individual drawn at random from the population, with particular observed characteristics, constraints and facing a particular set of options, is assumed to be drawn from a subpopulation of individuals with identical observed characteristics, constraints and options.

2) Each of these individuals is assumed to associate a net utility with each option, $U_i$, $i=1,\ldots,N$ , and to select that option with the highest value of U.

3) Individuals within the subpopulation with identical observed characteristics, etc., are assumed to vary in respect of some unobserved characteristics, in such a way that the net utilities $U_1,\ldots,U_N$ each vary randomly across the subpopulation; this variation can be described by a joint density function, $f(U_1,\ldots U_N)$ say. Drawing an individual at random from the subpopulation results in observing preferences generated by a vector of net utilities drawn at random from this joint distribution.

## 2.3 Models

2.3.1 We then postulate that the nature of the variation of each $U_i$ across the subpopulation of individuals with identical characteristics can be represented in the form

$$\begin{aligned} U_1 &= \overline{U}_1 \quad (\underline{\theta},\underline{Z}^1) + \varepsilon_1 \\ \vdots \quad & \quad \vdots \qquad\quad \vdots \qquad\quad \vdots \\ U_N &= \overline{U}_N \quad (\underline{\theta},\underline{Z}^N) + \varepsilon_N \end{aligned} \tag{1}$$

where $\overline{U}_i$, the 'representative utility', is fixed for all members of the subpopulation, and is a function of observed characteristics $\underline{Z}^i$ describing the option and the subpopulation, a number of unknown parameters $\underline{\theta}$. Here, $\underline{\varepsilon}$

is a vector of error terms drawn from a particular distribution G say, which itself contains unknown parameters, $\underline{\phi}$ say, and may also be a function of the observed characteristics $\underline{Z}^1; .. , \underline{Z}^N$.
We can thus write the distribution function of the disturbance terms as $G(\underline{\varepsilon}, \underline{\phi}, \underline{Z}^1, .., \underline{Z}^N)$.
The most popular of the models that can be generated by specific assumptions about the form of $\overline{U}$ and the form of G are described in Gunn et al (1980).

## 2.4 Data

2.4.1 Revealed preference data sets then consist of the vectors of observed characteristics for each option available, together with an indication of which option was selected. Stated preference data sets can also include a ranking of preferences extending over all or part of the set of options. For transfer pricing data sets, the data refers to comparisons of pairs of options : for transfer-price studies (such as those reported by Hensher, 1976, and Lee and Dalvi, 1969 and 1971) only the comparison between the option actually selected and the next-best of the available options are compared, although there appears to be no reason (other than decreasing credibility of the data) why comparisons should not be made between all possible pairs of options. For each pair, an estimate of the utility difference is collected, together with the two vectors of observed characteristics.

2.4.2 For the purposes of illustration, it is convenient to consider a simple case which can be presented graphically. Suppose we had a population of individuals with identical observed characteristics, choosing between two options each of which was characterised by only two observed dimensions. We can consider this as a highly simplified representation of the choice between two very similar modes of travel, differing only in respect of time and cost characteristics. For the two modes, let us make the usual distinction between the positive utility to be gained at the end of the trip and the disutility incurred during travel itself, and write the net utility expressions for individual j as

$$U_1^j = U^o - \theta_1 Z_1^{j1} - \theta_2 Z_2^{j1} + \eta_1^j$$
$$U_2^j = U^o - \theta_1 Z_2^{j2} - \theta_2 Z_2^{j2} + \eta_2^j \qquad \ldots (2)$$

where the terms $\eta_k^j$ are disturbance terms of a magnitude and sign usually unknown to the modeller, and about which we would usually only hypothesise that they were drawn from an underlying distribution whose general form could be specified, having unit variance.

The net utility difference between the options is then eq. 3

$$(U_1^j - U_2^j) = -\theta_1 (Z_1^{j1} - Z_1^{j2}) - \theta_2 (Z_2^{j1} - Z_2^{j2}) + (\eta_1^j - \eta_2^j) \qquad \ldots (3)$$

and as usual, we would assume that mode 1 would be selected if $U_1$ were larger than $U_2$, and $U_2$ only selected if $U_1$ were smaller than $U_2$. (We shall ignore the possibility of equality!)

2.4.3. For each individual, we can plot a position on a (Beesley-) graph with axes $(Z_1^{j1} - Z_1^{j2})$ and $(Z_2^{j1} - Z_2^{j2})$ corresponding to the net difference in observed characteristics in the options confronting the individual. Let us assume that $(Z_1^{j1} - Z_1^{j2})$ represents a difference in journey times and denote that axis by $\Delta T$. Similarly, let $(Z_2^{j1} - Z_2^{j2})$ denote difference in costs, and denote that axis by $\Delta C$.
Working Paper 6 has described the essential indeterminacy in the unit system appropriate for such expressions, and we have seen that, providing we take care to make consistent adjustments throughout, we can work in any unit we please. For example, if we choose to set $\theta_1$ to unity, leaving $\theta_2$ as a parameter to be estimated, we must also acknowledge the need to estimate the scale parameter in the distribution function for the disturbance term. (The more usual approach is to standardise that scale parameter to unity and express the problem as one of estimating both $\theta_1$ and $\theta_2$: in practice, of course, this amounts to exactly the same thing). To illustrate the relationship between the transfer price and revealed preference approach, it is useful to work throughout in money units, so let us rewrite eq 3 as

$$\Delta U^j = (U_1^j - U_2^j) = -\Delta C^j - \theta_3 \Delta T^j + (\varepsilon_1^j - \varepsilon_2^j) \qquad (4)$$

Where $\theta_3 = \theta_2/\theta_1$ and $\text{var}\,\varepsilon = 1/\theta_1^2$ In this form var $\varepsilon$ must now be estimated.

For one particular individual, suppose the centre of the circle on figure 1 denotes the point corresponding to the difference in characteristics of the options which confront him, and let L denote the distance of the centre of the circle from the line $(\Delta C) = -\theta_3(\Delta T)$, measured parallel to the $(\Delta C)$ axis as shown. Since we have hypothesised that our individual is 'rational', he will choose to take mode 2 if and only if the net utility he will gain is greater than that arising from a choice of mode 1. This will only happen if the net value of the unobserved factors in the utility expressions, $(\varepsilon_1^j - \varepsilon_2^j)$, is less than $-L$.

FIGURE 1

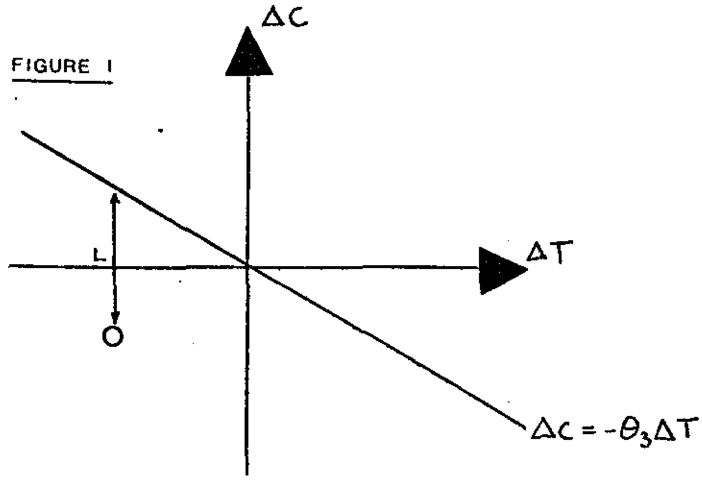



Figure 2a illustrates the sort of pattern we might observe in practice, denoting the choice of mode 1 by a hollow circle and the choice of mode 2 by a shaded circle. As a result of the presence of the unobservable factors, some individuals choose modes which are apparently inferior in their net time-and-cost characteristics, and indeed some choose modes which are apparently inferior in each of time-and-cost characteristics.

FIGURE 2a

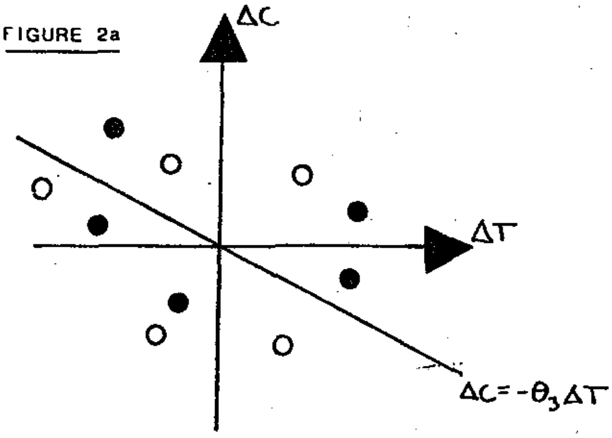
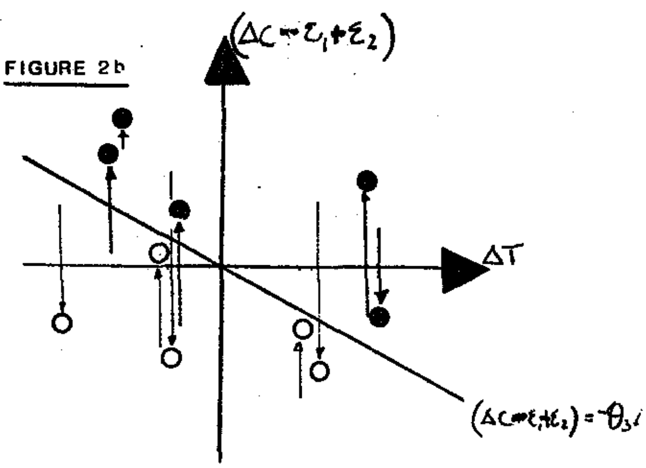



FIGURE 2b



Now let us suppose that the net effect of the unobservables $(\varepsilon_1 - \varepsilon_2)$ could be determined that we knew $\theta_3$, and that we could replot each individual on a graph with axes $\Delta T$ and $(\Delta C - \varepsilon_1 + \varepsilon_2)$. Figure 2b illustrates the expected result: every individual is now seen to be making a rational choice.

2.4.4 The transfer price questions described in Working Paper 6 are intended to discover the net utility difference between the options for each individual; in figure 2b, this would correspond to the distance from the 'location' of each individual to the line $(\Delta C - \varepsilon_1 + \varepsilon_2) = -\theta_3 \Delta T$ measured parallel to the $(\Delta C - \varepsilon_1 + \varepsilon_2)$ axis. It is easy to see that IF we did have a measure of $(\varepsilon_1 - \varepsilon_2)$, AND the transfer price data did truly represent the net difference in utility between the options, we could replot the points corresponding to each individual on a graph with axes $\Delta T$ and $(\Delta C - \varepsilon_1 + \varepsilon_2 - TP)$, and that we would then find that all the points lay on the line $(\Delta C - \varepsilon_1 + \varepsilon_2 - TP) = -\theta_3 \Delta T$, as illustrated in figure 3a.

FIGURE 3a

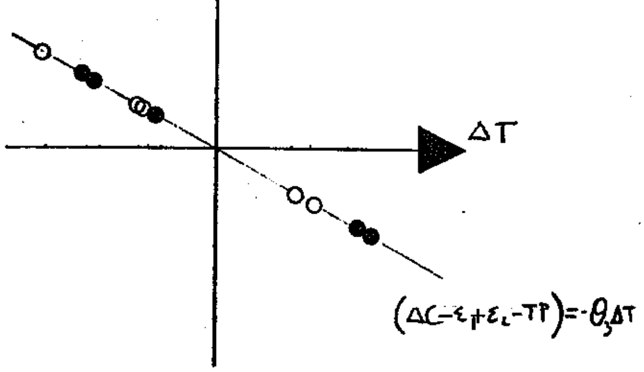
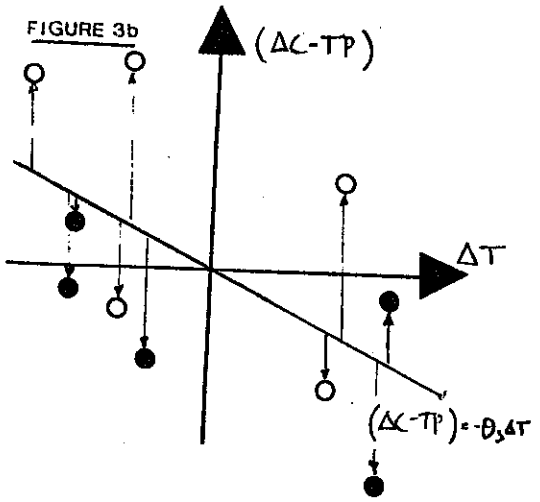



FIGURE 3b




In practice, we only know $\Delta T$ and $\Delta C - TP$ Plotting individuals' locations on these axes produces a set of points scattered about the $(\Delta C - TP) = -\theta_3 \Delta T$ line, as on Figure 3b, where the displacement from the line is just that which distinguishes figure 2b from figure 2a, i.e. the net effect of the unobservables.

