Working Paper No. 261

Random Utility Theory and the Structure of Travel Choice Models

H.C.W.L. Williams*
and
J.D. Ortuzar**

Prepared for submission to the Bulletin of the Institute of Mathematics and its Applications.

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

October 1979.

** Institute for Transport Studies,
University of Leeds,
Leeds LS2 9JT,
England.

RANDOM UTILITY THEORY AND THE STRUCTURE OF TRAVEL CHOICE MODELS.

Mathematical models have been employed in the analysis and projection of travel demand for over two decades. Although modified in a number of ways, the basic forecasting methodology implemented in both British and American urban transportation studies is still recognizably an outgrowth of that developed in the pioneering Detroit and Chicago studies of the late 1950s.

Models are developed for the purpose of predicting the response of travellers to specific policies, which correspond to identifiable changes in the transport system. The implications of these policies which are to be introduced in a design year - typically 5 to 15 years hence - are then assessed with respect to a reference strategy (perhaps a policy of 'doing nothing'). In the absence of direct information on behavioural response or suitable time series data, it has become conventional practice to infer demand elasticities - the propensity for individuals to modify their behaviour under specific changes in the transport system - from the observed variability in behaviour which exists in one cross sectional data set, collected in the base year. The forecasting approach thus relies heavily on statistical and econometric methods. It is often (wrongly) assumed that a good statistical fit of a cross-sectional model to existing (base year) travel behaviour is the sole requirement for an acceptable forecasting model. We shall note below that the model must obey certain theoretical restrictions in order to exhibit satisfactory response properties.

A typical passenger demand model is of the following form:

$$T(nijkr) = G_i^n \cdot M_{ij}^n \cdot M_{ij}^{nk} \cdot M_{ij}^{nkr} \qquad (1)$$

in which the number of trips T(nijkr) by persons of type n, between locations (zones) i and j on mode k by route r is expressed in terms of the attributes

$\underline{Z}$ of the transport system (times, costs, etc.). G_i^n is the total number of trips generated in zone i; M_{ij}^n is that proportion attracted to zone j, while M_{ij}^{nk} is the proportion of these $G_i^n \cdot M_{ij}^n$ trips which are associated with mode k (car, bus, or rail, for example). The route share M_{ij}^{nkr} is similarly defined. The four quantities, G_i^n, M_{ij}^n, M_{ij}^{nk}, M_{ij}^{nkr}, correspond to the widely employed sequence of generation (G), distribution, (D), modal split, (MS), and assignment (A) segments in the 'compound' demand forecasting model, the structure of which will be denoted G/D/MS/A. This conveniently summarizes the relative positions of the sub-models in the sequence.

In this G/D/MS/A structure the distribution and modal split models are frequently taken to be of the logit form. The modal share function, for example, is given by

$$M_{ij}^{nk} = \frac{e^{-\lambda^n C_{ij}^k}}{\sum_{k \epsilon H(n)} e^{-\lambda^n C_{ij}^k}} \tag{2}$$

in which C_{ij}^k is termed the generalized cost of travel by mode k between locations i and j, and $\sum_{k\ H(n)}$ denotes summation over those modes available to persons of type n. In British studies, since the late 1960s, C_{ij}^k has been expressed as a linear combination of attributes Z_{ijk}^η, η=1, 2, ...

$$C_{ij}^k = \sum_\eta a_\eta Z_{ijk}^\eta \tag{3}$$

in which in-vehicle time, out-of-vehicle time, and out-of-pocket costs are the usual constituents. The quantities λ and a_η, η=1, 2, ... in equation (2) are parameters to be estimated from observed patterns of behaviour. If C_{ij}^k is measured in time units, the coefficient a_ℓ of the attribute time $Z_{ijk}^\ell = t_{ij}^k$ is set equal to unity.

There are two alternative model structures which have been employed in transportation studies. These are denoted G/MS/D/A and G/D-MS/A in which the relative positions of the modal split and location models are reversed in the first case, and combined in a single sub-model in the second. In the early 1960s the G/MS/D/A form was favoured, to be replaced by the G/D/MS/A structure in the late 1960s and early 1970s. The relative position of the two sub-models was rationalised on the basis of rather imprecise scenarios of the trip decision process. It is currently fashionable to argue that the choice of mode and location (of residence or shop) are made 'simultaneously', and for this reason it is appropriate to combine the D and MS models.

