Mathematical Techniques for Estimation of the Value of Travel Time Savings: A Review

Occasional Paper

This Paper has been prepared as part of an investigation into the value of travel timesavings. It follows on from the work reported in the Bureau of Transport Economics Occasional Paper 51 which found that little confidence could be attached to the currently available values.

Mathematical Techniques for Estimation of the Value of Travel Time Savings:

A Review

G.W. King

AUSTRALIAN GOVERNMENT PUBLISHING SERVICE, CANBERRA 1983

ISBN 0 644 02525 5

Printed by Watson Ferguson and Co., Brisbane

FOREWORD

This paper has been prepared as part of an investigation into the value of travel time savings. It follows on from the work reported in the Bureau of Transport Economics Occasional Paper 51 which found that little confidence could be attached to the currently available values.

This paper reviews the methods available for estimating the value of travel time savings and modelling traveller behaviour. The assumptions and limitations are compared to identify a preferred method. This forms the basis for Bureau of Transport Economics fieldwork leading to the estimation of the value of travel time savings.

This paper was prepared by Dr G.W. King under the supervision of Mr D.R. Scorpecci.

R. W. L. WYERS
Assistant Director
Special Studies Branch

Bureau of Transport Economics
Canberra
May 1983

CONTENTS

TABLES

FIGURES

SUMMARY

A decrease in travel time is usually one of the benefits resulting from improvements to a transportation system. For cost benefit analysis of project proposals it is necessary to estimate the value of travel time savings. Methodologies which may be used for this purpose are reviewed in this paper with the aim of comparing the assumptions and limitations of the formulations and solution techniques used.

The methodologies are divided into aggregate and disaggregate models. The aggregate models discussed are the gravity, entropy and abstract mode models. These are limited by their restricted behavioural basis and use of data describing zonal and inter-zonal characteristics. Any values produced by these models are calculated at the mean of the sample.

The many disaggregate models are mostly based upon theories of individual decision making behaviour. This paper divides the models into simple, multinomial, general behavioural and functional measurement models.

The simple disaggregate models are obtained by describing the results of travel decisions in terms of choice functions that are linear combinations of explanatory variables. The models considered are the linear probability, linear logit, probit and discriminant analysis models.

The multinomial logit model is the most commonly used multinomial formulation. Its linear functional form restriction can be relaxed by the use of transformations. The necessity of obeying the independence from irrelevant alternatives axiom can be removed by the dogit model formulation. The multinomial logit model can be modified to describe more general decision structures giving rise to the conditional, sequential and recursive sequential models.

The multinomial probit model is a generalised alternative to the multinomial logit model. It allows taste variations between individuals and covariance between alternatives, but requires considerable computation.

The general behavioural models use different theories of decision making. The elimination-by-aspects model assumes a search across attributes in order of importance and the satisficing model a search across alternatives. The priority evaluator method is used to identify the trade-offs involved in multi-attribute decision making.

Functional measurement is a technique used to obtain interval scaled responses to alternative trips and to determine the functional form of the choice function used. This is the only approach which allows cross-product and power terms in the function.

When applying any model the sources of error, aggregation of disaggregate results and the updating and transfer of results should be considered. This paper discusses each of these issues and the precautions which may be taken.

The desirable properties of models for estimating the value of travel time savings are listed and used to compare the models reviewed. None of the models has all the properties, but functional measurement comes closest. For situations in which aggregate results are acceptable the elimination-by-aspects and satisficing models should be considered.

CHAPTER 1—INTRODUCTION

Among the benefits resulting from improvements to the road system is, usually, a decrease in travel time. BTE (1982) found that various authors identified these time savings as being between 29 and 80 per cent of the total benefits arising from transport investment.

The motivation for this paper arose from the need to estimate the value of travel time savings for use in cost benefit analysis of proposed transport projects, as BTE (1982) found considerable variation in the estimated values in an extensive literature search. There are a large number of potentially useful techniques available to estimate the value of travel time savings, but no comparative review of the assumptions in each methodology is available.

The two broad approaches to estimating the value of travel time savings resulting from changes to transport systems are:

- the value of the marginal productivity of working time; and
- the values implicitly used in consumer decision making behaviour.

The marginal productivity approach considers that travel is not an end purpose itself, but results from the pursuit of productive activities. Therefore there is an opportunity cost associated with travel since time savings are available for other economic activities. The opportunity cost of time savings is determined by measuring the value of any additional production resulting from decreases in travel time during working hours. This measure is only appropriate for time savings made in the employer's time, and assumes that the time savings are used to provide additional production. This requires the determination of the uses to which savings in travel time could be put, and an estimation of the economic value of the potential additional production. Often it is difficult to determine if any additional production results from travel time savings.

The usual rationale used to value additional production, and hence travel time savings, is that people will work and employers will hire labour as long as its value in production is greater than its cost. Therefore, at the margin, the average wage rate is a measure of the value of the additional production. This is only true in a perfect economy because minimum-wage and maximum-hours legislation, overtime payments, payroll tax etc distort the appropriateness of the wage rate. Also any production increases due to the impact of road improvements are generally confounded with increases due to changes in productivity and technology in the workplace, as well as institutional and maintenance changes in the road system.

Even when these difficulties are overcome, the value of marginal productivity approach is only suitable for working hours valuation of travel time savings. Time savings have a social value when not linked to production, due to the release of time for other activities which do not have a productive use.

The consumer decision making behaviour approach is concerned with the values individuals implicitly use when deciding between alternative trips where they can trade-off cost, time, comfort, privacy, safety, reliability, convenience etc and is not restricted to working hours. The time traded-off may consist of a number of components with different values. For example, waiting time, walking time, in-vehicle time and parking time. The cost includes not only the direct operating costs or fares, but could include parking costs, tolls and standing costs depending on the situation and attitude of the individual making the decision. The behavioural value of time is the market value of time as reflected by consumer behaviour and is inferred from models of decision making behaviour.

This paper addresses only the behavioural value of time because of the intrinsic problems that exist in trying to estimate the value of marginal productivity resulting from time savings, particularly when the time saved does not have a readily recognisable productive use. Also the behavioural value of time savings is not restricted to working hours situations and hence is more generally applicable.

This paper discusses the commonly available choice models and the assumptions upon which they are based. Consumer choice models can be developed via either aggregate or disaggregate formulations. Aggregate models use information which is only based on zonal characteristics and produce average (over all individuals) values of travel time savings for trip options. Disaggregate models are based upon theories of individual decision making behaviour, but can produce aggregate or disaggregate results depending on the model and solution technique chosen.

There are many disaggregate models which can be used. This paper divides the models into simple, multinomial, general behavioural and functional measurement models. These are discussed with particular emphasis on the assumptions and limitations of the formulation and solution technique used.

Chapter 9 discusses some of the items which should be considered when using disaggregate models. The main items are sources of error, the aggregation of disaggregate results and the updating and transfer of models over space and time.

The issues which should be considered when selecting a choice model for application are discussed in Chapter 10. The choice models described in this paper are then compared with respect to their important assumptions, limitations and characteristics.

CHAPTER 2—AGGREGATE AND DISAGGREGATE MODELS

It is common to divide behavioural models of travel decisions into aggregate and disaggregate models. A disaggregate choice model is 'a model which describes individual choice amongst a finite number of discrete alternatives as a function of a number of variables defined and measured at the same individual level' (Ruijgrok 1979). The term disaggregate is related to both the choice itself and the measurement of explanatory variables.

In constrast, aggregate models describe the overall behaviour of a group of people. The explanatory variables usually relate to the group under consideration rather than the individuals.

Even when disaggregate models are used, the final objective is to provide reliable estimates of the behaviour of the population being studied as a group rather than a lot of individual predictions. The terms aggregate and disaggregate are relative, as the level of division of a population for a disaggregate model depends upon the degree of detail required. It will be shown later that disaggregate models do not necessarily provide disaggregate parameters for a function describing the choices of individuals but, due to the solution techniques used, provide aggregate results.

Some disaggregate models can be used with aggregated data making them aggregate models when used this way. In general this will give results different from those obtained if a disaggregate model was used with the same data before it was aggregated. This is because the value of a parameter in average circumstances differs from the average value of the parameter over all circumstances.

Hence for a model to be useful in evaluating the value of travel attributes to individuals, it should model the choice function of the individual and provide parameter values for the individual. Then the value of travel attributes for each individual surveyed can be estimated and the results accumulated, with appropriate weights, to reflect the population of interest. Alternatively it could produce aggregate results, but with additional distribution information.

Hensher and Hotchkiss (1974) list the following as the basic weaknesses of aggregate models.

- Transport system characteristics are poorly handled at the level of aggregation used. Such characteristics can show wide variations within a zone and the use of an average figure does not account for that variation.
- When aggregate models are used to estimate elasticities and the value of time savings, these are calculated at the mean of the sample.
- Many aggregate models are extensions of the gravity model such that in the process of calibration an equilibrium relationship is explicitly established by assuming the stability of the transport system characteristics which the transport planner wishes to vary.

Aggregate models also require large amounts of data to be accurate. Gunn et al (1980) found that to provide reasonable estimates of the value of time savings, at least 50 per cent of the traffic in their study would need to be sampled.

Gordon et al (1979) describe disaggregate models as seeming to have the following six properties:

- individual behaviour is studied directly;

- an available body of choice theory can be invoked;
- a wide set of modes can be considered;
- modern estimation techniques are employed;
- the models are less cumbersome and more operational than traditional methods; and
- being sensitive to a wider variety of modal attributes, these models can test the impact of a wide variety of transportation policy issues.

This all is damning criticism of aggregate methods, but Gordon et al conclude that they do not believe that all six properties are actually fulfilled by disaggregate models. Also they believe that when the results of a disaggregate model are applied, the forecasting system suffers from a wide variety of faults so the greater accuracy of some parameters may not improve the overall system. Hence, there may be a place for aggregate models despite their disadvantages.

CHAPTER 3—A BRIEF DISCUSSION OF THREE AGGREGATE MODELS

The most commonly used aggregate models are the gravity, entropy and abstract mode models. Gravity and entropy models are used to describe aggregate spatial patterns of activities and the respective theories are analagous to the physical theories they are named after.

GRAVITY THEORY

Gravity theory describes the spatial interaction between two or more points in a manner analogous to the gravitational attraction between bodies in a physical system. The basic form is:

$$t_{ij} = \gamma P_i P_j S_{ij}^{-2} \tag{3.1}$$

where t_{ij} is the flow from point i to point j (degree of spatial interaction), γ is the constant of proportionality, P_i and P_j are the populations of i and j respectively and S_{ij} is the distance between i and j. This model can be generalised to:

$$t_{ij} = \gamma P_i P_j S_{ij}^{-\beta}, \ \beta > 0 \tag{3.2}$$

so that t_{ij} is an unweighted geometric average of P_i and P_j.

The additivity condition that the sum of the flows should equal the total flow is not satisfied by unmodified gravity models. This condition implies that:

$$\sum_j t_{ij} = o_i, \ \forall_i \tag{3.3}$$

and

$$\sum_i t_{ij} = d_j, \ \forall_i \tag{3.4}$$

where o_i and d_j are the flows at the origins and destinations respectively.

The gravity model, equation 3.2, only satisfies these conditions if:

$$t_{ij} = a_i b_j o_i d_j S_{ij}^{-\beta} \tag{3.5}$$

where:

$$a_i = (\sum_j b_j d_j S_{ij}^{-\beta})^{-1} \tag{3.6}$$

and

$$b_j = (\sum_i a_i o_i S_{ij}^{-\beta})^{-1} \tag{3.7}$$

Without the additivity condition a doubling of all o_i's and d_j's would quadruple the flows rather than double them.

With known values of t_{ij}, S_{ij}, o_i, and d_j the other parameters of the model (a_i, b_j, β) can be calculated by recursion techniques from equations 3.5 to 3.7 (Nijkamp 1979).

Although the gravity model would appear to have no behavioural basis, equation 3.2 can be derived from a simple hypothesis for spatial interactions (Isard 1960). Namely, that the expected number of trips of an individual from i to j is proportional to the population of j but that the ratio of the actual number of trips undertaken from i to j by the whole population of i to the expected number of trips by the population of i is an inverse function of distance.

