Published paper

Andrew Daly (2001) *Independence, Homoskedasticity and Existence in Random Utility Models.* Institute of Transport Studies, University of Leeds, Working Paper 557

ITS Working Paper 557

Independence, Homoskedasticity and Existence in Random Utility Models

Andrew Daly
RAND Europe and ITS Leeds
January 2001, revised April 2001

Contents

1. Introduction

Random utility models are often characterised by descriptions such as 'homoskedastic' or 'independent' in the utilities of the alternatives. However these descriptions do not have meaning in any absolute sense and must therefore be used with care. It is the main aim of this paper to demonstrate this point and discuss the issues it raises. In particular, the discussion leads into a consideration of the circumstances under which the models can be said to exist.

The paper gives a definition of random utility models and goes on the define a large sub-class of those models, the *additive stimulus* models, on which the main discussion of the paper is focussed. The area of discussion is further specified by relating the probability statement, which is the main form in which the model is estimated and used, to the utility and utility difference distributions. New concepts are introduced of indistinguishability and almost-indistinguishability, which can be used in assessing discrete choice models. The paper then shows how a reasonable notion of *model structure* can be interpreted in terms of utility difference distributions for a class of indistinguishable models.

The discussion of the independence of the utility distributions of the alternatives is based on the concepts introduced in the early parts of the paper. This discussion shows that many indistinguishable models exist for which the correlation of the utility functions is radically different. A following discussion goes on to show that the notion of heteroskedasticity is similarly incapable of clear definition, even within classes of indistinguishable models.

The final main section discusses the issue of existence, finding that it is quite difficult to ensure that models actually represent a 'real' situation, although it is seen as important that the models actually 'exist' in some sense.. An *error components* approach, whether using purely probit models or substituting a logit *kernel* appears a useful approach to maintaining the 'reality' of the model.

2. 'Random utility' models

Random utility models (RUM) are a method of predicting discrete choice behaviour, i.e. choice by a traveller or consumer from a mutually exclusive and finite set of alternatives. The method sets out a clear and consistent set of criteria by which choices in a wide range of situations may be analysed.

Random utility models are said to date back to the work of Thurstone in 1927 and were further developed by others including McFadden, who first introduced them into transportation analysis (Domencich and McFadden, 1975, but describing work done somewhat earlier). They appear first to have been used in this field in Britain by Harris and Tanner (1974), Daly and Zachary (1975) and Williams (1976). Over the succeeding 25 years these models have proved extremely successful as tools for the analysis of traveller behaviour and the forecasting of travel demand (McFadden, 2000).

RUM places the modelling of discrete choice consumer behaviour within the framework of consumer utility maximisation. The exact assumptions required to establish this framework are difficult to state concisely. Certainly they include consistency and transitivity of preference. The difficulties in accepting these assumptions lie in their simple validity, i.e. whether preference is always transitive, rather than in the technical details. However, without assumptions very like these it is difficult to imagine any constructive forecasting of human behaviour. It is also necessary to assume the existence of a cardinal measure of utility.

However, for the purposes of the present paper, it is not necessary to make any assumption of continuity of utility with respect to attributes of the alternatives, although this will usually be necessary in practical applications. Thus the utility functions considered here do not necessarily imply compensatory forms of behaviour. Nor is it necessary here to consider the implications of a budget constraint.

Specifically RUM implies the prediction of discrete choice by a model in which the consumer's probability of choice of alternative a from the choice set C is given by

$$p_i = \Pr \{ U_i \geq U_j, \forall j \in C \} \quad (1)$$

where U gives the (indirect) utility of each alternative and can always be written as[1]

$$U_i = V_i + \varepsilon_i \quad (2)$$

where V is the best available approximation to U; and
ε is the error in that approximation.

It is generally reasonable to take the mean of ε to be zero or some other standard value (for example, it is possible to include the mean (or base) value of the distribution of ε within V).

