



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/2403/>

---

**Monograph:**

Ortuzar, J.D. (1979) Testing the Theoretical Accuracy of Travel Choice Models Using Monte Carlo Simulation. Working Paper. Institute of Transport Studies, University of Leeds , Leeds, UK.

Working Paper 125

---

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



## White Rose Research Online

<http://eprints.whiterose.ac.uk/>

ITS

[Institute of Transport Studies](#)

**University of Leeds**

This is an ITS Working Paper produced and published by the University of Leeds. ITS Working Papers are intended to provide information and encourage discussion on a topic in advance of formal publication. They represent only the views of the authors, and do not necessarily reflect the views or approval of the sponsors.

White Rose Repository URL for this paper:

<http://eprints.whiterose.ac.uk/2403/>

---

### **Published paper**

Ortuzar, J.D. (1979) *Testing the Theoretical Accuracy of Travel Choice Models Using Monte Carlo Simulation*. Institute of Transport Studies, University of Leeds, Working Paper 125

---

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

*ITS Working Paper 125*

November 1979

**TESTING THE THEORETICAL ACCURACY OF  
TRAVEL CHOICE MODELS USING MONTE CARLO  
SIMULATION**

**BY J. D. ORTUZAR**

*ITS Working Papers are intended to provide information and encourage discussion on a topic in advance of formal publication. They represent only the views of the authors, and do not necessarily reflect the views or approval of the sponsors.*

## CONTENTS

	<u>Page</u>
Abstract	
1. Introduction	1
2. The Generation of Random Utility Models	3
3. Attribute Correlation and Model Structures	8
4. The Probit Model: Structure and Transformations	15
5. The Theoretical Accuracy of Alternative Logit Model Structures	24
6. Conclusions	29
References	31
Acknowledgements	
Figures	
Appendices	

### ABSTRACT

ORTUZAR, J. de D. (1979) Testing the theoretical accuracy of travel choice models using Monte Carlo simulation. Leeds: University of Leeds, Inst. Trans. Stud., WP 125 (unpublished)

In recent years a considerable advance has been made in the construction of micro-travel demand models from choice theoretic principles. Within random utility theory, the structure of models may be shown to relate to the perceived similarity between discrete choice alternatives, and this aspect may be interpreted mathematically in terms of the correlation between the components of random utility functions. Several possible model structures have now been proposed, varying from the multinomial logit model (uncorrelated) through the partly correlated structures (hierarchical and cross-correlated logit functions) to the most general form of probit model which allows an arbitrary variance-covariance matrix.

In this paper, these model structures are discussed using a geometric interpretation of random utility theory, and the possibility of invoking transformations on the general probit model is examined. Monte Carlo simulation methods are then used to investigate some aspects of the trade-off between the generality and accuracy of correlated structures (the cross-correlated logit model in particular) and the greater ease with which less consistent structures may be implemented. In this way, the theoretical accuracy of the multinomial logit model is assessed.

It is concluded that where the general probit model is too complex to implement, the practice of comparing the multinomial logit model with alternative hierarchical logit structures is unlikely to lead to significant errors in forecasting.

TESTING THE THEORETICAL ACCURACY OF TRAVEL  
CHOICE MODELS USING MONTE CARLO SIMULATION

---

1. INTRODUCTION

In recent years considerable interest has centred on the relationship between the structure of a travel demand model and the behavioural principles associated with its formation. This has arisen not only because of the need to underpin models with a consistent theoretical rationale, but also from the recognition of structural ambiguities in existing models - as, for example, with the relative positions of distribution and modal split models in the conventional planning system - which can give rise to significantly different results in policy analysis (Ben-Akiva, 1974; Williams and Senior, 1977). One particular framework within which this relationship has been sought is that provided by random utility theory (for a review, see Domencich and McFadden, 1975).

In this quantal choice theory individuals are considered to associate with each member  $A_n$ ;  $n=1, \dots, N$  of a discrete set of options  $A$ , a net utility  $U_n$ ;  $n=1, \dots, N$ , and to select that member with the highest value of  $U$ . To account for interpersonal variation in the value of attributes incorporated in the utility functions, and the influence of unobserved factors, the modeller considers the variables  $(U_1, \dots, U_n, \dots, U_N)$  to be randomly distributed over the population confronted by a choice. The probability  $P_n$  that an individual with particular characteristics selects an alternative  $A_n$  is then simply expressed in terms of the probability that  $U_n$  be greater than those values associated with all other options. A formal choice model may be derived when the density function  $f(\underline{U}) = f(U_1, \dots, U_N)$  of the utility components is specified.

It has recently been recognised that the analytic structure of a model is crucially related to the interdependency, or statistical correlation, between the utility functions associated with each alternative - that is, with the structure of  $f(\underline{U})$  (Williams, 1977; Langdon 1976; Daly and Zachary, 1978; McFadden, 1979). A set of formal models now exists which accommodates varying degrees of "similarity" or correlation between alternatives ranging from the widely used multinomial logit model (MNL),

generated by uncorrelated distributions, through the hierarchical logit model (HL), to the generalised probit function (GP) with arbitrary correlation, expressed in terms of a variance - covariance matrix.

Until recently application of the generalised probit model has been restricted to a small number (3 or 4) of choice options (Hausman and Wise, 1978). However by invoking the Clark approximation (Clark, 1961), Daganzo et al (1977) have extended its practical range. In spite of the advances in its applicability there appear to be many practical cases in which (a) the model cannot cope (Daganzo, 1979), or (b) there is a need for a compromise between the generality it can afford and the greater ease with which less consistent structures may be implemented. One such compromise is the cross-correlated logit (CCL) function (Williams, 1977), which is a closed-form model containing alternative HL functions as special cases.

In this paper we wish to examine some general themes such as the relationship between certain utility functions and the structure of travel choice models; the possibility of invoking transformations in order to simplify models and derive conceptual links between them; the theoretical accuracy of particular choice models, and the problems of misspecification associated with model structures and utility functions. More specifically, we wish to address the following questions:

- i) Is it possible to apply transformations in 'utility space' in order to simplify 'symmetric' probit models and enable conceptual links to be forged with the logit family?
- ii) How serious is the absence of 'similarity effects' in the multinomial logit model? In other words, how much is the well known 'independence from irrelevant alternatives' (IIA) property of the model an impediment in choice modelling?
- iii) How good an approximation to a general function is the cross-correlated logit model?
- iv) What is the effect of misspecification of choice models with respect to model structure and their utility components?
- v) Can any of the logit models display pathological response properties, and is it possible to recognise their symptoms at the calibration stage?

vi) Can we discriminate between contending model structures on the basis of goodness of statistical fit, and the character of their inherent elasticity parameters? In particular, does a good agreement to base year data necessarily imply good response characteristics?

It should be stressed at the outset that any reference to the accuracy of a model will refer to its consistency with the underlying theoretical rationale, and not necessarily to its appropriateness in choice modelling.

In Section 2, the basic principles of generating random utility models are reviewed, a geometric interpretation of the theory is presented, and the Monte Carlo method as a means for numerical evaluation of choice models is outlined. The existence and implications of correlation between the utility functions associated with different alternatives are then examined in Section 3 and the various approaches to its incorporation in choice models noted. In Section 4 we investigate the possibility of invoking transformations in utility space as a means of simplifying the general probit model. Although conceptionally appealing in terms of its links with MNL, HL and CCL structures, the potential for implementing such transformations does not, in general, appear practicable.

The numerical tests to determine the theoretical accuracy of the alternative logit structures in a general choice context are then described in detail in Section 5.

## 2. THE GENERATION OF RANDOM UTILITY MODELS

Formally, we can express the model generator equations of random utility theory as follows:

$$P_n = \text{Prob} (U_n > U_n, \forall A_n, \epsilon A) \quad (2.1)$$

$$= \int_{R_n} dU f(U) \quad (2.2)$$

in which  $f(U)$  is the joint distribution function of  $(U_1, \dots, U_N)$  and  $R_n$  is that region of utility space defined by

$$R_n: U_n \geq U_{n'}, \quad \forall A_{n'} \in A \quad (2.3)$$

$$U_n \geq 0 \quad (2.4)$$

In this paper we shall be concerned only with those cases in which a trip is actually made. The non-negativity restriction (2.4) will thus be considered inoperative. For the distribution functions considered later this will involve a negligible inconsistency, which does not affect the argument to be presented.

To derive an explicit probabilistic choice model we need to know both the form of  $f(\underline{U})$  and an expression for the utility functions in terms of the attributes of alternatives in the set  $A$ .

We shall take the components  $U_n$  to be of the following form:

$$U_n(\underline{\theta}, \underline{Z}_n) = \bar{U}_n(\underline{\theta}, \underline{Z}_n) + \varepsilon_n \quad (2.5)$$

in which  $\bar{U}_n$  is the so-called 'representative' utility of the population  $Q$  confronted by the choice, and  $\varepsilon_n$  is a stochastic residual.  $\bar{U}_n$  is normally taken to be linear in terms of the attributes  $Z_n^\mu$  characterising  $A_n$ . That is:

$$\bar{U}_n(\underline{\theta}, \underline{Z}_n) = \sum_{\mu} \theta_{\mu} Z_n^{\mu}, \quad \forall A_n \in A \quad (2.6)$$

$$= \underline{\theta} \cdot \underline{Z}_n \quad (2.7)$$

The vector of parameters  $\underline{\theta}$  is estimated from observed choices. It remains to specify the distribution function  $f(\underline{U})$  or equivalently that of the stochastic residuals  $\underline{\varepsilon}$ .

A geometric interpretation of the theory may readily be derived from expression (2.2). In the utility space  $\mathcal{U}$ , bounded by the components  $(U_1, \dots, U_N)$ , the probability  $P_n$  is, for normalised  $f(\underline{U})$ , the total density of points in the region  $R_n$  bounded by the hyperplanes defined by:

$$U_n = 0$$

and

$$U_n = U_{n'}, \quad \forall A_{n'} \in A \quad (2.8)$$

This can be more easily seen in the convenient cartesian space. In Figure 1(a) we illustrate the fundamental utility distributions associated with binary choice (Williams, 1977). It is important to

distinguish those distributions  $g_1(U_1)$  and  $g_2(U_2)$ <sup>(1)</sup> which are associated with the whole population  $Q$  confronted by the choice between alternatives  $A_1$  and  $A_2$ , from  $\tilde{g}_1(U_1)$  and  $\tilde{g}_2(U_2)$  which are the "choice specific" distributions of utility, for those members of  $Q$  who have selected options  $A_1$  and  $A_2$  respectively. The sum of these last two distributions is termed the distribution of maximum utility  $g_*(U)$ , and the three functions are formally defined as follows:

$$\tilde{g}_1(U_1) = \int_{R_1} dU_2 f(U_1, U_2) \quad (2.9)$$

$$\tilde{g}_2(U_2) = \int_{R_2} dU_1 f(U_1, U_2) \quad (2.10)$$

$$\tilde{g}_*(U) = \int_{R_1} dU_2 f(U_1, U_2) + \int_{R_2} dU_1 f(U_1, U_2) \quad (2.11)$$

We shall also write the distribution of maximum utilities in the form

$$g_*(U) = \max_U (U_1, U_2) \quad (2.12)$$

and we note here that the mean value of this distribution,  $\bar{U}_*$ , has great significance in the evaluation problem (Williams, 1977).

The geometric interpretation of this simple choice process, which is an extension of that provided by Robertson (1977), is given in Figure 1(b). For identical and independent distributions (IID),  $f(U)$  has a circularly symmetric shape centred on  $\bar{U}_1$  and  $\bar{U}_2$ . The line OZ divides the positive quadrant into the regions  $R_1$  and  $R_2$ , and  $P_1$  and  $P_2$  comprise of those corresponding portions of the distribution in these regions. The distinction between  $g(U_1)$  and the choice specific distribution  $\tilde{g}(U_1)$  can readily be seen in terms of the respective projections onto the  $U_1$  axis of the density function  $f(U)$ , and that portion of  $f(U)$  bounded by OZ and the  $U_1$  axis.

An important class of random utility models includes those generated by IID utility distributions for which we can decompose  $f(U)$  as follows:

$$f(U) = \prod_{n=1}^N g(U_n) \quad (2.13)$$

... ..

(1) Because they are identical in the figure, we have labelled them  $g(U_1)$  and  $g(U_2)$ , respectively.

Here  $g(U_n)$  is the distribution of the utility component associated with  $A_n$ . The expression for  $P_n$  can now be written

$$P_n = \int_{-\infty}^{\infty} dU_n g(U_n) \prod_{n' \neq n} \int_{-\infty}^{U_n} dU_{n'} g(U_{n'}) \quad (2.14)$$

Omission of the constraint (2.4) allows the lower limits of integration to be extended to minus infinity.

It is by now widely known that the much favoured multinomial logit model (MNL)

$$P_n = \frac{e^{\frac{\Delta \bar{U}_n}{\Delta \bar{U}_{n'}}}}{\sum_{n'} e^{\frac{\Delta \bar{U}_{n'}}{\Delta \bar{U}_{n'}}}} \quad (2.15)$$

is an IID model generated from Weibull (Gnedenko) probability distributions (Charles Rivers Associates, 1972) for which

$$g(U_n) = \Delta e^{-\Delta(U_n - \bar{U}_n)} e^{-e^{-\Delta(U_n - \bar{U}_n)}} \quad (2.16)$$

This is a skewed unimodal distribution, in which the dispersion parameter  $\Delta$  is inversely related to the standard deviation,  $\sigma$ , as follows (Cochrane, 1975):

$$\Delta = \frac{\pi}{\sqrt{6}\sigma} \quad (2.17)$$

Similarly simple probit models are generated from IID Normal distributions.