ENTROPY THEORY

Normally spatial systems are complex and show a high degree of uncertainty. Entropy theory states that the most likely equilibrium stage of a closed system is that which maximises the number of combinatorial possibilities. This theory identifies the most likely equilibrium of a spatial system in a manner analagous to physical systems.

Nijkamp (1979) shows that an entropy theory formulation yields the result for unknown t_{ij}'s:

$$t_{ij} = a_i\, b_j\, o_i\, d_j\, e^{-\beta C_{ij}},\ \beta > o \tag{3.8}$$

where:

$$a_i = (\sum_j b_j\, d_j\, e^{-\beta C_{ij}})^{-1} \tag{3.9}$$

$$b_j = (\sum_i a_i\, o_i\, e^{-\beta C_{ij}})^{-1} \tag{3.10}$$

The transportation unit cost from i to j, C_{ij}, is constrained to travel budget C so that

$$\sum_i \sum_j C_{ij}\, t_{ij} = C \tag{3.11}$$

The system of equations 3.8 to 3.10 is similar to equations 3.5 to 3.7 with S_{ij} and C_{ij} performing similar roles as distance/cost parameters. An increase in distance and hence cost for an origin-destination pair decreases the flow between them.

There is, *a priori* , no reason to believe that people behave like molecules in terms of gravitational attraction or achievement of entropy levels. Alternatively the two models can be derived from specific utility approaches (Nijkamp 1979) instead of from a physical analogy. This means there is some restricted behavioural foundation to the models.

ABSTRACT MODE MODEL

The abstract mode model (Quant and Baumol 1966) is based on the assumption that cross-elasticity between modes exists only with respect to 'best' variables. The abstract mode model has the form:

$$t^k_{ij} = F\,[L^k_{ij}, L^b_{ij}, SE_i, SE_j] \tag{3.12}$$

where t^k_{ij} is the number of trips between i and j by mode k, L^k_{ij} are the level-of-service variables for the k^{th} mode divided by the respective 'best' value between i and j, and L^b_{ij} are the 'best' values of each individual level-of-service variable between i and j, regardless of which mode contains the 'best' value for any given variable; this vector constitutes the 'abstract mode' as in general no such mode will exist. The socio-economic characteristics of zone i and zone j are described by SE_i and SE_j respectively.

As the level-of-service variables are normalised with respect to a 'best' value, any changes to variables which were not 'best' variables will have no effect on the travel demand predicted by the model for modes which were not changed. This can clearly limit the value of the model (Hensher and Hotchkiss 1974).

Gordon et al (1979) stress that two attributes a modified abstract mode model possesses, which disaggregate mode choice models do not have, are the ability to predict trip creation and trip diversion.

VALUE OF TRAVEL TIME SAVINGS USING AGGREGATE MODELS

The aggregate models discussed (except the abstract modes model) do not explicitly contain parameters relating to time. These are included by modifying the distance parameter S_{ij} in the gravity model or the cost parameter C_{ij} in the entropy model to be functions of time, cost and distance. For example, a possible functional form is:

$$\left.\begin{array}{l} S_{ij} \\ C_{ij} \end{array}\right\} = \alpha_1 \text{ time} + \alpha_2 \text{ distance} + \text{cost} \tag{3.13}$$

If the model fitted to the data assumes such a form, then an estimate of the value of travel time savings can be made.

CHAPTER 4—A DISAGGREGATE CHOICE THEORY

Many of the disaggregate choice models discussed in this paper are based upon a general theory of choice. This theory assumes rational behaviour by individuals who have perfect information on the alternative travel modes available to them, and the attributes of the modes. In most models the individual is assumed to consider all the information available to him and choose his 'best' alternative. Not all models make this assumption however, and several models with alternative choice theories are discussed in later chapters.

FUNCTIONAL REPRESENTATION OF ATTRIBUTE EVALUATION

Lerman and Louviere (1978) state that there is extensive support based on theoretical and empirical research in mathematical psychology and related fields for the following assumptions of functional measurement.

Assumption 1
For any observed travel behaviour there exists a set of attributes that are functionally related to its occurrence. The level of the k^{th} attribute of alternative i is denoted z_k^i. The levels of these attributes may be physically measurable, or alternatively a qualitative factor, for example, comfort or privacy, which takes several levels.

Assumption 2
Every individual associates a corresponding value x_k^i with the level of each attribute z_k^i, so that:

$$x_k^i = f_k^i(z_k^i) \tag{4.1}$$

Assumption 3
Individuals determine their net utility U_i for alternative i, consisting of a combination of levels of the attributes, by combining the associated values of the respective attributes. That is:

$$U_i = g_i\,(X_{1i}, \ldots, X_{ki}) = g_i(\underline{x}^i)$$

$$= g_i\,(\underline{f}\,(\underline{z}^i)) \tag{4.2}$$

The vector of overall utility for all alternatives $\underline{U} = (U_1, \ldots, U_n)$ is connected to the observed travel behaviour B by means of an algebraic function, that is:

$$B = h(\underline{U}) \tag{4.3}$$

Then by substitution

$$\begin{aligned} B &= h\,(\underline{U}) \\ &= h\,(\underline{g}(\,\underline{f}\,(\underline{z}))) \\ &= W(\underline{z}) \end{aligned} \tag{4.4}$$

The composite response function W relates the levels of each attribute to the observed travel behaviour. When investigating travel behaviour it is necessary to assume a form for the composite response function, or derive one from behavioural and statistical considerations.

By using the values (utilities) individuals assign to alternatives, equation 4.2 can be used to determine the parameters of an assumed functional form of an individual's net utility function. The implied values of the attributes of the alternatives can then be calculated. This technique has been used by Louviere (1981), Louviere et al (1981) and IMG (1981) and is discussed in detail in Chapter 8.

CHOICE MODELS

Many models of behavioural choice use the following general theory of choice (Gunn et al 1980):

- individuals in the market segment (same choices and constraints) select the option with the highest net utility U_i;
- to account for unobserved factors and interpersonal variation, the utilities U_i are considered to be randomly distributed over the population being considered; and
- the probability that a particular individual selects option i is simply:

$$P_i = \text{Prob}\ (U_i \geqslant U_j;\ \forall_j = 1, \ldots, N) \tag{4.5}$$

Choice models assume that each individual q defines the net utility U_{iq}, of alternative i, in terms of the levels of attributes according to a common functional form (Hensher and Johnson 1981). That is, as:

$$\begin{aligned} U_{iq} &= g_q\ (f_q\ (\underline{z}^i), \\ &= U_q\ (\underline{z}^i), \end{aligned} \tag{4.6}$$

then U_q has the same functional form for all individuals, but has parameter values associated with each individual.

To make the problem tractable it is usual to divide the net utility U_{iq} into two components: the representative utility V_{iq}, and a stochastic residual ϵ_{iq}. Then

$$U_{iq} = V_{iq} + \epsilon_{iq} \tag{4.7}$$

The representative utility V_{iq} of alternative i, as perceived by individual q, is a function of the values z_r^i taken by the attributes of alternative i and the parameters θ_k associated by individual q with the levels of the attributes. Therefore:

$$V_{iq} = V_{iq}\ (\underline{\theta}^q, \underline{z}^i) \tag{4.8}$$

There are a number of models used to determine $\underline{\theta}^q$ and hence V_{iq}, the representative utility perceived by individual q, in disaggregate models. The models arise from different assumptions concerning the stochastic residuals.

THRESHOLDS AND RESISTANCE TO CHANGE

Most models of individual behaviour assume a continuous response to changes in the value of an attribute. This is not realistic because there may be some threshold below which an individual cannot perceive the existence of change, or the change is unacceptably small for an individual to consider changing or reconsidering his alternative.

The resistance of travellers to changing their mode due to the formation of habits, delays in receiving information on changes, or to errors in perception, results in a hysteresis effect analogous to hysteresis in magnetism. That is, there will be a different response depending both on whether costs change upwards or downwards and on whether there is a previous history of mainly upwards or downwards trends (Goodwin and Hensher 1978).

These behavioural properties require the functional form of the net utility function (equation 4.6), or the representative utility function (equation 4.8), used by the model to be carefully considered when evaluating the effects of small changes. In particular, linear models would be very suspect if used to predict responses to small changes, as their form implies a constant rate of response to all changes.

CALCULATION OF THE VALUE OF TRAVEL TIME CHANGES

The behavioural choice methodologies that use a utility maximising approach allow the value of travel time savings to be calculated from the representative utility (equation 4.8) once the coefficients have been determined. The procedure used to calculate the values of a particular time attribute is to vary the time by the amount of interest but keep the representative utility constant by changing the cost variable. The value of the amount of time is the change in cost required to balance it.

If a model is based on a linear representative utility function the value of time changes is just the cost coefficient divided by the time coefficient multiplied by the time change. Hence the value of time savings in such a model is just a constant multiple of the time changes and is independent of the values taken by other parameters or the relative size of the time change.

Non-linear models require a separate calculation for each time change because the value of the time changes is then dependent on the values of other parameters. This is a more realistic representation of the value of time savings than a linear model.

CHAPTER 5—SIMPLE DISAGGREGATE MODELS

The simplest disaggregate choice models are obtained by describing the results of travel decisions in terms of choice functions whose coefficients are estimated by appropriate statistical techniques. It is assumed in these models that the choice functions are a linear combination of the explanatory variables of the alternative modes. The explanatory variables can be the values of the attributes or some algebraic function of the attributes. Care must be taken that the explanatory variables are not correlated if a technique sensitive to correlations is used to estimate the coeffients.

This chapter briefly discusses four simple disaggregate models.

LINEAR PROBABILITY MODEL

Consider a binary choice situation, such as the choice between two competing transport modes. If it is assumed that the choice is the result of a linear function of the explanatory variables, for individual q the model can be written as (Hensher and Johnson 1981):

$$f_q = \sum_k \beta_k X_{kq} + \epsilon_q \qquad (5.1)$$

where

$$f_q = \begin{cases} 0 \text{ if first option chosen} \\ 1 \text{ if second option chosen} \end{cases}$$

X_{kq} = k^{th} explanatory variable

β_k = coefficients

ϵ_q = a stochastic error assumed to be normally distributed with zero mean and constant variance.

This model is known as the linear probability model because the estimating equation is linear and the dependent variable has as its expected value the probability that the second option is chosen. That is, the expected value of f_q is the probability that $f_q = 1$.

Given observations for Q individuals on the values of the explanatory variables and the choice made, the coefficients of equation 5.1 can be estimated by ordinary least squares regression. The coefficients can then be used to investigate the relative importance (value) of the explanatory variables and to predict the overall behaviour of the sampled population if the values of some of the explanatory variables were changed by a new transport policy. This model is an example of a disaggregate model which provides an aggregate result for the coefficients of a choice function. The coefficients minimise the errors when used to explain the behaviour of all the sampled population, not each individual.

The binary form (0,1) of the dependent variable f_q leads to restrictions of the values that can be taken by stochastic error ϵ_q. Watson (1974) shows that for the expected value of the error term to be zero as assumed, the variance of the error is not constant, a condition known as heteroskedasticity. This means that the estimated values for the coefficients obtained by ordinary least squares regression are unbiased and consistent, but they are no longer minimum variance estimates. Because the error terms are not normally distributed, the estimates of the coefficients are also not normally distributed. This makes it difficult to test the significance of the values of the coefficients and of the regression as a whole, because no confidence can be placed in the computed F-test values.

The use of weighted least squares procedure (Hensher and Johnson 1981) can provide minimum variance estimates of the coefficients. However, there is no guarantee that the estimated values of the dependent variable f_q will be between 0 and 1. This is particularly disturbing as f_q may be interpreted as a probability.

LINEAR LOGIT ANALYSIS

Logit analysis is aimed at limiting the problems which occur in the linear probability model as a result of having a dichotomous variable as the dependent variable when using ordinary least squares regression.

Let P_{1q} be the probability that individual q chooses mode 1 in a binary choice situation. The logit of P_{1q} (the logarithm of the odds of individual q choosing mode 1) is defined as:

$$\log \frac{P_{1q}}{1-P_{1q}}$$

Logit analysis assumes that

$$\log \frac{P_{1q}}{1-P_{1q}} = \sum_k \beta_k X_{kq} \tag{5.2}$$

This formulation has the advantage that the dichotomous dependent variable of equation 5.1 has been transferred to a variable in the range $(-\infty,\infty)$.