The estimation of the slope of the line then becomes a matter for statistical resolution, given the error distributions of these unobservables. For example if they were all normally distributed with constant variance, we would fit the familiar 'least squares' regression line.

2.4.5 In practice, of course, it is desirable to allow for the possibility that the transfer price data does not give an exact measure of the utility difference between the options, but is itself subject to certain errors. This possibility is discussed by Daly (1978) who demonstrates that simple solutions are available for conveniently chosen error distributions. Daly* has also noted that an elegant distinction can be made between the distributions of transfer prices in the context of choices actually made and those for hypothetical options and contexts, in that the former can be restricted to take only positive values. Thus, at the expense of some extra complexity, we can ensure that the transfer price question can only <u>add</u> to our information if we actually know the outcome of the choice.

* private communication.

## 3. ESTIMATION

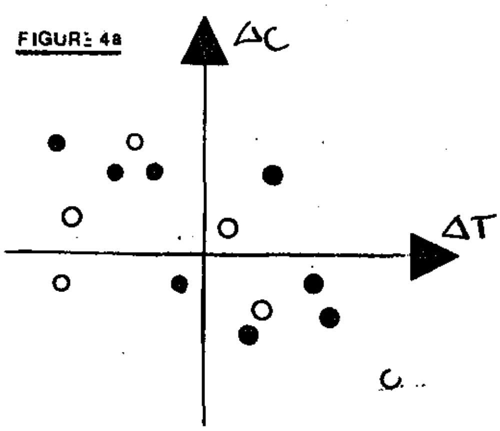
### 3.1 Introduction

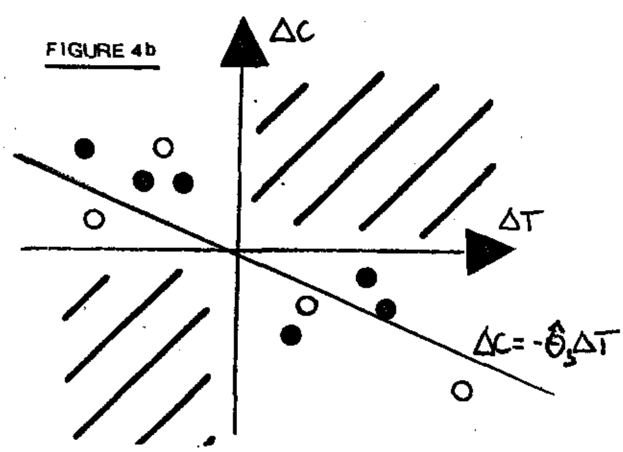
3.1.1 The relative advantages of the various methods by which probabilistic choice models can be estimated are by now well known (see for example Stopher and Meyburg, 1976) and the analysis of transfer price data calls only for the straightforward application of regression methods (Daly 1978). We shall not discuss these here.

3.1.2 Instead, in recognition of the historical importance of the 'Beesleygraph' approach in connection with the analysis of binary choice data for value of time measurement, the interest there is in the connection between this method and the more recent probabilistic choice analyses, and the importance of the related issues of use of data, we shall use this section to pursue the graphical illustration of the previous section in demonstration of the essential differences between the methods.

### 3.2. Beesleygraph versus Probabilistic Choice Analysis.

3.2.1 The 'Beesleygraph' technique can be simply illustrated as follows; obtain a sample of outcomes of choices between two options differing in respect of time and cost characteristics, and plot these on net time cost, net money cost axes as on figure 4a, distinguishing as before between those points at which the outcome of the choice was that



mode 1 was selected and those points at which mode 2 was selected. The problem is then to find that straight line drawn through the origin which minimises the number of points at which the mode chosen is apparently inconsistent with rational behaviour in terms of time and cost alone.

The data points in the first and third quadrants are redundant for this analysis; were trading to occur on time and cost alone, mode 2 should always be chosen by those in the first quadrant and mode 1 by those in the third quadrant, (one mode being better than the other in 'all' respects in these areas).

3.2.2 Figure 4b illustrates the process with a line drawn which results in only two apparently 'inconsistent' observations; 'consistency' would require that all decisions characterised as points plotted above the line led to mode 2 being selected, since in that area we have $\Delta C > \theta_3 \Delta T$ or $(C_1-C_2) > \theta_3 (T_1-T_2)$ i.e. the value of the time saving offered by mode 2 outweighs its extra cost (above the line in quadrant 4) or the cost saving offered by mode 2 outweighs the extra time taken (above the line in quadrant 2). Similarly, all 'decision points' below the line should result in the selection of mode 1.

3.2.3 Now let us consider the corresponding analysis provided by, for example, logit analysis, using exactly the same model of net utility. The probabilistic choice analysis supplies, for every point on the decision plane, a probability that mode 1 would be selected (and of course the probability that mode 2 would be selected is thus defined at the same time). We can illustrate the end result by drawing a series of iso-probability-choose-mode-1 lines one the decision plane, as in figure 5. It can now be seen quite vividly that the data must supply an extra piece of information, for the model requires not only an orientation for the iso-probability lines, but a rate of change also

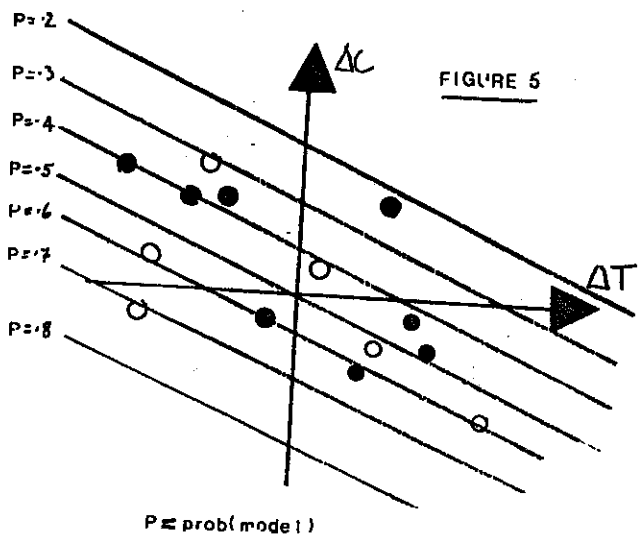



FIGURE 5

For example, in figures 6a and b we have iso-probability lines corresponding to differences in orientation(value-of-time) but not rate of change, and in figures 7a and b we show differences in rate-of-change for lines with the same orientation.

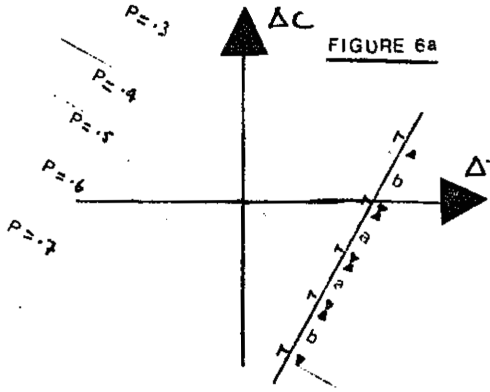



FIGURE 6a

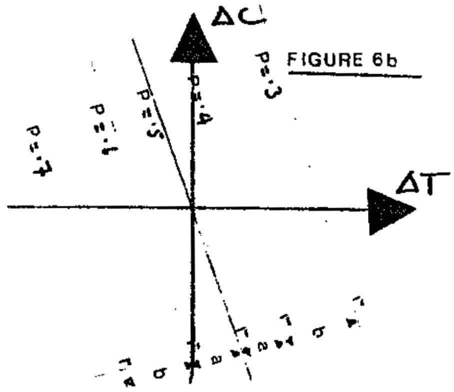



FIGURE 6b

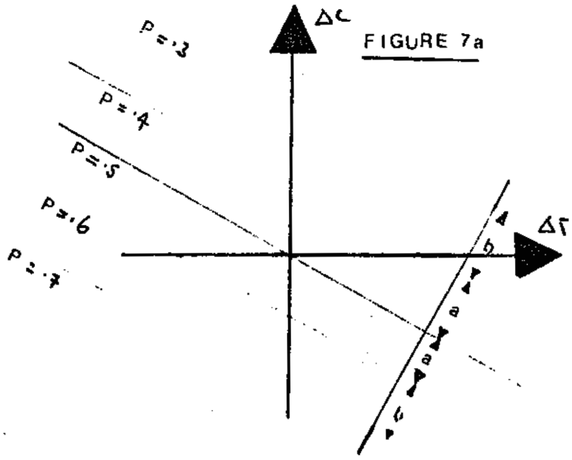



FIGURE 7a

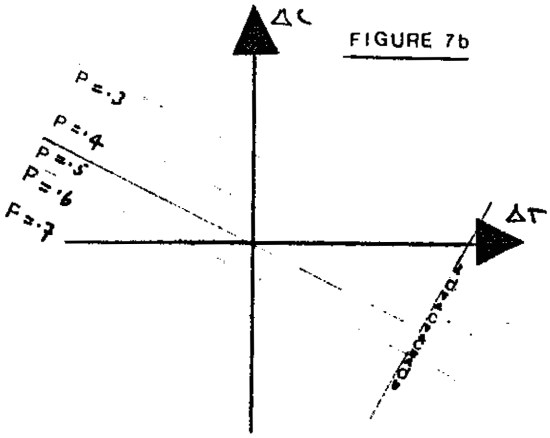



FIGURE 7b

3.2.4

We are free to choose a system of units for our utility expressions, provided that we remember that the dispersion of the random element must then be made parametric. The rate-of-change of the iso-probability lines is determined by just this dispersion. The requirement that the data determines this rate-of-change reflects our implicit choice of a particular unit system (money) in our example.

## 3.3. Conclusions

3.3.1 The example given above illustrates exactly why the probabilistic choice models are potentially more powerful in their use of data than the 'Beesleygraph' approach. The evidence for the orientation of the equiprobability lines is taken from ALL the data, regardless of its location on the plane. On the other hand, it is also clear that 'potential power' and 'lack of robustness' will go hand in hand, and that the probabilistic choice models will be more sensitive to miscoded data points, or sections of the data to which the model does not apply. Ignoring the evidence from observations in quadrants 1 and 3, as is inevitable with the 'Beesleygraph' approach as outlined and indeed as has been done in the past in specific transfer price experiments (see Lee and Dalvi, 1969) does result in the loss of information that could improve estimates of co-efficients in probabilistic choice models, or in transfer-price experiments, (at least if the model and data are both correct). However, such an omission should not bias the results, merely reduce their precision.

3.3.2 The 'Beesleygraph' approach illustrated here is a specific, highly simplified application of the 'Score Maximisation' technique developed by Manski (1975); the same principles can be used to extend the estimation procedure to more than two dimensions.

## 4. MODEL VALIDATION

### 4.1 Introduction

4.1.1 The general task of model appraisal can be considered under two headings. Firstly, there is the issue of the internal consistency of the complete model with the data from which it has been estimated. This aspect of appraisal includes the well-known tests of significance and examinations of residuals from standard statistical theory; for disaggregate choice models, the various tests that are commonly used are listed in Gunn et al (1980). To this list we would now add the Lagrangian Multiplier tests and the range of 'overfitting' tests described by Horowitz (1980). The second issue concerns the performance of the fitted models, and the description of behaviour and values that these embody, in the prediction of choice for data sets other than that from which the model has been estimated. This we shall call 'validation'.