Thus behaviour is represented as being determined exclusively by the maximisation of utility, but that the utility is not known exactly. It is perhaps useful to note that the model does not necessarily represent random behaviour. More plausible is to think of the individual's behaviour as being deterministic, but that the analyst does not have full information on the variables motivating behaviour, or even accurate measurement of those that are known, nor are the individual's preferences over the motivating variables known. For this reason an error term ε is introduced into the model. Equivalently, and perhaps even more plausibly, an individual consumer can be considered as representing a random sample from a population, each of whom has his or her own value of ε. Then, even if the analyst had perfect information about the distribution of preferences in the population, we would still not know the preferences of a randomly sampled individual. Either way, when we speak of random utility models, it's the model that's random, not necessarily the utility.

[1] This additive form is *not* an assumption, as is sometimes stated; all that is implied by the following equation is that ε is the difference between the true value of cardinal utility and the analyst's approximation of that value.

For the present discussion, the best approximation V to the utility function is assumed to be known. Generally V will involve a number of measured data items and a number of parameters which have to be estimated statistically from observations. This important aspect of the model is however not of interest in the present note. The approximation V is taken to include a specification of the set of alternatives available to each consumer, i.e. we can say which alternatives have $V = -\infty$.

Given V and the distribution of ε, the probability of making each choice can in principle be calculated by integration

$$p_i = \int_{\{Ui \geq Uj,\ \forall j\}} f(t).dt = \int_{\{\varepsilon i - \varepsilon j \geq Vj - Vi,\ \forall j\}} f(t).dt \qquad (3)$$

where f is the multivariate frequency function for ε. The probability of choosing i is simply the probability that the individual's value of ε lies within the region where the utility of i has a value higher than that of all the other alternatives.

Of course this integration presents numerous practical and theoretical problems, which are also not covered here. The aim is rather to discuss the ways in which the distribution of ε can be represented and how this affects the forecasts of behaviour derived from the models.

3. 'Additive stimulus' models

The objective of the model is usually to predict consumers' responses to changes in the function V, or perhaps their valuations with respect to potential changes in V. One aspect of their responsiveness is the distribution of ε, to which detailed consideration is given below. The other aspect is the specification of V, which must in any case contain all the variables for which responsiveness has to be predicted. Often, V will contain other background variables that are not of interest for forecasting but are important for explaining behaviour: socio-economic variables commonly appear, for example.

Separating these two classes of variables, U can be expressed as

$$U = X + Y + \varepsilon \qquad (4)$$

where X represents the 'stimulus', the variables of interest for forecasting, potentially interacted with background variables;
Y represents the other (background) variables, independent of the stimulus variables.

The class of *additive stimulus* (AS) models is then those for which the distribution of ε does not depend on X, although it may depend on Y. That is, changes to the stimulus variables add to the utility equally for all consumers, they do not change the distribution of ε which would imply different changes for some consumers than for others.

It is also possible that the distribution of ε is also conditioned by a third group of variables, say Z, which do not affect the utility directly. In the same way that the

assumption is made that the best specification of V is known, it will also be assumed that the best parametrisation of the distribution of ε in terms of Y and Z is known.

This formulation of the model represents a generalisation for a population of the 'location parameter' models covered by Zachary's theorem (Daly and Zachary, 1978), in that the AS formulation effectively defines a separate model (depending on Y and Z) for each individual or for each subset of the population with the same values of Y and Z. The calculated probabilities are left unchanged by an equal change to all the X's (i.e. these are location parameters) but may change when Y or Z changes uniformly because of changes in the ε distribution. Zachary's theorem applies to AS models with respect to changes within X, i.e. it applies separately to each class of the population defined by a given value of (Y, Z).

The AS models appear to fall within McFadden's (1981) Additive Income RUM class if (i) the costs of the alternatives appear only in the X component and (ii) the costs appear linearly with the same coefficient for each alternative and (iii) income appears in the model only as a fixed personal characteristic, i.e. in Y or Z, in which case its role is as an index of social class etc., not as a measure of spending power.[2] If these conditions are satisfied, McFadden's (1981) analysis applies in full; when these conditions are not satisfied the model will still meet the original Zachary specification[3], i.e. be consistent with some notions of rationality, but it will not necessarily permit the development of consistent social surplus measures.

The need to restrict the discussion to the additive stimulus class of models arises in the following section.