Equation 5.2 can be used with grouped (aggregated) data if observations are repeated for each given value of the explanatory variables. Ordinary least squares regression can be used to eliminate the values of the parameters by replacing P_{1q} by r_1/n, where r_1 is the number of observations choosing mode 1 out of a total of n observations. A large sample size is required to ensure approximation to a normal distribution. For small samples appropriate weights can be used with weighted least squares regression to reduce any sample induced bias (Hensher and Johnson 1981).

If grouped data is not available, or the values of the attributes of the alternative modes have different values for most individuals, equation 5.2 can be solved by maximum—likelihood estimation.

The linear logit model need not be restricted to a binary choice situation. Hensher and Johnson (1981) discuss the generalisation to more than two alternatives and the analysis of the model.

PROBIT ANALYSIS

Probit analysis is a generalisation of the linear probability model which allows threshold values for choosing alternatives. The threshold values are assumed to be normally distributed over the whole population. The probit model and the estimation of its parameters is discussed in detail in Watson (1974). A brief description of the formulation of the model follows.

The dependent variable f_q is postulated to take the values 0 or 1, depending upon the values taken by the explanatory variables X_{kq}. The index I_q, constructed for individual q, is a linear combination of the independent variables:

$$I_q = \sum_k B_k X_{kq} \tag{5.3}$$

If $\bar{I}_q$ is the threshold value of I_q for individual q

$$f_q = \begin{cases} 0 \text{ if } I_q \leqslant \bar{I}_q \\ 1 \text{ if } I_q > \bar{I}_q \end{cases} \tag{5.4}$$

The threshold values are assumed to be normally distributed over the population, N(0,1). This represents differences among individuals which are either random or the result of variables not included in the model.

DISCRIMINANT ANALYSIS

Discriminant analysis was designed to solve classification problems, with the aim of minimising misclassifications, but by a probabilistic extension of the basic technique it can be used to produce estimates of the probabilities of mode choice. The technique is only appropriate to binary choice situations where the population can be divided into two subpopulations on the basis of the choice they made.

It is usual to assume the attributes for which measurements are obtained are multivariate normally distributed and the variance-covariance matrices of the attributes are the same for both subpopulations. The discriminate function is defined as:

$$Z_{ij} = \sum_p \lambda_p X_{pij} \tag{5.5}$$

where λ_p is a weighting coefficient and X_{pij} is the value of the p^{th} explanatory variable of mode i for person j.

Discriminant analysis aims to find values of the weights λ_p such that the between-subpopulation variance is maximised relative to the within-subpopulation variance. This provides the greatest separation between the two subpopulations and hence is the 'best' contrast between the subpopulations. The 'best' values of the weights can be determined by a matrix equation (Watson 1974).

Watson shows the extension of discriminant analysis so that probability for an individual selecting a particular mode can be predicted, is no more than a special case of the logistic function in which the discriminant function is chosen as the most suitable linear combination of explanatory variables.

ASSESSMENT

In a comparison of the data requirements. statistical properties and accuracy of prediction, Watson concludes that 'on balance, logit analysis is the most appropriate tool for use in studies of travel mode choice'. Probit and logit analysis both require large samples to ensure significant tests and a normality assumption but the logit results are easier to interpret. It should be noted that when Watson drew his conclusion he had not considered the multinomial methods discussed in the following chapters.

CHAPTER 6—MULTINOMIAL MODELS

Multinomial models are disaggregate behavioural models derived from theories of consumer choice behaviour. They are also probabilistic because they assign a probability of choice to each possible outcome of a travel decision.

Multinomial models are mainly used for investigating mode choice behaviour, but as the coefficients of the representative utility function are estimated it can also be used to estimate the value of time savings. The most commonly used model is the multinomial logit model (MNL), but now that computational procedures are available, the multinomial probit model (MNP) can be used.

MULTINOMIAL LOGIT ANALYSIS

The MNL model is derived by assuming that:

- the representative utility of an alternative is a linear combination of its attributes, ie:

$$V_{iq}(\underline{\theta}^q, \underline{z}^i) = \sum_k \theta_k^q z_k^i \qquad (6.1)$$
$$= \underline{\theta}^q . \underline{z}^i$$

- the Independence-from-Irrelevant Alternatives (IIA) axiom is valid; and
- the residual terms in equation 4.7 obey the extreme value distribution, ie:

$$\text{Prob } (\epsilon_i < \epsilon) = \exp(-\exp(-\epsilon)) \qquad (6.2)$$

The assumption that the representative utility is a linear combination of the values of the attributes restricts the application of MNL to situations where higher order or non-linear terms are not significant. Also this assumes that the relative importance of the attributes in the representative utility is the same across the population. Clearly, this places a severe restriction on the applicability of the model unless homogeneous subpopulations are considered separately.

The IIA axiom states that the ratio of probabilities of choosing one alternative over another (where both alternatives have non-zero probability of choice) is unaffected by the presence or absence of any additional alternatives in the choice set (Hensher and Johnson 1981). This means that the model is not applicable if all the alternatives are not sufficiently discrete. The IIA axiom greatly reduces the complexity of estimation and forecasting procedures making it convenient to use when it is valid (McFadden et al 1977). However, undesirable consequences can result if MNL is used when IIA is invalid. McFadden et al discusses the application of diagnostic tests to determine the validity of the IIA axiom.

The residual term assumption makes this a random utility model. This does not mean that individuals maximise utility in a random manner, rather that the utility they are maximising contains some unobserved contributions.

With these assumptions, the probability that alternative i is selected by individual q:

$$P_{iq} = \frac{\exp V_{iq}}{\sum_j \exp V_{jq}} \qquad (6.3)$$

$$= \frac{\exp \sum_k \theta_k^q z_k^i}{\sum_j \exp \sum_k \theta_k^q z_k^i} \qquad (6.4)$$

which is referred to as the MNL model. Stopher and Meyburg (1975) show that this model can be derived from the IIA axiom with a fairly broad assumption concerning the representative utility function.

The MNL model is usually applied to revealed preference data. That is, data from choice situations where individuals have had alternatives to choose between and made a particular choice. A criticism with this approach is that often individuals do not know what the alternatives are, or what the values of the attributes of the alternatives are. Also, observers can have difficulty in determining what alternatives (if any) were considered by the individuals under observation unless a very clear choice situation exists.

Data can be gathered for MNL by a survey where individuals detail the values of the attributes of the choice they have made and, as they perceive or know, the values of the same attributes of the alternative(s). A difficulty can arise with this approach when the alternatives described by the individuals in the sample are not sufficiently different and the IIA axiom is violated. This survey procedure has been used successfully by Hensher and Johnson (1981).

A maximum likelihood procedure is used to determine the mean value of the attribute parameters θ_k in equation 6.4 over all individuals in the sample. The distribution of attribute parameters amongst the individuals is not determined, even though the model examines decisions by individuals. This aggregation during computation makes the transfer of results over time and space difficult, and at times uncertain (Hensher and Johnson 1981).

The MNL model has the advantage of being computationally convenient with a number of computer packages available, for example BLOGIT (Crittle and Johnson 1980).

The basic MNL model can be modified to remedy violation of IIA (McFadden et al 1977 and Hensher and Johnson 1981); this is briefly discussed later. It may not always be necessary to modify the MNL model as Horowitz (1980) has demonstrated that with moderate violations of the IIA property, the basic MNL procedure produces only small errors in the parameters it estimates.

FUNCTIONAL FORM TRANSFORMATIONS

Gaudry and Wills (1978) suggest the linearity restriction on the functional form of the representative utility can be relaxed by using the Box-Tukey transformation which is defined as:

$$(X+\mu)^{(\lambda)} = \begin{cases} ((X+\mu)^{\lambda}-1)/\lambda & \text{if } \lambda \neq 0 \\ \log\,(X+\mu) & \text{if } \lambda = 0 \end{cases} \qquad (6.5)$$

The location parameter μ is chosen to ensure that $X + \mu$ is greater than zero for all observations. When μ is equal to zero equation 6.5 reduces to the Box-Cox transformation.

These transformations are used to change the functional form of the representative utility function equation 6.1 so that

$$V_{iq} = \sum_k \theta_k^q (z_k^i + \mu_k)^{(\lambda_k)} \qquad (6.6)$$

This is then used in the MNL model instead of the linear form. Although this generalises the form of representative utility it does not allow interaction terms between attributes.

The results obtained by using the MNL model with equation 6.6 are sensitive to the values chosen for λ_k and μ_k (Hensher and Johnson 1981). Hence, if this transformation is used, an investigation of the effect of varying λ_k and μ_k would be required to find the maximum of the maximum log-likelihood functions. For a large number of parameters this could involve considerable computation. The Box-Tukey transformation is an option allowed in the BLOGIT Package (Crittle and Johnson 1980).

DOGIT MODEL

An alternative logit formulation is the dogit model proposed by Gaudry and Dagenais (1979). The model allows the IIA axiom to be violated by some pairs of alternatives but not necessarily all alternatives.

The model is a modification of equation 6.3 so that:

$$P_{iq} = \frac{\exp V_{iq} + \alpha_i \sum_j \exp V_{jq}}{(1 + \sum_j \alpha_j) \sum_j \exp V_{jq}} \tag{6.7}$$

where α_i is a non-negative parameter associated with the i^{th} alternative. When all the α_i's are zero this reduces to the MNL model. All pairs of alternatives with α_i equal to zero for each alternative obey the IIA axiom. Although the dogit model can overcome the IIA assumption of the MNL model, apart from tests by Gaudry and Wills (1979), there are no known applications to individual choice situations (Hensher and Johnson 1981). Box-Cox transformations can be used to generalise the functional form of the representative utility.

No software packages are known to exist for dogit, but Gaudry and Wills (1979) describe a maximum likelihood method which could be used. A disadvantage of the dogit formulation is that additional parameters α_i will need to be estimated. If a preliminary survey or other knowledge about the independence of alternatives is used, some of the α_i's can be set to zero hence diminishing the computational task.

MULTINOMIAL PROBIT MODEL

The multinomial probit model (MNP) is a generalised alternative to the MNL model. It is derived by assuming that residual terms in equation 6.7 are multivariate normal distributed across alternatives rather than independent and identically extreme value distributed as for the MNL model (Daganzo 1979). This allows for taste variations among individuals and for covariance between alternatives. Therefore, the error terms in the utility expressions are correlated (Hensher and Johnson 1981), so MNP relaxes the IIA axiom. MNP does not lift the restriction of the representative utility being a linear combination of the values of the attributes.

When formulated for a problem with N attributes, the MNP model requires the evaluation of an N-1 dimensional integral which cannot be reduced to an analytic expression. Although procedures for the estimation of the parameters in a MNP model exist, for example Daganzo et al (1977), Daganzo (1979) and Hausman and Wise (1978), with more than a few parameters they require excessive computer time (Gunn et al 1980). Therefore MNP provides a greater degree of flexibility than MNL but its application is limited by computational difficulties.

MORE GENERAL LOGIT MODELS

The MNL model described earlier in this chapter is a simultaneous model with the decision being made simultaneously on the values of all attributes. This section discusses generalisations which allow alternate decision structures and relax some of the assumptions of MNL.

The simplest modified structure assumes that the choices are independent, such that the joint probability of choice is a product of the individual probabilities. Separate utility functions exist for each choice and are additive to form the joint utility. Consider a three attribute decision process, IxJxK. Then the separability of the utility functions implies that:

$$U_{ijk} = U_i + U_j + U_k \quad (6.8)$$

where U_{ijk} is the joint utility of an ijk combination and U_i is the part of the utility due to i and similarly for U_j and U_k. The joint probability

$$P_{ijk} = P_i \, P_j \, P_k \quad (6.9)$$

where

$$P_i = \text{Prob}\ (U_i > U_i'; \ \forall i \epsilon \ I) \quad (6.10)$$

and similarly for P_j and P_k.

In many situations the attributes that define probabilities are not independent and mutually exclusive. In these cases a conditional model is more realistic.