4.1.2 In this section, we shall discuss the second of these issues in the context of one particular test described by Foerster (1979). In particular, we are interested in the question of the amount of data that is necessary to 'validate' a model. A separate question concerns the sort of data that should be used for validation. Most frequently, the validation data set is actually a randomly selected subsample of the estimation data set; certainly a validation procedure based on such a partitioning of the data will guard against some of the dangers of model misspecification. However, in many cases it will be clear from the purpose for which the model has been developed that there is a particular sort of context in which the model should be validated. For example, if the model is derived from data from one set of geographical areas for general application in other areas, it should be tested specifically for its performance in a sample of such other areas. Similarly, a forecasting model should be tested for its performance in other time periods.

4.1.3 The work reported in this section was undertaken to explore the issues in the context of a tractable example. It will be obvious that a completely general treatment of the problem is a task far beyond the scope of this project. On the other hand, the inference that can be made about 'values-of-time' from revealed preference data is all CONDITIONAL on the adequacy of the model used to represent behaviour. We should not underestimate the importance of establishing the adequacy of that representation.

## 4.2. The FPR Criterion for Model Validation and for Model Comparison Using Validation Data Sets

4.2.1 A disaggregate model specifies a set of probabilities attaching to each of a number of options available to an individual. The option associated with the maximum of these probabilties will be deemed the individual's 'first preference'. In application to a validation data set, the model may or may not indicate that the option actually selected was the 'first preference' for the individual. If it does, this is deemed to be a 'first preference recovery'.

Naturally we would not expect all individuals to select their 'maximum probability' option if the model were absolutely correct (unless of course the model was specifying probabilities of 1 for that option and thus 0 for all others). In general, the expected number of FPR's will depend on the actual sizes of the maximum probabilities, assuming a correctly specified model. This is discussed later; first we shall consider the comparison of two competing models.

4.2.2 Two different models may be compared in respect of their FPR's by a method described by Foerster (1979), due originally to McNemar and generalised by Cochran (1950) to apply to an arbitrary number of models or weighted averages of models. Only the simple case of two model comparisons will be considered here.

Consider a 2x2 table layout as shown in fig. 8; for each individual in the validation sample, a set of probabilities of choosing each option is calculated for each of the two models under investigation. The individual is assigned to one of the cells of the table according to the rules:
assign to cell (1,1) if the actual option chosen is not the 'maximum probability' option for either model;
assign to cell (1,2) if the actual option chosen is the max. prob. option for model 2 but not for model 1 (the numbering is of course arbitrary);

assign to cell (2,1) if the actual option chosen is the max. prob. option for model 1 but not for model 2;
assign to cell (2,2) if chosen option is the max. prob. option for each model.

FIGURE 8

| MODEL 1 \ MODEL 2 | no FPR | FPR |
|---|---|---|
| no FPR | $n_{11}$ | $n_{12}$ |
| FPR | $n_{21}$ | $n_{22}$ |

$n_{ij}$ = no. individuals assigned to cell $(i,j)$

$$e_{ij} = \frac{(n_{i1}+n_{i2})(n_{1j}+n_{2j})}{(n_{11}+n_{12}+n_{21}+n_{22})^2}$$

4.2.3 This sort of contingency table layout is most familiar in the context of a null hypothesis of independence of row and column classifications, which is tested with the $\chi^2$ distributed statistic $\chi^2 = \sum_{ij} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$ That hypothesis is not appropriate in this case, quite apart from being highly implausible for any sensible pair of models (which should be specifying broadly similar choice probabilities, thus concentrating the data in the (1,1) and (2,2) cells). Rather, we are interested in the null hypothesis that the probabilities with which individuals fall into the (2,1) and (1,2) cells are equal, for in that case the implication is that the two models are equivalent in terms of expected number of FPR's. We can test this hypothesis by considering the entries in the (1,2) and (2,1) cells alone. On the null hypothesis outlined, (after McNemar), the statistic Q,

$$Q = \frac{\left(n_{12} - \frac{1}{2}(n_{12}+n_{21})\right)^2}{\frac{1}{2}(n_{12}+n_{21})} + \frac{\left(n_{21} - \frac{1}{2}(n_{12}+n_{21})\right)^2}{\frac{1}{2}(n_{12}+n_{21})}$$

is $\chi^2$ distributed with 1 d.f. With some easy manipulation, we can show that $Q = \frac{(n_{12} - n_{21})^2}{(n_{12} + n_{21})}$

4.2.4 Thus, a test of the 'equivalence' of the two models, in terms of FPR's, is given by computing Q and comparing the result with $\chi^2_1$. If Q is not larger than the appropriate chosen critical value of $\chi^2_1$ (3.85 for the usual 95% confidence level) we conclude that the models are equivalent in these terms.

Cochran also gives a statistic 'corrected for continuity', $Q' = \frac{(n_{12} - n_{21} - 1)^2}{n_{12} + n_{21}}$ and demonstrates the correspondence of the general procedure with the simple 'sign' test.

Thus, given $n_{12}$ and $n_{21}$, we can simply consult tabulated values of the sign test (for example see Crow, Davis and Maxwell, Table 9) to test the hypothesis that the probabilities of an individual being assigned to the (1,2) cell, and to the (2,1) cell, are equal.

4.3 Some comments on the FPR Criterion

4.3.1 The FPR criterion has some intuitive appeal; it is easy to calculate and has an obvious sort of connection with model performance. However, it should be stressed that it is not in itself an unambiguous indicator of model reliability; too many FPRs should lead to rejecting the model as well as too few. This is discussed more fully later. A second point is that, even if the total numbers of FPRs are acceptable, a test which weights each correct prediction equally will not be suitable for circumstances where some options are more important than others. For example, a mode-split model might specify several modes but be particularly interesting in respect of its predictions of patronage of a minor mode such as car-pooling. We would not then judge two rival models equivalent even if they had exactly the same number of FPRs, if one model got the car-pooling patronage entirely 'wrong', and all other modes correspondingly slightly more 'right', than a rival model which performed adequately for all modes, including car-pooling.

4.3.2 The latter point is linked to the choice of sample for the validation exercise. This requires to be chosen randomly from the population being modelled, in order to allow the desired inference about the general suitability of the models in this population; however, if some options are more 'important' than others, it seems clear that this importance should be reflected somehow in the composition of the validation sample. This question is not explored here, but in passing it can be seen that the need for a validation process emphasises the importance of considerations of sample size and design, especially if both estimation and validation data is gathered at the same time. The accuracy of one stage may thereafter only be increased at the expense of the accuracy of the other.

4.4 Sample size for the comparison of models by the Q statistic

4.4.1 Given the procedure outlined above, based on the number of FPRs, we can choose whichever level of confidence seems appropriate for the assertion that the two models under comparison differ in respect of expected number of FPRs. We thus have control over the fraction of times that we will incorrectly assert a difference between similar models. As usual, the aim of selecting a particular sample size is to ensure a corresponding control over the proportion of times we will make the other sort of error, namely incorrectly concluding that there is no difference between different models.

4.4.2 The actual calculation of the probability of an error of the second kind depends on the exact difference between the models, which of course, will not be known at the outset. One way around this problem is available, if we are able to decide on a minimum difference that we should like to be able to detect. If we then calculate the sample size needed to reduce the chance of errors of the second kind to an acceptable level for models which differ by exactly this minimum amount, then we have ensured that there will be even less chance of such an error for discriminating between models which differ by more than the minimum of interest.

4.4.3 The actual size of the minimum difference that we should aim to detect will vary from application to application, and may in many cases be a matter for judgement rather than for hard and fast rules, although it may be possible to develop a decision-theoretic approach for problems in which the 'cost' of wrong predictions can be estimated. For the purpose of illustration, table 1 lists the probabilities of an error of the second kind corresponding to various sample sizes when the criteria Q and Q' are used to assess the statistical significance of the difference between FPRs of two models, for the particular case when Pr(1,2) = 0.05 and Pr(2,1) = 0.00. (Pr(i,j) denotes the probability that an individual drawn at random from the validation data set will be assigned to cell (i,j) in the table in fig.1.) This particular case corresponds to two models such that, on average, model 2 produces 5 extra FPRs per 100 individuals modelled as compared to model 1. For the purposes of the test, it does not matter whether this arises as a result of model 1 having 0% FPRs and model 2 10% FPR, or model 1 80% and model 2 90% FPRs.

4.4.4 For this case, $n_{21}$ will always be 0, so Q simply becomes $n_{12}$, and Q' becomes $\frac{(n_{12}-1)^2}{n_{12}}$. If we are ensuring 95% confidence that any difference we establish could not have arisen by chance from identical models, we will be comparing Q and Q' respectively with the appropriate value of $X_1^2$ , namely 3.85. Thus, if we consider the test based on Q for example, the probability of an error of the second kind, namely accepting the null hypothesis of no difference between the two models, is the probability of three or fewer individuals being classified to the (1,2) cell. For any given sample size, n say, the probability that r individuals will be assigned to the (1,2) cell is simply the Binomial probability $^nc_r\ p^r(1-p)^{n-r}$, where p denotes the probability of an individual chosen at random being assigned to the (1,2) cell. Given n, and taking p = 0.05, we can calculate the probabilities of 0, 1, 2 and 3 individuals being assigned, and sum these to give the total probability of accepting the null hypothesis, which is, in this case, an error of the second kind. The calculation for Q' is similar, except that we must also add the probabilities of exactly 4 and exactly 5 individuals in the (1,2) cell, since the null hypothesis is rejected only for $n_{12} \geq 6$.

4.4.5 Q' is a more 'conservative' statistic than Q, in the sense that it requires stronger proof of any difference between models. Correspondingly, it is more prone to make errors of the second kind, failing to detect differences when they do occur. It is clear from table 1 that the required validation sample size needs to be relatively quite large, given that the estimation data sets are usually only a few hundred data points, to allow us to discriminate between the two models under consideration with any degree of certainty.

Table 1. Probabilities of an error of the second kind for given sample size, and stated test, test size, and models as defined.

| Sample size | Pr(error II) Q | Q' |
|---|---|---|
| 50 | .76 | .96 |
| 100 | .26 | .62 |
| 150 | .05 | .24 |
| 200 | .01 | .06 |
| 250 | .00 | .01 |

The method outlined here can be extended to indicate required validation sample sizes for other levels of minimum difference, including cases where both (1,2) and (2,1) cells have non-zero probabilities.

4.5 Remarks about FPRs and 'validation'

Given a model M which specifies a preferred option for each of n individuals in a given data set, and supposing that the $i^{th}$ individual has $c_i$ options to choose among, and that the calculated (maximum) probability associated with his preferred option is $p_i$, we can easily derive the following:

a) the expected number of FPRs in the whole data set which would be returned by a random prediction of preferred options would be $N_r = \sum_{i=1}^{n} \frac{1}{c_i}$. The variance of $N_r$ is $\sum_{i=1}^{n} \frac{1}{c_i} (1- \frac{1}{c_i})$ (for individual i, a FPR is an independent random event occurring with probability $\frac{1}{c_i}$.)

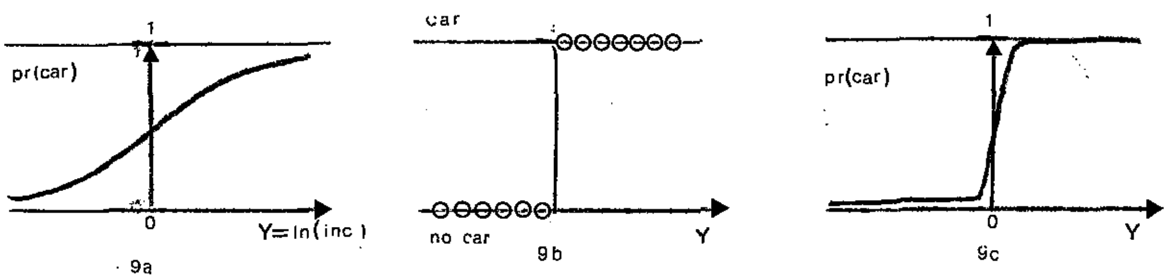
b) the expected number of FPRs from the specified model M is $N_s = \sum_{i=1}^{n} p_i$. The variance of $N_s$ is $\sum_{i=1}^{n} p_i (1-p_i)$ (in the same way as above).