In this paper we shall describe and develop a theory of model structure which includes a consideration of the relation between alternative sub-models. Attention will be confined to the location and mode choice contexts as their relative position poses the greatest source of structural ambiguity. We discuss, in turn, the generation of models within the framework or random utility theory; and finally consider some practical implications of the theoretical issues.

CORRELATION AND MODEL STRUCTURES : THE MATHEMATICS OF RANDOM UTILITY THEORY.

In random utility theory, individuals are considered to identify with each choice option A_ρ, out of a set A containing elements $A_1, \ldots, A_\rho, \ldots A_N$, a net utility U_ρ (benefit minus cost) and to select the most desirable (maximum utility) alternative. The utility U_ρ, is expressed in terms of a set of attributes $\underline{Z}_\rho = Z^\mu_\rho$, $\mu=1, \ldots m$ and parameters $\underline{\theta} = \{\theta_1, \ldots, \theta_m\}$. A general form :

$$U_\rho = U_\rho(\underline{Z}_\rho, \underline{\theta}) \qquad \forall A_\rho \epsilon A. \qquad (4)$$

may be specified.

Because the valuation of attributes may vary over the choice making population, and because some of the attributes entering the choice process may be unobserved or unmeasured by the modeller, the utility function (4) is considered to be a random variable. By invoking this notion of interpersonal differences in utility, it is possible to explain why individuals, who are associated with the same values of $\underline{Z}_\rho$, $\rho=1, \ldots, N$, may select different choice options. A common form of random utility function is the linear form

$$U_\rho = \bar{U}_\rho(\underline{Z}_\rho, \underline{\theta}) + \varepsilon_\rho \tag{5}$$

$$= \sum_\mu \theta_\mu Z_\rho^\mu + \varepsilon_\rho \tag{6}$$

in which the mean, or 'representative', utility $\bar{U}_\rho$ is expressed as a linear function of the attributes. ε_ρ is the residual error term.

The formal characteristics of random utility models may now be summarized as follows :

$$P_\rho = \text{Prob}\,\{U_\rho \geq U_{\rho'} \;:\; \forall A_{\rho'} \varepsilon A\} \qquad \forall A_\rho \varepsilon A \tag{7}$$

$$= \int_{R_\rho} d\underline{U}\, f(\underline{U}) \tag{8}$$

in which the probability that an individual will select alternative A_ρ is determined as the integral over the joint probability density function $f(\underline{U}) = f(U_1, \ldots, U_\rho, \ldots, U_N)$ in an N-dimensional utility space. R_ρ is that portion of the space defined by

$$R_\rho \;:\; \quad U_\rho \geq U_{\rho'} \qquad \forall A_{\rho'} \varepsilon A. \tag{9}$$

P_ρ is, in general, a function of the attributes $\underline{Z}$ of <u>all</u> choice options and this expression may be found when the function $f(\underline{U})$ is specified.

We shall be particularly interested in the covariance structure of the residuals ε_ρ, $\rho=1, \ldots, N$ which represents the <u>similarity</u> or <u>structure of dependence</u> between the alternatives. The variance-covariance matrix $\underline{\underline{\Sigma}}$ which embodies this dependence is specified in the usual way, with elements

$\Sigma_{\rho\rho'}$ given by

$$\Sigma_{\rho\rho'} = E\ (\varepsilon_\rho, \varepsilon_{\rho'}) \quad , \tag{10}$$

E( ) denoting the expectation value.

For general forms of $f(\underline{U})$ it is necessary to appeal to numerical integration of Equation (8), and one approach to this problem is through Monte Carlo simulation. However, analysts have gratefully resorted to those distributions $f(\underline{U})$ which are consistent with either analytic expressions for P_ρ or which result in a manageable computational task. An important class of models is that derived from <u>independent</u> distributions which are associated with a diagonal matrix $\underline{\underline{\Sigma}}$. A sub-class contains so-called <u>identically and independently distributed</u> (IID) models for which

$$f(\underline{U}) = \prod_\rho f_\rho(U_\rho, \bar{U}_\rho) \tag{11}$$

$$= \prod_\rho f(U_\rho, \bar{U}_\rho) \tag{12}$$

in which $f(U_\rho, \bar{U}_\rho)$ is the probability density function of the utility distribution associated with A_ρ. This class of models is characterised by a variance-covariance matrix of the form

$$\underline{\underline{\Sigma}} = \sigma^2 \underline{\underline{I}} \tag{13}$$

in which σ is the (common) standard deviation of the distributions of residuals ε_ρ, $\rho = 1, \ldots N$, and $\underline{\underline{I}}$ is the unit matrix.