SEQUENTIAL STRUCTURED MODEL

In a sequential structured model an individual is assumed to make a choice on some subgroup of attributes independent of any other choice, and then to make subsequent choices conditional on previous choices. There is an implied hierarchical structuring to the decision process with no feed back or mutual interactions between decisions. Such structures are also known as nested models. Figure 6.1 shows the possible decision structures for three attributes: the first structure is the simultaneous MNL model. Sobel (1980) shows that for 4 attributes there are 26 different decision structures.

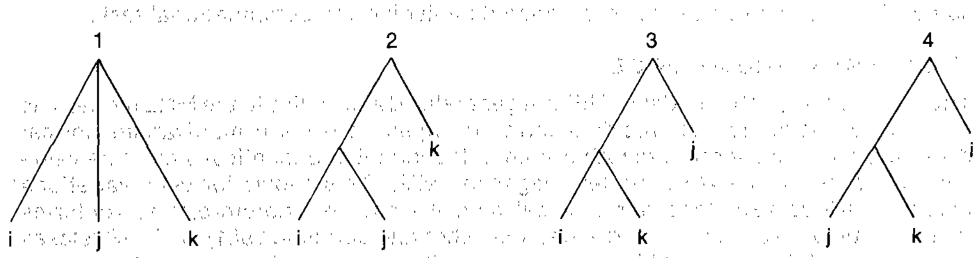

Figure 6.1. Decision structures for three attributes

The assumption of a choice hierarchy for the selection of an alternative from a three attribute decision process, IxJxK, implies the joint utility function is additive separable:

$$U_{ijk} = U_i + U_{j|i} + U_{k|ij} \quad (6.11)$$

where

U_{ijk} is the joint utility of an ijk combination.

U_i is the part of the joint utility that is independent of j and k.

$U_{j|i}$ is the part of the joint utility of a choice of j, which is independent of k, given that i has been selected.

$U_{k|ij}$ is the part of the joint utility for a choice of k given that i and j have been selected.

Therefore the hierarchy of the decision process described by equation 6.11 is a decision i followed by a decision on j conditional on i and a decision on k conditional on i and j.

The joint utility:

$$U_{ijk} = V_{ijk} + \epsilon_{ijk} \tag{6.12}$$

where the joint representative utility:

$$V_{ijk} = V_i + V_{j|i} + V_{k|ij} \tag{6.13}$$

and ϵ_{ijk} is the random component of the joint utility function. If the ϵ_{ijk}'s are independently and identically extreme value distributed, the joint logit model:

$$P_{ijk} = \frac{\exp V_{ijk}}{\sum\limits_{i}\sum\limits_{j}\sum\limits_{k} \exp V_{ijk}} \tag{6.14}$$

is obtained. Ben-Akiva and Lerman (1979) give expressions for the conditional and marginal probabilities $P(k|ij)$, $P(j|i)$ and $P(i)$ which are used in the sequential estimation process starting from the lowest level of choice. Although a particular sequence of decisions is assumed, the probabilities are not mutually interactive so:

$$P_{ijk} = P(i)\ P(j|i)\ P(k|ij) \tag{6.15}$$

RECURSIVE SEQUENTIAL STRUCTURE

The recursive sequential structure incorporates feedback into a sequential structure. At each decision level the choice is assumed to be dependent upon the previous choices in totality rather than a single alternative in the previous decision. The feedback is incorporated into the utility function as an additional variable. This model, which has a complex form, is described in Hensher and Johnson (1981).

From these structures three models called the basic multinomial-logit, the sequential nested-logit and the generalised-extreme-value nested-logit can be formulated (McFadden 1979, Hensher and Johnson 1981). The basic multinomial-logit is the MNL generalised to a sequential recursive structure. The sequential nested-logit is a modification to the basic multinomial-logit that permits pairwise attribute correlation. The final model, generalised-extreme-value nested-logit incorporates a measure to allow for the relative dissimilarity between the unobserved attributes of alternatives and non-equal variances. McFadden (1978) shows that the multinomial and nested-logit are special cases of the generalised-extreme-value nested-logit.

Although these models have been formulated, the nested-logit is the least restricted form implemented (Sobel 1980).

CHAPTER 7—SOME OTHER BEHAVIOURAL MODELS

The models discussed in Chapter 6 assume a perfectly rational utility maximising individual. Apart from the sequential logit models, the individual is assumed to base his decision on a model which is linear-in-parameters and with simultaneous consideration of the values of all attributes. Such models are compensatory in that trade-offs among attributes are possible with changes in some attributes being compensated by specific changes in one or more other attributes. That is, a high level of satisfaction with one attribute compensates for low levels of satisfaction with others. In contrast to compensatory models, non-compensatory models do not allow trade-off behaviour. This chapter discusses two models which use a non-compensatory process and one which is only concerned with identifying trade-offs.

THE ELIMINATION-BY-ASPECTS MODEL

The elimination-by-aspects model is an attribute search model in which attributes are compared in a sequence from the attribute considered most important by the individual, through to the least important. The principle used for comparing the values of attributes is that the value must exceed some standard (threshold) to be considered acceptable. The sequential processing of information is referred to as elimination-by-aspects (EBA) (Tversky 1972).

The EBA process as described by ARRDO (1981) is summarised by:

- the attributes are compared starting with the most important through to the least important;
- if the value of the attribute being considered is less than the minimum acceptable standard, the alternative that value belongs to is eliminated from further consideration; and
- steps 1 and 2 are repeated until only one alternative remains.

The second step is the EBA step. If all alternatives are eliminated rather than obtaining a single preferred alternative, the process is taken back an iteration and the alternatives remaining at that stage are chosen between by some other decision criterion.

The EBA process can be developed as either a deterministic binary choice model (Recker and Golob 1979) or a probabilistic model (Tversky 1972, Richardson 1978 and ARRDO 1981). The probabilistic model has the advantage of not requiring repeated simulations of the decisions as would be needed if a deterministic model was used.

PROBABILISTIC EBA MODEL

The following description of the probabilistic EBA model is derived from the three alternative choice problem described in ARRDO (1981). The three alternative choice problem can be represented by the Venn diagram in Figure 7.1. Each alternative is represented by a circle encompassing the attributes for which the alternative provides a minimum level of acceptability. The area of the circle for each alternative corresponds to the sum of the importance of the attributes. The areas of overlap between the circles represent attributes which are satisfactory for two or more alternatives.

The set notation used is that $\overline{x}$ is the set of attributes which are satisfactory for alternative x alone, $\overline{xy}$ is the set of attributes which are satisfactory for alternatives x and y but not z, and $\overline{xyz}$ is the set of attributes which are satisfactory for all three

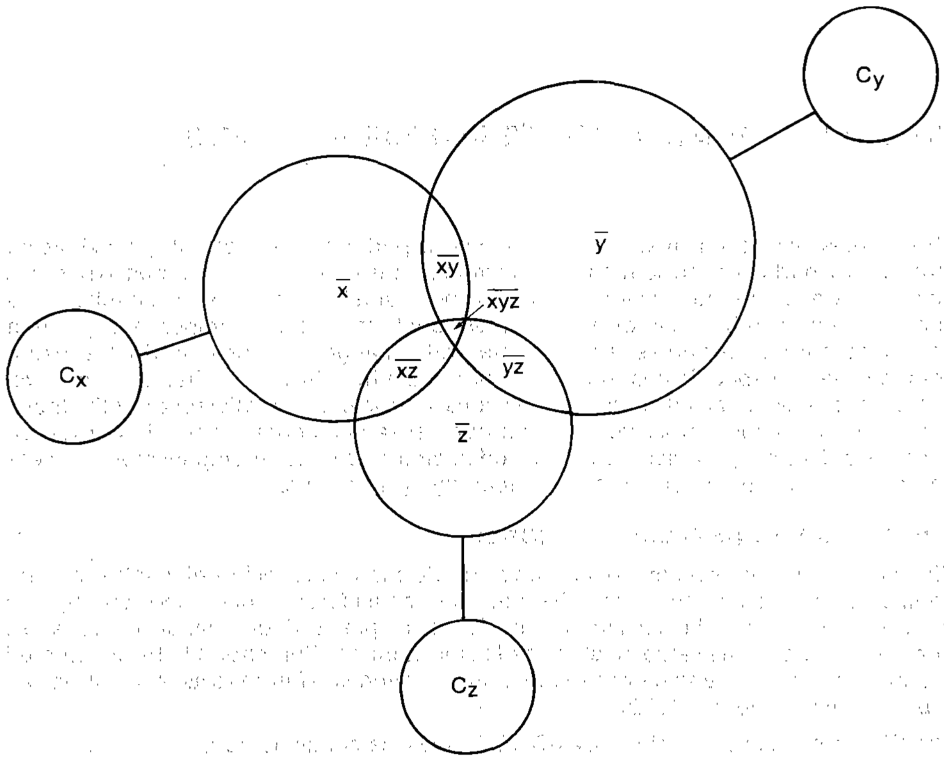

Figure 7.1. Venn diagram of alternative sets

alternatives. The sets C_x,C_y,C_z represent alternative-specific constants which are unique attributes for the respective alternatives. The alternative-specific attribute sets are mutually exclusive and must contain a non-zero value.

The area of each set I (set), is given by the sum of the importances over relevant attributes. To enable the standardisation of the importances the measure:

$K = \Sigma\ I\ (\text{set})$

is used, where the summation is over all sets (including C_x,C_y,C_z) except $\overline{xyz}$. The set is excluded as it contains attributes which are satisfactory for all alternatives, and therefore cannot eliminate any alternatives.

To derive the EBA model consider the probability of selecting alternative x. There are three ways in which x can be selected. Firstly on the first attribute examined, secondly by initially considering an attribute in $\overline{xy}$ and then choosing x over y in subsequent comparisons and thirdly by initially considering $\overline{xz}$ and then selecting x over z in later comparisons. The respective probabilities are:

$$P_1\ (x) = \frac{C_x + I(\overline{x})}{K} \tag{7.1}$$

$$P_2(x) = I(\overline{xy})\,\frac{P(x|xy)}{K} \tag{7.2}$$

$$P_3(x) = I(\overline{xz})\,\frac{P(x|xz)}{K} \tag{7.3}$$

where $P(x|xy)$ is the probability of selecting x in a comparison of x and y and $P(x|xz)$ is the probability of selecting x in a comparison of x and z. Therefore:

$$P(x|xy) = \frac{C_x + I(\bar{x}) + I(\overline{xz})}{I(\bar{x}) + I(\overline{xz}) + I(\bar{y}) + I(\overline{yz}) + C_x + C_y} \tag{7.4}$$

$$P(x|xz) = \frac{C_x + I(\bar{x}) + I(\overline{xy})}{I(\bar{x}) + I(\overline{xy}) + I(\bar{z}) + I(\overline{yz}) + C_x + C_z} \tag{7.5}$$

The total probability of selecting x is given by

$$P(x|xyz) = P_1(x) + P_2(x) + P_3(x) \tag{7.6}$$

$$= \frac{C_x + I(\bar{x}) + I(\overline{xy})\,P(x|xy) + I(\overline{xz})\,P(x|xz)}{K} \tag{7.7}$$

An attribute is considered satisfactory if it lies within specific percentage tolerance of the maximum satisfaction level for that attribute over all alternatives (Recker and Golob 1979). That is, the satisfaction S_{jkq} associated with attribute j of alternative k as perceived by individual q is acceptable when:

$$S_{jkq} \geqslant (1 - T_j)\,\underset{k'}{\mathrm{Max}}\,[S_{jk'q}] \tag{7.8}$$

where T_j is the tolerance for attribute j.

The EBA model can be solved by maximum likelihood methods. The alternative-specific attributes are required so that the probability of an alternative being chosen is non-zero. A zero probability of an alternative being chosen can occur if the chosen alternative is unsatisfactory on any attribute and either of the non-chosen attributes is unsatisfactory in all alternatives.

The EBA model is fitted to the data in two stages (ARRDO 1981). Firstly, the program tests a range of tolerances and determines the most likely tolerance for each attribute in turn. Then, once all the most likely tolerances have been determined, the optimum alternative-specific constants are calculated. The entire process is then repeated until the model is considered to have converged.

The EBA model can be used to predict the effect of changes by the calculation of attribute elasticities. The elasticities are calculated by making a change in the satisfaction level of an attribute and determining the resultant change in the predicted probability of selecting an alternative. By calculating both the cost and travel time elasticities the value of travel time savings can be calculated.

