Additive Random Utility Models of Probabilistic Choice

Stan Zachary

Heriot-Watt University

Introduction

Many models of probabilistic choice assume that the individual attaches, at least implicitly, a utility to each available option and then chooses that for which this utility is maximized. However these utilities, which may vary from one individual to another, are only partially known to the observer and are thus modelled as containing random elements. Such models are frequently used in econometric analysis to reflect decision making (see, for example, McFadden, 1978, 1981, and Williams, 1977), and are further found extensively in the psychological literature (Strauss, 1979, 1981; Yellott, 1977, 1980).

A *random utility model* is a specification of the utility of each choice as a function both of observed and of unknown, or random, components, the joint distribution of which is given. Any such model then determines, for each option, the probability that it has maximal utility and is thus chosen. This probability is a function of the observed components of the utilities. (We assume here that the probability that two options simultaneously have maximal utility is zero.) The *choice probability functions* thus determined define a model of probabilistic choice, and we then no longer require the underlying random utility model, either for choice prediction or for any associated data analysis. Indeed, as shown below, different underlying random utility models may determine the same model of probabilistic choice. It is nevertheless very important to understand the mathematical relation between the random utility model and the choice model it generates. This facilitates the construction of good choice models and permits economic interpretation of the results of fitting such models to data. Some useful general discussion is given by McFadden (1981).

This note considers the relationship between a often used class of random utility models, frequently known in the econometric literature as *additive random utility models* (ARUMs), and the associated models of choice. Versions of some of the results below were given, without proof, by Daly and Zachary (1978). The present note aims to give a mathematically more coherent treatment of its subject matter, and also takes the opportunity to supply simple proofs of these earlier results.

We do not consider here the detailed statistical problems of fitting such models to data. For some discussion of these, see Daly and Zachary (1978), McFadden (1981), Bunch (1987) and Stern (1990a, 1990b).

ARUMs and their surplus functions

Consider the common random utility model for choice and ranking within a finite set of possible *options* $\{1, \dots, n\}$, $n \geq 2$, in which the utility associated by the individual with each option i is given by

$$U_i(z_i) = z_i + X_i\,, \qquad 1 \leq i \leq n. \tag{1}$$

Here each $z_i \in \mathbb{R}$ (the set of real numbers) represents the observable component of the utility of the option i, and $X_1, \dots, X_n$ are *not necessarily independent* random variables defined on some underlying probability space, the probability measure of which we shall denote by $\mathbf{P}$. (When the model is used in a population, as in statistical applications, $z_1, \dots, z_n$ typically vary according to the individual, while the underlying joint distribution of $X_1, \dots, X_n$ is assumed to remain the same for all members of the population.) The individual is assumed to choose the option of maximum utility. (Note that the model also permits the set of options available to an individual to be restricted: for each unavailable option i we simply let $z_i \to -\infty$.) We shall further assume that the joint distribution of $X_1, \dots, X_n$ is absolutely continuous. This assumption will be satisfied in almost all applications. It follows in particular that the individual's choice is well-defined with probability one. We refer to such models as *additive random utility models* (ARUMs)—the additive form clearly requires that the utility measures z_i be constructed in such a way that this formulation is appropriate. In an econometric context such models are often known as *additive income random utility maximising models*—see McFadden (1981), while in a psychological context, Strauss (1979) and Yellott (1980) refer to them as *generalized Thurstone* models.

For each $\boldsymbol{z} = (z_1, \dots, z_n) \in \mathbb{R}^n$, define the random variable

$$U_{\max}(\boldsymbol{z}) = \max_{1 \leq i \leq n} U_i(z_i),$$

and for each i, $1 \leq i \leq n$, let

$$p_i(\boldsymbol{z}) = \mathbf{P}(U_{\max}(\boldsymbol{z}) = U_i(z_i)).$$

Thus $p_i(\boldsymbol{z})$ is the probability that the individual chooses option i, and in particular

$$\sum_{i=1}^{n} p_i(\boldsymbol{z}) = 1. \tag{2}$$

Further the absolute continuity of the joint distribution of $X_1, \dots, X_n$ ensures that each of the functions $p_1, \dots, p_n$ is continuous. These functions are referred to as the *choice probability functions* generated by the ARUM (1).