4. Probability statements and utility differences

It is important to remember that utility is a latent variable, introduced into the discussion only to help explain behaviour. It is not possible to make direct observations of utility; to be able to do so would *imply* the existence of a cardinal utility. While considerable progress can be made by careful questioning of consumers in interview contexts, the ultimate test of theories developed in these contexts is always 'revealed preference' observations of actual behaviour and in that context utility is latent.

In dealing with models explaining human behaviour, it is unreasonable to expect that a complete explanation of every individual's actions can be given; the best that can be hoped for is that probability statements can be set up which will explain and predict the behaviour of most people, while allowing for the fact that unusual behaviour will often be encountered (Daly, 1982). The form of the model that can be directly tested on revealed preference data is then the probability statement. This is the form in which the scientific notion of falsifiability can be applied.

---

[2] Income may also appear in X, providing it appears linearly and with a coefficient that is the negative of the cost coefficient, but in this case it has no impact on choice as it is effectively a constant added to each of the alternatives.

[3] The Zachary theorem is expressed in terms of the utility of each alternative, i.e. (X + Y) in the present notation, and does not refer to the cost of an alternative at all.

Quite a few behavioural paradigms can be set up which can yield testable probability statements (e.g. Daly, 1982). For this note discussion has been limited to the utility maximisation paradigm, but even within that paradigm it has long been known that many different conceptual formulations can lead to identical probability statements.

To give precision to this point the concept of indistinguishability can be defined:

models are <u>indistinguishable</u> if they predict the same choice probabilities over the full range of the stimulus and background variables.

No test using observed choices can be used to reject a model in favour of one that is indistinguishable, i.e. a discussion as to which model is better can be conducted only on theoretical grounds. For practical purposes the models are exactly the same, both in terms of explaining observed behaviour and in predicting future choices.

Indistinguishability relates closely to the identifiability of the models on choice data. Models that are indistinguishable cannot be separately identified on *any* choice data and will give identical forecasts. That is, indistinguishability is a slightly stronger property.

Because of the utility maximisation basis of these models, it is clear that only the utility differences are relevant to generating distinguishable models. Subtracting one utility value from all the utilities does not change the choice probabilities. Referring to the integration over ε to calculate the probabilities (equation 3), it can be seen that the integration is *defined* in terms of utility differences. Thus once the utility differences are specified, in terms of their distribution as well as mean values, the model is defined up to the level of indistinguishability.

Note that if the models were not defined as being within the AS class, then a change in stimulus variables might also change the utility variances, which could induce further changes in the probabilities. The model could not then be specified exclusively by the utility differences and their distributions; it would also have the property that the addition of a constant to the utility of each alternative might well change the choice probabilities.

5. Second-moment properties

For the remainder of the note, attention will be focussed on the distribution of ε and the differences of ε.

It has already been noted that the first moment (mean) of ε can reasonably be taken as zero. Further, the distribution of ε will often be symmetrical or near-symmetrical, implying that its third moment is zero (indeed all the odd-numbered moments are zero); even more often, the distribution of ε differences will be symmetrical[4]. Thus

[4] A symmetrical distribution of differences will occur whenever (a) ε is distributed symmetrically (e.g. normally) or (b) ε is distributed identically (i.e. with the same functional form and equal variance, but not necessarily independently) across the alternatives. There may be other cases.

the second moment (variance-covariance matrix) of the distributions of ε across the alternatives becomes the main topic of interest.

The discussion will be set out for models based on the normal distribution (probit models). However, the findings will also apply to models based on other distributions of ε, such as models of the GEV family or hybrid models, with inaccuracies that depend on the differences of those distributions from normal in the fourth and higher moments. For a probit model, of course, the utilities and their covariance define the model entirely.

The concept of almost-indistinguishability can be defined:

> *models are <u>almost indistinguishable</u> if the differences between the choice probabilities they predict are due to differences in the fourth and higher-numbered moments of the distributions of utility differences.*

The simple probit model is almost indistinguishable from the simple logit model. In fact, all GEV models are almost indistinguishable from a suitable AS probit model. In practice, experience indicates that noticeable differences can be found between models that are 'almost indistinguishable' by the definition above; care is needed. Nevertheless, almost indistinguishable models can be taken as having a general similarity in structure.