For a number of special distributions, it is possible to evaluate the integral (2.2) to produce analytical expressions for  $P_n$ , such as the MNL in equation (2.15). In general, however, we have to resort to some form of numerical method. One such approach involves Monte Carlo simulation. As far as we are aware the first application of this method to the solution of random utility models is that of Albright, Lerman and Manski (1977), in the development of an estimation program for the general probit model. However, and in most cases independently, the power of the approach has attracted numerous applications recently (Bonsall, 1979; Chicago Area Transportation Study, 1979; Horowitz, 1978; Kreibich, 1979; Manski and Lerman, 1978; Ortuzar, 1978; Robertson, 1977; Robertson and Kennedy, 1979; Williams and Ortuzar, 1979) but clearly its roots can be traced back to certain stochastic assignment methods (Burrell, 1968).

In this approach we follow tradition (Hammersley and Handscomb, 1965); a sample of size  $S$  is created, and each 'individual' member  $t$ , of  $S$ , is confronted by the choice between  $A_1, \dots, A_N$ . Using a random number generator a set of utility values  $(U_1, \dots, U_N)$  is drawn from  $f(\underline{U})$ , and the member  $t$  is assigned to that option with the maximum associated utility. For large  $S$ , the proportion  $S_n$  of 'individuals' assigned to option  $A_n$  will approximate to  $P_n$ , which is given by

$$P_n = \lim_{S \rightarrow \infty} \frac{S_n}{S} \quad (2.18)$$

In the simple cartesian utility plane examined before, the method involves random sampling of points from  $f(U_1, U_2)$ . For a given sampled observation  $(U_1^t, U_2^t)$ , the corresponding 'individual' will be assigned to  $A_1$  or  $A_2$  according to the region in which the 'point' may be found. That is

$$\begin{aligned} &\text{if } U_1^t > U_2^t, \text{ i.e. } (U_1^t, U_2^t) \in R_1, \text{ assign to } A_1 \\ &\text{if } U_1^t < U_2^t, \text{ i.e. } (U_1^t, U_2^t) \in R_2, \text{ assign to } A_2 \end{aligned} \quad (2.19)$$

To test the accuracy of the method with sample size, the numerical solution of a 4-option logit model was compared with the analytic solution. For a sample of size  $S$ , the choice probabilities  $P_n^S$  were determined by drawing random values from four IID Weibull functions, with given means  $(\bar{U}_1, \bar{U}_2, \bar{U}_3, \bar{U}_4)$  and standard deviation  $\sigma$ . These numerically derived probabilities were then fitted by a logit function

$$P_n^S = \frac{e^{\Delta^S \bar{U}_n}}{\sum_{n'} e^{\Delta^S \bar{U}_{n'}}} \quad (2.20)$$

in which the parameter  $\Delta^S$  was estimated by the usual maximum likelihood method (Domencich and McFadden, 1975). In Figure 2, we show the empirically derived relationship between the variance of  $\Delta^S$  with the size of the sample  $S$ . In order to examine the accuracy of the numerical solution under different conditions, we repeated this procedure for a binary logit model and different values of the difference between mean utilities. As it can be seen, the closer the options (smaller difference in mean utilities), the less stable the simulation becomes. In the

numerical tests described in later sections the sample size was fixed at  $S = 30,000$ .<sup>(2)</sup>

We now proceed to consider more complex choice contexts in which the presence of correlation between utility functions is central to the structural development of the models.

### 3. ATTRIBUTE CORRELATION AND MODEL STRUCTURES

For the utility distributions  $U_n$ ;  $n=1, \dots, N$  we can define a variance-covariance matrix  $\underline{\Sigma}$  with elements  $\Sigma_{nn'}$  given by:

$$\begin{aligned} \Sigma_{nn'} &= E (U_n - \bar{U}_n; U_{n'} - \bar{U}_{n'}) \\ &= E (\epsilon_n \epsilon_{n'}) \quad \forall A_n, A_{n'} \in A \end{aligned} \quad (3.1)$$

in which  $E(\cdot)$  denotes an expectation value. In the case of IID utility components the matrix has, by construction, a simple diagonal form

$$\underline{\Sigma} = \sigma^2 \underline{I} \quad (3.2)$$

where  $\underline{I}$  is the unit matrix of dimension  $N$ , and  $\sigma$  the common standard deviation of the distributions  $g(U)$ , that is

$$\sigma^2 = E(\epsilon_n \epsilon_n) \quad \forall A_n \in A \quad (3.3)$$

It is one of the intentions of this work to determine the extent to which this very simple structure constitutes a real restriction to choice modelling.

The multinomial logit (MNL) model (2.15) generated from IID Weibull distributions, which is therefore characterised by a matrix with the diagonal structure (3.2), has been very widely applied in mode choice, and more recently destination choice modelling (for a review, see Spear 1977). It is now well known, however, that the model suffers a restrictive property of cross-substitution, the 'independence from

... ..

(2) Manski and Lerman (1978) have examined the simulation approach carefully and have proposed less naive stopping rules for more efficient programs.

irrelevant alternatives' (IIA) property, whereby the ratio

$$\frac{P_n}{P_n'} = e^{\Delta(\bar{U}_n - \bar{U}_n')} \quad \forall A_n, A_n' \in A \quad (3.4)$$

is independent of the utility values associated with other options. The IIA property, once seen as a positive advantage to be exploited in 'new option' situations, is now recognised to be a potential hazard when certain alternatives are more 'similar' than others in the set A. In random utility theory this notion of 'similarity' is interpreted in terms of the presence of off-diagonal elements in the matrix  $\underline{\Sigma}$ .

In certain applications, specific forms for the utility functions tend to suggest themselves. Consider 'two dimensional' choice contexts involving, for example, combinations of destination (D) and mode (M). Alternatives in each dimension will be denoted by  $(D_1, \dots, D_n, \dots, D_N)$  and  $(M_1, \dots, M_m, \dots, M_M)$ , respectively, and the combination of dimensions produces the NM discrete choice options  $(D_1 M_1, \dots, D_n M_m, \dots, D_N M_M)$ , which comprise the set A. The general element  $A_n$  is now  $D_n M_m$  which might be a specific destination-mode combination for the purpose of performing an activity.

For such choice contexts we shall be particularly interested in utility functions of the form

$$U(n,m) = U_n + U_m + U_{nm} \quad \forall D_n M_m \in A \quad (3.5)$$

here  $U_n$  and  $U_m$  may, for example, correspond to destination and mode specific utilities, respectively, while  $U_{nm}$  might be the travel disutility associated with  $D_n M_m$  combination. This form was used in the shopping model developed by Ben Akiva (1974), and in a number of other applications in the United States since that time.

Writing  $U(n,m)$  in terms of a 'representative' term  $\bar{U}(n,m)$  and a stochastic residual  $\varepsilon(n,m)$  we have

$$U(n,m) = \bar{U}(n,m) + \varepsilon(n,m) \quad (3.6)$$

in which

$$\bar{U}(n,m) = \bar{U}_n + \bar{U}_m + \bar{U}_{nm} \quad (3.7)$$

and

$$\varepsilon(n,m) = \varepsilon_n + \varepsilon_m + \varepsilon_{nm} \quad (3.8)$$

We shall now assume that the residuals  $\varepsilon_n$ ,  $\varepsilon_m$  and  $\varepsilon_{nm}$  are separately IID, with

$$\begin{aligned} E(\varepsilon_n \varepsilon_{n'}) &= \delta_{nn'} \sigma_D^2 \\ E(\varepsilon_m \varepsilon_{m'}) &= \delta_{mm'} \sigma_M^2 \\ E(\varepsilon_{nm} \varepsilon_{n'm'}) &= \delta_{nn'} \delta_{mm'} \sigma_{DM}^2 \\ E(\varepsilon_n \varepsilon_m) &= E(\varepsilon_n \varepsilon_{nm}) = E(\varepsilon_m \varepsilon_{nm}) = 0 \quad \forall D, M \in A \end{aligned} \quad (3.9)$$

in which  $\delta$  is the Kronecker delta. The elements of  $\underline{\Sigma}$  now become

$$\Sigma_{nm, n'm'} = \delta_{nn'} \sigma_D^2 + \delta_{mm'} \sigma_M^2 + \delta_{nn'} \delta_{mm'} \sigma_{DM}^2 \quad (3.10)$$

and the matrix is expressed in Figure 3, together with those corresponding to the residual structures

$$\varepsilon(n,m) = \varepsilon_{nm} \quad (3.11)$$

$$\varepsilon(n,m) = \varepsilon_n + \varepsilon_{nm} \quad (3.12)$$

$$\varepsilon(n,m) = \varepsilon_m + \varepsilon_{nm} \quad (3.13)$$

which are clearly special cases of that defined in Equation (3.8). It is readily seen that the source of correlation in 'multiple dimension' cases is the existence of a common term or 'dimension specific' element ( $U_n$  or  $U_m$ ) in the utility function. For the four cases (3.8), (3.11) - (3.13)

we have developed in Figure 3, a pictorial representation of the structure of the  $\underline{\Sigma}$  matrix with correlation between alternatives incorporated through common bonds as shown. This is the basis for a representation of the choice model itself (Williams, 1977). In the first case both  $\sigma_D$  and  $\sigma_M$  are zero and a diagonal  $\underline{\Sigma}$  matrix results. This case which is consistent with Equation (3.11) will correspond to the MNL model (2.15) if the utility functions are drawn from IID Weibull distributions. It is clear that the use of the utility function (3.5) in a MNL model of the form (2.15) will be inconsistent because the appropriate  $\underline{\Sigma}$  matrix, corresponding to that utility function, is not of the diagonal form involved in the generation of the model.

Before treating the more general case (3.8), which is consistent with the utility function (3.5) and which corresponds to the fourth  $\underline{\Sigma}$  matrix of Figure 3, we shall consider the derivation of a hierarchical or nested model from a function consistent with the residual structure (3.12),

$$\begin{aligned} U(n,m) &= U_n + U_{nm} \\ &= \bar{U}_n + \bar{U}_{nm} + \epsilon_n + \epsilon_{nm} \end{aligned} \quad (3.14)$$

and which corresponds to the second representation in Figure 3. In this case the component  $\sigma_M$  vanishes and the two parameters  $\sigma_D$  and  $\sigma_{DM}$  allow different degrees of cross-substitution between intra and inter-branch alternatives in the 'tree' form shown in Figure 3(b); that is, between  $D_n M_m$  and  $D_n M'_m$ , in the former case, and between  $D_n M_m$  and  $D_n M'_m$ , in the latter. It may be shown (Williams, 1977) that  $P(n,m)$ , the probability of selecting  $D_n M_m$  can be written

$$P(n,m) = P_n \cdot P_{nm} \quad (3.15)$$

in which

$$P_{nm} = \text{Prob} (U_{nm} > U_{nm'}, \forall M_m' \in M) \quad (3.16)$$

and

$$P_n = \text{Prob} (U_n + U_{n^*} > U_{n'} + U_{n'^*}, \forall D_{n'} \in D) \quad (3.17)$$

with

$$U_{n^*} = \max_m (U_{n1}, \dots, U_{nm}, \dots, U_{nM}) \quad (3.18)$$

If the components  $U_{nm}$  are Weibull distributed variables  $W(U, \bar{U}_{nm}, \Delta)$  with mean  $\bar{U}_{nm} + \gamma/\Delta$  (where  $\gamma$  is Euler's constant), and standard deviation  $\pi/(\sqrt{6}\Delta)$ , then it is readily shown (Cochrane, 1975) that  $U_{n^*}$  is also Weibull distributed, with mean

$$\bar{U}_{n^*} = \frac{1}{\Delta} \ln \sum_m e^{\Delta \bar{U}_{nm}} + \gamma/\Delta \quad (3.19)$$

and standard deviation given by

$$\sigma_{D^*} = \frac{\pi}{\sqrt{6}\Delta} \quad (3.20)$$

The marginal distribution  $P_n$  is then derived from the sum of Weibull distributed variables  $U_{n^*}$  and variables  $U_n$ , derived from some distribution  $\Gamma(U, \bar{U}_n)$ ,  $n=1, \dots, N$  to be specified.

Now the hierarchical logit (HL) model (Williams, 1977; Daly and Zachary, 1978; McFadden, 1979)

$$P(n, m) = \frac{e^{\beta(\bar{U}_n + \bar{U}_{n^*})}}{\sum_{n'} e^{\beta(\bar{U}_{n'} + \bar{U}_{n'^*})}} \cdot \frac{e^{\Delta \bar{U}_{nm}}}{\sum_{m'} e^{\Delta \bar{U}_{nm'}}} \quad (3.21)$$

can be generated by specifying that  $\Gamma(U, \bar{U}_n)$  be that distribution of a variate which is formed from the difference between random variables drawn from Weibull functions  $W(U, \bar{U}_n + \bar{U}_{n^*}, \beta)$  and  $W(U, \bar{U}_{n^*}, \Delta)$ .<sup>(3)</sup>

Because  $U_n$  and  $U_{n^*}$  are independent, the variance of their sum is given by

$$\frac{\pi^2}{6\beta^2} = \sigma_D^2 + \frac{\pi^2}{6\Delta^2} \quad (3.22)$$

or

$$\frac{\beta}{\Delta} = \left(1 + \frac{6\sigma_D^2 \Delta^2}{\pi^2}\right)^{-\frac{1}{2}}$$

... ..

(3) A logistic distribution (Domencich and McFadden, 1975).

When  $\sigma_D = 0$ , the model collapses to the MNL, characterised by the single parameter  $\Delta$ . It can be seen that for a consistent model (and for  $\Gamma(U, \bar{U}_n)$  to have a non-negative variance), we require (Williams, 1977)

$$\beta \leq \Delta \tag{3.23}$$

This condition is of particular importance, and its violation may imply cross-elasticities of the wrong sign. Violation has, in fact, been observed in conventional transport models (Williams and Senior, 1977). We will come back to this concept later when discussing the pathological response properties of certain mis-specified models.

In the simulation tests to be described in Section 5, in which the model corresponding to Equation (3.8) is derived numerically,  $\epsilon_n$ ,  $\epsilon_m$  and  $\epsilon_{nm}$  will themselves be taken as Weibull functions, and it is necessary to know what approximations are made if the resultant model is assumed to be of HL form. In fact, the only approximation is involved in the marginal probability  $P_n$ , because the sum of the two Weibull variates, drawn from the distributions  $W(U, \bar{U}_{n*}, \Delta)$  and  $W(U, \bar{U}_n, \pi/(\sqrt{6}\sigma_D))$  is not itself distributed Weibull.