Thus any actual out-turn total number of FPRs associated with a given model can be compared with $N_r$ and $N_s$; if all three are reasonably close (given the estimated standard errors) the model is reasonable but uninformative; if N and $N_s$ are similar and larger than $N_r$, the model is reasonable and informative; if N and $N_s$ are not similar, the model does not explain the variation in the data and should be rejected (whether N is larger than or smaller than $N_s$).

A simple illustration of this point may be made by considering a model of a binary choice made on the basis of a single variable (such as the conventional car-ownership/income models).

Figure 9a shows a hypothetical postulated model; fig. 9.b shows the data points that would correspond to a 100% FPR.

FIGURE 9




Far from confirming the validity of the model, the data suggests that the income co-efficient in the model is far too low, and that a model like that shown in fig. 9c would be more appropriate.

4.6 Conclusions

4.6.1 One suggested procedure for the comparison of a number of models on the basis of a validation sample is thus:

a) For each model, compare the expected and observed number of FPRs, rejecting models which are inconsistent with the data;

b) If more than one model is left, compare models pairwise in respect of their total numbers of FPRs by the McNemar/Cochran Q and Q' tests, rejecting models which can be shown to produce fewer FPRs on average.

c) If more than one model remains, all have been shown consistent with the data and indistinguishable in respect of expected number of FPRs. Choose one at random, unless another criterion (theoretical elegance, ease of application ...) seems relevant.

d) Compare the chosen model with the 'random choice of options' model, by means of $N_r$ and N in the light of var($N_r$) to assess usefulness (Hauser's (1978) statistics are more informative, assuming that the model is indeed better than a random choice).

4.6.2. Finally, we note that other criteria of model performance have been suggested- see for example in Gunn and Bates, 1980. Much work remains to be done on this aspect of model scrutiny. However, for our present purposes, we recommend the use of the conventional F P R statistic, as interpreted by the rules we have supplied above.

# 5. EFFICIENT DESIGN

## 5.1 Introduction

5.1.1 The general principles of survey design and sample size assessment can be described in simple terms as follows. We suppose that a survey is to be conducted in which observations of a variable Y are to be made at N points corresponding to different values of a variable X - say $X_1$, $X_2$, .., $X_N$ - and that a model is to be fitted in which Y is to be related to X by a relationship involving unknown parameters. Suppose that the distribution of Y, given X and the true values of the unknown parameters, is known. Then, in anticipation of the results of the survey/experiment, we can write down general forms for the estimators of the unknown parameters, and hence of the fitted model, and also write down general forms for the variance-covariance matrix of the parameters and thus also of function of these parameters, including the fitted model. If our intention is to maximise the accuracy of estimation of some function of the fitted parameters for given sample size, or to minimise sample size for some required accuracy of estimation, we can refer to these general forms to indicate the relationship between the amount of data, the location of this data and the consequent accuracy of the fitted parameters.

5.1.2 Note that we will for now ignore problems of model validation, and make the assumption that we can correctly specify the model form and the distribution of Y from the outset. In practice we may have to adopt sequential procedures, and also build in validation requirements when designing the survey. Two simple examples may help to establish the main points of the approach.

5.1.3 Example 1.

Suppose we have the model $Y = \alpha + \beta X$ X, and know that the distribution of Y given $\alpha$, $\beta$ and X is Normal, with known variance $\sigma^2$. Given fixed sample size N, and the opportunity to observe the N values of Y corresponding to N selected values of X, how should these values ($X_1$ ... $X_N$ say) be chosen to minimise the estimation error associated with the maximum likelihood estimates of $\beta$? We can write down the log-likelihood function in general terms as

$$\ell = K_1 + K_2 \sum_{i=1}^{N} (Y_i-(\alpha + \beta X_i))^2 \qquad (K_1, K_2 \text{ constants})$$

The maximum likelihood estimators of $\alpha$ and $\beta$ are obtained as solutions of the equations $\frac{\partial \ell}{\partial \alpha} = \frac{\partial \ell}{\partial \beta} = 0$ : call the solutions a and b. This leads to the familiar form $b = \sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y}) \{ \sum_{i=1}^{N} (X_i - \overline{X})^2 \}^{-1}$ with estimated variance $\text{var } b = (N-1)\sigma^2 \{ N \sum_{i=1}^{N} (X_i - \overline{X})^2 \}^{-1}$

Thus we should pick the points $X_1 \ldots X_N$ to maximise $\{ \sum_{i=1}^{N} (X_i - \overline{X})^2 \}$ in order to minimise the estimation error associated with b. If we are restricted to experimentation within a particular range of X, say in the interval $(X_{min}, X_{max})$, then we should take half our observations at $X_{min}$ and half at $X_{max}$.

5.1.4 Example 2.

Suppose we have the model $Y = \alpha X$, and know that the distribution of Y, given $\alpha$ and X is Poisson, how many observations $Y_i$ at chosen points $X_i$ should be taken, and how should the $X_i$ be selected, in order to have 95% confidence that the maximum likelihood estimate of $\alpha$ lies within $\pm$ 10% of the true value?

In this case we can write down the log-likelihood function corresponding to a sample of size N as $\ell = K - \sum_{i=1}^{N} (-\alpha X_i + Y_i \log(\alpha X_i))$

The maximum likelihood estimator of $\alpha$ is a, given as the solution of $\frac{\partial \ell}{\partial \alpha} = 0$, i.e. $\alpha = \sum_{i=1}^{N} Y_i \, ( \sum_{i=1}^{N} X_i )^{-1}$

with associated estimate of the variance of a given by $\{ -E(\frac{\partial^2 \ell}{\partial \alpha^2}) \}^{-1} = \alpha \{ \sum_{i=1}^{N} X_i \}^{-1}$

Thus in this case the error associated with the estimate is reduced by taking all observations at as high as possible values of X - i.e. all at $X_{max}$, if observation is only possible within a restricted interval $(X_{min}, X_{max})$.

Thus the $\pm$ 95% confidence limits around the mean $\alpha$ within which a is expected to lie are given by $\{ \alpha - 1.96S, \alpha + 1.96S \}$

where $S = \sqrt{\alpha \{ N X_{max} \}^{-1}}$ obviously depends on $\alpha$.

The form of the confidence interval is based on the (asymptotic) Normal distribution of M.L. estimators.

5.1.5 Thus, without knowledge of the value of $\alpha$ the quantity we are setting out to estimate we cannot choose the required sample size. There are two ways to approach the problem. Firstly, if a sequential procedure is permissable, we may form a first estimate of $\alpha$ on the basis of $N_1$ observations (all at $X_{max}$) and then estimate how many more would be needed for the required accuracy on the assumption that the estimate of $\alpha$ from the $N_1$ observations is the true value. Secondly, we might have a reliable estimate of the region in which $\alpha$ is expected to lie, and could form a "pessimistic" estimate of the required sample size on the basis of the maximum sample needed for any value of $\alpha$ in that region.

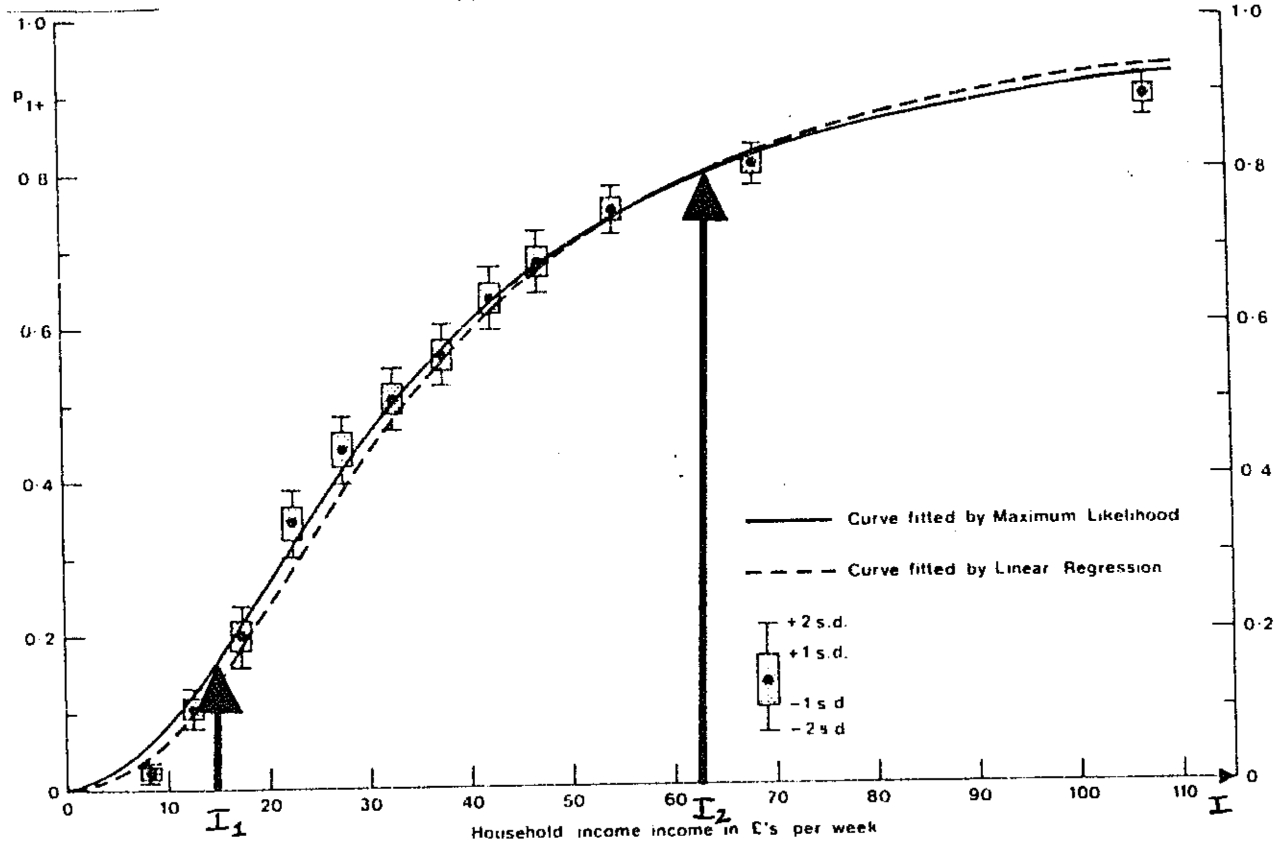
5.1.6 Both of these examples demonstrate the absolute reliance of the results on the assumption of known distributions/models. Obviously it would be impossible to reject the hypothesis of a linear relationship between Y and X on the basis of the experimentation at just two points advocated in example 1, or the proportional model of example 2 on the basis of experimentation at a single point. It would be dangerous in the extreme to interpret the guidelines for 'optimal design' too literally in most practical applications. However, if we are able to specify some acceptable test of model validation, the same methods may be used to prescribe the most efficient design and minimum sample size requirements for both validation and estimation. This problem is not one which appears in the literature, and will involve careful thought as to the appropriate criteria for model validation. The remainder of this section is concerned with 'design for estimation' alone.

## 5.2 'Optimal Design' and Value-of-Time

5.2.1 An overview of the various approaches that have been taken to the design problem is given by Silvey (1980). In the specific context of disaggregate models, see also Daganzo (1980). In general, the solutions depend on the objective. In the examples above, we have considered the problem of maximising the accuracy of a single parameter. However, the same approach could be used for any general function of parameters, provided that it is single valued. This does raise difficulties when there is no 'natural' choice of such a function. According to Silvey, the most commonly adopted (or at least, for theoretical exposition, most frequently postulated) is the 'criterion of D-optimality' which amounts to minimising the determinant of the variance-covariance matrix of the parameters in the model. This objective is equivalent to minimising the area of any given confidence region for the parameters, thus in some sense maximising the joint accuracy of the parameter estimates.