In a probit model, if the covariance matrix of the utilities is defined, it is easy enough to calculate the covariance matrix of the utility differences and hence the probability statement. However, there will in general be many utility covariance matrices that will lead to indistinguishable models. To be clear, for n alternatives the covariance matrix of the utilities, which will be denoted by Σ, has n.(n+1)/2 components, while the covariance matrix of utility differences, which will be denoted by Δ, has n.(n–1)/2 components, n fewer, so that there are n degrees of freedom in defining the utility covariances that will all lead to indistinguishable models.[5]

The above finding with respect to the number of degrees of freedom in probit models was presented by Bunch (1991), who also indicated that it was known to previous authors. Bunch further showed an additional range of degrees of freedom of variance which are (effectively) parametrised by characteristics of individuals and alternatives. In the present context, variance parametrised by variables appearing in Y or Z is handled by specifying effectively different models for each individual, while the possibility of variance being parametrised by variables appearing X is excluded in the additive stimulus form.

6. Model structure

In consequence of the preceding sections, the discussion is restricted to probit models of the AS class, with the covariance matrix of the utility functions effectively constant if we consider one individual at a time.

[5] A further excess degree of freedom may arise because the model may be indistinguishable from one in which utility means and standard deviations, and utility difference means and standard deviations, are all scaled by a common factor. This point is discussed in more detail subsequently.

In additive stimulus models, the probabilities of choice respond to changes in the stimulus variables through changes in the utility functions. Thus, a change in one of the stimulus variables for an alternative – e.g. its cost – will have an impact on the choice probability of that alternative, and consequently the choice probabilities of other alternatives, through its impact on the utility function of that alternative. However, the connection between the stimulus variable and the utility is more properly described as the specification of the utility function than the structure of the model.[6]

Rather, by 'model structure' is meant the impact that a change in the utility function of a specified alternative has on the choice probabilities. Do all the other choice probabilities change proportionately, as in a multinomial logit model? If not, how do the relative changes in the probabilities of the other alternatives differ? Which alternatives are close substitutes and which serve relatively separate markets?

For AS probit models, the notion of 'structure' relates to the specification of the utility covariance matrix. Alternatives that are close substitutes will be characterised by having high correlation of their utilities, while alternatives serving differing markets will have low correlation of their utilities.

This notion of model structure is consistent with the concept of indistinguishability that has been introduced. Models that are indistinguishable have the same choice probabilities for all values of the stimuli, so they respond in the same way to utility changes. That is, they have the same structure. In the same way, the results shown here can be extended to other models that are almost indistinguishable from probit models, subject to the practical reservation mentioned above.

Also, because of the way in which AS models calculate probabilities from utility differences, models with the same distributions of utility differences are indistinguishable, i.e. they have the same structure. However, there are very many ways of specifying the distribution of utilities to derive the same distribution of utility differences. Specifically, as noted above, there are n degrees of freedom to define models indistinguishable from a specified model over n alternatives.

7. Independence

Given a model with structure defined by a matrix of *utility difference* covariance, there exist many indistinguishable models of *utility* covariance which can underlie that structure. These models, indistinguishable in practice, may however have different interpretations in theory.

---

[6] There is no reason, within the framework discussed in this paper, that the connection between stimulus variable and utility should be continuous. The continuity of the connection between utility and probability depends on the specification of the distribution of ε. Zachary's theorem requires infinite differentiability of this connection. In any case, we can claim, whatever the distribution of ε, providing differences of ε's are distributed symmetrically, that the model is almost indistinguishable from a probit model, i.e. almost indistinguishable from a model with a continuous utility–probability connection.