For an example with  $N = M = 2$ , the parameters  $\beta$  and  $\Delta$  (the latter should be exact) were estimated from the logit function (3.21) by Maximum Likelihood, and their ratio was plotted against the standard deviation  $\sigma_D$  associated with the residuals  $\epsilon_n$ , and compared with the theoretical values in Equation (3.22). The results of this exercise are shown in Figure 4. It can be seen that a reasonably good approximation is obtained.

We now turn to consider the choice model generated from the utility function (3.5). Because of the form of the random residuals, (3.8), we can say immediately that this model should contain as special cases the MNL and alternative HL functions. As far as the author is aware no explicit analytic function has been obtained for such a structure. Clearly one could appeal to the probit form and exploit the Clark approximation (Clark, 1961), but this would for medium size problems still be unmanageable. Alternatively, we could try to exploit the very symmetric structure of  $\Sigma$  (as shown in Figure 3(d)) and attempt to transform the probit model into an equivalent MNL model. In fact, this will be the subject of the next section.

The cross-correlated logit function (CCL) was an ad-hoc model proposed by Williams (1977)<sup>(4)</sup> as a closed form approximation which corresponded to the utility function (3.5). It is defined by the equations

$$P(n, m) = \frac{e^{\beta U_n^\dagger + \lambda U_m^\dagger + \Delta \bar{U}_{nm}}}{\sum_{n', m'} e^{\beta U_{n'}^\dagger + \lambda U_{m'}^\dagger + \Delta \bar{U}_{n', m'}}} \quad (3.24)$$

where

$$U_n^\dagger = \bar{U}_n + \frac{(\beta - \Delta)}{\beta} \bar{U}_{n^*} \quad (3.25)$$

$$U_m^\dagger = \bar{U}_m + \frac{(\lambda - \Delta)}{\lambda} \bar{U}_{m^*} \quad (3.26)$$

$$\bar{U}_{n^*} = \frac{1}{\Delta} \ln \sum_{m'} e^{\Delta \bar{U}_{nm'} + \lambda \bar{U}_{m'}} \quad (3.27)$$

$$\bar{U}_{m^*} = \frac{1}{\Delta} \ln \sum_{n'} e^{\Delta \bar{U}_{n'm'} + \beta \bar{U}_{n'}} \quad (3.28)$$

and

$$\frac{\beta}{\Delta} = \left(1 + \frac{6\sigma_D^2 \Delta^2}{\pi^2}\right)^{-\frac{1}{2}} \quad (3.29)$$

$$\frac{\lambda}{\Delta} = \left(1 + \frac{6\sigma_M^2 \Delta^2}{\pi^2}\right)^{-\frac{1}{2}} \quad (3.30)$$

It may be checked that as  $\sigma_D^2$  and  $\sigma_M^2$ , the variances of the residuals  $\epsilon_n$  and  $\epsilon_m$ , tend to zero the respective hierarchical logit models are formed. If both variances are zero, the CCL collapses to the multinomial logit form (2.15).

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 underlying 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 first three cases in Figure 3 involve utility maximisations in which the variance-covariance

... ..  
 (4) In that paper (section 5.3.2, pp 321-323), the function was denoted General Choice Model. More recently, and in deference to the general probit model and to the class of General Extreme Value (GEV) models (McFadden, 1979), the function has been rechristened appropriately.

matrices  $\underline{\Sigma}$  are special cases of the cross-correlated structure, with a  $\underline{\Sigma}$  matrix and pictorial representation summarised in Figure 3(d). In Section 5 we will present a set of simulation tests on structural misspecification designed to examine some specific questions concerning how good an approximation to (3.5) is the three parameters CCL model, and what potential errors can be introduced by using the single parameter MNL and two parameters HL models instead. First, however, we will examine the general probit model and the scope for applying transformations in order to produce more tractable models.

#### 4. THE GENERAL PROBIT MODEL, STRUCTURE AND TRANSFORMATIONS

In random utility theory, the density function which generates the general probit model (GP), for choice between N alternatives is given by:

$$f(\underline{U}) = \frac{1}{(2\pi)^{N/2} |\underline{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{U} - \bar{\underline{U}})^T \underline{\Sigma}^{-1} (\underline{U} - \bar{\underline{U}})\right\} \quad (4.1)$$

We shall immediately transform Equation (4.1) from  $\underline{U}$ - space into  $\underline{\epsilon}$ - space using Equation (2.5), giving

$$f(\underline{\epsilon}) = \frac{1}{(2\pi)^{N/2} |\underline{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2} \underline{\epsilon}^T \underline{\Sigma}^{-1} \underline{\epsilon}\right\} \quad (4.2)$$

If we define

$$\bar{U}_{nn'} = \bar{U}_{n'} - \bar{U}_n \quad (4.3)$$

then resorting to Equation (2.2) the model can be stated as

$$P_n = \int_{-\infty}^{\bar{U}_{1n} + \epsilon_n} \int_{-\infty}^{\bar{U}_{2n} + \epsilon_n} \dots \int_{-\infty}^{\bar{U}_{Nn} + \epsilon_n} f(\underline{\epsilon}) d\underline{\epsilon} \quad (4.4)$$

Although the GP(4.4) is more general in its theoretical statement, it is considerably more cumbersome than the MNL or HL to implement. The difficulties of achieving a solution to the GP by direct numerical integration for other than 'small' problems, involving 3 or 4 options (Hausman and Wise, 1978) are well known, and have led to the formulation of approximate solution schemes. One method involves Monte Carlo simulation directly to evaluate the model (Albright et al, 1977).

The method is elegant, theoretically appealing and has the advantage of being completely general, in the sense that in principle any function can be integrated. However, it is not well suited for optimisation purposes near the neighbourhood of the optimum, it is biased, and very slow and expensive to use. (Bouthelier, 1978).

The second method, due to Daganzo et al (1977) invokes the Clark (1961) approximation, which essentially involves the replacement of the maximum of bivariate normal variables by one normally distributed variable. By repeated application of the Clark approximation, the multiple integral in Equation (4.4) may be reduced to a particular univariate integral. When the correlation between variables is non-negative, this approximation which has been extensively examined by Manski and Lerman (1978), using Monte Carlo simulation, is apparently accurate to a few per cent, for up to 20 alternatives. However, problems with the possible existence of multiple optima associated with the likelihood function of GP models, for more than 2 alternatives, have recently been reported (Daganzo, 1979). These imply that in general, there is no guarantee that the model can be calibrated. The program and documentation of a powerful algorithm for calibrating the GP model, using this method, are now widely available (Daganzo and Schoenfeld, 1978).

When encountering normally distributed variables, it has often been the case that a transformation to a co-ordinate system in which the structure of variation in a data set is more appropriately described, has provided not only insight into the nature of factors giving rise to the variation, but has also formed the basis for approximation schemes. Principal component analysis is perhaps the best such example. (For a very didactic treatment of transformation theory in multivariate analysis, see Green and Carroll, 1976). Moreover, it is well known that the MNL and an uncorrelated, equal variance probit model (with suitably normalised standard deviation) are almost indistinguishable. That is, if we could transform general probit models into equivalent functions with diagonal variance-covariance matrices, it might be possible to establish conceptual links with the logit family, and in the process erase the burden of numerical integration.

If this were not enough motivation, consider that the hierarchical logit (3.21) may be written as a MNL function with transformed utilities and parameters of the underlying distributions:

$$P(n,m) = \frac{e^{\beta(\bar{U}_n + \frac{\beta-\Delta}{\beta} \bar{U}_{n*} + \frac{\Delta}{\beta} \bar{U}_{nm})}}{\sum_{n'm'} e^{\beta(\bar{U}_{n'} + \frac{\beta-\Delta}{\beta} \bar{U}_{n'*} + \frac{\Delta}{\beta} \bar{U}_{n'm'})}} \quad (4.5)$$

The reader can check this by reorganising and using Equation (3.19). In fact, this should not be too surprising, since McFadden (1979) has shown that any model derived from extreme value (Weibull) or generalised extreme value functions may be written as an equivalent multinomial logit model. In this group follow, for example, the hedonic demand models developed recently by Charles Rivers Associates (Cordell and Reddy, 1977).

Let us first examine and illustrate the power of transformations in the convenient cartesian two dimensional space.

Equation (4.4) reduces in two dimensions simply to:

$$P_1 = \int_{-\infty}^{\infty} d\varepsilon_1 \int_{-\infty}^{\bar{U}_{21} + \varepsilon_1} \frac{\exp\{-\frac{1}{2(1-\rho^2)} [(\frac{\varepsilon_1}{\sigma_1})^2 - \frac{2\rho\varepsilon_1\varepsilon_2}{\sigma_1\sigma_2} + (\frac{\varepsilon_2}{\sigma_2})^2]\}}{2\pi \sigma_1\sigma_2 \sqrt{1-\rho^2}} d\varepsilon_2 \quad (4.6)$$

where  $\rho$  is the coefficient of correlation, and  $\underline{\Sigma}$  is given by

$$\underline{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad (4.7)$$

Figure 5(a) presents a pictorial representation of  $f(\underline{U})$  in  $\underline{U}$ -space for the general variance-covariance matrix (4.7). OZ, the iso-utility line, is defined by

$$OZ: U_1 = U_2 \quad (4.8)$$

and the region of integration  $R_1$  is defined by

$$R_1: \begin{aligned} -\infty &\leq U_1 \leq \infty \\ -\infty &\leq U_2 \leq U_1 \end{aligned} \quad (4.9)$$

Figure 5(b) presents a pictorial representation of the density function in  $\hat{\underline{U}}$ -space defined by the transformation

$$\begin{aligned} \hat{U}_1 &= \frac{1}{\sigma_1} (U_1 - \bar{U}_1) = \frac{\varepsilon_1}{\sigma_1} \\ \hat{U}_2 &= \frac{1}{\sigma_2} (U_2 - \bar{U}_2) = \frac{\varepsilon_2}{\sigma_2} \end{aligned} \quad (4.10)$$

In this transformed space the line OZ, which defines the region of integration is given by

$$OZ: \sigma_2 \hat{U}_2 + \bar{U}_2 = \sigma_1 \hat{U}_1 + \bar{U}_1 \quad (4.11)$$

and consequently the region of integration is:

$$R_1: \begin{aligned} -\infty &\leq \hat{U}_1 \leq \infty \\ -\infty &\leq \hat{U}_2 \leq \frac{\sigma_1}{\sigma_2} \hat{U}_1 + \frac{(\bar{U}_1 - \bar{U}_2)}{\sigma_2} \end{aligned} \quad (4.12)$$

Recall that we are searching for transformations that will restore the symmetry of the independent case. The next move possible is to apply a rotation in order to have the ellipse-shaped density function oriented along the new axes of the co-ordinate system. Figure 5(c) presents a pictorial representation of the probability density function in this new space  $\hat{U}$  defined by

$$\begin{aligned} \hat{U}_1 &= \frac{1}{\sqrt{2}} (\hat{U}_1 + \hat{U}_2) \\ \hat{U}_2 &= \frac{1}{\sqrt{2}} (\hat{U}_2 - \hat{U}_1) \end{aligned} \quad (4.13)$$

Notice that this transformation requires algebraic operations which entangle the previous axes. The region of integration is this time defined by:

$$R_1: \begin{aligned} -\infty &\leq \hat{U}_1 \leq \infty \\ -\infty &\leq \hat{U}_2 \leq \frac{\sigma_1 - \sigma_2}{\sigma_1 + \sigma_2} \hat{U}_1 + \frac{\sqrt{2} (\bar{U}_1 - \bar{U}_2) \sigma_2}{\sigma_1 + \sigma_2} \end{aligned} \quad (4.14)$$

The symmetry will be restored by a further compression of the axes. It can easily be seen that this is achieved by:

$$\begin{aligned} \tau_1 &= \frac{\hat{U}_1}{\sqrt{1 - \rho}} \\ \tau_2 &= \frac{\hat{U}_2}{\sqrt{1 + \rho}} \end{aligned} \quad (4.15)$$

The corresponding pictorial representation is given in Figure 5(d), and the region of integration is defined by

$$R_1: \begin{matrix} -\infty \leq \tau_1 \leq \infty \\ -\infty \leq \tau_2 \leq \frac{\sqrt{1-\rho}(\sigma_1-\sigma_2)}{\sqrt{1+\rho}(\sigma_1+\sigma_2)}\tau_1 + \frac{\sqrt{2}(\bar{U}_1-\bar{U}_2)\sigma_2}{\sqrt{1+\rho}(\sigma_1+\sigma_2)} \end{matrix} \quad (4.16)$$

Appendix 1 gives a matrix treatment of the problem, and shows how the inverse of the variance-covariance matrix defining the quadratic form of the bivariate normal is gradually transformed from the general expression

$$\underline{\Sigma}^{-1} = (1-\rho^2)^{-1} \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix} \quad (4.17)$$

to the simple expression of the independent equal variance model

$$\underline{\Sigma}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (4.18)$$

after the three transformations defined by Equations (4.10), (4.13) and (4.15).

Notice that because all the transformations are linear the iso-utility line OZ remains a line throughout; indeed it is given by

$$OZ: \tau_2 = \frac{\sqrt{1-\rho}(\sigma_1-\sigma_2)}{\sqrt{1+\rho}(\sigma_1+\sigma_2)}\tau_1 + \frac{\sqrt{2}(\bar{U}_1-\bar{U}_2)\sigma_2}{\sqrt{1+\rho}(\sigma_1+\sigma_2)} \quad (4.19)$$

Now, the structural and numerical characteristics of the variance-covariance matrix  $\underline{\Sigma}$  are dependent on the coordinate system (utility space) in which it is measured. It is natural, therefore, to enquire about the

form of the basic problem in the new space in which the transformed matrix  $\tilde{\Sigma}$  is diagonal. We should not expect the benefit of an analytically simpler density function to be obtained at zero cost, for the region of integration  $R_n$ , over which  $A_n$  is preferred in Equation (2.2), will also be transformed.