Note that we may rewrite the definition of the choice probability functions $p_1, \dots, p_n$ as

$$\begin{aligned} p_i(\boldsymbol{z}) &= \mathbf{P}(z_i + X_i \geq z_j + X_j \;\forall\, j \neq i) \\ &= \mathbf{P}(X_j - X_i \leq z_i - z_j \;\forall\, j \neq i), \qquad \boldsymbol{z} \in \mathbb{R}^n, \quad 1 \leq i \leq n. \end{aligned} \tag{3}$$

Thus, for each i, the function p_i determines, and is determined by, the joint distribution of the random variables $X_j - X_i$, $j \neq i$; further, for every other $k \neq i$, the joint distribution of the random variables $X_j - X_k$, $j \neq k$, is easily seen to be determined by that of the random variables $X_j - X_i$, $j \neq i$. It follows that any one of the functions p_i determines the remainder.

We shall say that two ARUMs are *equivalent* if and only if they generate the same set of choice probability functions, which, for every i, is the same as the requirement that the joint distribution of the random variables $X_j - X_i$, $j \neq i$, is the same for both models. If the only observables of a ARUM (1) are the vector $\boldsymbol{z}$ and the option actually chosen, then equivalent ARUMs are essentially the same. Indeed this remains true even if rankings of options are observable, since the probability of each ranking again depends solely on the joint distribution of the differences of the random variables $X_1, \ldots, X_n$. Thus, in either of these circumstances, it will be sufficient for applications to identify choice models up to equivalence. Specification of a particular ARUM model within an equivalence class is primarily a matter of mathematical convenience. (Some other authors have used broader concepts of equivalence of ARUMs, allowing two models to be called equivalent if their choice probability functions are suitably related to each other, for example through a linear transformation of their arguments. The definition of equivalence used here is made for clarity of mathematical exposition.)

Given a ARUM of the form (1), define its associated *surplus function* $u : \mathbb{R}^n \to \mathbb{R}$ by

$$u(\boldsymbol{z}) = \mathbf{E}(U_{\max}(\boldsymbol{z})),$$

where the random variable $U_{\max}$ is as defined earlier, and $\mathbf{E}$ denotes expectation with respect to the probability measure $\mathbf{P}$. (In an econometric context, the surplus function is essentially a consumer surplus function, with the understanding that, in the model, some option must be chosen—again see McFadden (1981).) Lemma 1 below is more or less standard. In the present context it is due to Harris and Tanner (1974), although we give here a more careful proof.

Given a real-valued function f, say, of one or more real variables, let $D_i f$ denote its partial derivative (where this exists) with respect to its i^{th} argument, and let $D_{ij} f = D_i(D_j f)$, etc.

Lemma 1. *For a ARUM with surplus function u, each of its choice probability functions p_i, $1 \leq i \leq n$, is given by $p_i = D_i u$.*

Proof. For each i, $1 \leq i \leq n$, define the vector $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{in})$ by $e_{ij} = 1$ if $j = i$, $e_{ij} = 0$ otherwise. Then, for every $\boldsymbol{z}$ and every $h \geq 0$,

$$U_{\max}(\boldsymbol{z} + h\boldsymbol{e}_i) = \max(U_1(z_1), \ldots, h + U_i(z_i), \ldots, U_n(z_n)).$$

On taking expectations both for $h > 0$ and $h = 0$, it follows easily that

$$hp_i(\boldsymbol{z}) \leq u(\boldsymbol{z} + h\boldsymbol{e}_i) - u(\boldsymbol{z}) \leq hp_i(\boldsymbol{z} + h\boldsymbol{e}_i).$$

A similar result holds with $\boldsymbol{z}$ replaced by $\boldsymbol{z} - h\boldsymbol{e}_i$. The lemma now follows on dividing by h, letting $h \to 0$ and using the continuity of p_i. □

It follows that a ARUM is identified up to equivalence by its surplus function. Conversely, from the relation

$$u(\boldsymbol{z}) = E(X_1) + E(\max_{1\le i\le n}(z_i + X_i - X_1)),$$

it follows that equivalent ARUMs have surplus functions differing by at most an additive constant.