SATISFICING MODEL

The satisficing model is a sequential search model across alternatives, rather than attributes, as done by the EBA model. The theory of satisficing (Richardson 1978) involves the sequential examination of alternatives while comparing the attributes of the alternative against a set of minimum acceptable standards. As soon as an alternative which is satisfactory in all attributes is found, the search is discontinued and that alternative is chosen.

If at a later time one or more of the attributes of the chosen alternative is considered to have become unsatisfactory, a search for a new alternative will begin. The level at which the set of minimum acceptable standards is set can be lowered when satisfactory alternatives are hard to find, or raised when satisfactory alternatives are easy to find.

The randomness of the satisficing model is introduced by the order in which alternatives are searched because those searched early have a greater probability of being selected. The satisficing model can be broadened by considering the reasons why a search could be discontinued. Richardson suggests three major reasons:

- if the search costs time, money and effort the search will stop when the expected gain from examining another alternative is less than the cost of the extra examination;
- the decision maker is forced to accept the last alternative considered because he is faced with not only the cost of extra comparisons, but also the cost of rejecting the most recently considered alternative; and
- the existence of a limit on the number of alternatives which can be examined.

These rules can be extended to give a formulation of a probabilistic satisficing model from which the elasticities and the relative values of the attributes can be calculated.

Richardson (1978) compares the results of a basic EBA model, a satisficing model and the multinomial logit. He found that they gave different results but he was unable to decide which was the more realistic choice process. He concluded that EBA models are appropriate when many characteristics are considered, or when there are likely to be dependent choices (IIA disobeyed). Whereas, the satisficing model may be more appropriate when there are many alternatives to be considered and a complete search is not expected.

PRIORITY EVALUATOR

The priority evaluator method is concerned with identifying the trade-offs individuals are willing to make by examining the priorities assigned to competing and costed alternatives (Hoinville and Berthoud 1970).

The basic procedure is to present subjects with a range of values (usually three) of attributes with a cost associated with each value. The subjects are given a limited amount of money (tokens) to spend in order to choose their best combination of values of the attributes that they can achieve within the limited budget. This method has been used to identify the relative importance of environmental factors such as noise and pollution, and social factors such as accessibility (Hoinville and Berthoud 1970), as well as, the importance of cost, time in different activities and characteristics such as type of journey, reliability and seat availability (Hoinville and Johnson 1971, Wildermuth 1976).

Although this method has been applied to several transport situations, a difficulty associated with it is the necessity to provide realistic relative costs between the values taken by all attributes. Even when obtained these relative costs may substitute for costs as perceived by individuals, or are based upon information subjects would not usually consider. Its advantage is that it provides travellers with a restricted travel budget and identifies what they are willing to pay to change the attributes and their relative importance. The changes in the transport system presented to subjects should include values which describe their current journey but it would appear impractical to make the budget allocated reflect the amount the subject is willing to pay to change the values taken by some travel attributes.

CHAPTER 8—FUNCTIONAL FORM OF UTILITY FUNCTIONS

In the discussion of disaggregate choice theory in Chapter 4 the basic assumptions of functional measurement and the resulting functional relationships (equations 4.1 to 4.4) were introduced. The disaggregate models discussed above generally assume a functional form and then go on to estimate the parameters of the assumed form. Most travel demand models, as a result of computational difficulties, are restricted to specifications which have linear parameters. The Box-Cox transformations discussed in Chapter 6 allow insight to be obtained into whether linear or logarithmic transformations are appropriate.

The arguments concerning function form are summarised in the following statement by Lerman and Louviere (1978):

> This restriction (of a linear functional form) is not particularly burdensome if one already knows that a particular nonlinear specification is appropriate, since, by judicious use of piece-wise linear forms and nonlinear transformations of the dependent and independent variables, one can approximate most nonlinear parameter functions fairly well. However, lacking guidance as to the appropriate functional form, and given, with existing techniques, the virtually infinite number of candidate transformations, choosing among specifications on goodness-of-fit criteria is far more likely to lead to one of the numerous incorrect specifications than the correct one.

The determination of the functional form of decision functions is the province of mathematical psychology (psychometrics). The areas which are of potential interest for transport modelling are functional measurement, information integration theory, conjoint analysis and direct utility assessment.

FUNCTIONAL MEASUREMENT

The aim of functional measurement is to provide methods for measuring cognitive quantities on interval scales. Anderson (1976) describes an interval method for functional measurement which evolved from a general theory of information integration. The following discussion is a summary of Anderson's work; a complete description is contained in Anderson (1981).

The linear model

Suppose two stimulus variables are thought to combine in an ordinary Row x Column factorial design, so that each cell of the design corresponds to a pair of stimuli. Subjects judge the product on overall desirability by assigning a number to the stimulus combination in each cell. These numbers are assumed to obey the simple linear model:

$$R_{ij} = C_o + R_i + C_j \tag{8.1}$$

where R_{ij} is the response to the stimulus of the i^{th} subjective level of the rows R_i, and the j^{th} subjective level of the columns C_j. C_o is a constant which allows for an arbitrary zero in the response scale.

If the linear model is correct, and if the response is an interval scale, then a two-way graph of the data will appear as a set of parallel lines. If the parallelism is observed, it not only supports the linear model, but also indicates that the response measure is an interval scale. It can also be shown that the row means of the design are interval scale estimates of the subjective values of the stimuli, and similarly for the *column* stimuli.

The multiplying model

In some situations a multiplying rule for the integration of two stimulus variables may seem appropriate. A Row x Column factorial design is again used. The numbers assigned by subjects are assumed to obey the multiplying model:

$$R_{ij} = C_o + R_i C_j \tag{8.2}$$

where the parameters have the same meaning as for the linear model.

Suppose that the subjective values C_j are known, and are plotted on the horizontal axis. If the multiplying model is correct and the response measure is an interval scale, then the data in row *i* of the design will plot as a straight line against C_j with slope R_i. Thus, the two way graph of data will plot as a linear fan of lines.

In general the subjective values C_j will not be known, but if the model is correct and the response measured is an interval scale, then the column means are interval scale estimates of the subjectives values of column stimuli. Therefore, the column means may be used as provisional values of C_j to test the linear fan prediction.

There are three qualifications on functional measurement:

- parallelism may be obtained even though the linear model is invalid;
- it depends entirely on the empirical validity of the integration function; and
- more than one integration model could imply the same pattern.

The first qualification arises because if both (not just one) of the premises are incorrect, then parallelism could be obtained, but only in very special cases. If only one premise is wrong parallelism will not be obtained.

Although the assumption that the response is interval scaled can be overcome, by linearising a rank-order scale, it is necessary that the integration model be empirically valid.

An important instance of two models having the same pattern occurs under certain conditions when the linear adding and averaging models both imply parallelism. So observed parallelism would not discriminate between the models.

The averaging model

The averaging model is used in the information integrating theory of attitudes. The hypothesis is that when an individual receives additional information, this is averaged with his prior attitude A_o to produce the new attitude:

$$A = w_o A_o + \sum_i w_i s_i \tag{8.3}$$

where s_i is the scale value of the i^{th} piece of information. The weights w_o and w_i sum to one and represent the relative importance of the prior attitude and the new pieces of information respectively.

The distinguishing test between the linear adding and averaging models is called the cross-over test by Anderson. Basically the cross-over test uses the same two-way graph as the parallelism test for the linear adding model, but looks for some of the lines crossing over. The occurrence of a cross-over indicates the suitability of the averaging model for describing the responses being considered.

THE RATING METHOD

The rating method is designed so that subjects respond in a manner suitable for use with the above theory. The rating scale is limited to a fixed range of responses with experimental precautions to restrict bias in the response.

Louviere (1978) and Anderson (1981) recommend the inclusion of end anchors and possibly filler combinations in the experimental design to restrict extreme responses. End anchors are treatment combinations that are presented prior to the start of the experiment and are intended to give the subjects a frame of reference within which to respond. They represent extreme combinations which are more extreme than those the subject is to judge. Hence, they set the ends of the response scale.

Floor and ceiling effects refer to the propensity of subjects to learn the most and least desirable items on a judgement scale, and to adjust their responses up or down accordingly. This is a problem for experimental designs in which subjects judge combinations of item values. This can be overcome by the use of filler combinations which are treatment combinations more extreme than those of interest in the study. The inclusion of filler combinations allows subjects to learn them and respond up or down to them and not the items of interest.

An alternative to end anchors is to provide a standard treatment combination for subjects to judge against. In a carefully controlled experiment, it may be possible to allow subjects to choose as a standard the treatment combination which corresponds most closely to their usual experience and to assign an arbitrary value to this treatment combination.

It is also desirable to randomise the order of treatment and the combinations of treatments between subjects, or when more than one factorial experiment is administered to each subject, among the lists of treatment combinations presented. This reduces the biases arising from the order in which treatment combinations are presented but, it can complicate the administration of the survey and increase the cost.

When the rating method is used together with functional measurement theory a two stage fitting process is indicated. Firstly functional measurement theory is used to determine the appropriate model form, then this model has its parameters determined by ordinary least squares regression. In practice the first step may be difficult if more than a few attributes are involved and sometimes higher order terms such as square terms may be considered applicable. In these cases only the regression step is performed.

This technique has been successfully used by Anderson (1976), Norman (1976), Levin et al (1977), Lerman and Louviere (1978), Meyer et al (1978), Louviere et al (1979), Louviere (1981), IMG (1981) and ABT (1982). Not all these papers used both stages of the method. Some were more concerned with finding the form of the model whereas others used the interval scale property of the rating method to fit a general function.

The use of experimental designs, functional measurement and the rating method is more a laboratory experiment, although it may not be conducted in a laboratory, than a field measurement of behaviour. The real world validity of these models has been confirmed in several cases by Lerman and Louviere (1978), Louviere et al (1979), and Levin et al (1982). In particular they show that functional forms derived in the laboratory can be transferred outside for application.

MAGNITUDE ESTIMATION

Magnitude estimation and the rating method both yield a direct numerical response to a combination of treatments. Magnitude estimation allows responses to be arbitrarily large, whereas the rating method has a fixed response scale. Unlike the rating method, magnitude estimation does not obey simple algebraic models for stimulus integration.

Anderson (1976) concludes 'that the rating method can yield true interval scales and that the method of magnitude estimation is biased and invalid.'

Horowitz (1978, 1981) has applied magnitude estimation techniques with some success. The statistical methods used to analyse the trip rating relied on:

- Ekman's Law of Psychophysics which states that variability in subjective units tends to grow as a linear function of subjective magnitude; and
- a change in the standard stimulus will cause a constant multiplicative change in all ratings.

The simplest linear statistical model of one-way experimental designs that has these multiplicative effects is:

$$\log R_{ij} = S_i + A_j + SA_{ij} + \epsilon_{ij} \qquad (8.4)$$

where R_{ij} is the subject's magnitude rating of a treatment combination with S_i as the standard stimuli, A_j as the j^{th} value of the treatment, SA_{ij} is an interaction term and ϵ_{ij} the error term. If more than one main effect is found to be significant a multiplicative model results, but with antilogs of the values of the standard stimuli and treatment being parameters.

The two studies by Horowitz (1978, 1981) are the only applications to transport problems known to the author. Due to the limited use of magnitude estimation and the criticism of Anderson (1976), it would appear that the rating method is to be preferred.

CONJOINT MEASUREMENT

The ordinary rating method and magnitude estimation both use the basic principle of having subjects assign numerical responses to treatment combinations in a factorial experiment. The numerical responses either form an interval scale or can be transformed to one. Conjoint measurement uses ordinal responses to determine the functional form used by the subjects. Krantz and Tversky (1971) describe the theory of conjoint measurement in detail for experiments with three variables.

The analysis of responses by conjoint measurement is a complex task but it can distinguish between the following functional forms in three variables:

$$\left.\begin{array}{l} A + P + U \\ (A + P)\,U \\ AP + U \\ APU \end{array}\right\} \qquad (8.5)$$

This can be computerised (Louviere 1978 refers to two packages) and the results can be used to derive an interval scale related to the response obtained in the experiment if the data is sufficiently rich in values (Krantz and Tversky 1971). This can be carried out only when the composition rule is known.

Louviere (1978) states that although conjoint measurement could be applied to travel behaviour modelling, he knows of no applications.