For example, consider the four-alternative model with utility covariance matrix Σ given by

$$\begin{matrix} \sigma^2 & \rho\sigma^2 & 0 & 0 \\ & \sigma^2 & 0 & 0 \\ & & \sigma^2 & \rho\sigma^2 \\ & & & \sigma^2 \end{matrix}$$

For this model we can calculate Δ

$$\text{var}(U_1 - U_2) = \text{var}(U_3 - U_4) = 2\sigma^2.(1 - \rho)$$
$$\text{var}(U_1 - U_3) = \text{var}(U_1 - U_4) = \text{var}(U_2 - U_3) = \text{var}(U_2 - U_4) = 2\sigma^2$$

When $\rho \geq 0$, this model is almost indistinguishable from a tree logit model with two nests (each of two alternatives) with the structural parameters both equal to $\sqrt{(1 - \rho)}$. It would usually be considered that the utilities in this structure are homoskedastic[7] *and that alternatives 1 and 2 were independent of alternatives 3 and 4.*

The model can be transformed by subtracting the utility of the first alternative from the others, obtaining the covariance matrix

$$\begin{matrix} 0 & 0 & 0 & 0 \\ & \sigma^2_1 & \rho_1\sigma_1\sigma & \rho_1\sigma_1\sigma \\ & & \sigma^2 & \rho\sigma^2 \\ & & & \sigma^2 \end{matrix}$$

and this gives exactly the same utility difference variances as in the previous case – i.e. the model is indistinguishable – if

$$\sigma_1^2 = 2\sigma^2.(1 - \rho) \text{ and } \rho_1 = (\sigma^2 - 2\sigma_1^2) / 2\sigma\sigma_1$$

Note that in this indistinguishable model, both independence and homoskedasticity have vanished. The matrix of utility covariance is also no longer positive definite but only semi-definite.

An alternative transformation gives the matrix

$$\begin{matrix} \sigma^2_1 & 0 & 0 & 0 \\ & \sigma^2_1 & 0 & 0 \\ & & \sigma^2 & \rho\sigma^2 \\ & & & \sigma^2 \end{matrix}$$

which again gives the same difference covariances Δ if $\sigma_1^2 = \sigma^2.(1 - \rho)$.

Here there is another variant on the independence interpretation, with alternatives 1 and 2 being independent of each other as well as from 3 and 4. This matrix shows the 'maximum independence' that can be achieved in this model, i.e. it is not possible

[7] Homoskedastic is used to mean 'of equal variance'.

to eliminate the off-diagonal elements entirely without losing indistinguishability. Bunch (1991) gives another example with three alternatives.

Thus statements such as 'independence' of utility distributions must be treated with care, since models with apparently different structures of independence – i.e. with a different pattern of zeroes in the utility correlations – can in fact be indistinguishable, because the choice probabilities are identical. Similarly, notions of symmetry in the model can be misleading.

However, it is the case that not all models can be expressed with independent utility functions, as the last example illustrates.[8] On the other hand, homoskedasticity is a more general property.

8. Homoskedasticity

The main result of this section is the following somewhat chimeral lemma.

Homoskedasticity Lemma: Any multinomial probit AS model is indistinguishable from a homoskedastic model; *the homoskedastic model may however be imaginary*.

Proof

Represent the model by its utility differences and their covariance matrix $\Delta = \{\delta_{ij}\}$, which can be derived from the utility covariance matrix Σ by

$\delta_{ij} = \sigma_{ii} + \sigma_{jj} - 2\sigma_{ij}$, i.e. $\delta_{ij} \geq 0$, since Σ must be positive semi-definite

This matrix is also the matrix of utility differences of the homoskedastic model whose utility covariance matrix is k.A, where A is derived from Δ by the transformation h:

i) $\alpha_{ij} = h_k(\delta_{ij}) = 1 - \delta_{ij} / (2.k)$,

i.e. $\alpha_{ii} = 1 \; \forall$ i, since $\delta_{ii} = 0$; and $\alpha_{ij} \leq 1$ since $\delta_{ij} \geq 0$;

ii) k is chosen so that $k \geq \frac{1}{4}.\max_{ij} \delta_{ij}$.

The definition of k ensures that $-1 \leq \alpha_{ij}$, while the presence of 1's on the diagonal implies that A is a matrix of correlations. kA is obviously what one would normally describe as a homoskedastic matrix.