In general, under the transformation

$$\underline{U} \longrightarrow \underline{T} \quad (4.20)$$

the expression for  $P_n$  given in Equation (2.2)

$$P_n = \int_{R_n} f(\underline{U}) d\underline{U}$$

becomes

$$P_n = \int_{R_{n*}} h(\underline{T}) |J| d\underline{T} \quad (4.21)$$

in which  $h(\underline{T})$  is the transformed density function,  $J$  is the Jacobian and  $R_{n*}$  the new region of integration.

In the probit model (4.4), the algebraic manipulations and geometric interpretations of the required transformations are essentially those of principal component analysis. The surfaces of constant density in  $\underline{g}$ -space are this time ellipsoids, given by the quadratic form.

$$Q_F = \underline{\varepsilon}^T \underline{\Sigma}^{-1} \underline{\varepsilon} = \text{constant} \quad (4.22)$$

We wish to invoke an orthogonal transformation

$$\underline{T} = \underline{\Lambda} \underline{\varepsilon} \quad (4.23)$$

such that the vectors  $\underline{V}_1, \underline{V}_2, \dots, \underline{V}_N$ , which are the columns of  $\underline{\Sigma}$ , are the the principal axes of the ellipsoid. In the new coordinate system, the transformed matrix  $\tilde{\Sigma}$ , is written

$$\begin{aligned} \tilde{\Sigma} &= \underline{\Lambda} \underline{\Sigma} \underline{\Lambda}^T \\ &= \begin{pmatrix} \lambda_1 & 0 & - & - & - & 0 \\ 0 & \lambda_2 & - & - & - & 0 \\ 0 & 0 & - & - & - & 0 \\ \vdots & & & & & \\ 0 & - & - & - & - & 0 \lambda_N \end{pmatrix} \end{aligned} \quad (4.24)$$

in which  $\lambda_1, \dots, \lambda_N$  are the eigenvalues of  $\underline{\Sigma}$ . The eigenvalues and corresponding eigenvectors are determined from the usual equation

$$\underline{\Sigma} \underline{V}^r = \lambda_r \underline{V}^r \quad r = 1, \dots, N \quad (4.25)$$

The quadratic form (4.22) may now be written

$$QF = \sum_r \frac{T_r^2}{\lambda_r} \quad (4.26)$$

and the transformed probit model becomes

$$P_n = \int_{R_{n^*}} \prod \exp\left(-\frac{1}{2} \frac{T_r^2}{\lambda_r}\right) \cdot d\underline{T} \quad (4.27)$$

the Jacobian of the orthogonal<sup>(5)</sup> transformation being unity.

The transformed region of integration becomes

$$R_{n^*}: \bar{U}_n + (\underline{\Lambda}^T \underline{T})_n \geq \bar{U}_n + (\underline{\Lambda}^T \underline{T})_n, \quad \forall A_n \in A \quad (4.28)$$

which is quite an un hospitable region involving all components of  $\underline{T}$  on both sides of the inequality without possibilities of simplification, and therefore rendering useless the effort to decompose the multivariate density function (4.1) into the product of univariate functions (4.27).

Notice that this is not the case in the binomial context. Consider equation (4.6) and define

$$\begin{aligned} \eta_1 &= \frac{\sigma_2 - \rho\sigma_1}{\sigma_1\sqrt{1-\rho^2}} \epsilon_1 + \frac{\sigma_1 - \rho\sigma_2}{\sigma_2\sqrt{1-\rho^2}} \epsilon_2 & ) \\ & & ) \\ & & ) \\ \eta_2 &= \epsilon_2 - \epsilon_1 & ) \end{aligned} \quad (4.29)$$

... ..

(5) Variance covariance matrices are especially well-behaved. They are square, symmetric and positive semidefinite. All their eigenvalues are real and non-negative, the transformations that diagonalise them are orthogonal, and further, their inverse is equal to their transpose. (Green and Carroll, 1976).

then it can be seen that Equation (4.6) reduces to

$$P_1 = \int_{-\infty}^{\infty} \frac{\exp(-\eta_1^2/2(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2))}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)}} d\eta_1 \int_{-\infty}^{\bar{U}_{21}} \frac{\exp(-\eta_2^2/2(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2))}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)}} d\eta_2 \quad (4.30)$$

that is, by means of the transformation (4.29) the multiple integral (4.6) has been separated in its two components, and  $\epsilon_1$  has been eliminated from the upper limit of integration of the second integral. Note also that

$$\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

is precisely  $\sigma_{n_2}^2$ , the variance of the newly defined variable  $n_2 = \epsilon_2 - \epsilon_1$ .

By making another transformation, namely

$$\begin{aligned} t_1 &= \eta_1 / \sigma_{n_2} \\ t_2 &= \eta_2 / \sigma_{n_2} \end{aligned} \quad (4.31)$$

equation (4.30) further reduces to:

$$P_1 = \int_{-\infty}^{\infty} \frac{\exp(-t_1^2/2)}{\sqrt{2\pi}} dt_1 \int_{-\infty}^{\bar{U}_{21}/\sigma_{n_2}} \frac{\exp(-t_2^2/2)}{\sqrt{2\pi}} dt_2 \quad (4.32)$$

where the first integral equals 1 and the second is none other than the standardised normal cumulative distribution  $\Phi(\cdot)$ , with tabulated values.

In this case then, the binary probit model

$$\begin{aligned} P_1 &= \Phi \left( \bar{U}_{21} / \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \right) \\ P_2 &= 1 - P_1 \end{aligned} \quad (4.33)$$

is very simple and efficient to use while still being completely general, both in terms of correlation among alternatives and standard deviations of the marginal distributions. A workable version of the model, along these lines, but for three alternatives has recently been put forward by Hausman and Wise (1978). Unfortunately, this method is also non-generalizable.

Before abandoning the transformation theory, let us examine probit models corresponding to symmetric variance-covariance matrices, appropriate to the utility functions.

$$U(n,m) = U_n + U_{mm}$$

and 
$$U(n,m) = U_n + U_m + U_{nm}$$

as depicted in Figure 3. The block diagonal structure of  $\underline{\Sigma}$  in these cases imply that the eigenvalues and eigenvectors of the matrix, will not mix many utility components from the 'branch' associated with  $D_n$  and from other branches. Although this occurs, there is also considerable degeneracy in the system characteristics, some eigenvalues being not unique. Consider the model (3.14) in the simplest 2 x 2 case. The  $\underline{\Sigma}$  matrix is given by

$$\underline{\Sigma} = \begin{pmatrix} \sigma_D^2 + \sigma_{DM}^2 & \sigma_D^2 & 0 & 0 \\ \sigma_D^2 & \sigma_D^2 + \sigma_{DM}^2 & 0 & 0 \\ 0 & 0 & \sigma_D^2 + \sigma_{DM}^2 & \sigma_D^2 \\ 0 & 0 & \sigma_D^2 & \sigma_D^2 + \sigma_{DM}^2 \end{pmatrix} \quad (4.34)$$

Solving for the eigenvalues  $\lambda$  yields the equation

$$(\sigma_D^2 + \sigma_{DM}^2 - \lambda)^4 - 2\sigma_D^4(\sigma_D^2 + \sigma_{DM}^2 - \lambda)^2 + \sigma_D^8 = 0 \quad (4.35)$$

which simply reduces to

$$\pm\{(\sigma_D^2 + \sigma_{DM}^2 - \lambda)^2 - \sigma_D^4\} = 0 \quad (4.36)$$

the degeneracy already apparent. The solution of Equation (4.35) is simply:

$$\begin{aligned} \lambda_1 &= 2\sigma_D^2 + \sigma_{DM}^2 \\ \lambda_2 &= \sigma_{DM}^2 \\ \lambda_3 &= \lambda_1 \\ \lambda_4 &= \lambda_2 \end{aligned} \quad (4.37)$$

In conclusion, it has been shown that it is possible to define suitable transformations that allow one to restore the simplicity of the integrand of independent equal variance models, to any more general function, although in the case of models incorporating correlation among many alternatives, the method does not commend itself because the limits of integration of Equation (4.4) become a function of the utilities of several, if not all, the options. Neither does the method work for simpler symmetric matrix structures, the problem this time being highlighted by the high degeneracy of the eigenvalues of the matrix.

#### 5. THE THEORETICAL ACCURACY OF ALTERNATIVE LOGIT MODEL STRUCTURES

In Sections 2 and 3, we outlined a theory of choice behaviour, random utility theory. Within its framework the behaviour of individuals is governed by rational decision-making among discrete alternatives, ('homo economicus'), the structure of models 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.

If it is accepted that individuals select alternatives and respond to changes in a manner which approximates the assumptions of this theory, there are two immediate practical consequences. Firstly as the three model structures in Figure 3(a), (b) and (c), (MNL and two alternative HL models) are all special cases of the more general structure in Figure 3(d) (in which  $\sigma_D$ ,  $\sigma_M$  and  $\sigma_{DM}$  are all non-zero), any structural ambiguity, as referred to earlier in the paper, may be obviated if the later model is implemented.

Secondly, if a particular model, say the hierarchical structure in Equation (3.21) is adopted for forecasting demand response, the composite utilities (3.19) and estimated elasticity parameters  $\beta$  and  $\Delta$ , must be consistent with the theoretical conditions underpinning the model i.e. satisfy inequality (3.23). It has been found in British Transport Studies which have employed a HL of this form, that either condition (3.19) or the parameter relation  $\beta \leq \Delta$  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 evaluation measures (Williams, 1977)) 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 general probit model appropriate to the utility structure (3.5) have yet to be implemented, and it has been suggested that one should implement all three special structures (3.11), (3.12) and (3.13) and select that which yields the best statistical fit and is consistent with the theoretical conditions outlined in previous sections. (Ben Akiva, 1977; 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, Monte Carlo methods provide a very handy tool.<sup>(6)</sup>

We are now in a position to present a set of simulation tests on structural mis-specification which are designed to examine the following questions:

- (i) How good an approximation is the three parameter CCL model to the exact model generated from Equation (3.5) through utility maximisations?
- (ii) What potential errors are made by invoking the single parameter MNL and two parameter HL models, which accommodate restricted degrees of similarity between alternatives, to an appropriate three parameter specification?

Figure 6 depicts the experimental scheme. Data was generated by direct simulation from utility functions of the form (3.5) for a simple 2 x 2 case. A whole range of models was tested, which can be conveniently divided into two classes:

- theoretical, i.e. with specified parameters based on knowledge of the values of the underlying standard deviations;
- calibrated, i.e. with parameters fitted by maximum likelihood.

The first class contains the four logit models discussed before (MNL, two alternative HL structures and CCL) and the second only the first three.<sup>(7)</sup> Because the 'calibrated' versions always performed better

... ..

(6) Williams and Ortuzar (1979) have used the method outlined here to test the effects of theoretical mis-representation allowed by the relaxation of some of the assumptions associated with the decision process of 'homo-economicus'.  
(7) In fact, it is precisely the difficulty of calibrating a CCL model that has prevented its implementation.

than the 'theoretical' versions<sup>(8)</sup>, we will consider only the former from now on.

The simulated data sets consisted of the mean utilities and aggregate shares of each alternative. The MNL<sup>(9)</sup> (4 options) parameter  $\Lambda$  was estimated by maximum likelihood (using a Newton Raphson procedure described in Appendix 2) and the HL<sup>(10)</sup> models were calibrated heuristically and similarly as a series of binary logit models; the appropriate composite utilities providing a link between the two levels in the hierarchy.

Having estimated or theoretically determined the parameters of the models for a given data set (base data), a second set of data was generated for a particular change in the values of the mean utilities (say  $\bar{U}_m^{\text{new}} = \bar{U}_m^{\text{old}} + 1$ ) consistent with the effects of a particular policy.

This second set, the 'design year data' was compared with the predictions of the models for the same change in mean utility values. By this means, the response properties of the models were also assessed. The complete mechanism is depicted in Figure 6; it can be seen that it can easily be adapted to test not only structural mis-specification as we did here, but more profound problems of theoretical mis-representation (Williams and Ortuzar, 1979).

The simulation tests involved variation of the co-ordinates ( $\sigma_D, \sigma_M, \sigma_{DM}$ ). A standardisation or 'normalisation' condition to bound the joint variation of these quantities, of the following form was used:

$$\sigma_D^2 + \sigma_M^2 + \sigma_{DM}^2 = \text{constant} \quad (5.1)$$

and a particular co-ordinate ( $\sigma_D^a, \sigma_M^a, \sigma_{DM}^a$ ) corresponds to a particular simulation test. To illustrate the possible combinations of these three quantities, we appealed to that property of equilateral triangles whereby the sum of perpendicular distances to the three sides from an interior

- ... ..
- (8) The approximations involved in the models preventing the specified parameters to replicate the data as closely as the fitted parameters.
  - (9) In a first set of tests the mean utility in Equation (2.15) was taken as  $\bar{U} = \bar{U}_n + \bar{U}_m + \bar{U}_{nm}$ . In a second set of tests, we examined the effect of omitting a particular component, say by putting  $\bar{U}_m = 0$ .
  - (10) In the first set of tests, the mean utility in the lower hierarchy of Equation (3.21) was taken as  $\bar{U} = \bar{U}_m + \bar{U}_{nm}$  (and correspondingly  $\bar{U}_n + \bar{U}_{nm}$  for the alternative form). In the second set of tests the effect of omitting a particular component, say  $\bar{U}_m = 0$ , could easily be tested.

point is a constant, equal to the height of the triangle. Any test point may thus be identified with a point in or on the boundary of the triangle, as shown in Figure 7(a). At interior points a three parameter model (such as the CCL model) is necessary to capture the full range of cross substitution implied by the utility function (3.5). On the boundaries CB and CA the alternate HL models for which  $\sigma_M = 0$  and  $\sigma_D = 0$  respectively, are appropriate (see Figures 3(b) and (c)). It is only at the vertex C (i.e.  $\sigma_D = \sigma_M = 0$ ) that the MNL is an appropriate specification.