FIGURE 10

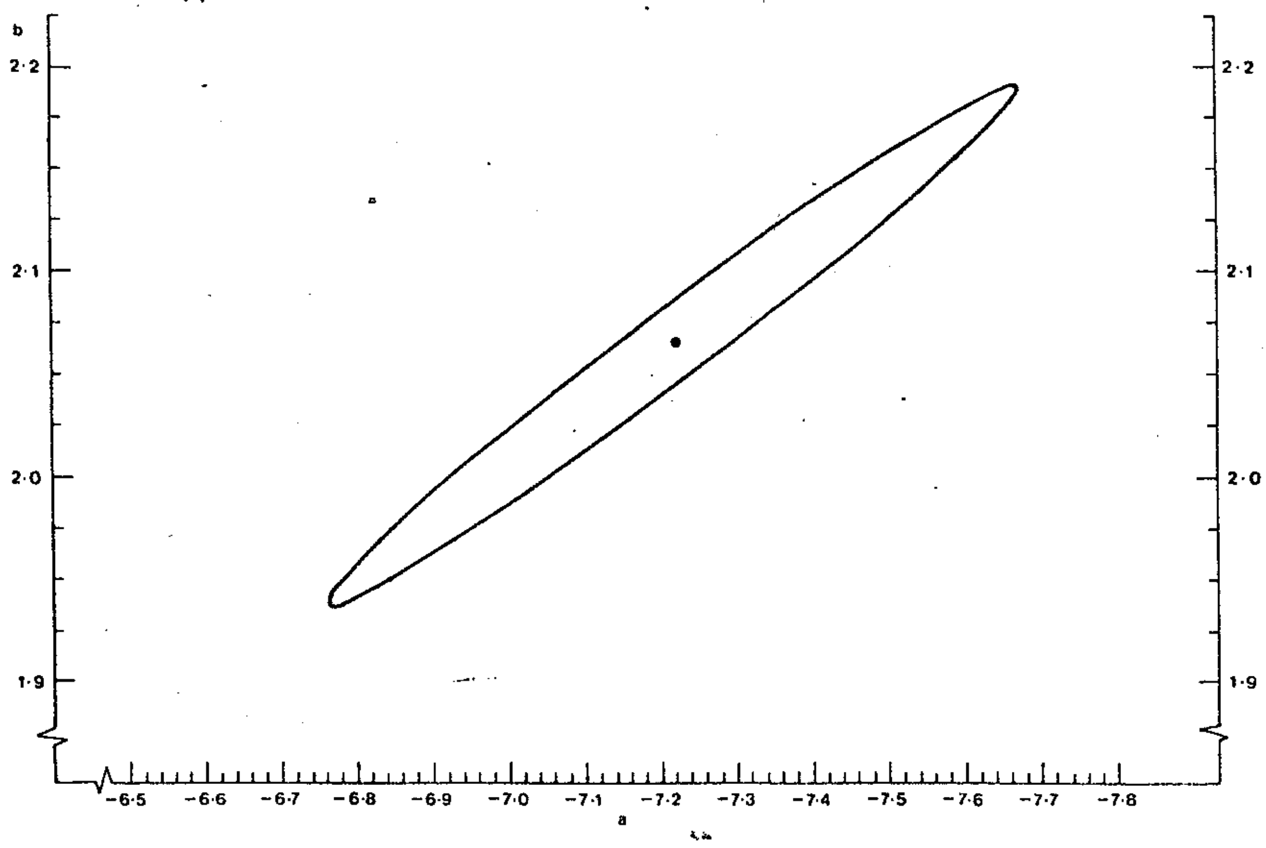
1972 FES DATA FOR $p_{1+}$ SHOWING CONFIDENCE INTERVALS AND FITTED CURVES




$$P_{1+} = (1 + \exp(-a - b.\log I))^{-1} \qquad \hat{a} = -7.2 \qquad \hat{b} = 2.0$$

FIGURE 11

$p_{1+}$ RELATIONSHIP FOR 1972 FES: 95% CONFIDENCE REGION FOR $(\hat{a}, \hat{b})$



5.2.2 Figures 10 and 11 illustrate this concept with reference to a logit model framed in terms of two parameters, a and b. The model is taken from Bates et al (1978) and refers to the proportion of households owning at least one car as a function of gross household income. Figure 10 shows the data and the fitted model. Figure 11 shows the 95% confidence region associated with the estimated parameters, using the maximum likelihood estimators, for the given data set. One use of the 'optimum design' approach would be to determine 'where' (i.e. at which income points) the data should be collected for populations with similar expected relationships between car ownership and income in order to minimise the error of the fitted parameters, as described by the size of the corresponding confidence regions.

5.2.3 The solution to this problem is given by Silvey, quoting from Ford (1976); for any given sample size, half the observations should be taken at one income point, and half at another. The points are given by a general formula; in the case of the relationship described these turn out to be approximately £15 and £62 for the 1972 data, $I_1$ and $I_2$ in figure 10 Once again we see how crucial is the assumption that the model is correct! However, this sort of information does provide valuable insights into the relative values of taking observations at different points, providing we are reasonably cautious about the policies it advocates.

5.2.4 Inference about 'values-of-time' has usually involved models with a particular form of parameter structure; typically, there has been a function relating 'utility' to observed variables by an expression such as

$$U_i = (\theta_o^i - \theta_1 M_i - \theta_2 T_i - \theta_3 Z_i) \qquad (5)$$

in which $Z$ is some variable like comfort, 'M' denotes a money cost and 'T' denotes time in an activity, and the θ's are constants.

In certain cases, there may be advantages in interpreting the fitted co-efficient of cost variables in probabilistic choice models based on random utility theory with the dispersion parameter $\Lambda$, which is inversely related to the standard deviation of the random component of the utility function (the effect of the 'unobservables'). This corresponds to a choice of money units for the utility expression.

For logit models the relationship is $\Omega = \frac{\Pi}{\sqrt{6}} \frac{1}{\sigma}$

where $\sigma$ denotes the standard deviation of the random component.

5.2.5 Adopting this convention, we can write the general form of the linear utility function which is used in many empirical studies as

$$U_i^j = \theta_o^i - (\Omega)M_i^j - (\Omega V_T)T_i^j - (\Omega V_Z)Z_i^j + \varepsilon_i^j \qquad (6)$$

where $\theta_o^i$ refers to the mean of the 'unobservables', $V_T$ to the value of time and $V_Z$ to the value of some other variable, all now measured in money terms. $\varepsilon$ is assumed Weibull, standard deviation 1, for 'logit' models, ('i' refers to option, 'j' to individual.)

5.2.6 There is no asymmetry introduced by this convention. The estimate of the dispersion parameter in <u>time</u> terms, for example, is

(value of $\Omega$ in money terms). (value of time) = $\Omega V_T$

Thus the co-efficient corresponding to the time variable could also be described as measuring the dispersion parameter, this time in units of time. The bracketed expressions in eq. 6., are the estimated coefficients recovered from (eg) logit models. From 'transfer price' experiments, we obtain data to fit the equation(1)

$$(TP - \Delta M^j) = C + V_T \, \Delta T^j + V_Z \, \Delta Z^j + \varepsilon^{1j} \qquad (7)$$

5.2.7 Note that the dispersion parameter is also measured in the usual fitting procedure, (since we also estimate $var(\varepsilon') \fallingdotseq 2\sigma^2$), and that the smaller the dispersion parameter (the larger the standard deviation of the 'unobservables') the worse the statistical precision of the fitted co-efficients $V_T, V_Z$ In fact, equation 7 has usually been estimated by least squares regression, in which with negligible (and removable) inconsistency, the error term that was conveniently assumed Weibull for the logit analysis is conveniently assumed Normal for the regression. With the data in this form, the V coefficients are estimated independently from the $\sigma$ co-efficient.

5.2.8 Returning to the notation of equation 5. In this case, there is a single function of parameters that is of paramount importance, namely the ratio $\theta_2/\theta_1$, the 'value-of-time' if circumstances are appropriate, the accuracy with which a particular design estimates this ratio forms a natural criterion of optimality.

---

(1) 'C' a constant. Note $\Delta$ M etc. now refer to <u>differences</u> between options.

5.2.9 The Taylor series approximation for the variance of a function gives

$$\mathrm{Var}\left(\underline{f}(\underline{X})\right) \doteq \left[\frac{\partial \underline{f}}{\partial \underline{x}}\right]^T \mathrm{Var}\ (\underline{X}) \left[\frac{\partial \underline{f}}{\partial \underline{x}}\right]$$

Thus if we have a general expression for the variance-covariance matrix of the fitted parameters we can approximate the variance of a function of the parameters. If the estimates have been derives as likelihood maximising solutions, such an estimate is provided by the inverse of the expectation of the matrix of second derivatives of the log-likelihood function. In the case of the hypothetical example given above, denote this by $\underline{\underline{V}}$ where

$$\underline{\underline{V}} = \begin{bmatrix} v_{00} & v_{01} & v_{02} & v_{03} \\ v_{01} & v_{11} & v_{12} & v_{13} \\ v_{02} & v_{12} & v_{22} & v_{23} \\ v_{03} & v_{13} & v_{23} & v_{33} \end{bmatrix}$$

For $f(\underline{\theta}) = \theta_2/\theta_1$, we have

$$\begin{bmatrix} \frac{\partial f}{\partial \theta_0} \\ \frac{\partial f}{\partial \theta_1} \\ \frac{\partial f}{\partial \theta_2} \\ \frac{\partial f}{\partial \theta_3} \end{bmatrix} = \begin{bmatrix} 0 \\ -\theta_2/\theta_1^2 \\ 1/\theta_1 \\ 0 \end{bmatrix}$$

Thus the criterion to be minimised is

$$VR = \left[\frac{1}{\theta_1^2} v_{22} - 2\frac{\theta_2}{\theta_1^3} v_{12} + \frac{\theta_2^2}{\theta_1^4} v_{11}\right]$$

Note that the design which is optimal from the viewpoint of minimising the variance of the 'value-of-time' estimate will not in general be that which is D-optimal, or optimal under any other criterion.

5.2.10 To recap then, we can tackle the design problem if (a) we can write down expressions for the $v_{ij}$ in terms of sample size and location, and if (b) we have some idea about the 'true' values of the parameters. Assuming that condition (b) can be met from previous studies, we shall next consider (a). Note in passing that, if there are different costs attaching to experimentation at different points, the same approach can given 'maximal accuracy for given survey expenditure' rather than for given sample size. Similar arguments will lead to 'optimal designs' to augment existing data sources.

## 5.3 Approximations to the variance-covariance terms

5.3.1 We have the criterion VR

where the $\Theta$ coefficients are those fitted in a model of the form

$$\Delta U = \Theta_o - \Theta_1 \Delta M - \Theta_2 \, \Delta T - \Theta_3 \, \Delta Z \; + \varepsilon$$

and $V_{ij}$ is the covariance of $\Theta_i$ and $\Theta_j$.

5.3.2 To make the VR expression useful for design purposes, we need approximations to the $V_{ij}$, which will in general be functions of sample <u>location</u> (in terms of ( $\Delta M$, $\Delta T$, $\Delta Z$)) and sample <u>size</u> as well as of the unknown coefficients $\underline{\Theta}$.

For simplicity, we shall consider the problem on the assumption that replicated observations will be taken. (It is interesting to note that optimal design considerations would indeed lead to such designs. For the practical purposes of evaluating feasible 'non-optimal' designs, the conclusions reached in this way should be broadly similar to more detailed analyses).

5.3.3 A suitable approximation to the variance covariance matrix of the fitted coefficients for (aggregate) logit models is given in Gunn and Whittaker (1981) for the case of Poisson errors. A similar approximation [1] for the Binomial case (we shall assume a binary choice here) is as follows:

Write $m = \Delta M$, $t = \Delta T$

Define $\bar{m} = \sum_i W_i m_i \,/\, \sum_i W_i$

$\bar{t} = \sum_i W_i m_i / \sum_i W_i$

$i = 1, \ldots, n$, the no. of points at which observations are taken

where

$W_i = n_i p_i \, (1 - p_i)$

$n_i$ = no. of observations taken at point i, defined by $(m_i, t_i)$

and $p_i = \left[1 + \exp(-\Theta_1 m_i - \Theta_2 t_i)\right]^{-1}$

(assuming $\Theta_o = \Theta_3 = 0$ for illustration)

set $V(m,m) = \sum_i W_i \, (m_i - \bar{m})^2$

$V(t,t) = \sum_i W_i \, (t_i - \bar{t})^2$

$V(m,t) = V(t,m) = \sum_i W_i \, (m_i - \bar{m}) \, (t_i - \bar{t})$

---

(1) NOTE we assume $\Theta_o = \Theta_3 = 0$ here, for illustration: if $\Delta Z$ is independent of $\Delta M$, $\Delta T$ by design, the same result holds.