The utility difference variances implied by the model k.A are

$$\begin{aligned} \text{var}(U_i - U_j) &= k \,.\, (\alpha_{ii} + \alpha_{jj} - 2.\alpha_{ij}) \\ &= (-\delta_{ii} - \delta_{jj} + 2.\delta_{ij}) / 2 \end{aligned}$$

[8] Of course there are n.(n–1) distinguishable degrees of freedom for a model over n alternatives, while only n degrees of freedom for the diagonal elements. For this reason it is obviously not possible for a model with independent heteroskedastic utilities to give a complete representation of all models over n alternatives.

$$= \delta_{ij},$$

thus proving the Lemma.

The problem is that there is no guarantee that kA is positive semi-definite, which is necessary (and sufficient) for it to be described as a covariance matrix. The limit on the correlations implies that, if there is a problem, it involves more than two variables.

But the lemma applies in reality in some important cases. For example, consider the 'independent heteroskedastic' model defined by the utility covariance matrix

$$\begin{matrix} \sigma_1 & 0 & 0 \\ & \sigma_2 & 0 \\ & & \sigma_3 \end{matrix}$$

with $\sigma_1 \geq \sigma_2 \geq \sigma_3 > 0$[9]. This model has the covariance matrix of utility differences Δ with

$$\delta_{ij} = \sigma_i + \sigma_j, \text{ if } i \neq j; \ \delta_{ii} = 0$$

which can be converted to an indistinguishable homoskedastic form A by the transformation above. The determinant of the matrix A is

$$D = 1 - \alpha_{12}^2 - \alpha_{23}^2 - \alpha_{31}^2 + 2.\alpha_{12}.\alpha_{23}.\alpha_{31}$$

If we choose $k = \frac{1}{2}.\max_{ij} \delta_{ij} = \frac{1}{2}.(\sigma_1 + \sigma_2)$, we get

$$\begin{aligned} \alpha_{12} &= 0; \\ \alpha_{23} &= 1 - \delta_{23}/\delta_{12} = (\sigma_1 - \sigma_3) / (\sigma_1 + \sigma_2); \\ \alpha_{31} &= 1 - \delta_{31}/\delta_{12} = (\sigma_2 - \sigma_3) / (\sigma_1 + \sigma_2). \end{aligned}$$

$$\begin{aligned} D &= 1 - \alpha_{23}^2 - \alpha_{31}^2 \\ &= 2 \, . \, (\sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_3\sigma_1 - \sigma_3^2) / (\sigma_1 + \sigma_2)^2 \end{aligned}$$

which must be positive under the assumptions made. Thus in this case the matrix A is positive definite[10] and is a valid covariance matrix.

However, if k is set to $\frac{1}{4}.\max_{ij} \delta_{ij} + \iota = \frac{1}{4}.(\sigma_1 + \sigma_2) + \iota$, for some small ι, we get $\alpha_{12} = -1 + \iota^*$ (for a corresponding small ι^*) and

$$D = -(\alpha_{23} + \alpha_{31})^2 + \varepsilon.(2 - \alpha_{23}.\alpha_{31}) - \varepsilon^2$$

---

[9] That the variances are positive implies non-degeneracy. Then the alternatives can always be permuted to get the conditions required.

[10] A matrix is positive definite if and only if each of its principal submatrices has a positive determinant. The principal submatrices (2 * 2) clearly have positive determinants, because of the limit on the α variables, so the demonstration that D is positive is sufficient to show that A is positive definite (Noble and Daniel, 1988, p. 413, Theorem 10.19).

and unless $\alpha_{23} + \alpha_{31} = 0$ this is negative for small enough ι, so that the matrix A is *not* positive definite and thus not a valid covariance matrix[11].

Thus any heteroskedastic-independent model for three alternatives is indistinguishable from a valid homoskedastic-correlated model, by choosing a suitable k. However, k values exist for which the homoskedastic model derived does not have a valid covariance matrix.

Thus, the chimeral quality of the lemma means that it not clear in advance whether a model derived by applying the transformation for a given value of k will yield a valid covariance matrix. But it is clear that the concept of heteroskedasticity, like that of independence treated in the previous section, can be made to disappear from many models without affecting the probabilities predicted.

The transformation h of the homoskedasticity lemma has two notable features.