In addition to test points randomly sampled from within the triangle, four particular co-ordinate test points, as shown in Figure 7(b), were selected for the presentation of results. Recall that in all tests two alternatives were taken in each of the D and M dimensions, allowing a four alternative choice model to be generated.

The general performance of the four models for these test points<sup>(11)</sup> is shown in Figure 7(c). We have restricted ourselves to a comparative quality assessment of the fit, in that it is the relative performance of the models in which interest lies.<sup>(12)</sup> A visual display of the meaning of this informal assessment, in the form of a particular set of base and response results for the four models used to fit data generated from test point 2 in Figure 7(b), is shown in Figure 7(d).

Under conditions of change, points in the second and fourth quadrants of the right hand side of Figure 7(d) are deemed pathological because the change in behaviour predicted by the model is opposite to that simulated. We found that this behaviour is associated with the violation of the condition (3.23) in hierarchical logit specifications. If, for example,  $\sigma_D \gg \sigma_M$ , then the HL specification M/D corresponding to  $\sigma_M > \sigma_D$  will involve pathological behaviour. The condition, however,

... ..

(11) Results are shown for the first set of test only, i.e. when it was assumed complete knowledge of the mean utilities.

(12) The performance measures used were average relative errors (ARE) defined as:

$$ARE = \frac{1}{N} \sum_{i=1}^N \{ |P_i^{sim} - P_i^{mod}| / P_i^{mod} \}$$

and a  $\chi^2$  measure, defined as:

$$\chi^2 = \sum_{i=1}^N \{ (P_i^{sim} - P_i^{mod})^2 / P_i^{mod} \}$$

where N = number of options (e.g.4)

$P_i^{sim}$  = simulated share of option i, i = 1, ..., N

$P_i^{mod}$  = modelled share of option i, i = 1, ..., N

appears also to depend on the underlying representative utility values.<sup>(13)</sup>

If we further examine the performance of the models as reported in Figure 7(c), we note that the CCL appears to be a good approximation to the general function (3.5). Its superiority to the other logit forms was especially apparent when the three coordinates  $\sigma_D$ ,  $\sigma_M$  and  $\sigma_{DM}$  were different from zero and from each other. The most surprising and welcome outcome, however, was the remarkably robust performance of the MNL, even at interior points of the triangle and especially when  $\sigma_D = \sigma_M$ . As expected the IIA property was a considerable impediment near the sides of the triangle, except in the immediate region of point C.

The last point to note in this section is, that out of the estimated models (MNL and two HL forms), that with the best base fit consistently provided a good estimate of the response to change; this would seem to lend some theoretical/numerical support for the suggestion of Ben Akiva (1977) that results of alternative HL and MNL models could be compared and the appropriate model selected according to the estimated value of the 'similarity' parameter, which in our notation is  $\beta/\Delta$ .

We offer no apologies for the fact that the simulation tests were confined to a 2 x 2 example (4 options). We believe that the results and conclusions in next section, would not be qualitatively modified when the number of alternatives are increased, because the structure of the variance-covariance matrix itself was the focus of the mis-specification tests. The dependence of the results on the number of options could, of course, be tested.<sup>(14)</sup>

... ..

(13) In the second series of tests the effect of omitting a particular utility component - by putting  $\bar{U}_m = 0$ , for example - was determined. All results were inferior to their counterparts in the first series, as would be expected because the number of 'degrees of freedom' of the model specifications had been reduced. The performance of HL (asymmetric in nature) was particularly suspect and pathological behaviour became more prevalent.

(14) More important, in exercises of this kind, is to make sure that the process converges. We found it was necessary to sample 30,000 observations to get consistent results. The reader is referred to Section 2.

## 6. CONCLUSIONS

In this paper we have presented the essential features of random utility theory, both formally and in a geometric framework which provides a clear interpretation of the nature of choice.

We have studied the role of correlation and how the form of the utility function uniquely determines the structure of models, their complexity being dictated precisely by the structure of similarity or correlation between the alternatives. In this way, we have been able to present the basic assumptions of the most popular models, multinomial logit, general probit and hierarchical logit, and discuss their implications.

The general probit model is the more conceptually appealing of model forms, although unfortunately the least tractable of them for more than small problems, and even then with some yet unknown properties. We have investigated the possibility of invoking transformations in utility space as a means of simplifying it. We have shown that it is indeed possible to define suitable transformations that allow one to restore the simplicity of the integrand of independent 'equal variance' models to any more general model, although in the case of models incorporating correlation among many alternatives, what is gained on the roundabouts is lost on the swings because of the non-separability of the multiple integral (4.27) which in turn is due to the unhospitable form of the region of integration (4.28). Although transformations certainly give more insight into the problem, the potential for implementing them, even in the case of models with symmetric variance-covariance matrices, does not appear practicable.

The results of the simulation tests are consistent with the following conclusions:

- i) The cross-correlated logit model is a good theoretical approximation to the three parameter utility function (3.5). The superiority to other logit forms is especially apparent when the standard deviations  $\sigma_D$ ,  $\sigma_M$  and  $\sigma_{DM}$  are rather different from zero and from each other. However its estimation is very complex (Williams, 1977) and for this reason it does not commend itself.

- ii) The multinomial logit model performs reasonably well at interior points of the triangle (indeed it is considerably more robust than we had anticipated) and this is particularly true when  $\sigma_D \approx \sigma_M$ . Its maximum error occurs near the sides of the triangle (except in the immediate region of point C) where the 'independence from irrelevant alternatives' property is a considerable impediment.
- iii) A good base fit does not necessarily imply a good response model, after all every model 'calibrates'. However, out of the three alternative logit structures (multinomial logit and two hierarchical logit forms), the model which provides the best base fit provides a good estimate of the response to change.
- iv) A mis-specified model (typical case is the inappropriate hierarchical logit form) will tend to display pathological behaviour in a response context. However it is possible to recognise the symptoms at the calibration stage, by examining the consistency conditions (eg. rule (3.23)).

In these tests we examined the capability of extended members of the logit family to accommodate the structure of similarity between alternatives embodied in the utility function (3.5). The general probit function would be appropriate in this case, and indeed to more general utility forms. Horowitz (1978) has, in fact, examined the potential mis-specification problems of the multinomial logit model when compared with a 3- alternative probit model. His results are complementary and consistent with ours, in the sense that in most practical cases, the greater ease with which less consistent structures -- provided they are robust enough to cope relatively well with not serious mis-specification -- may be implemented, will win the day.

Finally the powerful tool employed in our analysis, Monte Carlo simulation, deserves an especial word of praise. We noted the increasing application of the method in the transport field, but we believe that our particular use here, creating artificial data sets on which the effects of specific model mis-specifications can be tested in a controlled manner, indicates one way ahead to attack the problem of model evaluation. Indeed, it has already been used in a more ambitious project to test for theoretical misrepresentation and to assess the validity of cross-sectional models in general (Williams and Ortuzar, 1979).

REFERENCES

- ALBRIGHT, R.L., LERMAN, S.R. and MANSKI, C.F. (1977) Report on the development of an estimation program for the multinomial probit model, prepared for the Federal Highway Administration, US Department of Transportation, by Cambridge Systematics Inc.
- BEN-AKIVA, M.E. (1974) Structure of passenger travel demand models. Transportation Research Record, 526, 26-42.
- BEN-AKIVA, M.E. (1977) Passenger travel demand forecasting: applications of disaggregate models and directions for research. Proceedings, World Conference on Transportation Research, Rotterdam.
- BONSALL, P.W. (1979) Microsimulation of mode choice - a model of organised car sharing. PTRC Summer Annual Meeting, University of Warwick.
- BOUTHELIER, F.J. (1978) An efficient methodology to estimate and predict with multinomial probit models: applications to transportation problems. PhD Thesis, Department of Ocean Engineering, MIT.
- BURRELL (1968) Multipath route assignment and its application to capacity restraint. Proceedings, Fourth International Symposium on the Theory of Traffic Flow, Karlsruhe, Germany.
- CARDELL, N.S. and REDDY, B.J. (1977) A multinomial logit model which permits variations in tastes across individuals. Charles Rivers Associates Inc.
- CHARLES RIVERS ASSOCIATES Inc. (1972) A disaggregate behavioural model of urban travel demand. Federal Highway Administration, U.S. Department of Transportation.
- CHICAGO AREA TRANSPORTATION STUDY (1979) Travel forecasting process. A staff technical report, CATS 2344-01, Chicago.
- CLARK, C. (1961) The greatest of a finite set of random variables. Operations Research, 9, 145-62
- COCHRANE, R.A. (1975) A possible economic basis for the gravity model. Journal of Transport Economics and Policy 9, 34-49
- DAGANZO, C.F. (1979) Calibration and prediction with random utility models: some recent advances and unresolved questions, presented at the Fourth International Conference on Behavioural Travel Modelling 1979, Eibsee, West Germany.
- DAGANZO, C.F., BOUTHELIER, F. and SHEFFI, Y. (1977) Multinomial probit and qualitative choice: a computationally efficient algorithm. Transportation Science 11, 338-58.
- DAGANZO, C.F. and SCHOENFELD, S. (1978) CHOMP User's Manual. Report UCB-ITS-RR-78-7. Institute of Transportation Studies, University of California at Berkeley.

- DALY, A.J. and ZACHARY, S. (1978) Improved multiple choice models. In Hensher, D.A. and Dalvi M.Q. (eds.), Determinants of Travel Choice, Saxon House, Sussex.
- DOMENCICH, T. and MCFADDEN, D. (1975) Urban Travel Demand - a behavioural analysis, North Holland, Amsterdam.
- GREEN, P.E. with contributions by CARROLL, J.D. (1976) Mathematical Tools for Applied Multivariate Analysis, Academic Press, Inc., London.
- HAMMERSLEY, J.M. and HANDSCOMB, D.C. (1965) Monte Carlo Methods, Chapman and Hall, London.
- HAUSMAN, J.A. and WISE, D.A. (1978) A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. Econometrica, 46, 403-26.
- HOROWITZ, J. (1978) The accuracy of the multinomial logit as an approximation to the multinomial probit model of travel demand, unpublished manuscript, U.S. Environmental Protection Agency, Washington D.C.
- KREIBICH, V. (1979) Modelling car availability, modal split and trip distribution by Monte Carlo simulation: a short way to integrated models. Transportation, 8, 153-66.
- LANGDON, M. (1976) Modal split models for more than two modes. In Urban Traffic Models. PTRC Summer Annual Meeting, University of Warwick.
- MANSKI, C.F. and LERMAN, S.R. (1978) On the use of simulated frequencies to approximate choice probabilities, Cambridge Systematics Inc.
- 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.
- ORTUZAR, J.D. (1978) Mixed-mode demand forecasting: an assessment of current practice. PTRC Summer Annual Meeting, University of Warwick.
- ROBERTSON, D.I. (1977) A deterministic method of assigning traffic to multiple routes of known cost. TRRL Report LR 757, Crothorne.
- ROBERTSON, D.I. and KENNEDY, J.V. (1979) The choice of route, origin and destination by calculation and simulation. TRRL Report LR 877, Crothorne.
- 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.

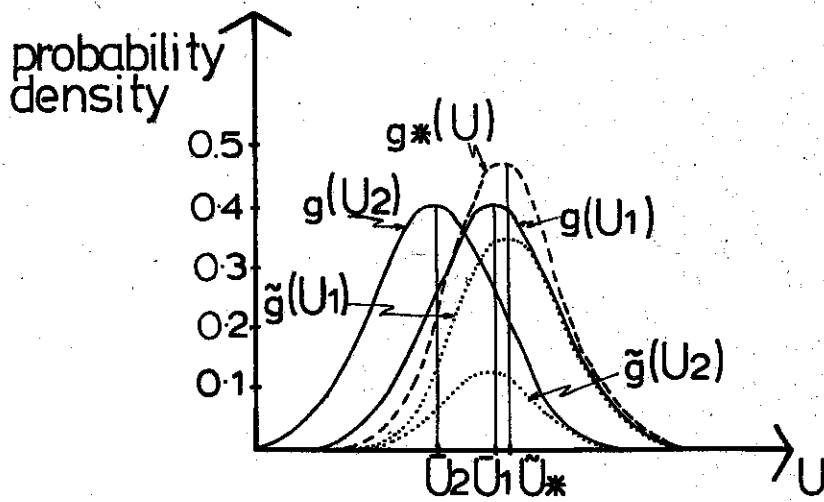
- SPEAR, B.D. (1977) Applications of new travel demand forecasting techniques to transportation planning: a study of individual choice models. Federal Highway Administration, U.S. Department of Transportation, Washington D.C.
- WILLIAMS, H.C.W.L. (1977) On the formation of travel demand models and economic evaluation measures of user benefit. Environment and Planning A, 9, 285-344.
- WILLIAMS, H.C.W.L. and ORTUZAR, J.D. (1979) Behavioural travel theories model specification and the response error problem. PTRC Summer Annual Meeting, University of Warwick, (Available as Working Paper 116, ITS, University of Leeds).
- 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), 464-9.