Theorem 2 below is an improved version of a result published, without proof, by Daly and Zachary (1978). (The original version becomes the present corollary.) A proof of a variation of this result was subsequently given by McFadden (1981), but that given below is considerably simpler. The revised statement of Theorem 2 also owes much to McFadden. For every vector $\boldsymbol{z} = (z_1, \dots, z_n) \in \mathbb{R}^n$ and every $t \in \mathbb{R}$, let $\boldsymbol{z} + t$ denote the vector $(z_1 + t, \dots, z_n + t)$.

Theorem 2. *A function $u : \mathbb{R}^n \to \mathbb{R}$, for which the n^{th} mixed partial derivative $D_{1,\dots,n}u$ exists everywhere, is the surplus function of a ARUM if and only if*

(S1) $u(\boldsymbol{z} + t) = u(\boldsymbol{z}) + t$ *for all* $\boldsymbol{z} \in \mathbb{R}^n, t \in \mathbb{R}$,

(S2) *for all i, j with $i \neq j$,* $\lim_{z_j\to\infty} D_i u(\boldsymbol{z}) = 0$ *(for all fixed z_k, $k \neq j$),*

(S3) $(-1)^{n-1} D_{1,\dots,n}u$ *is everywhere non-negative.*

Proof. Consider first a ARUM (1) with surplus function u such that $D_{1,\dots,n}u$ exists everywhere. The condition (S1) is then immediate from the definition of $U_{\max}$ and u. The condition (S2) is immediate from Lemma 1 and (3) (since $\mathbf{P}$ is a probability measure), while, also from Lemma 1 and (3), the condition (S3) simply states that the joint density of the random variables $X_i - X_n$, $1 \le i \le n-1$, should be non-negative.

Conversely, suppose that the function u is such that $D_{1,\dots,n}u$ exists everywhere and satisfies the conditions (S1)–(S3). Note that from (S1),

$$D_n u(\boldsymbol{z} + t) = D_n u(\boldsymbol{z}), \qquad \boldsymbol{z} \in \mathbb{R}^n, t \in \mathbb{R}. \tag{4}$$

Note also that, by differentiating (S1) with respect to t,

$$\sum_{i=1}^{n} D_i u(\boldsymbol{z}) = 1. \tag{5}$$

Define the function $F : \mathbb{R}^{n-1} \to \mathbb{R}$ by

$$F(z_1, \dots, z_{n-1}) = D_n u(-z_1, \dots, -z_{n-1}, 0). \tag{6}$$

We show that F is a distribution function. From (6) and the condition (S3), $D_{1,\dots,n-1}F$ exists everywhere and is non-negative. Further, from (6) and the condition (S2) with $i = n$,

$$F(z_1, \dots, z_{n-1}) = \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_{n-1}} D_{1,\dots,n-1}F_n(\zeta_1, \dots, \zeta_{n-1})\, d\zeta_1 \dots d\zeta_{n-1} \tag{7}$$

Finally, from (4)–(6) and the condition (S2),

$$\begin{aligned}\lim_{t\to\infty} F(z_1+t,\dots,z_{n-1}+t) &= \lim_{t\to\infty} D_n u(-z_1-t,\dots,-z_{n-1}-t,0)\\ &= \lim_{t\to\infty} D_n u(-z_1,\dots,-z_{n-1},t)\\ &= 1-\sum_{j=1}^{n-1}\lim_{t\to\infty} D_j u(-z_1,\dots,-z_{n-1},t)\\ &= 1.\end{aligned}$$

Thus F is a distribution function with density $D_{1,\dots,n-1}F$.