– The factor k introduced into the matrix kA is a convenient and natural way to deal with the excess degree of freedom introduced by common scaling of the covariance matrix and the mean utility values. For example, k can be constrained to 1 if it is required to estimate a scale for utilities; alternatively, if the utilities themselves are scale-constrained, k may be estimated.

– The number of degrees of freedom in the homoskedastic form is exactly that available, i.e. ½.n.(n–1) – 1, in the covariance matrix. Moreover, because the transformation h is invertible and of 'full rank' (in the sense of Bunch, 1991, who calls transformations from utility difference variances to utility variances *normalisations*), there is a 1-to-1 correspondence between the original models and the transformed homoskedastic models.

Thus the homoskedasticity lemma appears to offer a 'canonical form' in which probit models can be expressed to permit greater comparability between them. However, unless a proof can be found that these homoskedastic models always exist, the application of the canonical form cannot be carried through.

9. Existence and Parametrisation

The fact that indistinguishable models can be set up in a wide range of ways raises the issue of how this should be done. The key step in setting up a model is to estimate the unknown parameters from observations. The 'fundamental' way is to conduct the estimation based on utility differences. But how can we be sure that a model that has been estimated actually exists, i.e. has a covariance matrix of utilities that is positive semi-definite?

A first issue is whether this question is worth answering. It could be thought that the existence or otherwise of a model of utility variation corresponding to the matrix of utility differences was not important; after all the predictions of the model depend only on the variances of utility differences. However, the whole thrust of RUM theory is to ensure that predictions are made on the basis of a sound underlying theory

[11] A has to be positive *semi*-definite to be a valid covariance matrix. However, it is clear in this context that the zero value – i.e. the points of singularity of A – are not the issue.

which corresponds to real-valued utility functions. The introduction of imaginary values (i.e. negative variances) departs from this basis and is likely to lead to counter-intuitive prediction. It may be concluded that the maintenance of reality in the distributions is highly desirable.

When estimating a model it is very difficult to specify in advance the constraints that need to be applied to maintain positive definiteness in the utility covariance matrix. This issue is discussed extensively by Bunch (1991) who advises working, rather than with utility differences, with a matrix C_J in which the utility of one alternative (the last, J, in Bunch's work) is subtracted from the utilities of all the other alternatives. To derive a model with a matrix of utility variances Σ, Bunch advises that "the best course to follow is to first estimate C_J, then explore the various possible normalisations by performing transformations between C_J and Σ". However, this approach has three drawbacks.

- As has been shown above, some normalisations lead to valid matrices, some do not. The search for a valid normalisation might be time-consuming.
- Often, it is not required to estimate a full matrix of covariance of all the utilities. A reduced matrix is often required, because the data is insufficient to support estimates of the full matrix or because of the nature of the problem being studied. Or it may be required that the matrix is determined by a smaller set of parameters which have to be estimated. Restrictions of this type are often difficult to apply in estimating the full matrix Σ, and may be even more difficult when estimating C_J.
- Bunch does not explain how the matrix C_J should be kept positive semi-definite.

An alternative approach, inspired by Daly and Zachary (1975) is to parametrise C_J by an underlying matrix Y from which C_J is derived by

$$C_J = Y^2$$

which guarantees that C_J is positive definite. Daly and Zachary show that there is a correspondence between Y and C_J which is 1-to-1, except for possible sign changes on the rows or columns of Y. Further, C_J can be restricted to be diagonal, or diagonal and constrained to fixed values, by imposing parallel restrictions on Y, allowing at least some degree of flexibility in the parametrisation. However, this approach does not solve all the problems of offering suitable parametrisation, nor does it indicate how to derive matrices Σ from C_J.

A further approach might be to work with the matrix Δ of utility differences. However, this does not solve many of the problems listed above, while when working with the matrix Δ, constraints are also needed to ensure existence. If i is closely correlated with j, and j is closely correlated with k, it cannot be the case that the correlation of i and k has a large negative value. A kind of 'triangle constraint' applies:

$$\sqrt{\delta_{ij}} \leq \sqrt{\delta_{jk}} + \sqrt{\delta_{ki}}, \text{ for all i, j and k.}$$

However this constraint, while necessary, is far from sufficient to ensure positive semi-definiteness.