With this notation, an approximation to the variance-covariance matrix of the fitted $\theta_1$ and $\theta_2$ coefficients is

$$\text{var}(\theta_1, \theta_2) = \frac{1}{\left[V(m,m)V(t,t) - [V(m,t)]^2\right]} \begin{bmatrix} V(t,t) & -V(m,t) \\ -V(m,t) & V(m,m) \end{bmatrix}$$

We can now write

$$VR = \frac{1}{\theta_1^2 \left[V(m,m)V(t,t) - (V(m,t)^2\right]} \left[V(m,m)+2\left[\frac{\theta_2}{\theta_1}\right] V(m,t) + \left[\frac{\theta_2}{\theta_1}\right]^2 V(t,t)\right] \quad (8)$$

Note that $\left(\frac{\theta_2}{\theta_1}\right)$ is our estimate of the value of time.

5.3.4 Equation 8 allows us to say a number of things about the conditions necessary for 'value of time' measurement, as well as providing actual quantitative information about accuracy for any proposed design (i.e. selection of points at which to experiment), and determining the relative trade-off between the number and location of the experimental points and the survey effort to apportion to each.

5.3.5 Firstly, we can see that the larger is $\theta_1^2$, the more accurate our measurement (other things being equal). Having identified $\theta_1$ as being inverseley related to the standard deviation of the random component of the utility function, in money terms , we can interpret this as saying that conditions in which the 'representative' component (i.e. that which is made explicit) dominate the total utility expression will be most favourable for accurate value-of-time measurement. In other words, where the model explains little of the variation, measurement will be poor.

5.3.6 Secondly, it is easy to see that the term $\left[V(m,m)V(t,t)-(V(m,t))^2\right]$ will be zero if M and T are linearly related. In other words, in such conditions VR would be infinite: no measurement is possible if 'time' and 'cost' are perfectly correlated, and the less they are correlated the better.

5.3.7 Thirdly, we can see in general terms that VR contains a term linear in the V(.,.) divided by one quadratic in the V(.,.). Broadly speaking, accuracy will come from maximising the V(.,.). From the definitions of the V terms we can see that such a maximum will occur as a compromise between two opposing trends: terms such as $(m_i - \bar{m})$ and $(t_i - \bar{t})$ will suggest placing the experimental points as far apart as possible to maximise the expression; however, at extreme points the $w_i$ will tend to zero ($p_i$ will tend to zero or unity, so $p_i(1 - p_i)$ will tend to zero) and so a compromise will occur. (The one dimensional example given above produced a solution roughly at the points of inflection, and this may generalise. Using eq 8 together with symmetry arguments [1] should lead to a straightforward, if tedious, solution for the optimal design in the general case.)

5.3.8 Finally, we can see that the optimal design/accuracy of measurement depend on the level of the value-of-time. It is more sensible to consider the ratio of the standard error of measurement of the vot to its absolute level in this case: $\sqrt{VR}/\left[\theta_2/\theta_1\right] =$ RSE, say

We obtain

$$RSE = \sqrt{\frac{1}{\theta_2^2\left[V(m,m)V(t,t) - V(m,t)^2\right]}\left\{\left[\frac{\theta_1}{\theta_2}\right]^2 V(m,m) + \left[\frac{\theta_1}{\theta_2}\right]2V(m,t) + V(t,t)\right\}}$$

For very small values of time, the expression in curly brackets is dominated by V(m,m), whereas for large values of time the V(t,t) expression dominates. Different design strategies will be appropriate for different values.

---

(1) Note that Silvey (1980) uses Caratheodory's Theorem to demonstrate that the optimal design will involve experimentation at less than 5 distinct points, which must in this case be sited symmetrically about the p = 0.5 line in the (m,t) space.

## 5.4 Conclusions

5.4.1 The actual solutions to the 'optimal design' problems are not easy to derive, even for the highly simplified examples we have considered. The requirement to parameterise a model large enough to measure variations in time values (as between modes, for example), and the possibility that non-linear functions and randomly distributed parameter models will be needed, demonstrates that even more difficult areas remain to be tackled.

5.4.2 In practice, however, we will probably be restricted to consideration of a smallish number of design options. This was the experience gained in the application of these methods in Gunn et al (1980), where the aim was to design a survey from which an aggregate O/D matrix would be formed as a basis for value-of-time inference from aggregate distribution patterns. Practical considerations of how to implement the design greatly restricted the 'feasible region' in which a maximum was sought.

# 6. TRANSFER PRICING AND REVEALED PREFERENCE - A COMPARISON OF MAXIMAL EFFICIENCY

## 6.1 Introduction

6.1.1 From a statistical point of view, the difference between 'revealed preference' approaches and 'transfer pricing' approaches can be reduced to a question of the information content of the data. For the purposes of illustration it is helpful to use the context of a data set relating to a choice between options whose outcome has been observed, thus avoiding for the moment the difficulties associated with eliciting future intentions. Suppose we have observed a sample of travellers choosing between two options with differing time and money characteristics, noted the characteristics, the outcome of the choice AND asked a Transfer Price question. Assuming that the TP responses will all be of the correct sign (i.e. if we ask by how much the cost of the option actually selected would have to rise in order to make the traveller indifferent between options, the answers will all be positive) then the TP data contains at least as much information as the 'revealed' behaviour. We could throw away the estimate of the magnitude of the TP and estimate the unknown coefficients in a representative utility function assuming only that the utility difference between chosen and rejected modes was positive: in fact this is what happens in the usual models of discrete choice.

6.1.2 More speculatively, we might postulate that the TP estimates were related to the utility difference between options, and write (after DoE Economic and Statistical Note 22)

$$TP_i = \lambda \Delta U_i + \alpha_o + {}_1\varepsilon_i \qquad (9)$$

together with

$$\Delta U_i = \underline{\theta} . \underline{z}_i + \varepsilon_i \qquad (10)$$

Different assumptions about the distributions of ${}_1\varepsilon_i$ , the error introduced by the difficulty of responding to the TP question, would lead to estimation problems of varying degrees of complexity. The simplest of these assumptions would be that ${}_1\varepsilon_i$ could be neglected altogether; this, together with Normality assumptions about $\varepsilon_i$ would allow the use of ordinary least squares to estimate the model (or iteratively weighted least squares if the coefficients are assumed random.)

6.1.3 Finally, we might assume from the outset that $\lambda = 1$, i.e. that the response to the TP question was indeed in the same money units as those in which the travellers value other cost items.

6.1.4 As always, as the strength of the assumptions on which the analysis is based is increased, so the apparent information content of the data rises, and the uncertainty associated with the fitted model appears to decrease. For the purposes of this note we shall assume that each successive stage is justified in order to illustrate the reasons for the increase in precision that TP data can allow.

## 6.2 Incremental information in relative magnitude of utility difference over sign

6.2.1 Let us assume that we have decided that time and cost alone are sufficient to explain choice between the options, and that our representative utility function is simply

$$\bar{U}_q^i (\underline{\theta}, \underline{Z}_q) = \theta_o - \theta_1 M_q^i - \theta_2 T_q^i \qquad \begin{cases} \text{individual } q \\ \text{option } i \end{cases}$$

Suppose that the standard deviation of the random component of each utility function is unity* , so that the difference between utilities has expectation

$$\Delta\bar{U}_q = -\theta_1 \Delta M_q - \theta_2 \Delta T_q$$

and variance 2

where $\Delta X$ is defined as $(X_{\text{chosen mode}} - X_{\text{alternative mode}})$.

Let D1 refer to a data set consisting of the signs of the utility differences together with the characteristics of the options, $\underline{Z}_1$ and $\underline{Z}_2$. Assuming the random components are Weibull distributed, we can write the likelihood function for D1 as

$$L(D1, \underline{\theta}, \Omega) = \prod_{q=1}^{Q} P(Cq, \underline{\theta}) \qquad (11)$$

where (from the Weibull assumption)

$$P(Cq, \underline{\theta}) = \left[ \sum_{i=1}^{2} \delta_{qi} \exp(\Omega \bar{U}_q^i) \right] \left[ \sum_{i=1}^{2} \exp(\Omega \bar{U}_q^i) \right]^{-1}$$

where $C_q$ denotes the choice made by individual q;

Q denotes the number of decisions observed;

$\delta_{qi}$ = 1 if individual q chooses option i, and

$\delta_{qi}$ = 0 otherwise, and $\Omega = \pi/(\sqrt{6})$

* to resolve the indeterminacy. The same results follow if we set $\theta_1$ to unity, and treat the standard deviation as a parameter.

6.2.2 Now let D2 refer to a data set consisting of the absolute magnitudes of the TP responses, and assume that we have the model

$$TP_q = \lambda \Delta U_q = \lambda(\Delta \bar{U}_q + \varepsilon_q)$$

$$= -(\lambda\theta_2)\ \Delta T_q - (\lambda\theta_1)\ \Delta M_q + \varepsilon^1_q$$

where $\varepsilon^1_q \sim N(0, 2\lambda^2)$

6.2.3 Note that we have interchanged Weibull and Normal distributions for convenience; in this case the differences will be slight. We can now write the likelihood function of D2 as

where $$L(D_2, \underline{\theta}, \lambda) = \prod_{q=1}^{Q} \frac{1}{2\lambda\sqrt{\pi}} \exp\{-\frac{(TP_q - \lambda\Delta\bar{U}_q)^2}{2.2\lambda^2}\} \qquad (12)$$

Note that the choices are assumed to be independent.

6.2.4 From inspection of the form of the functions, it is clear that we should write

$L(D1, \underline{\theta}, \Omega)$ as $L(D1, \Omega\underline{\theta})$, and

$L(D2, \underline{\theta}, \lambda)$ as $L(D2, \lambda\underline{\theta}, \lambda)$

6.2.5 The problem of units has been discussed above; in this case we are really interested in estimating the ratio $\theta_2/\theta_1$, which may be equivalently expressed as either $\Omega\theta_2/\Omega\theta_1$ or $\lambda\theta_2/\lambda\theta_1$

Accordingly, let us rewrite (3) and (4) as

$$L(D1, \underline{\theta}^1) = \prod_{q=1}^{Q} \{[\sum_{i=1}^{2} \delta_{qi}\ e^{-\theta'_1 M^i_q - \theta'_2 T^i_q}] / [\sum_{i=1}^{2} e^{-\theta'_1 M^i_q - \theta'_2 T^i_q}]\} \qquad (13)$$

where $\theta'_j = \Omega\theta_j \quad \frac{\pi.\theta_j}{\sqrt{6}}$, and

$$L(D2, \underline{\theta}'', \sigma'') = \prod_{q=1}^{Q} \frac{1}{2\lambda\sqrt{\pi}} \exp\left\{-\frac{(TP_q + \theta''_1\ \Delta M_q + \theta''_2\ \Delta T_q)^2}{4\lambda^2}\right\} \qquad (14)$$

where $\theta''_j = \lambda\theta_j$

6.2.6 Note also that (13) can be written as

$$L(D1, \underline{\theta}') = \prod_{q=1}^{Q} \{1 + \exp(-\theta_1' \Delta M_q - \theta_2' \Delta T_q)\}^{-1} \qquad (15)$$

6.2.7 Using (14) and (15), and assuming that Maximum Likelihood estimators will be adopted, we can now give a simple illustration of the increase in precision that can be gained by using the TP estimates of the magnitude of the utility difference between options (without assuming this to be in the same units as the cost elements associated with each option) instead of merely the sign of the utility difference between options.

6.2.8 As usual, we would estimate the variance-covariance matrix of the fitted coefficients by the negative of the inverse of the matrix of expectations of second derivatives of the log-likelihood function, $-(J)^{-1}$ say.