By substituting the density function (12) into Equation (8) an expression for the probability P_ρ appropriate to an IID model can be obtained as follows

$$P_\rho = \int_{-\infty}^{\infty} dU_\rho\, f(U_\rho, \bar{U}_\rho) \prod_{\rho' \neq \rho} \int_{-\infty}^{U} dU_{\rho'}\ f(U_{\rho'}, \bar{U}_{\rho'}) \tag{14}$$

It can readily be shown that the multinomial logit model

$$P_\rho = \frac{e^{\lambda \bar{U}_\rho}}{\sum_{\rho'} e^{\lambda \bar{U}_{\rho'}}} \tag{15}$$

is an IID model generated from Weibull (Gnedenko) probability distributions for which

$$f(U_\rho, \bar{U}_\rho) = \lambda\ e^{-\lambda(U_\rho - \bar{U}_\rho)}\ e^{-e^{-\lambda(U_\rho - \bar{U}_\rho)}} \tag{16}$$

This is a skewed uni-modal distribution which has standard deviation given by

$$\sigma = \frac{\pi}{\sqrt{6}\,\lambda} \tag{17}$$

It has been found that, for given means and appropriate standardization of variances, the resultant discrete choice probabilities $\{P_\rho,\ \rho=1, \ldots, N\}$ are rather similar for all bell-shaped distributions. There is, for example, a close correspondence between the logit and probit models, generated from Weibull and Normal distributions, respectively (Domencich and McFadden, 1975).

The simplicity and flexibility of the multinomial logit model have led to its wide application in studies of individual choice in a transportation context (Spear, 1977) and in other disciplines. The limitations of the model are also being increasingly recognized. One such limitation results from the so-called 'independence from irrelevant alternatives' property of the model which is embodied in the expression for the ratio of probabilities given by

$$\frac{P_\rho}{P_{\rho'}} = e^{\lambda(\bar{U}_\rho - \bar{U}_{\rho'})} \qquad \forall A_\rho,\ A_{\rho'} \varepsilon A \tag{18}$$

This ratio is seen to depend only on the representative utilities associated with A_ρ and $A_{\rho'}$ and is independent of other components. It will be unaffected by the expansion or contraction of the choice set A.

The weakness of the multinomial logit (MNL) model become particularly apparent when 'similarities' exist between some of the alternatives. It would, for example, be inappropriate to employ a 3-alternative MNL model for choice between car and bus journeys operating on two different routes, as the latter clearly have rather similar characteristics.

The similarity between alternatives may be accommodated in random utility theory by introducing correlation between the residuals associated with different alternatives, that is, by generating models which have off-diagonal elements in the matrix $\underline{\underline{\Sigma}}$. One possible approach is to appeal to the multivariate normal distribution (Domencich and McFadden, 1975)

$$f(\underline{\varepsilon}) = (2\Pi)^{-\frac{N}{2}} \; |\underline{\underline{\Sigma}}|^{-\frac{1}{2}} \; e^{-\frac{1}{2}\underline{\varepsilon}' . \underline{\underline{\Sigma}}^{-1} . \underline{\varepsilon}} \qquad (19)$$

in which $\underline{\varepsilon}$ is a vector with the elements ε_ρ, $\rho=1, \ldots, N$, and $\underline{\varepsilon}'$ is its transpose. The structure of dependence in this multinomial probit model may be arbitrarily specified in the matrix $\underline{\underline{\Sigma}}$. The generality of the form is both appealing and restrictive. For small problems, with up to five alternatives the numerical integration, which is performed iteratively in parameter estimation, is manageable. Much effort has recently been devoted to the development of approximate solutions to the multinomial probit model for applications involving a larger number of alternatives. These have included : moment expansion techniques ; Fourier analysis; Monte Carlo simulation; coordinate transformation methods; and special 'Extreme value' approximations.