FRACTIONAL EXPERIMENTAL DESIGN

The basic principles of experimental design are described in Winer (1971). The most commonly used multi-variable plan is the full factorial experiment which permits the estimation of the effects of all interactions among the treatments. However, if it is desired to investigate the effect of more than three or four treatments at several levels, a full factorial experiment quickly grows to a large number of combinations. It is not practical to present subjects with the task of responding to a large number of treatment combinations.

The number of treatment combinations can be reduced by using a fractional factorial design rather than a full plan. In choosing a fractional factorial design it may be necessary to assume that the effect of some interactions is negligible as their effect can be confounded with some of the main effects. Hahn and Shapiro (1966) and Webb (1971) provide plans for fractional factorial designs which describe the degree of confounding (if there is any) in the designs. Although they provide a readily available source of designs, the designs presented are not necessarily the most efficient for a given experimental situation.

Information overload can occur if a subject is required to consider seven or more treatments (IMG 1982). When the subject wants to combine the effect of the treatments, overload can occur with as few as four treatments. If presented with too much information subjects may ignore all the treatments except the one or two attributes that are most important. This may not be an inaccurate representation of decision making when the levels of most treatments are changed and the subject only considered the changes most important to him. A more common situation would be when only a few of the treatment levels were changed and a decision made on this. If only treatments of lesser importance were changed their relative effects on the decision would need to be known. Therefore, for a model to be useful it must not only include the effects of important attributes, but also their relation with lesser attributes and their inter-relation.

A suitable method for avoiding information overload, while providing details on all effects is the method of partial (or differential) information (Norman 1976 and IMG 1982). Consider three treatment factors S, T and U. To determine the interaction between these factors an S x T x U three-way factorial design could be used. An alternative would be to consider three two-way factorial designs, S x T, S x U and T x U, so that all pairwise combinations of each information type are presented. Each sub-design allows an independent test of the related two-way interactions. The results from each sub-design can be combined to give the overall three-way interactions.

IMG (1982) discusses the generalisation of the method of partial information to more than three attributes. Each sub-design is assumed to be a full factorial design although a fractional factorial design could be used. The method of partial information simplifies the response task into combinations of fewer treatments but it does this at the expense of having a greater number of treatment combinations to be rated.

CHAPTER 9—ISSUES IN THE APPLICATION OF CHOICE MODELS

This chapter discusses issues which should be considered when using disaggregate choice models. The issues of concern are:

- sampling, specification and data errors;
- aggregation of results; and
- updating and transferring results.

Each of these can result in biased predictions from a model. A balanced approach is required so that it is not attempted to minimise one source of error, when others dominate the predictions.

SAMPLING, SPECIFICATION AND DATA ERRORS

Horowitz (1981) discusses these errors in detail. A brief synopsis of his exposition is given here.

Standard statistical procedures for the estimating of errors in the values of parameters generally do not apply to choice models. This is because the correct functional form of the model is not known *a priori* and the data used is not error free.

Statistical sampling errors in probabilistic choice models arise from estimating the models parameters from finite data sets. Although the computation of confidence regions for parameters is not difficult (see Horowitz), converting them into a confidence region for choice probabilities is complicated by the non-linear relationship between the parameters and choice probabilities. The choice probability confidence regions can be estimated by linearising the model by a first order Taylor series expansion in the parameters or by using a non-linear programming formulation.

The five types of specification errors considered by Horowitz are:

- inclusion of an irrelevant explanatory variable in the model;
- random utility components that are not independently and identically distributed (IID);
- random taste variations;
- omission of a relevant explanatory variable from the model; and
- random utility components that are correlated with explanatory variables.

These are errors resulting from an incorrect description of the model or the use of a model inappropriate for the data.

The use of erroneous data will produce a biased model. A commonly occurring error of this type is the use of group-mean values (eg average income) for explanatory variables. The use of non random surveys can produce biased parameter values, but corrections can be made to produce unbiased parameter estimates.

Horowitz calculated examples of the errors which can occur in choice probabilities. He concluded that:

> the errors caused by random taste variations, omission of a relevant explanatory variable and grouping of data all are large enough to seriously degrade or destroy the practical value of a model. The errors caused by non-IID random components of utility are smaller, although these errors, as well as sampling errors, clearly can impair a model's practical value.

AGGREGATION OF RESULTS

Koppelman (1976) describes two procedures to obtain aggregate predictions from disaggregate models. The first approach is to aggregate the model to obtain a model that can be estimated with aggregate data (Figure 9.1). These results are used to make aggregate predictions. When non-linear choice models are used with heterogeneous aggregate groups, a consistent aggregate function will include parameters of both the choice function and the distribution of the independent variables. The estimation procedure is based upon the distribution of the independent variables as well as their mean values.

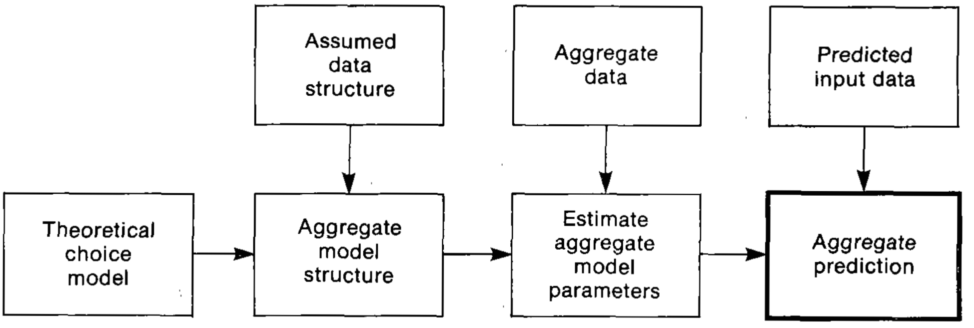

Figure 9.1. Aggregation of model structure prior to estimation

The second procedure is to estimate a disaggregate choice model using disaggregate data and then aggregate the results when the model is used for prediction (Figure 9.2). The advantage of this approach is that no assumptions about the future distribution of the independent variables are required until a prediction is made.

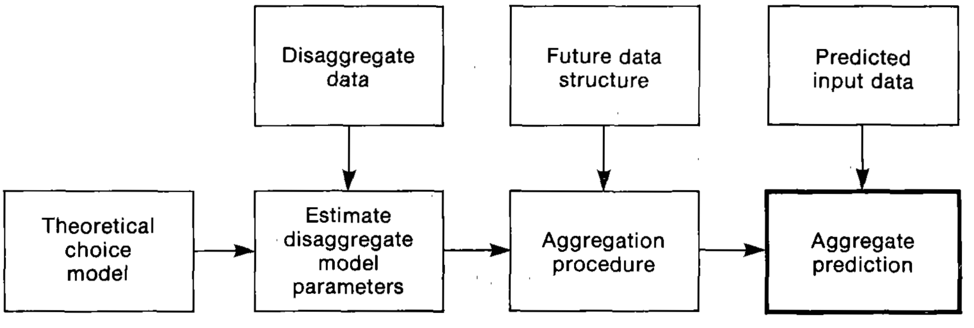

Figure 9.2. Aggregation of model structure after estimation

An intermediate procedure exists between these two extremes. Figure 9.3 shows the aggregation which commonly occurs when calculating the parameters of disaggregate models due to the solution techniques used. In these cases, although the models describe individual rather than group behaviour and use disaggregate data, the parameters calculated apply to the average of the group.

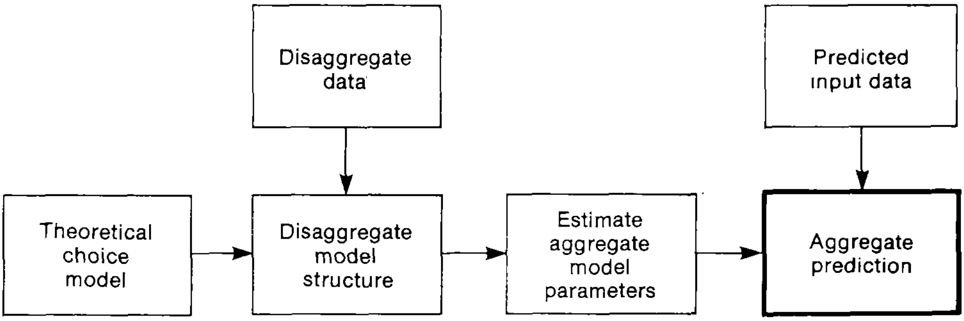

Figure 9.3. Aggregation of model structure during estimation

There are many potentially useful techniques for producing aggregate results from disaggregate models. Figure 9.4 shows a taxonomy of aggregation procedures adapted from Koppelman (1976).

The naive procedures are the simplest approaches but potentially produce the greatest errors in the prediction. The procedure is to substitute the mean values of the independent variables into the disaggregate demand function. In general this prediction will be incorrect, because a function of a mean is usually not equal to the mean value of the function. The magnitude of the error depends upon the form of the demand function and the distribution of the dependent variables. The predictions made by naive procedures can be adjusted to account for differences in choice set availability when such differences exist.

Enumeration procedures use the disaggregate nature of the model and the data. Complete enumeration uses the characteristics of *all* individuals and the alternatives. The expected behaviour of all individuals is predicted and the results averaged to give the aggregate prediction. In general this information will not be known or the computational task will be too large for experimentation with different transport policies. To restrict the amount of information required and the computation, sample enumeration can be used. This averages the choice probability for a sample of the individuals in the prediction.

Procedures for summation or integration weight the probability density function for the independent variables (Koppelman and Ben-Akiva 1977). This is done by integration when the density function is continuous, or by summation when the distribution is discrete. For a large number of variables the integration or summation tasks require considerable computation.

Statistical differentials procedures obtain an aggregate function by linearising the disaggregate choice function by the use of a Taylor series expansion and then calculating the expectation over the aggregate prediction group (Talvitie 1973). The resulting series is truncated so that aggregate demand is expressed in terms of means, variances and covariances of the distribution of dependent variables.

Classification procedures assign individuals belonging to the prediction group to identifiable classes and use the average value of each to predict aggregate results for each class. A weighted average summation procedure is used to compute the overall aggregate decisions. The classification may be on the basis of the choice made, individual characteristics, geographical location or some combination.

Reid (1978) advocates the use of a utility classification scale which can be used for simply scalable models such as the logit form. The principle of utility classification is that the information needed to predict each individual's choice is contained in the

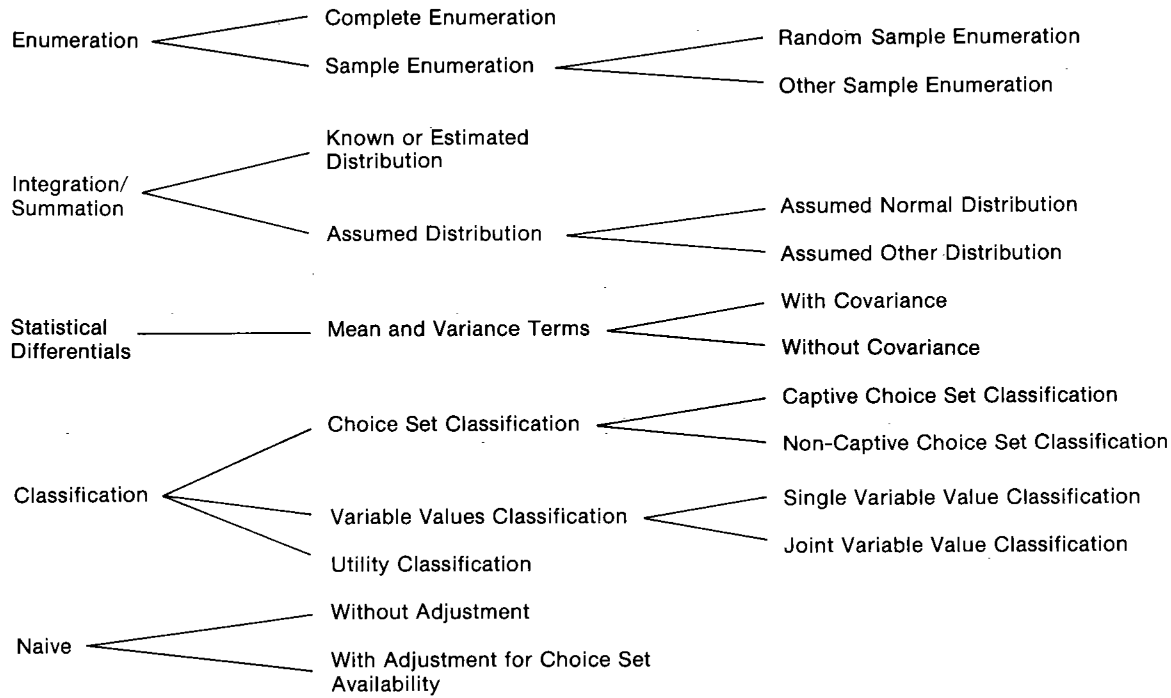

Figure 9.4. Taxonomy of aggregation procedures

utility scales of the attributes of the alternatives of choice. Cross-classification between the utilities of the different alternatives picks up full-scale variances and between-scale covariances, thus describing the full distribution of the individual choice factors in an aggregate sample. Classification on a total utility scale is more efficient than a classification by a limited number of variables because it includes the effects of minor variables. Reid describes the full procedure in detail.