### ACKNOWLEDGEMENTS

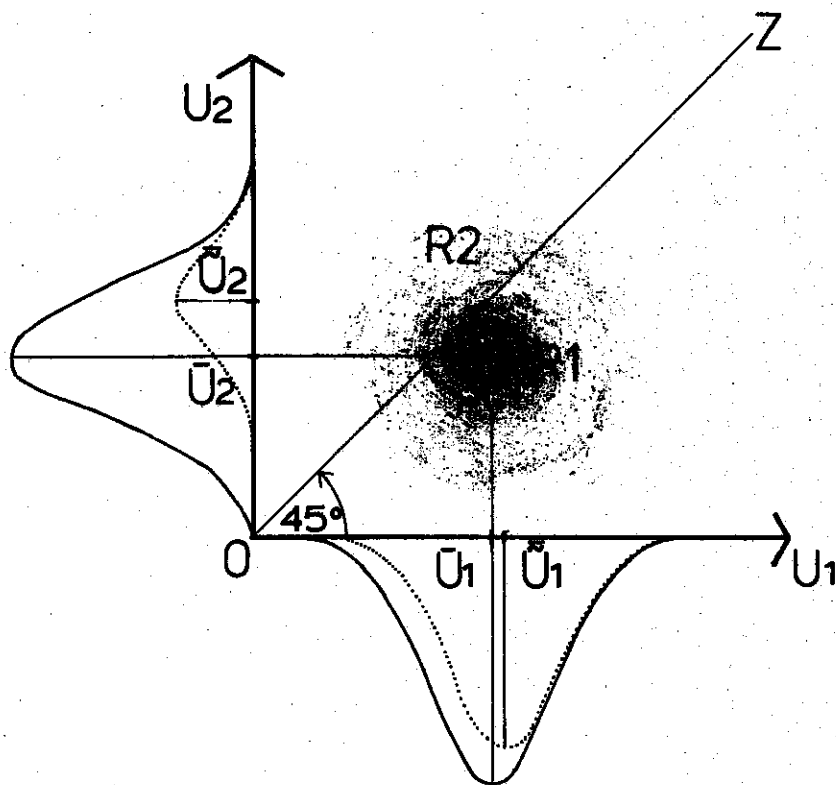
This work owes considerably to Dr. Huw Williams. His imprint on the paper's organization, ideas and exposition goes well beyond that of draft commentator. He is, however, to be held blameless for any parts or words that did not come out quite right.

I would like to express my appreciation also to Margarita Greene for the skill and effort necessary to produce the art work in the paper. The ITS team of secretaries were as usual instrumental in producing the paper and also deserve my thanks.

FIGURES



a.) Fundamental distributions.



b.) Geometric interpretation.

FIGURE 1: Utility distributions for independent model with equal standard deviations.

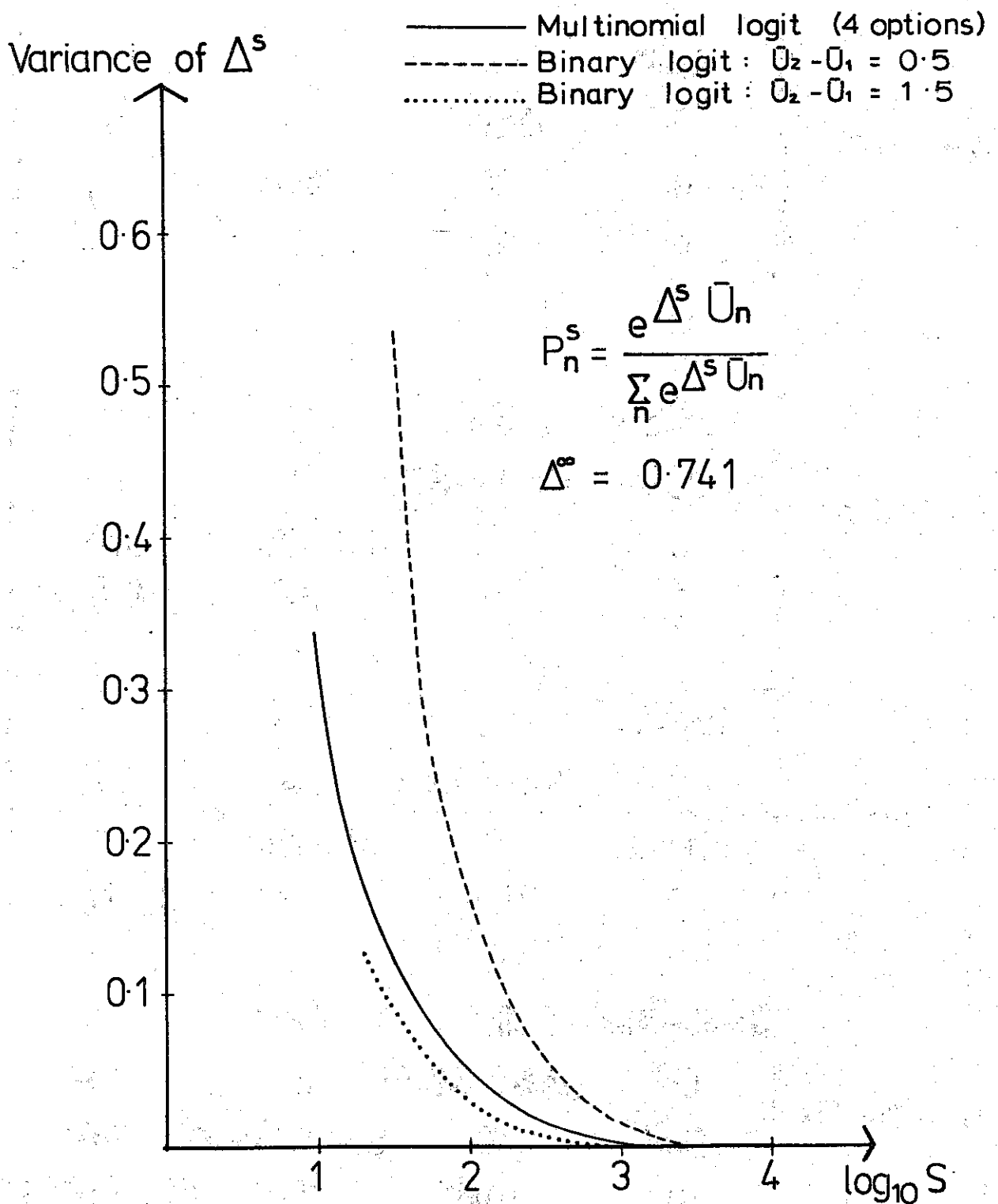


FIGURE 2: Sampling errors in parameter estimation.

**STRUCTURE**

**VARIANCE - COVARIANCE MATRIX ( $\Sigma$ )**

$\epsilon(D,M) = \epsilon_{DM}$

$$\begin{bmatrix} \sigma_{DM}^2 & 0 & 0 & 0 \\ 0 & \sigma_{DM}^2 & 0 & 0 \\ 0 & 0 & \sigma_{DM}^2 & 0 \\ 0 & 0 & 0 & \sigma_{DM}^2 \end{bmatrix}$$

$\epsilon(D,M) = \epsilon_D + \epsilon_{DM}$

$$\begin{bmatrix} (\sigma_D^2 + \sigma_{DM}^2) & \sigma_D^2 & 0 & 0 \\ \sigma_D^2 & (\sigma_D^2 + \sigma_{DM}^2) & 0 & 0 \\ 0 & 0 & (\sigma_D^2 + \sigma_{DM}^2) & \sigma_D^2 \\ 0 & 0 & \sigma_D^2 & (\sigma_D^2 + \sigma_{DM}^2) \end{bmatrix}$$

$\epsilon(D,M) = \epsilon_{DM} + \epsilon_M$

$$\begin{bmatrix} (\sigma_{DM}^2 + \sigma_M^2) & 0 & \sigma_M^2 & 0 \\ 0 & (\sigma_{DM}^2 + \sigma_M^2) & 0 & \sigma_M^2 \\ \sigma_M^2 & 0 & (\sigma_{DM}^2 + \sigma_M^2) & 0 \\ 0 & \sigma_M^2 & 0 & (\sigma_{DM}^2 + \sigma_M^2) \end{bmatrix}$$

$\epsilon(D,M) = \epsilon_D + \epsilon_{DM} + \epsilon_M$

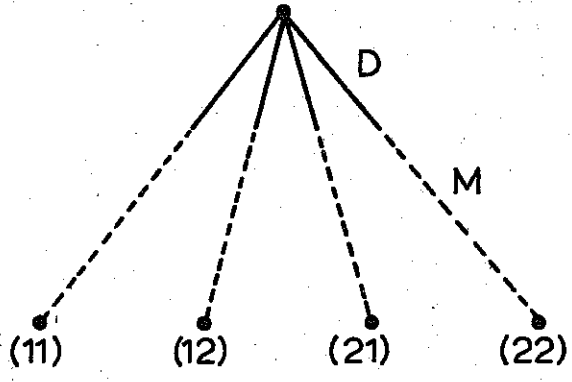
$$\begin{bmatrix} (\sigma_D^2 + \sigma_{DM}^2 + \sigma_M^2) & \sigma_D^2 & \sigma_M^2 & 0 \\ \sigma_D^2 & (\sigma_D^2 + \sigma_{DM}^2 + \sigma_M^2) & 0 & \sigma_M^2 \\ \sigma_M^2 & 0 & (\sigma_D^2 + \sigma_{DM}^2 + \sigma_M^2) & \sigma_D^2 \\ 0 & \sigma_M^2 & \sigma_D^2 & (\sigma_D^2 + \sigma_{DM}^2 + \sigma_M^2) \end{bmatrix}$$

FIGURE 3 : The structure of choice models :

# ALTERNATIVE NOTATION

# REPRESENTATION

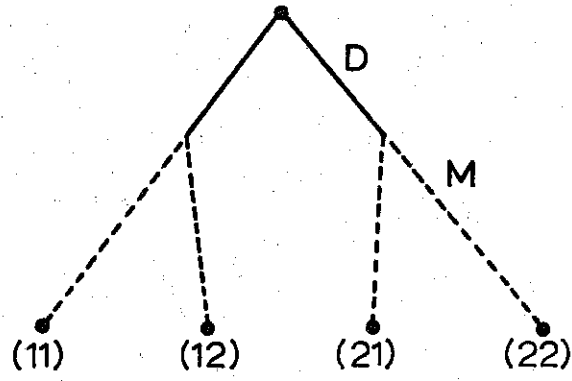
$$\underline{\underline{\Sigma}} = (\underline{\underline{D}} \underline{\underline{M}}^2 \underline{\underline{I}})$$



$$\underline{\underline{\Sigma}} = \begin{bmatrix} \underline{\underline{\Sigma}}_1 & 0 \\ 0 & \underline{\underline{\Sigma}}_1 \end{bmatrix}$$

$$\underline{\underline{\Sigma}}_1 = (\underline{\underline{D}}^2 (\underline{\underline{D}} \underline{\underline{M}}^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}))$$

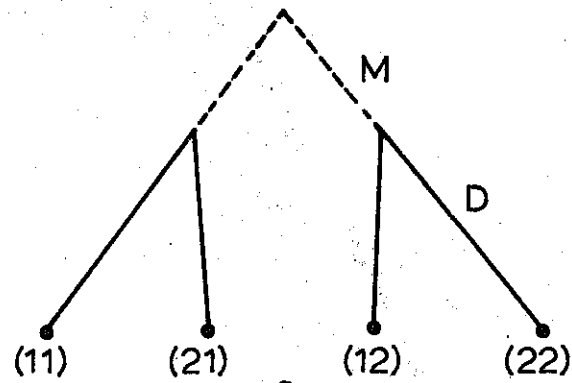
$$\rho = (\underline{\underline{D}}^2 / ((\underline{\underline{D}}^2 (\underline{\underline{D}} \underline{\underline{M}}^2)))$$



$$\underline{\underline{\Sigma}} = \begin{bmatrix} \underline{\underline{\Sigma}}_1 & \underline{\underline{\Sigma}}_2 \\ \underline{\underline{\Sigma}}_2 & \underline{\underline{\Sigma}}_1 \end{bmatrix}$$

$$\underline{\underline{\Sigma}}_1 = ((\underline{\underline{D}} \underline{\underline{M}}^2 (\underline{\underline{M}}^2) \underline{\underline{I}})$$

$$\underline{\underline{\Sigma}}_2 = (\underline{\underline{M}}^2 \underline{\underline{I}})$$

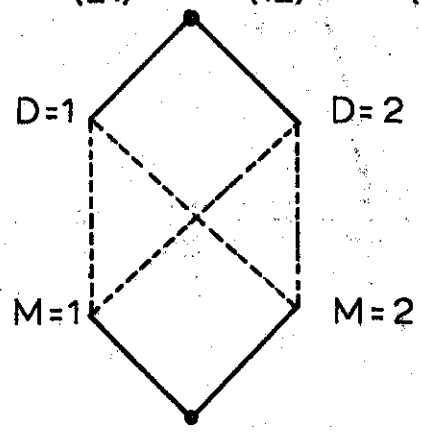


$$\underline{\underline{\Sigma}} = \begin{bmatrix} \underline{\underline{\Sigma}}_1 & \underline{\underline{\Sigma}}_2 \\ \underline{\underline{\Sigma}}_2 & \underline{\underline{\Sigma}}_1 \end{bmatrix}$$

$$\underline{\underline{\Sigma}}_1 = ((\underline{\underline{D}}^2 (\underline{\underline{D}} \underline{\underline{M}}^2 (\underline{\underline{M}}^2) \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}))$$

$$\rho = (\underline{\underline{D}}^2 / ((\underline{\underline{D}}^2 (\underline{\underline{D}} \underline{\underline{M}}^2 (\underline{\underline{M}}^2)))$$

$$\underline{\underline{\Sigma}}_2 = (\underline{\underline{M}}^2 \underline{\underline{I}})$$



a set of special - cases.