There are many examples for which the generality of the multinomial probit model even if it could be implemented is an un-necessary luxury. Consider 'two-dimensional' choice contexts for which the set A contains the elements $\{X_{\mu\nu} : \mu=1, \ldots M; \nu=1, \ldots N\}$.

The index μ might refer, for example, to a location, and ν to a mode, and the alternative $X_{\mu\nu}$ to a journey by mode ν to location μ. The problem now is to determine, from a knowledge of the random utility functions $\{U(\mu,\nu);\ \mu=1, \ldots M;\ \nu=1, \ldots N\}$, the probabilities $P_{\mu\nu}$ of selecting the various alternative combinations of locations and modes. An appropriate utility function for the two dimensional context is of the form

$$U(\mu,\nu) = \bar{U}_{\mu} + \bar{U}_{\nu} + \bar{U}_{\mu\nu} + \varepsilon_{\mu} + \varepsilon_{\nu} + \varepsilon_{\mu\nu} \qquad \mu=1, \ldots M;\ \nu=1, \ldots N \tag{20}$$

If now the set members $\{\varepsilon_{\mu},\ \mu=1, \ldots M\}$, $\{\varepsilon_{\nu},\ \nu=1, \ldots N\}$, and $\{\varepsilon_{\mu\nu} : \mu=1, \ldots N;\ \nu=1, \ldots M\}$ are each <u>separately</u> identically and independently distributed with standard deviations σ_D, σ_{MS} and σ_{DM} respectively, then the matrix $\underline{\Sigma}$ contains the following elements

$$\Sigma_{\mu\nu,\mu'\nu'} = \sigma_D^2\, \delta_{\mu\mu'} + \sigma_{MS}^2\, \delta_{\nu\nu'} + \sigma_{DM}^2\, \delta_{\mu\mu'} \cdot \delta_{\nu\nu'} \tag{21}$$

in which δ is the Kronecker delta. With σ_D and σ_{MS} <u>both</u> zero the matrix $\underline{\Sigma}$ is diagonal and the multinomial logit model is an appropriate model form. If, however, $\{\bar{U}_{\nu} = 0,\ \nu=1, \ldots N\}$ and $\sigma_{MS} = 0$, we may write

$$P_{\mu\nu} = \text{Prob}\ \{U(\mu, \nu) \geq U(\mu', \nu') : \forall\ \mu', \nu'\} \tag{22}$$

$$= \text{Prob}\ \{U_{\mu} + \max_{\nu^*} U_{\mu^*} \geq U_{\mu'} + \max_{\nu^*} U_{\mu'^*} \ : \forall_{\mu'}\} \cdot \text{Prob}\ \{U_{\mu\nu} > U_{\mu\nu'} \ : \forall_{\nu'}\} \tag{23}$$

in which $\max_{\nu^*} U_{\mu^*}$ is a utility function distributed according to the <u>maximum</u> of the random variables $U_{\mu 1} \ldots U_{\mu\nu} \ldots U_{\mu N}$.

When appropriate distributions are selected for the utility functions the <u>nested logit model</u> (NLM) is formed (Williams, (1977); Daly and Zachary, (1978) McFadden, (1979))

$$P_{\mu\nu} = \frac{e^{\beta(\bar{U}_{\mu} + \bar{U}_{\mu^*})}}{\sum_{\mu} e^{\beta(\bar{U}_{\mu} + \bar{U}_{\mu^*})}} \cdot \frac{e^{\lambda \bar{U}_{\mu\nu}}}{\sum_{\nu} e^{\lambda \bar{U}_{\mu\nu}}} \,. \tag{24}$$

The so-called <u>composite utilities</u> $\tilde{U}_{\mu*}$ are given by

$$\bar{U}_{\mu*} = \frac{1}{\lambda} \ln \sum_{\nu} e^{\lambda \bar{U}_{\mu\nu}} \tag{25}$$

and the parameters β and λ by

$$\lambda = \frac{\pi}{\sqrt{6}\,\sigma_{DM}} \quad ; \quad \beta = \frac{\pi}{\sqrt{6}} \left(\sigma_D^2 + \frac{\pi^2}{6\lambda^2}\right)^{-\frac{1}{2}} \tag{26}$$

Note that $\beta \leq \lambda$. When the structure (24) is adopted for probabilistic choice modelling the conditions (25) and (26) must be satisfied if we wish to underpin the model with the notion of optimal choice between discrete alternatives. It may readily be checked that when σ_D tends to zero the model collapses to a multinomial logit form.