To determine and compare aggregation errors the enumeration method is considered to be the correct method for obtaining aggregate results from the sum of individual choices. The usual measure of aggregation error is the percentage root mean square (RMS) of the proportion choosing different alternatives. This is expressed as:

$$\epsilon_{rms} = [\sum_j ((\hat{P}_j - P_j)/P_j)^2 P_j]^{1/2} \tag{9.1}$$

where $\hat{P}_j$ is the proportion choosing alternative j estimated by the aggregation method being tested and P_j is the proportion obtained by the enumeration method.

Reid (1978) compares several aggregation methods for a sample of 771 workers in the San Francisco Bay Area using a multinomial logit model. The RMS errors he obtained are shown in Table 9.1. The classification by auto ownership is the method recommended by Koppelman (1976) but the error is much larger than he obtained with different data. Reid suggests that this is a result of Koppelman's data being for a simpler choice situation with less disaggregate level-of-service data.

TABLE 9.1—RMS ERRORS FROM DIFFERENT AGGREGATION PROCEDURES

Method	*No of classes*	*Per cent error*
Naive	—	40.0
Statistical differentials	—	121.0
Classification by city	17	17.9
Classification by auto ownership	4	21.7
Classification by utility scale	4	3.1

Source: Adapted from Reid (1978).

By using the utility classification procedures Reid reduced the aggregation error from 38.4 per cent in the naive case to 2.3 per cent with four utility classes and 0.5 per cent with eight classes. He concluded that although the utility classification method is directly applicable only to simple scalable models, it may be useful for non-simple scalable models as well, eg probit models.

UPDATING AND TRANSFERRING RESULTS

Large surveys to estimate disaggregate choice models are expensive but are required for accurate predictions from the model. Often when using the model in later years or at a different location there is neither the time nor the money to conduct a large survey for the development of a new choice model. Therefore, procedures for adapting the results of previous large surveys are required.

All the factors which affect the reliability of predictions will also affect the updating and transfer of a model. If a model is unsuccessful in the area for which it was developed, there is no reason for it to be better in a different location. To be transferable, it is not enough for the model to merely fit existing data, it must represent the causal relationship between attributes (Atherton and Ben-Akiva 1976). This means that aggregate models are not reliably transferable because aggregate model coefficients are bound to a particular zonal structure.

Atherton and Ben-Akiva (1976) propose the following procedures as being available for transferring models.

Naive approach

The naive approach is to use the existing model with its original coefficients and substitute parameter values describing the new situation. This assumes that the choice process is fully explained by the model although, in most models, the constant terms account for factors not explicitly explained by the model.

Adjustment of constant terms

If the validity of model coefficients other than constant terms is accepted, then aggregate data can be used to adjust the constant terms so the model replicates existing aggregate data.

Transferring with a small disaggregate sample

This procedure uses a small sample of observations of individual choices assuming that the sample is representative of behaviour in the study area. The small disaggregate sample could be used to re-estimate the coefficients of the original model specification. Then only the new small sample coefficients need to be used in the model, but the use of a small sample is a potential source of error. Also because a model specification is good statistically on one particular data set, it does not guarantee that using the same specification will result in reliable coefficients with a different data set from another area.

Other ways in which a small disaggregate sample could be used are to re-estimate only the constant terms using the other original coefficients, or to combine the small sample coefficients with the original coefficients to produce modified coefficients.

Bayesian updating

Bayesian updating uses the original and small sample coefficients to estimate updated values for the coefficients. For a single parameter model, with the assumption that the coefficient is normally distributed, the appropriate formula for the value of the updated coefficient θ_u is:

$$\theta_u = \sigma_u^2\left(\frac{\theta_o}{\sigma_o^2} + \frac{\theta_s}{\sigma_s^2}\right)$$

with:

$$\sigma_u^2 = \left(\frac{1}{\sigma_o^2} + \frac{1}{\sigma_s^2}\right)$$

where: θ_o is the value of the original coefficient;

σ_o^2 is the variance of the original coefficient;

θ_s is the value of the small sample coefficient;

σ_s^2 is the variance of the small sample coefficient; and

σ_u^2 is the variance of the updated coefficient.

So for a one parameter model the variances of the original and small sample coefficients are required for this updating process. Lerman et al (1976) give the formula for the generalisation of this updating process to multi-parameter models where the variance-covariance matrices for the coefficient estimates are needed.

Atherton and Ben-Akiva (1976) tested some of the different updating and transferring procedures with three sets of disaggregate data of mode choice for work trips and concluded that the Bayesian updating procedure using a small disaggregate sample is the most effective procedure for transferring well-specified models. Hensher and Johnson (1981) are less optimistic due to the infrequent use of updating procedures and reported poor performance when used. It appears that the problem results from the suitability of the range of explanatory variables when applied to a new situation. So a well-specified model is required for the realistic transfer or updating of results.

CHAPTER 10—COMPARISON OF METHODOLOGIES

The preceding discussion has identified the available behavioural decision models and shown that they are dependent on assumptions as to the manner in which decisions are made, the functional form of the decision rule assumed, the solution technique used to solve the problem, the data required, and, if appropriate, the distribution of residual terms.

Heggie (1976) discusses a number of issues which should be considered when selecting a choice model.

- Human behaviour cannot be represented by a continuous linear function. A function which allows non-linearities and thresholds is required.
- Individuals exhibit habitual behaviour and their responses may be restricted by either a lack of knowledge of the available alternatives, or by only considering alternatives when their chosen mode deteriorates.
- The perceived values of attributes are not the same as those which are measured or observed. This is a result of the way humans perceive attributes in a particular context.
- The decision to make a journey is dependent upon the modes available. If a particular mode was unavailable a decision to cancel the journey may be made. Conversely, the existence of a mode (for example a steam train) may create a journey.
- The value of travel time savings during working hours is not only dependent on marginal productivity but should also include the effect of business travellers working while travelling and the possibility of this being more productive as a result of fewer distractions.
- Preference functions represent the difference between the opportunity cost of time and the disutility of travelling. Linear preference functions are uncompensated because they assume the marginal utility of money remains constant.
- The value of travel time savings is dependent upon the availability of substitute activities and their nature, as well as the size of the time saving. For example, there may be few substitute activities for small time savings on a journey to work.
- Utility is dependent on both the person and the circumstances which can give aggregation problems. Variations between people can result in behaviour that, while rational, is not homogeneous and leads to errors in the predictions made by models which produce aggregate results.
- Often a model uses revealed preference data. This suffers from the problem that in a study which investigates the effect of time and cost, the values are multicollinear as they both will in general depend on the distance of the journey.
- Values of travel attributes supplied by travellers may be biased due to question design, perception problems, post justification of choice and the difference between what they state they will do and what they actually do.
- For the prediction of traveller decisions in response to changes to be correct, the changes must be publicised so that travellers are aware of their existence and value. Also some changes may lead to a decision not to travel rather than a change in mode.

Goodwin and Hensher (1978) discuss the following additional issues.

- However rational people are in general, there will always remain a certain unpredictability and random element at the individual level. This may be a result of the attributes of the individual or the transport modes which are not included in the model, or irrational behaviour.
- To be a useful experimental tool a model should be easily manipulable; the resources, computing facilities and manpower needed to generate outputs should be neither so great nor so specialised that sensitivity testing or the investigation of alternative futures is impractical.
- If time is an important consideration in a travel choice situation the total journey time should be divided into its components. For example, in-vehicle, waiting, walking and transfer time.
- Delays in obtaining or reacting to information can result in a dependence on the time since a change was made.

It is not necessary for a model to satisfy all the above issues as, in particular contexts, some may not be relevant or particular assumptions may not be considered important. The overall effect of these issues is the need for a flexible model which does not use certain restrictive assumptions or, at least, allows its assumptions to be tested.

For an investigation of the value of travel time savings a model with the following properties is desirable:

- the choice function (if the model uses one) is non-linear with discontinuities;
- the individual's knowledge of the available alternatives is accounted for;
- the perceived and observed values of attributes coincide;
- the journey context can be distinguished;
- it provides disaggregate results;
- it makes efficient use of the data; and
- does not require excessive computation.

When using any model it is necessary to accept a number of assumptions or limitations. Table 10.1 compares the important characteristics and assumptions of the models so that properties which can be traded-off may be examined.

All the behavioural models in Table 10.1 assume a decision structure which is, at best, an approximation to actual human decision making. The most appropriate decision process will depend on the use to which the results are to be put. For example, the MNL model has had considerable success in predicting responses to transport policies which do not go outside the range of limitations of the data used to calibrate the model despite its assumption of decisions being made simultaneously on the basis of a linear utility function.

Each functionally based model is limited in the forms it can take by the available solution techniques. In general a linear relationship is required, but like ordinary least squares, powers of the values of the attributes can be used for some transformations such as Box-Tukey. Maximum likelihood methods require that the terms in the expression are not pairwise correlated otherwise biased estimates results. Hensher and Johnson (1981) discuss how to correct the biased estimates.

To investigate the value of travel time savings a non-linear utility function with cross product terms is desirable so that the value of the savings depends not only on the size of the saving but the context within which it is made.

Behavioural models do not contain parameters which describe all the attributes of the alternatives or of an individual's socio-economic situation that can influence the person making the decision. Residual error terms can be used to compensate for omitted or unobtainable information and behaviour which would be considered

TABLE 10.1. COMPARISON OF MODEL PROPERTIES

	Simple Models					Multinomial Models								
	Aggregate models	*Linear probability*	*Linear logit*	*Linear probit*	*Discriminant analysis*	*MNL*	*Dogit*	*MNP*	*Nested and other generalised logit*	*Elimination by aspects*	*Satisficing*	*Priority evaluator*	*Functional measurement*	
Decision structure assumed?		S	S	S		S	S	S	SE/H	H	R	H	S	H = Hierarchical, S = Simultaneous SE = Sequential, R = Random
Functional form of decision function?		L	L	L	L	L	L	L	L				NC	L = linear, N = Non-linear C = cross product terms
IIA required						Y								Y = yes
Distribution of residuals in utility expression				N		E	E	N	E					E = extreme value N = normal
Taste variations allowed								Y				Y	Y	Y = yes
Results—aggregate or disagregate?	A	A	A	A	A	A	A	AD	A	A	A	D	D	A = Aggregate, D = Disaggregate AD = aggregate but with distribution information.
Computation?	A	A	A	A	A	A	M	I	N	M	M	A	A	A = available software or not difficult to write M = may be obtained by modifying existing software N = not generally available I = computationally inefficient

irrational by the decision making process assumed in the model. Some models only require the assumption of a distribution for the error terms when testing the significance of the parameters estimated by the model.

It can be expected that there will be taste variations among individuals, but many models do not allow this desirable property. Those that do, have the advantage of explicitly allowing for variations in the population without relying on residual terms to account for it.

Although many models are based on theories of individual behaviour, the techniques used to solve the model may produce only mean aggregate results with no distribution information, thereby losing information on individual variations. Although the ultimate aim of any model is to provide aggregate predictions, information concerning the distribution of behaviour in the studied population is required to determine the impact of changes.