6.2.9 In this notation we can write down the uninverted matrices corresponding to data sets D1 and D2 as $J_1$ and $J_2$ where

$$J_1 = \begin{bmatrix} -\sum_q p_q(1-p_q)\Delta M_q^2 & -\sum_q p_q(1-p_q)\Delta M_q \Delta T_q \\ -\sum_q p_q(1-p_q)\Delta M_q \Delta T_q & -\sum_q p_q(1-p_q)\Delta T_q^2 \end{bmatrix} \quad J_2 = -\frac{1}{2\lambda^2} \begin{bmatrix} \sum_q \Delta M_q^2 & \sum_q \Delta M_q \Delta T_q \\ \sum_q \Delta M_q \Delta T_q & \sum_q \Delta T_q^2 \end{bmatrix}$$

where

$$p_q = \{1 + \exp(-\theta_1' \Delta M_q - \theta_2' \Delta T_q)\}^{-1}$$

6.2.10 For simplicity in illustration we shall assume that $\theta_1$ and $\theta_2$ are uncorrelated by design, so that

$$-(J_1)^{-1} = \begin{bmatrix} \dfrac{1}{\sum_q p_q(1-p_q)\Delta M_q^2} & 0 \\ 0 & \dfrac{1}{\sum_q p_q(1-p_q)\Delta T_q^2} \end{bmatrix}$$

i.e. $\mathrm{var}(\theta_1') = \dfrac{1}{\sum_q p_q(1-p_q)\Delta M_q^2}$ $\qquad \mathrm{var}(\theta_2') = \dfrac{1}{\sum_q p_q(1-p_q)\Delta T_q^2}$

$$-(J_2)^{-1} = 2\lambda^2 \begin{bmatrix} \dfrac{1}{\sum_q \Delta M_q^2} & 0 \\ 0 & \dfrac{1}{\sum_q \Delta T_q^2} \end{bmatrix}$$

i.e. $\text{var}(\theta_1'') = \dfrac{2\lambda^2}{\sum_q \Delta M_q^2}$ $\qquad$ $\text{var}(\theta_2'') = \dfrac{2\lambda^2}{\sum_q \Delta T_q^2}$

6.2.11 A simple example of such a design would be to choose experimental points $(\Delta M_q, \Delta T_q)$ such that when $\Delta M_q \neq 0$ then $\Delta T_q = 0$ and conversely when $\Delta T_q \neq 0$ then $\Delta M_q = 0$. . In other words, to look for options which differ in speeds, but not in costs, and options which differ in costs but not in speeds. Note that this is not necessarily the best strategy, merely the easiest to illustrate here. In practice, it is clear from the VR criterion that there is advantage in ensuring that the $\theta_1$ and $\theta_2$ coefficients are positively correlated - so that when our data set leads to too high an estimate of one it also leads to too high an estimate of the other, working to stabilise the ratio $\theta_2/\theta_1$ .

6.2.12 Returning to the illustrations here, we can write the VR approximation to the variance of the ratio $\theta_2/\theta_1$ (the estimated 'value of time') as

$$VR = \left[\frac{\text{var}\theta_2}{\theta_1^2} + \frac{\theta_2^2\ \text{var}\theta_1}{\theta_1^4}\right] \qquad \text{since } \text{cov}(\theta_1, \theta_2) = 0$$

Thus for data set D1 we obtain

$$VR1 = \left[\frac{1}{\theta_1'^2 \sum_q p_q(1-p_q)\Delta T_q^2} + \frac{\theta_2'^2}{\theta_1'^4 \sum_q p_q(1-p_q)\Delta M_q^2}\right]$$

and for data set D2 we obtain

$$VR2 = \left[\frac{2\lambda^2}{\theta_1''^2 \sum_q \Delta T_q^2} + \frac{2\theta_2''^2 \lambda^2}{\theta_1''^4 \sum_q \Delta M_q^2}\right].$$

6.2.13 The relative precision of the estimates provided by the two data sets can thus be related to the ratio VR1/VR2.
This is of the form $\frac{A1 + B1}{A2 + B2}$: consider the ratios $\frac{A1}{A2}$ and $\frac{B1}{B2}$ separately.

These are

$$\frac{A1}{A2} = \frac{\theta_1''^2 \sum_q \Delta T_q^2}{2\lambda^2 \theta_1'^2 \sum_q p_q(1-p_q)\Delta T_q^2} \quad \text{and} \quad \frac{B1}{B2} = \frac{\theta_1''^4 \theta_2'^2 \sum_q \Delta M_q^2}{2\lambda^2 \theta_2''^2 \theta_1'^4 \sum_q p_q(1-p_q)\Delta M_q^2}$$

6.2.14 We can simplify these to

$$\left.\begin{aligned} \frac{A1}{A2} &= \frac{\sum_q \Delta T_q^2 \quad 3}{\sum_q p_q(1-p_q)\Delta T_q^2 \, \pi^2} \\ \frac{B1}{B2} &= \frac{\sum_q \Delta M_q^2 \quad 3}{\sum_q p_q(1-p_q)\Delta M_q^2 \, \pi^2} \end{aligned}\right\} \quad \text{substituting for } \theta_j', \theta_j''$$

6.2.15 Now

$$0 \leq p_q \leq 1, \text{ so } p_q(1-p_q) \leq {}^1/_4$$

thus

$$\sum_q p_q(1-p_q)\frac{\pi^2}{3}\Delta T_q^2 \leq \sum_q \frac{\pi^2}{12}\Delta T_q^2 < \sum_q \Delta T_q^2$$

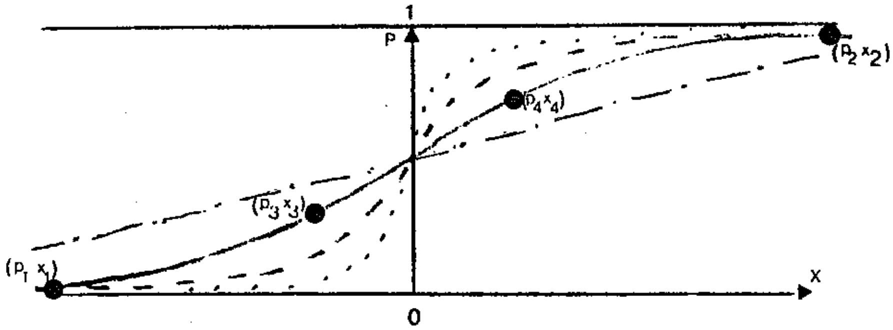
and similarly

$$\sum_q p_q(1-p_q)\frac{\Pi^2}{3}\Delta M_q^2 < \sum_q \Delta M_q^2$$

Hence A1 is greater than A2, B1 is greater than B2, all As and Bs are positive, so that (A1 + B1) is larger than (A2 + B2). In other words, the precision of the estimate of $\theta_2/\theta_1$ from data set D1 is <u>less</u> (the variance is higher) than of data set D2.

6.2.16 The form of the ratio VR1/VR2 gives some qualitative indication of the importance of the TP information over and above the sign of the utility difference. At points (decisions) where the representative utility difference is large (compared to the standard deviation of the random element) the product $p_q(1-p_q)$ is very small. Such points contribute little or nothing to improving the accuracy of the coefficient estimates in the context of data set D1, but are especially powerful in increasing precision with data set D2. We can note in passing that this implies that a design (choice of $(\Delta M, \Delta T)$ points at which to observe decisions) aimed at optimising accuracy via an analysis of the choices (cf D1) will be very inefficient from the point of view of an analysis based on the TP responses (cf D2). Figure 12 illustrates

FIGURE 12




6.2.17 The points $(p_1, x_1)$ and $(p_2, x_2)$ may be seen to be in reasonable agreement with a large number of logistic curves of the form $P = (1 + \exp(a + bX))^{-1}$ as compared to the points $(p_3, x_3)$ and $(p_4, x_4)$. This issue was discussed above; we noted that 'optimal design' in the context of linear regression led to placing experimental points as far apart as possible, whereas the estimation of a logistic curve from data on proportions choosing options led to experimentation around the points of inflection of the curve. The same sort of results will hold for 'value of time' estimation.

6.2.18 The result that 'extreme' points contribute little or nothing to the accuracy of the estimate of $\theta_2/\theta_1$ has been emphasised in connection with experimental design; it should also be borne in mind when examining the results and conclusions of previous studies which were based on data sets collected without any such consideration. It is not sensible to extrapolate the relationships between sample size and accuracy of estimate from such studies without making some attempt to correct for survey design.

## 6.3 Incremental information in assuming that TP response is in the same units of cost as other cost items

6.3.1 This assumption involves the assertion that $\theta_1=1$ (var $\theta_1 = 0$)

In this case, we would have

$$\text{var}(\theta_2/\theta_1) = \text{VR3} = \frac{\lambda^2}{\sum_q \Delta T_q^2}$$

Assuming the assertion to be true, the increase in precision may be gauged by the difference VR2 - VR3, which is a function of $\theta_1$ and $\theta_2$.

6.3.2 When the experimental points have been chosen so that $\theta_1$ and $\theta_2$ are uncorrelated in VR2, it can easily be seen that VR3 is less than VR2; in the more general case we must also consider the covariance terms.

## 6.4 Conclusions

6.4.1

1) Even when stated in non-money units, TP data can greatly improve the precision of coefficient estimates.
2) The increase in precision is a function of survey design.
3) The change in precision on assuming that TP data is in money units (perhaps subsequent to a test of a counter-hypothesis) also depends on survey design.

4) Design 'points' (e.g. groups of individuals facing options with a unique difference in time and in money characteristics) which best support inference about 'value-of-time' from revealed preference analysis will be very suboptimal for TP analysis.

# 7. AGGREGATE Methods

## 7.1 Introduction

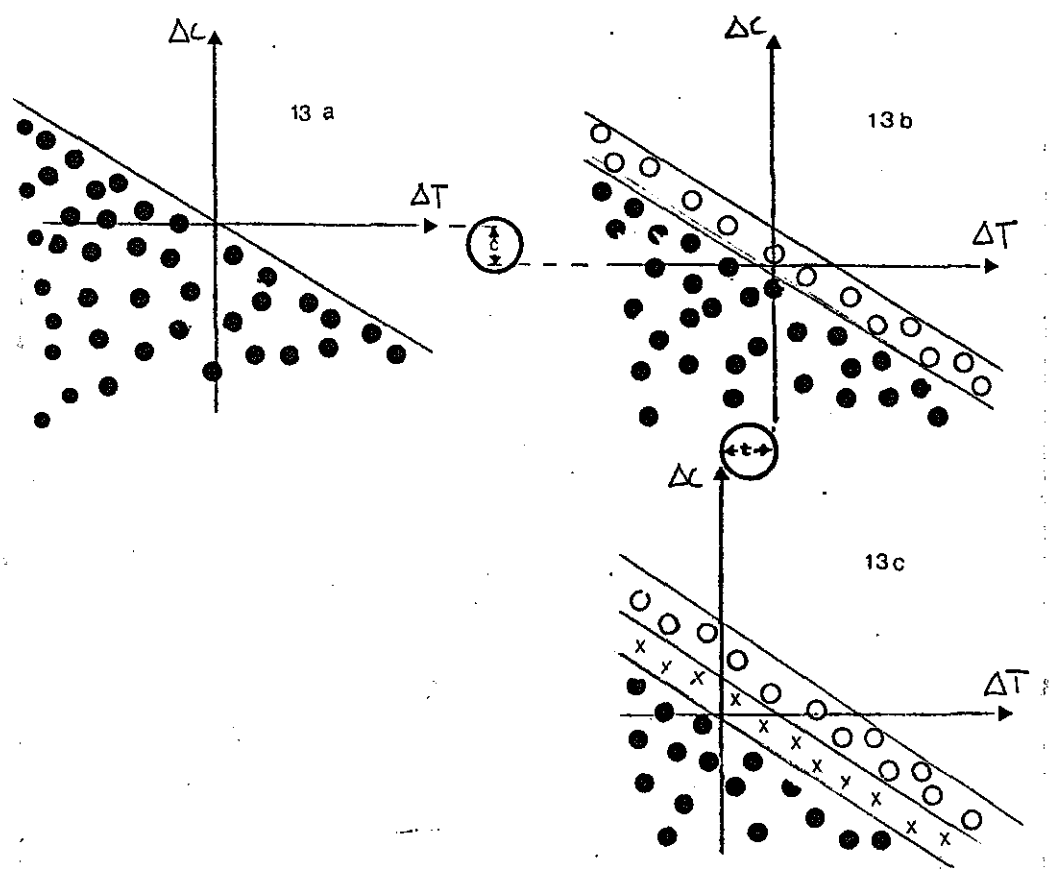
7.1.1 None of the approaches to VOT estimation based on aggregate data (discounting the conventional aggregate choice data, such as interaction data or aggregate mode split data) are sufficiently well defined to permit the sort of standard statistical appraisal that we have attempted for revealed preference data and transfer price data. The analyses of such data is normally based on the usual regression techniques, and the usual checks made for any departures from the basic assumptions about error structure that would call for adaptation of the approach.