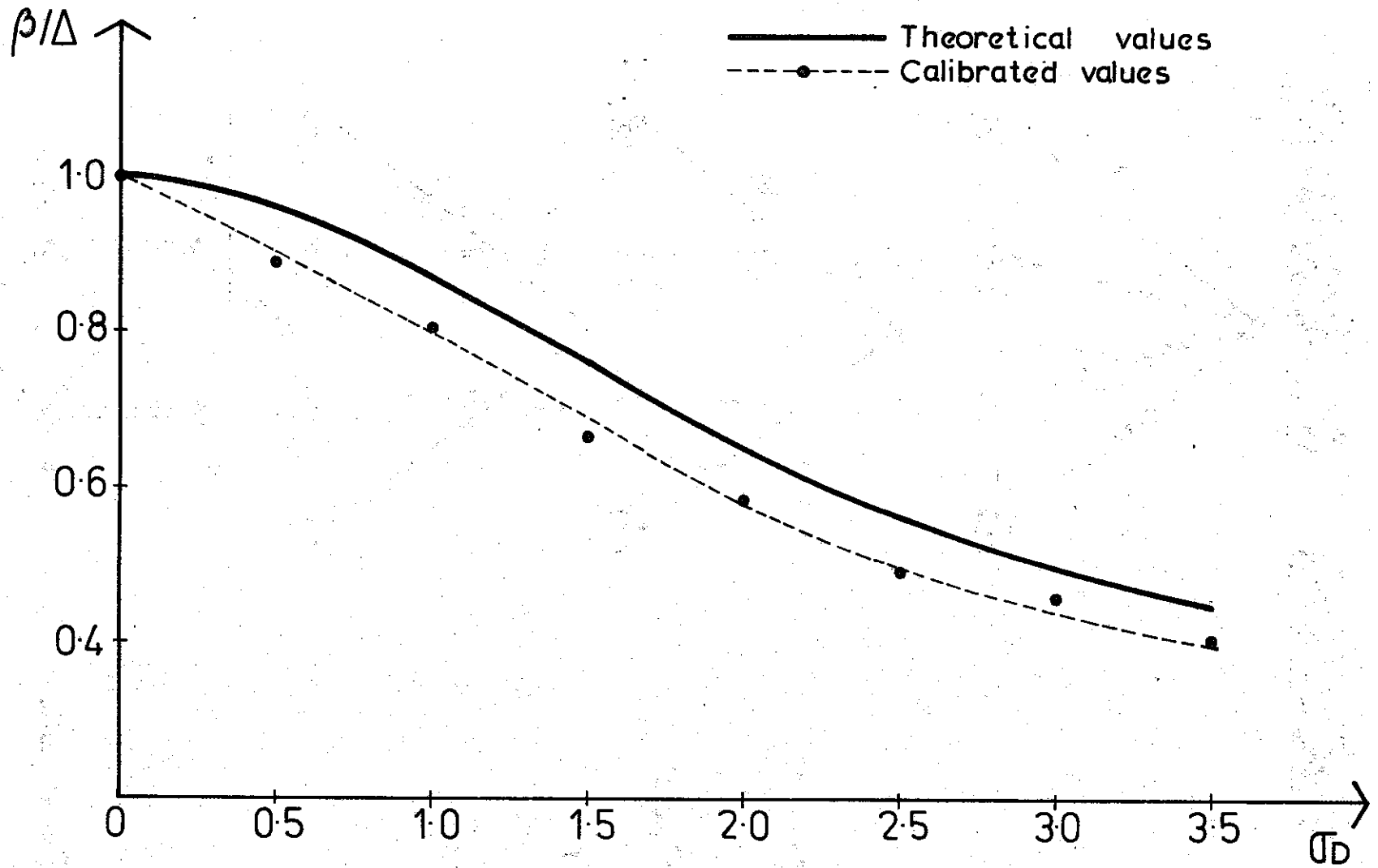
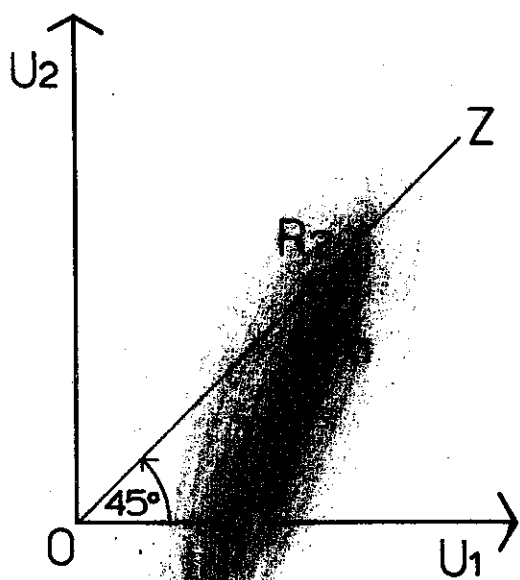
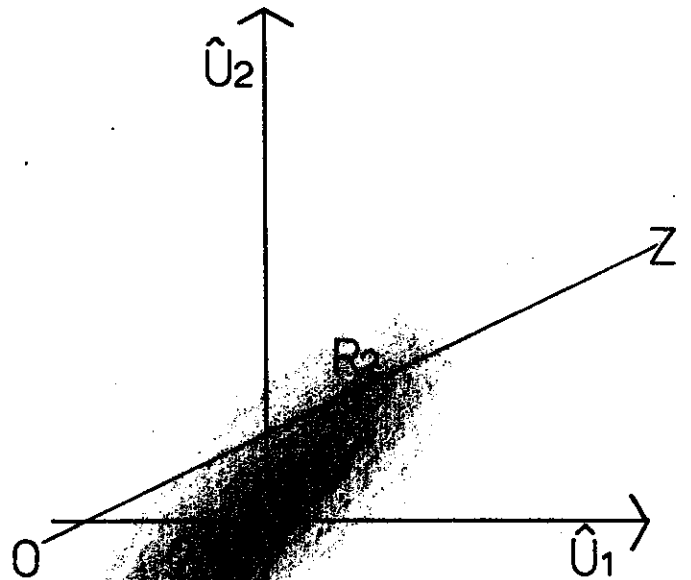


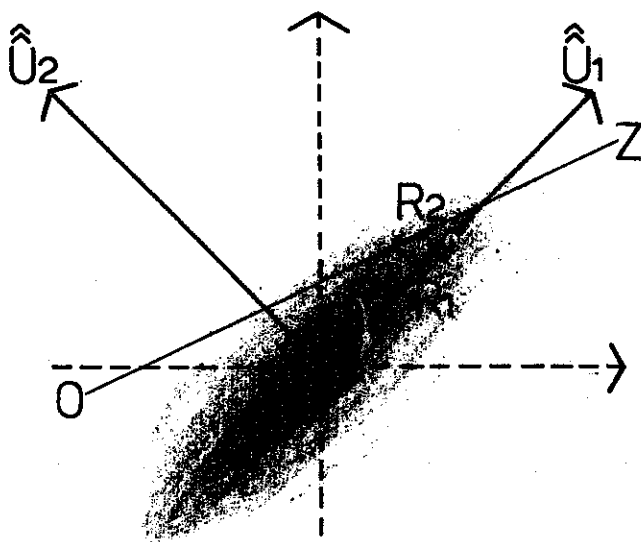
FIGURE 4 : Relationship between dispersion parameters in hierarchical logit structures.



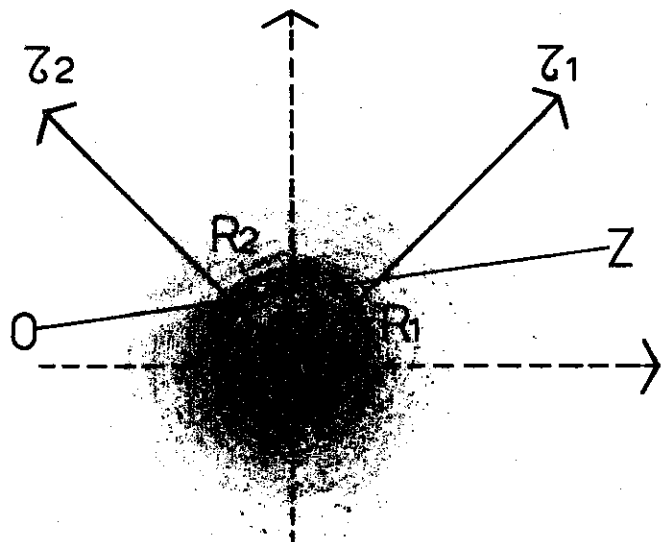
a)  $(U_1, U_2)$  plane



b)  $(\hat{U}_1, \hat{U}_2)$  plane



c.)  $(\hat{U}_1, \hat{U}_2)$  plane



d.)  $(Z_1, Z_2)$  plane

FIGURE 5: Transformations of the general probit model in two dimensions.

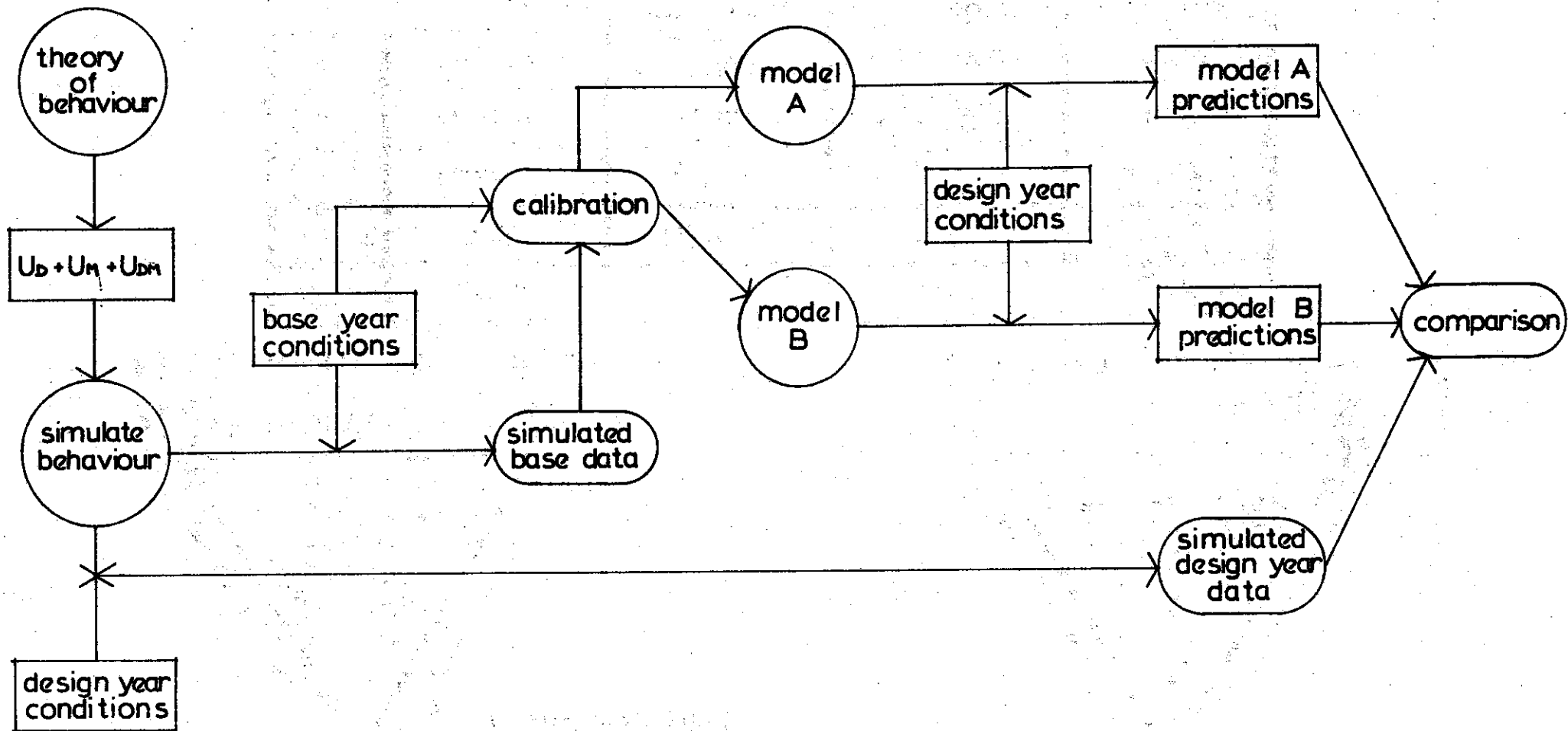
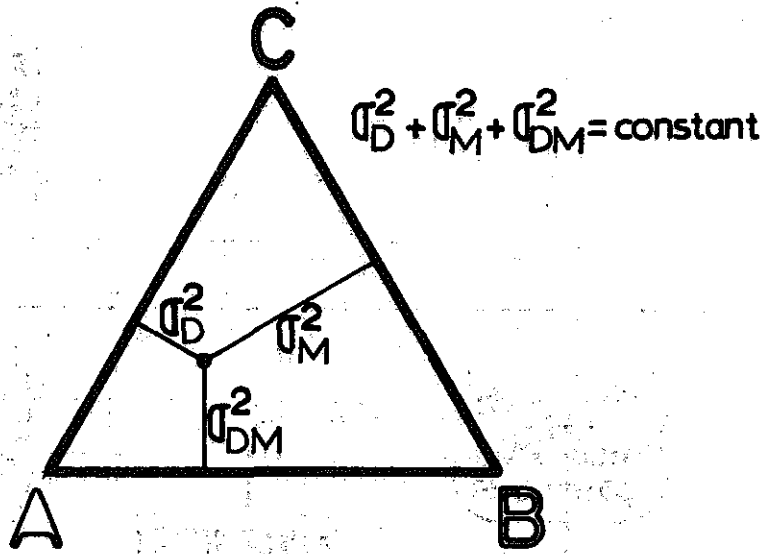
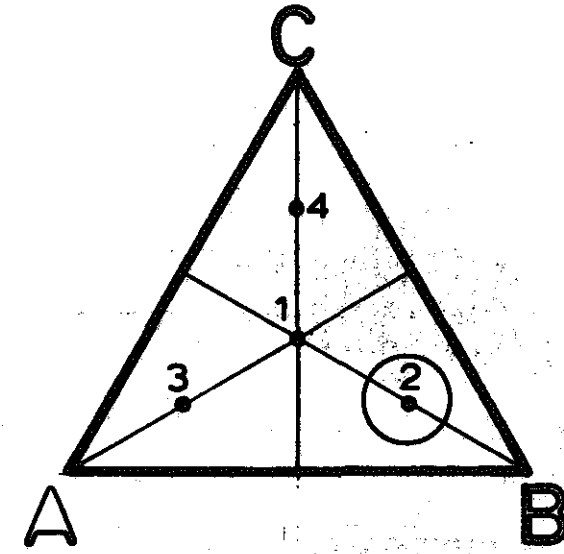


FIGURE 6: Model comparison using simulation.