Now consider any ARUM (1) in which the random variables $X_i - X_n$, $1 \le i \le n-1$, have joint distribution function F. (For example we may take $X_n = 0$.) Let u^* be its corresponding surplus function, and note that, by that part of the theorem already proved, the conditions (S1), (S2) are also satisfied when u is replaced by u^*. We show that u and u^* differ by at most an additive constant. It will then follow that u is the surplus function of the equivalent ARUM in which each of the random variables X_i is adjusted by this constant. From Lemma 1 and (3), (6) and (4),

$$\begin{aligned}D_n u^*(\boldsymbol{z}) &= F(z_n-z_1,\dots,z_n-z_{n-1})\\ &= D_n u(z_1-z_n,\dots,z_{n-1}-z_n,0)\\ &= D_n u(\boldsymbol{z}).\end{aligned}$$

Thus, for some function $g:\mathbb{R}^{n-1}\to\mathbb{R}$,

$$u^*(\boldsymbol{z}) = u(\boldsymbol{z}) + g(z_1,\dots,z_{n-1}). \tag{8}$$

For each i, $1 \le i \le n-1$, by differentiating (8) with respect to i, letting $z_n \to \infty$ (keeping $z_1,\dots,z_{n-1}$ fixed) and using the condition (S2), both as it is stated and with u replaced by u^*, we obtain that $D_i g$ is identically zero. Hence the function g is identically equal to a constant as required. □

Remark 1. The requirement in Theorem 2 that $D_{1,\dots,n}u$ should exist everywhere may be dropped if the condition (S3) is re-expressed in terms of a suitable Radon-Nikodym derivative. We have given the theorem in the form most useful for the majority of applications.

Theorem 2 has the following corollary, which re-expresses its results in terms of choice probability functions.

Corollary 3. *Suppose the functions $p_i:\mathbb{R}^n\to\mathbb{R}$, $1\le i\le n$, are such that, for each i, $D_{1,\dots,i-1,i+1,\dots,n}p_i$ exists everywhere. Then they form the choice probability functions of a ARUM if and only if*

(C1) *for all i, $p_i(\boldsymbol{z}+t) = p_i(\boldsymbol{z})$,* $\quad \boldsymbol{z}\in\mathbb{R}^n, \quad t\in\mathbb{R}$,

(C2) $\sum_{i=1}^n p_i(\boldsymbol{z}) = 1, \quad \boldsymbol{z}\in\mathbb{R}^n$,

(C3) *for all i, j with $i \ne j$, $\lim_{z_j\to\infty} p_i(\boldsymbol{z}) = 0$ (for all fixed z_k, $k \ne j$),*

(C4) *for all i, $(-1)^{n-1}D_{1,\dots,i-1,i+1,\dots,n}p_i$ is everywhere non-negative,* (C5) *for all* i, j, $D_j p_i = D_i p_j$.

Proof. The *'only if'* part of the result is immediate from Lemma 1 and Theorem 2. Conversely, suppose that the functions $p_1, \dots, p_n$ satisfy the conditions (C1)–(C5). Then (C4) and (C5) ensure that there exists a function $u : \mathbb{R}^n \to \mathbb{R}$ such that, for all i, $p_i = D_i u$ (see Apostol, 1957, p. 296). Now consider the function of $\boldsymbol{z} \in \mathbb{R}^n$ and $t \in \mathbb{R}$ given by $u(\boldsymbol{z} + t) - u(\boldsymbol{z}) - t$. The conditions (C1) and (C2) ensure that the partial derivatives of this function, both with respect to $z_1, \dots, z_n$ and with respect to t, are all zero, so that this function is identically equal to a constant, which, by putting $t = 0$, is seen to be zero. Thus, using also the conditions (C2)–(C4), the function u satisfies the conditions (S1)–(S3) of Theorem 2, and so is the surplus function of a ARUM with choice probability functions $p_1, \dots, p_n$. □

Logit and nested logit models

This section considers the properties of what is perhaps the most commonly used class of ARUMs. For the reasons discussed in the previous section we identify these up to equivalence, so that each is most concisely described by its surplus function. The ARUMs considered all have the property that, for each pair of distinct options i and j, the distribution of the random variable $X_i - X_j$ is logistic with mean zero, i.e. has a distribution function F given by

$$F(x) = [1 + \exp(-\lambda x)]^{-1},$$

for some $\lambda > 0$. (There is clearly no loss of generality in the requirement that the mean be zero.) We refer to this as the logistic distribution with parameter λ. Its variance is $\pi^2/3\lambda^2$ (see Cox and Snell, 1989, p. 200). In the well-known simple multinomial logit model of choice the parameter λ is the same for all such pairs of options. In the nested multinomial logit model, considered in detail below, the parameter λ varies between different pairs of options.