7.1.2 Of particular interest, by virtue of being perhaps the most soundly based in theory, is the 'ratio of elasticities' approach. It is interesting to speculate on the likely sources of error in such an approach, and how these can best be controlled.

## 7.2 The ratio of elasticities approach

7.2.1 This approach can be illustrated on the 'Beesleygraph' diagram as shown by figures 13a, b and c.

FIGURE 13



The basic assumption is that the population of people choosing some given transport option (in practice, usually a particular bus route or rail line) have a decision that is affected by the cost and time characteristics of that option. Taking any individual, were the cost to rise, there would come a point at which he would no longer use that optio n (whatever his alternative might be. Note that for simplicity we shall ignore the possibility that 'patrons' can vary the amounts of the option they purchase - for example, by getting off sooner and walking.) Similarly, were the time taken to rise, there would come a similar point at which he would no longer take that option. Thus a point on a plane with axes $\Delta C$ and $\Delta T$ could be plotted for that individual, if we knew these quantities (which of course we do not). Figure 6a illustrates the notional representation that might then be formed of the existing 'patronage' of the option. We might similarly represent a set of 'shadow' points for all possible future 'patrons', locating each at points of cost and time decrease in the option just sufficient to induce them to choose it (instead of whatever else they may be doing).

7.2.2 Figure 13b illustrates what happens when the cost of the option is increased by c units: all 'patrons' in the area marked with 'o's' now stop using the option. Similarly, figure 13c illustrates what happens when the time characteristic of the option increases by t units: all 'patrons' in the area marked with 'x's' now cease to choose the option.

7.2.3 Working Paper 6 has shown that the areas of the regions excluded at each step are in proportion to the 'value of time' in this simple example, and that more generally with a mixed population with a range of values of time, the ratio of the areas is some sort of weighted average of the various values of time in the population. The 'ratio of elasticities' method is normally based on either the absolute NUMBER of 'patrons' excluded (or attracted, since an identical argument follows for the shadow 'future patrons') or on a WEIGHTED SUM corresponding to these excluded patrons. Commonly, 'miles travelled' by the option or 'fare paid' on the option might be used in the case of a public transport service.

7.2.4 Thus the condition for accurate estimation of values-of-time from this method is that the densities of either 'patrons', or 'patron x miles travelled' or 'patronsx fare paid' should be identical in the two exclusion regions, depending on which definition is used. None of these seem especially likely, although there seems no reason to expect any particular systematic variation in densities either. In other words, this problem will be a source of error in the measurement of elasticities, and without any information about the likely size of that error, no one elasticity measure would be very useful. On the other hand, it seems likely that if enough of these are available, we shall be able to form consistent estimates by forming an average.

# 8. CONCLUSIONS AND RECOMMENDATIONS

## 8.1 Introduction

8.1.1 Amongst his nine conditions for accurate measurement of values-of-time, Harrison (1974) gives the following four

a) The variables considered relevant must not be too closely correlated.

b) The variables affecting choice must show a fair amount of variation in the sample.

c) The sample analysed must show a reasonable proportion choosing each of the relevant options.

d) As a check on validity, the number of choices explained by the analysis must be high.

8.1.2 In sections 5, 6 and 7, working on the basis of a simple model, we have shown that conditions a), b) and c) are all interconnected, and outlined how we may set about quantifying the 'not too closely' of a), the 'fair amount' of b), and the 'reasonable proportion' of c), all taken simultaneously. In section 4 we have considered what the criteria should be to judge what is a 'high' enough level of explanation to give assurance on model validity, and outlined a general approach in the context of one particular criterion.

8.1.3 The factors that are under our control in the conduct of any choice experiment are

a) context

b) sample size

c) composition of sample.

The VR criterion provides a guide on all these aspects; for example, it has already been remarked in 5.3 that contexts in which time and money differences are the most important factors in the choice (i.e. where they dominate any random variation) will be most suitable and that no measurement is possible if time and money differences are perfectly correlated. These sort of considerations militate against trip distribution (and for mode split) as an experimental context, even before any considerations of the difficulties in properly specifying the model.

8.1.4 As far as the composition of the sample is concerned, it seems clear from the VR expression that different decision 'points' make very different contributions to the estimation accuracy of the model parameters and thus the value-of-time. There appears to be scope for increasing precision or reducing sample size requirements by the use of survey techniques which will result in a sample rich in observations in the appropriate regions of the decision plane, (using the Beesleygraph representation of the problem). We have not discussed survey techniques in this paper; however, it is by the use of such techniques as variable sampling within strata defined in terms of access to different modes (i.e. on a geographical basis) or choice-based sampling strategies (see Manski and McFadden 1980) that we would try to achieve this 'sample enrichment'.

8.1.5 In Section 5, we have commented on the fact that, under the simple model considered, a design which would provide a sample optimal for revealed preference analysis would not be optimal for transfer pricing experimentation. In the one example we have considered, revealed preference data seems to be most powerful in establishing values of time at points at which the proportions choosing each option split around 20% to 80% (although we must be cautious about placing too much reliance on this observation until it has been verified as a general rule). Transfer price data, on the other hand, would increase in power as the experimental points became more extreme.

8.1.6 Note here that economic realism would probably discourage us from using any one form of function, let alone a simple linear one such as we have been considering, over too wide a range of characteristics of options. The potential advantages of the transfer price approach would be restricted by the range over which our model might reasonably apply.

## 8.2 Sample size

8.2.1 It will be clear from the preceding sections that there is no simple rule for calculating the sample size that is required to give a preset degree of confidence that a measurement of value-of-time can be made to a specified precision. The VR criterion could serve such a purpose for the simple model we have been considering, but that could not be used without prior information about model coefficients and information about the joint distribution of time differences and cost differences over the population (the 'locations' on the Beesleygraph).

8.2.2 Both of these items of information could be estimated from previous studies; however, such a process would take resources which are beyond those currently available to us. Accordingly, we would recommend that such a study be undertaken at an early stage of any further work that is undertaken on value-of-time estimation. For our current purposes, we can use the qualitative insights provided by this approach to guide our interpretation of such information as is available about sample size requirements from other sources.

8.2.3 Firstly, we note that Daly and Zachary (1975) estimate the sample size requirement to measure values-of-time to ± 35% with 90% confidence as 2000 individuals, extrapolating from the accuracy they achieved with some 542 individuals. Very similar levels of accuracy were achieved by Ortuzar (1980) on a similiar sized data set, so we shall tentatively assume that this sort of accuracy is representative, at least of U.K. mode split surveys on the basis of random samples of travellers.

8.2.4 We can then use this assumption to make rough estimates of accuracy for different sample sizes; for example, at the 90% confidence level, we would have measurement to within

± 70% from a sample of size 500
± 50% from a sample of size 1000
± 35% from a sample of size 2000

8.2.5 The statement of the confidence limits at the 90% level follows the lead of Daly and Zachary : using the more familiar 95% level, the figures would be

± 85% from a sample of size 500
± 60% from a sample of size 1000
± 45% from a sample of size 2000

8.2.6 Measurement with this sort of imprecision would make the task of distinguishing between values of time in different contexts very difficult indeed, unless we have extremely large samples of the variation in the values is very large. However, as we have seen, there is scope for improving the precision of the estimates for a given sample size by careful choice of survey method, in such a way as to concentrate on those individuals facing options of a sort most informative about rate at which time and money would be traded. We have now established a way to estimate the extent of the possible improvement given access to existing data sets.

8.3 Revealed preference and transfer pricing

8.3.1 In this paper we have contrasted the information content in revealed preference data sets with that in transfer pricing data sets, glossing over the basic difficulty that the transfer price data may simply not represent the continuous measure of utility difference that we have described. It is certainly clear from the literature that there may be grave problems in framing transfer price questions in an unambiguous way which can be sensibly answered. However, from the viewpoint of statistical power, it is obvious that this sort of data offers the possibility of reducing the sample sizes required from thousands to hundreds; were this the only advantage of the method, this would still be a sufficiently compelling reason to give priority to the tasks of developing the art of collecting the data and establishing the degree of reliance that could be put on the results. As we have seen, when applied to real choices the method can be seen as adding information to the basic data of the revealed preference. This additional information may be worthless, or it may be absolutely reliable, but most probably it will be at neither of these extremes; we recommend that priority be given to establishing

a) sound ways to collect transfer price data
b) the reliability of that data
c) the resulting requirements in terms of methods of statistical analysis.

# REFERENCES


BATES, J.J., Gunn, H.F. and Roberts, M. (1978). "A disaggregate model of household car ownership". Dept. of the Environment, Research Report 20.

BRUZELIUS, N. (1980) "The Value of Travel Time".

COCHRAN, W.E, (1950) "The comparison of percentages in matched samples". Biometrika 37.

CROW, E.L., Davis, F.A. and Maxfield, M.W. "Statistic manual" Dover Publications Inc., New York.

DAGANZO, C. (1979) "Multinomial Probit". Academic Press, New York.

DALY, A. (1977) "Cost influences on car-ownership and use". P.T.R.C. Summer Meeting, P21.

DALY, A. (1980) "Applicability of disaggregate behavioural modelling". Paper presented at Disaggregate Model Conference, I.T.S., University of Leeds, Leeds.

DALY, A., and Zachary, S. (1975) "Commuters' Values of time". L.G.O.R.U. Report T55.

D.O.E. (1976) "The Valuation of Travel Time".

ECONOMICS and Statistics Note 22, Dept. of the Environment, London.

FOERSTER, J. (1979). "Mode choice decision process models : a comparison of compensatory and non-compensatory structures". Transpn. Rev. 13A.

FORD, I. (1976) Ph.D. Thesis, University of Glasgow.

GUNN, H.F., Kirby, H.R., Murchland, J.D. and Whittaker, J.C. (1980). "R.H.T.M. Trip Distribution Investigation". Department of Transport London.

GUNN, H.F. and Bates J.J. (1980) "Statistical Considerations in fitting travel demand models". Paper presented at Disaggregate Model Conference, I.T.S., University of Leeds, Leeds.

GUNN, H.F. and Whittaker J.C. (1981) "Estimation errors for well-fitting gravity models". W.P. 149, I.T.S. - University of Leeds, Leeds.

HARRISON, A.J. (1974) "The economics of transport appraisal". Croom Helm, London.

HAUSER, J.R. (1978) "Testing the accuracy, usefulness and significance of probabalistic choice models : an information theoretic approach". Operation Research, 26(3).

HENSHER, D.A. (1976) "The Value of commuter travel time savings". Journal of Transport Economics and Policy, 10.

HOROWITZ, J. (1980) "Specification tests for probabalistic choice models". Paper presented at Disaggregate Model Conference, I.T.S., University of Leeds, Leeds.

LEE, N. and Dalvi, Q. (1969 and 1971). "Variations in the value of travel time". Manchester School, vol. 37.

MANSKI, C.F. and McFadden, D. (eds.) (1980). "Structural analysis of discrete data with econometric applications". M.I.T. Press, Cambridge, Mass.

MANSKI, C.F. (1975) "Maximum score estimation of the stochastic utility model of choice". Journal of Econometrics, 3.

ORTUZAR, J. de D. (1980) "Modal choice modelling with correlated alternatives" Ph.D. Thesis, University of Leeds, Leeds.

SILVEY, S.D. (1980) "Optimal Design" Chapman & Hall, London.

STOPHER, P.R. and Meyburg, A. (1976) "Behavioural travel demand models". Lexington, Mass. Lexington Books, D.C. Hecith & Co.

WILLIAMS, H.W.C.L. and Ortuzar, J. de D. (1980) "Travel demand and response analysis - some integrating themes". Paper presented at Disaggregate Model Conference, I.T.S. University of Leeds, Leeds.