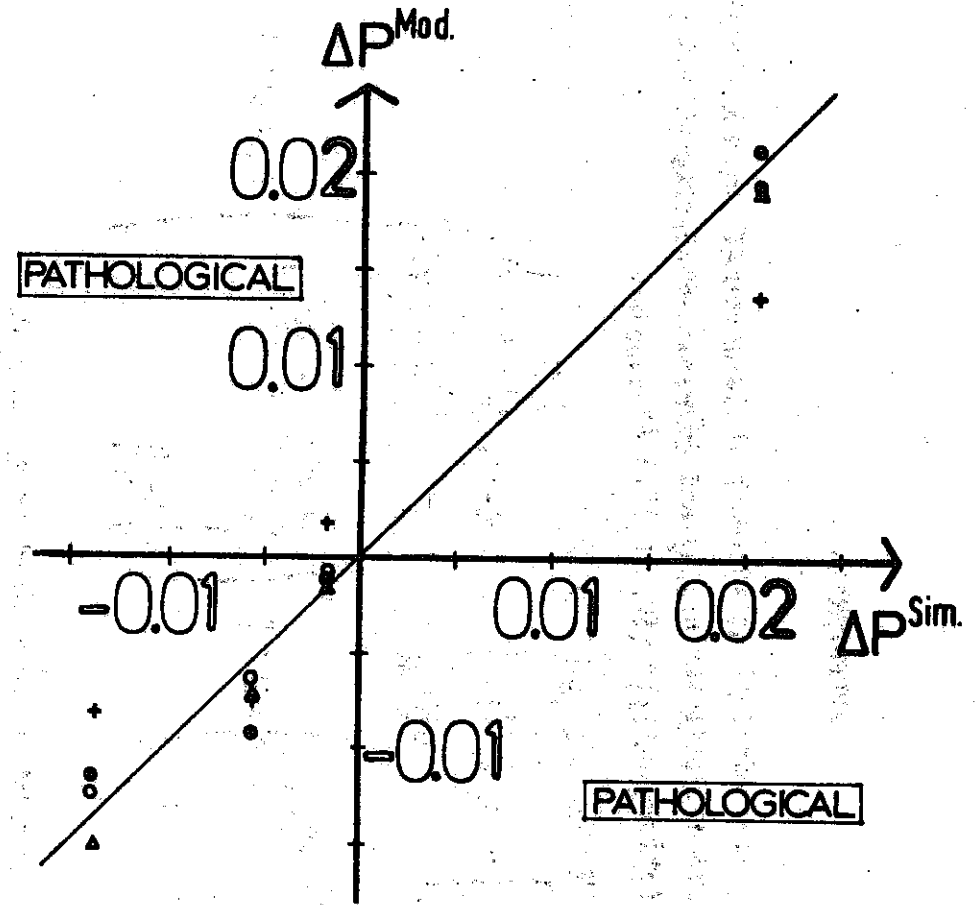
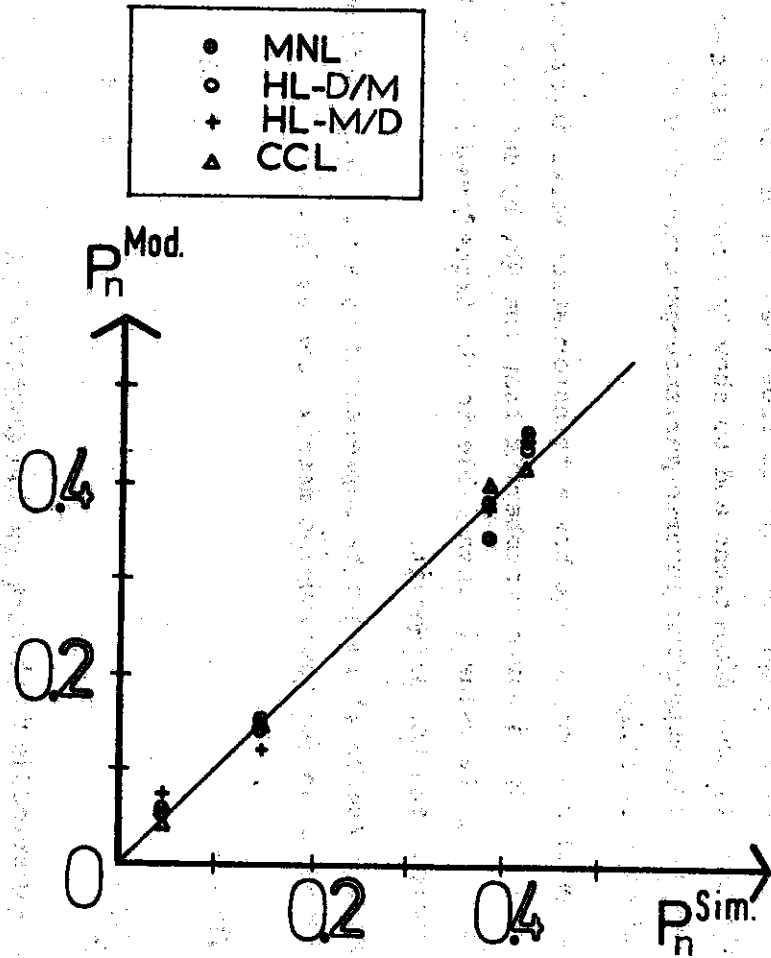
a.) The normalization condition.



b.) The test point coordinates  $(\sigma_D, \sigma_M, \sigma_{DM})$

POINT	MNL	HL-D/M	HL-M/D	CCL
1	good	good $\Delta \sim \beta$	good $\Delta \sim \beta$	very good
2	regular	very good	pathological $\Delta < \beta$	good
3	regular	pathological $\Delta < \beta$	very good	good
4	very good	very good $\Delta = \beta$	very good $\Delta = \beta$	very good

c.) Model performance at different test points.



d.) An illustration of base performance and response errors at test point 2.

FIGURE 7 : The design and results of simulation tests to investigate model structure variation.

## APPENDIX 1

If  $U_1$  and  $U_2$  are jointly distributed bivariate normal, with means  $\bar{U}_1$  and  $\bar{U}_2$ , standard deviation  $\sigma_1$  and  $\sigma_2$  and correlation coefficient  $\rho$ , then the quadratic form (QF) of  $f(U_1, U_2)$  is given by

$$QF = \left\{ -\frac{1}{2} (\underline{U} - \underline{\bar{U}})^T \underline{\Sigma}^{-1} (\underline{U} - \underline{\bar{U}}) \right\} \quad (A1)$$

$$\text{where } \underline{U} = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}; \quad \underline{\bar{U}} = \begin{pmatrix} \bar{U}_1 \\ \bar{U}_2 \end{pmatrix}; \quad (\underline{U} - \underline{\bar{U}}) = \begin{pmatrix} U_1 - \bar{U}_1 \\ U_2 - \bar{U}_2 \end{pmatrix}$$

$$\underline{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad \text{and} \quad \underline{\Sigma}^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix} (1 - \rho^2)^{-1}$$

the region of integration is defined by  $R_1: U_2 < U_1$

In this appendix a matrix treatment of the general transformation indicated in the text will be given, both to show that the process is easily generalizable to more dimensions and to show how each transformation in turn, affects the corresponding inverse variance-covariance matrix defining the QF at each stage.

First a general statement about how a transformation works in matrix terms will be given. It is worth remembering that the QF, by definition, is a scalar, that is, its value is invariable to the transformations, only changing the components that define it.

In general a transformation can be represented by a matrix  $\underline{\Lambda}$  acting upon a vector such as in (A2) and over a matrix such as in (A3).

$$\underline{\tau} = \underline{\Lambda} (\underline{U} - \underline{\bar{U}}) \quad (A2)$$

$$\underline{\hat{\Sigma}} = \underline{\Lambda} \underline{\Sigma} \underline{\Lambda}^T \quad (A3)$$

The QF in the new space defined by  $\underline{\tau}$  has the general form:

$$QF' = \left\{ -\frac{1}{2} \underline{\tau}^T \underline{\hat{\Sigma}}^{-1} \underline{\tau} \right\} \quad (A4)$$

We will show now that  $QF'$  is indeed equal to  $QF$

$$\underline{r}^T = \{ \underline{\underline{\Lambda}} (\underline{U} - \underline{\bar{U}}) \}^T = (\underline{U} - \underline{\bar{U}})^T \underline{\underline{\Lambda}}^T \quad (\text{A5})$$

$$\underline{\hat{\Sigma}}^{-1} = \{ \underline{\underline{\Lambda}} \underline{\underline{\Sigma}} \underline{\underline{\Lambda}}^T \}^{-1} = (\underline{\underline{\Lambda}}^T)^{-1} \underline{\underline{\Sigma}}^{-1} \underline{\underline{\Lambda}}^{-1} \quad (\text{A6})$$

Replacing (A2), (A5) and (A6) into (A4) we get

$$QF' = \{ -\frac{1}{2}(\underline{U} - \underline{\bar{U}})^T \underline{\underline{\Lambda}}^T (\underline{\underline{\Lambda}}^T)^{-1} \underline{\underline{\Sigma}}^{-1} \underline{\underline{\Lambda}}^{-1} \underline{\underline{\Lambda}} (\underline{U} - \underline{\bar{U}}) \} \quad (\text{A7})$$

and noting that  $\underline{\underline{\Lambda}}^T (\underline{\underline{\Lambda}}^T)^{-1} = \underline{\underline{I}} = \underline{\underline{\Lambda}}^{-1} \underline{\underline{\Lambda}}$ , the unit matrix, and that by definition any matrix or vector  $\underline{M}$  when pre- or post-multiplied by  $\underline{\underline{I}}$  remains unaffected, we show that

$$QF' = QF \quad (\text{A8})$$

In the remainder of the appendix we will present each transformation of the general case in the text and the form of the inverse of the variance-covariance matrix involved.

a) The first transformation: (4.10)

$$\underline{\underline{\Lambda}} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{pmatrix}, \text{ in this case } \underline{\underline{\Lambda}}^T = \underline{\underline{\Lambda}} \quad (\text{A9})$$

$$\text{and } \underline{\underline{\Lambda}}^{-1} = (\underline{\underline{\Lambda}}^T)^{-1} = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \quad (\text{A10})$$

$$\text{so } \underline{\hat{\Sigma}}^{-1} = \begin{pmatrix} \frac{1}{1-\rho^2} & \frac{-\rho}{1-\rho^2} \\ \frac{-\rho}{1-\rho^2} & \frac{1}{1-\rho^2} \end{pmatrix} \quad (\text{A11})$$

b) The second transformation: (4.13)

$$\hat{\underline{\Lambda}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$
$$\hat{\underline{\Lambda}}^T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad (\text{A12})$$

therefore  $\hat{\underline{\Lambda}}^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$  (A13)

and  $(\hat{\underline{\Lambda}}^T)^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$

so  $\hat{\underline{\Sigma}}^{-1} = \begin{pmatrix} \frac{1}{1-\rho} & 0 \\ 0 & \frac{1}{1+\rho} \end{pmatrix}$  (A14)

c) The last transformation: (4.15)

$$\hat{\underline{\Lambda}} = \begin{pmatrix} (1-\rho)^{-\frac{1}{2}} & 0 \\ 0 & (1+\rho)^{-\frac{1}{2}} \end{pmatrix} = \hat{\underline{\Lambda}}^T \quad (\text{A16})$$

therefore  $\hat{\underline{\Lambda}}^{-1} = (\hat{\underline{\Lambda}}^T)^{-1} = \begin{pmatrix} \sqrt{1-\rho} & 0 \\ 0 & \sqrt{1+\rho} \end{pmatrix}$  (A17)

so  $\hat{\underline{\Sigma}}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  (A18)

APPENDIX 2

The multinomial logit model has an analytic closed form given by:

$$P_j^{\text{mod}} = \frac{e^{\Delta \bar{U}_j}}{\sum_j e^{\Delta \bar{U}_j}} \quad (\text{A19})$$

where

$$\begin{aligned} P_j^{\text{mod}} &= \text{modelled share of alternative } j, j = 1, \dots, N. \\ \bar{U}_j &= \text{mean or representative utility of} \\ &\quad \text{alternative } j, j = 1, \dots, N. \\ \Delta &= \text{parameter to be estimated} \end{aligned}$$

In this Appendix we will consider the maximum likelihood estimation of  $\Delta$  given the values of  $\bar{U}_j$  and the simulated shares  $P_j^{\text{sim}}$ , of each alternative  $j$ . We will use a standard Newton Raphson search mechanism.

In this simple case, the log-likelihood function of  $\Delta^{(k)}$ , at iteration  $k$ , is given by

$$L(\Delta^{(k)}) = \sum_i \bar{U}_i (P_i^{\text{sim}} - \frac{e^{\Delta^{(k)} \bar{U}_i}}{\sum_j e^{\Delta^{(k)} \bar{U}_j}}) \quad (\text{A20})$$

and the first derivative with respect to  $\Delta^{(k)}$  is

$$L'(\Delta^{(k)}) = \frac{(\sum_i \bar{U}_i e^{\Delta^{(k)} \bar{U}_i})^2}{(\sum_i e^{\Delta^{(k)} \bar{U}_i})^2} - \frac{\sum_i \bar{U}_i^2 e^{\Delta^{(k)} \bar{U}_i}}{\sum_i e^{\Delta^{(k)} \bar{U}_i}} \quad (\text{A21})$$

Now the Newton Raphson solution involves defining a next best estimate for the parameter,  $\Delta^{(k+1)}$ , as

$$\Delta^{(k+1)} = \Delta^{(k)} - L(\Delta^{(k)})/L'(\Delta^{(k)}) \quad (\text{A22})$$

and the maximum likelihood estimate will be that found when the process converges (which always does, Domencich and McFadden, 1975), that is when

$$\Delta^{(k+1)} = \Delta^{(k)} \quad (\text{A23})$$