Expressions corresponding to Equations (24)-(26) exist when $\sigma_{MS} \neq 0$, $\sigma_D = 0$. If, however, σ_D <u>and</u> σ_{MS} are non zero a more general choice model D*MS is appropriate. For this case Williams (1977) has proposed a <u>cross correlated logit model</u> (CCL) which collapses in appropriate limits to nested- or multinomial logit forms.

In summary, we note that within the framework of random utility theory in which behaviour is governed by rational choice between discrete alternatives, the structure of the model is determined uniquely by the underpinning utility functions, and the structure of correlation or similarity between alternative choices is the essential feature which dictates the complexity of the model. Varying degrees of similarity may be accommodated within the logit family, the multinomial logit model being appropriate when the variance-covariance matrix is of the form (13).

SOME PRACTICAL IMPLICATIONS OF THE THEORY

If it is accepted that individuals select alternatives and respond to changes in a manner which approximates to the assumptions of the above theory, there are two immediate practical consequences. Firstly, as the three model structures D/MS, MS/D and D-MS are all special cases of the more general structure D*MS (in which σ_D, σ_{MS} and σ_{DM} are all non-zero), the structural ambiguity refered to earlier may be obviated if the latter model is implemented.

Secondly, if a particular model, say the nested structure D/MS (a 'post-distribution modal split model') is adopted for forecasting demand response, the composite utilities and estimated elasticity parameters β and λ of the distribution and modal split logit structures must be consistent with the theoretical conditions underpinning the model. It has been found that in those British Transport Studies which have employed D/MS nested logit models either the condition (25) and/or the parameter relation $\beta \leqslant \lambda$ have been violated. These violations can give rise to highly unrealistic response properties of the models, as discussed by Williams and Senior (1977).

While a theory of model structure (and corresponding economic evaluation measures) now exists which is consistent with rational choice behaviour, there are many theoretical and practical issues which remain to be resolved. The cross-correlated logit model or multinomial probit model appropriate to the D*MS form have yet to be implemented, and it is presently desirable to implement all three special structures D/MS, MS/D and D-MS and select that which yields the best statistical fit and is consistent with the theoretical conditions outlined in the previous section (Ben-Akiva, 1974; Senior and Williams, 1977). It remains to assess the extent of mis-specification involved in the implementation of a particular model in circumstances for which a more general representation is appropriate. In this context the authors have employed Monte Carlo simulation methods to investigate the conditions under which simple and easily implemented model structures - such as the multinomial logit model - are likely to give rise to large theoretical errors in practice.

In terms of theoretical developments, the relaxation of a number of assumptions associated with the decision process of 'homo economics' is of high priority. The incorporation of habit, learning, limited information, search and satisficing in models of behaviour awaits full development within an extended framework of random utility theory.

REFERENCES

Ben-Akiva, M.E. 1974. Structure of Passenger Travel Demand Models. Transportation Research Record, 526, 26-42.

Daly, A.J. and Zachary, S. 1978. Improved Multiple Choice Models, in Hensher, D.A. and Dalvi, Q. (eds.) Determinants of Travel Choice, Saxon House, England.

Domencich, T. and McFadden, D. 1975. Urban Travel Demand : A Behavioural Analysis. North Holland, Amsterdam.

McFadden, D. 1979. Quantitative methods for analysing travel behaviour of individuals : some recent developments, in Hensher, D.A. and Stopher, P.R. (eds.) Behavioural Travel Modelling, Croom Helm, London.

Senior, M.L. and Williams, H.C.W.L. 1977. Model based transport policy assessment I : The use of alternative forecasting models. Traffic Engineering and Control, 18(9), 402-406.

Spear, B.D. 1977. Applications of New Travel Demand Forecasting Techniques to Transportation Planning : A study of individual choice models. U.S. Department of Transport, FHWA, Office of Highway Planning.

Williams, H.C.W.L. 1977. On the formation of travel demand models and economic evaluation measures of user benefit. Environment and Planning A, 9, 285-344.

Williams, H.C.W.L. and Senior, M.L. 1977. Model based transport policy assessment II : removing fundamental inconsistencies from the models. Traffic Engineering and Control, 18(10), 464-469.