An alternative approach which seems to have several advantages is to specify the covariance of the utilities *constructively*, i.e. by specifying variance components *a priori*, as is done, for example, in the error components approach. With this approach, each component of the covariance is specified and explicitly attached to the alternatives to which it applies. The issues of existence are solved automatically, while parametrisation and normalisation also do not arise as problems.

The specification of error components is often connected with the use of a logit 'kernel', in which a component with a limiting value distribution is added to the utility of each alternative. This approach is then often called mixed logit or random parameters logit. However, the use of error components can also be applied within a purely probit framework, with advantages for the parametrisation and ensuring that the final model will have a positive definite covariance matrix.

The difficulty in the pure-probit error components approach, however, is that existing software does not support the approach and further software development may be necessary to apply the method. Software does exist for the logit kernel variant, which in the terms of this paper is almost indistinguishable from the pure-probit approach, as well as being much quicker and easier to apply.

10. Conclusions

A class of Random Utility models, the Additive Stimulus models, has been defined which allows a minor extension of Zachary's theorem. The results obtained in the paper are based on analysis of this class of models.

The concept of indistinguishability has been defined, grouping models into classes that are effectively identical. These classes can be extended to models with similar structure but different detailed functional form by the concept of almost-indistinguishability. Any AS model with a symmetrical distribution of utility differences is almost indistinguishable from an AS probit model.

The concept of independence of the utility functions is not well defined for probit models. Retaining indistinguishability, it is possible to change the apparent extent of independence of the utilities. Such changes also change the extent to which the variances of the utilities are equal, i.e. the homoskedasticity of the model.

The concept of homoskedasticity is also not well defined. Retaining indistinguishability, it is possible to convert *any* probit model to a homoskedastic model. This change alters the independence characteristics of the utilities also, i.e. independence and homoskedasticity must be defined, if at all, together. Changes to a model to make it homoskedastic may also cause it to lose reality, i.e. the covariance matrix of the utilities may cease to be positive semi-definite.

Maintaining positive semi-definiteness of the covariance matrix of utilities is not easy. A general estimation procedure, or a 'normalisation' from a space of lower

dimension may cause the matrix to lose semi-definiteness or at least to make a check necessary. Constructive model specifications, such as those employing error components, can be a useful way to maintain a positive semi-definite matrix.

Acknowledgements

I am grateful for comments on a previous draft of this paper by my colleagues on the 'Ecology' project team at ITS. However, I claim full intellectual property rights on any remaining errors and misinterpretations.

References

D. S. Bunch, Estimability in the multinomial probit model, *Transportation Research 25B*, No. 1, pp. 1–12, 1991.

A. J. Daly, Applicability of Disaggregate Behavioural Modelling: a Question of Methodology, *Transportation Research*, 16A, 5-6, 1982.

A. J. Daly and S. Zachary, *Commuters' Values of Time*, Local Government OR Unit Report T 55, January 1975.

A. J. Daly and S. Zachary, Improved multiple choice models, in D.A. Hensher and M.Q. Dalvi (eds.), *Determinants of Travel Choice*, Saxon House 1978; previously presented to *P.T.R.C. Summer Annual Meeting*, 1976.

T. Domencich and D. McFadden, *Urban Travel Demand*, North Holland, 1975.

A. J. Harris and J. C. Tanner, *Transport Demand Models Based on Personal Characteristics*, TRRL Supplementary Report 65UL, 1974.

D. L. McFadden, Econometric Models of Probabilistic Choice, in *Structural Analysis of Discrete Data with Econometric Applications*, Manski and McFadden (eds.), MIT Press, Cambridge, 1981.

D. L. McFadden, Disaggregate Behavioural Travel Demand's RUM Side: A 30-year Retrospective, presented to *IATBR Conference*, Gold Coast, Queensland, 2000.

B. Noble and J. W. Daniel, *Applied Linear Algebra*, Prentice-Hall International, 1988.

H. C. W. L. Williams, The Formation of Travel Demand Models and User Benefit Measures, presented to *Conference on Urban Transport Planning*, Leeds, 1976.