Regression based models such as the linear probability and linear logit models yield aggregate results because only one observation for each individual is used. Usually these data come from revealed preference situations. In contrast functional measurement uses ordinary least squares regression with data on multiple decisions by each individual to give a model for each individual.

The amount of computation required to yield results can vary greatly between models. Some use software which is either readily available or not difficult to write, others may possibly be obtained by modifying existing software. Unless there is a special reason for using a model which requires extensive software development or is computationally inefficient, it should be avoided.

The data requirements of the models have not been addressed here because the amount of data will not only depend on the model chosen and how efficiently it uses the data, but also the situation to which it is applied.

None of the models discussed in this paper can be selected to estimate the value of travel time savings without accepting some assumptions or limitations. For example, none of the functionally based models can explicitly allow for discontinuities in the decision rule, but those with non-linear functions may be able to approximate a discontinuity.

The two models which come closest to meeting the desirable properties are the priority evaluator method and functional measurement. Both models provide the subjects with alternative trip descriptions to choose between and produce disaggregate results. The perceived and observed values coincide because the values of the attributes are provided in the trip descriptions.

The choice function used by functional measurement does not explicitly allow for discontinuities. The priority evaluator method has the relative values of the attributes imposed on the model and is not suitable for determining the value of travel time savings.

If it is acceptable to have aggregate results, then the Elimination-by-Aspects or Satisficing models should be considered. Alternatively, if revealed preference data is obtainable, MNL may be acceptable.

The model least restricted by its assumptions and which allows most of them to be tested is the method of functional measurement. Functional measurement produces disaggregate results in that a model is built for each individual and the individual results are aggregated by either full enumeration or producing an average choice function. The Bayesian updating technique could be used to investigate seasonal effects on the value of time as well as providing an updating mechanism for the model.

REFERENCES

ABT (1982), Cable and Subscription Television Services for Australia, *Australian Broadcasting Tribunal Report,* Vol 4: Demand Models, Australian Government Publishing Service, Canberra.

Anderson, N.H. (1976), How Functional Measurement Can Yield Validated Interval Scales of Mental Quantities, *Journal of Applied Psychology,* Vol 61, No 6.

Anderson, N.H. (1981), *Foundations of Information Integration Theory,* Academic Press, New York.

ARRDO (1981), Freight Mode Choice: An Application of the Elimination by Aspects Model, *ARRDO Report 81/30,* Australian Railway Research and Development Organisation, Melbourne, Victoria.

Atherton, T.J. and Ben-Akiva, M.E. (1976), Transferability and Updating of Disaggregate Travel Demand Models, *Transportation Research Record,* No 610.

Ben-Akiva, M.E. and Lerman, S.R. (1979), Disaggregate Travel and Mobility—Choice Models and Measures of Accessibility, in Hensher, D.A. and Stopher, P.R. (Editors), *Behavioural Travel Modelling,* Croon Helm, London.

BTE (1982), *The Value of Travel Time Savings in Public Sector Evaluation,* Bureau of Transport Economics, Occasional Paper 51, Australian Government Publishing Service, Canberra.

Crittle F.J., and Johnson, L.W. (1980), *Basic Logit (BLOGIT)—Technical Manual,* Australian Road Research Board, *Technical Manual,* No 9.

Daganzo, C.F. (1979), 'Multinomial Probit' The Theory and Its Application to Demand Forecasting, Academic Press, New York.

Daganzo, C.F., Bouthelier, F. and Sheffi, Y. (1977), Multinomial Probit and Qualitative Choice: A Computationally Efficient Algorithm, *Transportation Science,* Vol 11, No 4.

Gaudry, M.J.I. and Dagenais, M.G. (1979), The Dogit Model, *Transportation Research,* Vol 13B, No 2.

Gaudry, M.J.I. and Wills, M.J. (1978), Estimating the Functional Form of Travel Demand Models, *Transportation Research,* Vol 12, No 4.

Gaudry, M.J.I. and Wills, M.J. (1979), Testing the Dogit Model with Aggregate Time-Series and Cross-Sectional Travel Data, *Transportation Research,* Vol 13B, No 2.

Goodwin, P.B. and Hensher, D.A. (1978), The Transport Determinants of Travel Choices: An Overview, in Hensher, D.A. and Dalvi, Q. (Editors), *The Determinants of Travel Choice* , Saxon House, England.

Gordon, P., Williams, C.S. and Theobald, P.M. (1979), Intrametropolitan Travel Demand Forecasting Using an Abstract Modes Approach, *Transportation Research,* Vol 13A.

Gunn, H.F., Mackie, P.J. and Ortuzar, J.D. (1980), Assessing the Value of Travel Time Savings—A Feasibility Study on Humberside, *Institute for Transport Studies,* The University of Leeds, Working Paper 137.

Hahn, G.J. and Shapiro, S.S. (1966), A Catalogue and Computer Programme for the Design and Analysis of Orthogonal Symmetric and Assymetric Factorial Experiments, *Technical Information Series* (66-C-165).

Hausman, J.A. and Wise, D.A. (1978), A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognising Interdependence and Heterogeneous Preferences, *Econometrica,* Vol 46, No 2.

Heggie, I.G. (1976), *Modal Choice and the Value of Travel Time,* Clarendon Press, Oxford.

Hensher, D.A. and Hotchkiss, W.E. (1974), Choice of Mode and the Value of Travel Time Savings for the Journey to Work, *The Economic Record,* Vol 50, No 129.

Hensher, D.A. and Johnson, L.W. (1981), *Applied Discrete-Choice Modelling* Croon Helm, London.

Hoinville, G. and Berthoud, R. (1970), Identifying and Evaluating Trade-Off Preferences: An Analysis of Environmental/Accesibility Priorities, *Social and Community Planning Research,* London.

Hoinville, G. and Johnson, E. (1971), The Importance and Values Commuters Attach to Time Savings, *Social and Community Planning Research,* London.

Horowitz, A.J. (1978), The Subjective Value of the Time Spent in Travel, *Transportation Research,* Vol 12.

Horowitz, A.J. (1981), Sources of Error and Uncertainty in Behavioural Travel Demand Models, in Stopher, P.R., Meyberg, A.H. and Brög, W. (Editors), *New Horizons in Behavioural Travel Research,* Lexington Books, Lexington, Mass.

Horowitz, A.J. (1981), Subjective Value of Time in Bus Transit Travel, *Transportation,* Vol 10.

IMG Consultants (1981), *Value of Time of Airport Users* prepared for the Department of Transport Australia.

IMG Consultants (1982), *Value of Travel Time Study: A Guide to Research,* prepared for the Bureau of Transport Economics, Canberra, Australia.

Isard, W. (1960), *Methods of Regional Analysis,* MIT Press, Cambridge, Mass.

Koppelman, F.S. (1976), Guidelines for Aggregate Travel Prediction Using Disaggregate Choice Models, *Transportation Research Record,* No 610.

Koppelman, F.S. and Ben-Akiva, M.E. (1977), Aggregate Forecasting with Disaggregate Travel Demand Models Using Normally Available Data, in Visser, E.J. (Editor), *Transport Decisions in an Age of Uncertainty,* Martinus Nijhoff, The Hague.

Krantz, D.H. and Tversky, A. (1971), Conjoint-Measurement Analysis of Composition Rules in Psychology, *Psychological Review,* Vol 78, No 2.

Lerman, S.R. and Louviere J.J. (1978), Using Functional Measurement to Identify the Form of Utility Functions in Travel Demand Models, *Transportation Research Record,* No 673.

Lerman, S.R., Manski, C.F. and Atherton, T.J. (1976), Non-Random Sampling in the Calibration of Disaggregate Choice Models, *Cambridge Systematics Inc.*

Levin, I.P., (1979), The Development of Attitudinal Modelling Approaches in Transport Research, in Hensher, D.A. and Stopher, P.R. (Editors) *Behavioural Travel Modelling,* Croon Helm, London.

Levin, I.P., Louviere, J.J., Schepanski, A.A. and Norman, K.L. (1982), *Validity Tests and Applications of Laboratory Studies of Information Integration,* mimeo.

Levin, I.P., Mosell, M.K., Lamka, C.M., Savage, B.E. and Gray, M.J. (1977), Measurement of Psychological Factors and Their Role in Travel Behaviour, *Transportation Research Record,* No 649.

Louviere, J.J. (1978) Psychological Measurement of Travel Attributes, in Hensher, D.A. and Dalvi, Q. (Editors), *Determinants of Travel Choice,* Saxon House, England.

Louviere, J.J. (1981), On the Identification of the Functional Form of the Utility Expression and its Relationship to Discrete Choice, Appendix B of Hensher, D.A. and Johnson, L.W., *Applied Discrete-Choice Modelling,* Croon Helm, London.

Louviere, J.J., Henley, D.H., Woodworth, G., Meyer, R.J., Levin, I.P., Stoner, J.W., Curry, D. and Anderson, D.A. (1981), Laboratory-Simulation Versus Revealed-Preference Methods for Estimating Travel Demand Models , *Transportation Research Record,* No 794.

Louviere, J.J., Wilson, E.M. and Piccolo, M. (1979), Applications of Psychological Measurement and Modelling to Behavioural Travel Demand Analysis, in Hensher, D.A. and Stopher, P.R. (Editors), *Behavioural Travel Modelling,* Croon Helm, London.

McFadden, D. (1978), Modelling the Choice of Residential Location, *Transportation Research Record,* No 673.

McFadden, D. (1979), Quantitative Methods for Analysing Travel Behaviour of Individuals: Some Recent Developments, in Hensher, D.A. and Stopher P.R. (Editors), *Behavioural Travel Modelling,* Croon Helm, London.

McFadden, D., Train, K. and Tye, W. (1977), An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model, *Transportation Research Record,* No 637.

Meyer, R.J. Levin, J.P. and Louviere, J.J (1978), Functional Analysis of Mode Choice, *Transportation Research Record,* No 673.

Norman, K.L. (1976), A Solution for Weights and Scale Values in Functional Measurement, *Psychological Review,* Vol 83, No 1.

Nijkamp, P. (1979), Gravity and Entropy Models : the State of the Art, in Jansen, G.R.M., Bovy, P.H.L., van Est, J.P.J.M. and le Clercq, F. (Editors), *New Developments in Modelling Travel Demand and Urban Systems,* Saxon House, England.

Quandt, R.E. and Baumol, W.J. (1966), The Demand for Abstract Transportation Modes: Theory and Measurement, *Journal of Regional Science* 6.

Recker, W.W. and Golob, T.F. (1979), A Non-compensatory Model of Transportation Behaviour Based on Sequential Consideration of Attributes, *Transportation Research,* Vol 13B, No 4.

Reid, F.A. (1978), Minimising Error in Aggregate Predictions from Disaggregate Models, *Transportation Research Record,* No 673.

Richardson, A.J. (1978), *A Comparative Analysis of Transport Choice Models,* Monash University, Department of Civil Engineering Working Paper 78/11, Melbourne, Australia.

Ruijgrok, C.J. (1979), Disaggregate Choice Models: An Evaluation, in Jansen, G.R.M., Bovy, P.H.L., van Est, J.P.J.M. and le Clercq, F. (Editors), *New Developments in Modelling Travel Demand and Urban Systems,* Saxon House, England.

Sobel, K.L. (1980), Travel Demand Forecasting by Using the Nested Multinomial Logit Model, *Transportation Research Record,* No 775.

Stopher, P.R. and Meyburg, A.H. (1975), *Urban Transportation Modelling and Planning,* Lexington Books, Lexington, Mass.

Talvitie, A.P. (1973), Aggregate Travel Demand Analysis with Disaggregate or Aggregate Travel Demand Models, *Transportation Research Forum Proceedings,* XIV(1).

Tversky, A. (1972), Elimination by Aspects: A Theory of Choice, *Psychological Review,* Vol 79, No 4.

Watson, P.L. (1974), *The Value of Time; Behavioural Models of Modal Choice,* Lexington Books, Lexington, Mass.

Webb, S.R. (1971), Small Incomplete Factorial Experimental Designs for Two- and Three-level Factors, *Technometrics,* Vol 13.

Winer, B.J. (1971), *Statistical Principles in Experimental Design*, McGraw-Hill, Kogakusha, Japan.

Wildermuth, H.K. (1976), *Transport Policies for Central Perth: Review and Formulation,* Director-General of Transport, Western Australia.