For every integer $n \geq 1$ and every $\lambda > 0$, define the function $\theta_{n,\lambda} : \mathbb{R}^n \to \mathbb{R}$ by

$$\theta_{n,\lambda}(\boldsymbol{z}) = \frac{1}{\lambda} \log\left(\sum_{i=1}^{n} \exp \lambda z_i\right).$$

Note that

$$\theta_{n,\lambda}(\boldsymbol{z} + t) = \theta_{n,\lambda}(\boldsymbol{z}) + t, \qquad \boldsymbol{z} \in \mathbb{R}^n, \quad t \in \mathbb{R}, \tag{9}$$

that

$$\begin{aligned} D_i\theta_{n,\lambda}(\boldsymbol{z}) &= \frac{\exp \lambda z_i}{\sum_{j=1}^{n} \exp \lambda z_j} \\ &= \exp \lambda(z_i - \theta_{n,\lambda}(\boldsymbol{z})), \qquad \boldsymbol{z} \in \mathbb{R}^n, \quad 1 \leq i \leq n, \end{aligned} \tag{10}$$

and that

$$\sum_{i=1}^{n} D_i\theta_{n,\lambda}(\boldsymbol{z}) = 1, \qquad \boldsymbol{z} \in \mathbb{R}^n. \tag{11}$$

For a set of options $\{1, \dots, n\}$, the *simple multinomial logit model* (SMLM) with parameter $\lambda > 0$ is the ARUM identified, up to equivalence, by the surplus function $\theta_{n,\lambda}$. The corresponding choice probability functions are thus given by $p_1, \dots, p_n$, where

$$p_i(\boldsymbol{z}) = \frac{\exp \lambda z_i}{\sum_{j=1}^{n} \exp \lambda z_j}. \tag{12}$$

While it may be verified directly from Theorem 2 that there do exist ARUMs with these surplus and choice probability functions, this is well-known. (For a useful discussion of this and related theory, see Yellott (1977).) In particular Holman and Marley (cited in Luce & Suppes, 1965) have shown that one member of the equivalence class of such models is given by taking the joint distribution function F of the random variables $X_1, \dots, X_n$ in (1) to be given by

$$F(\boldsymbol{x}) = \exp\{-\sum_{i=1}^{n} \exp(-\lambda x_i)\},$$

where $\boldsymbol{x} = (x_1, \dots, x_n)$. In this case the random variables $X_1, \dots, X_n$ are independent, each with the same double exponential distribution. For every ARUM (1) in this equivalence class, and for every pair of distinct options i and j, it follows from (3) and (12) (by considering p_i and letting $z_k \to -\infty$ for $k \neq i$, $k \neq j$), that the distribution of $X_j - X_i$ is logistic with parameter λ. Thus the variance in the relative evaluation of each such pair of options is the same for all pairs.

It is again well-known that, for $n \geq 3$, the SMLM is the *only* ARUM (identified up to equivalence) with strictly positive choice probability functions $p_1, \dots, p_n$ satisfying the condition that $p_i(\boldsymbol{z})/p_j(\boldsymbol{z})$ is independent of z_k for all k with $k \neq i$, $k \neq j$. (This condition is commonly known as Luce's Choice Axiom (Luce, 1959).) This result was proved by Strauss (1979, Theorem 2). His proof may be simplified as follows. Suppose that the choice probability functions $p_1, \dots, p_n$ satisfy the above condition. Then, using also the condition (C1) of Corollary 3, it follows that for all i, $2 \leq i \leq n$, there exists some function $f_i : \mathbb{R} \to \mathbb{R}$ such that $p_i(\boldsymbol{z})/p_1(\boldsymbol{z}) = f_i(z_i - z_1)$ for all $\boldsymbol{z}$. Thus, from the condition (C2) of Corollary 3,

$$p_i(\boldsymbol{z}) = \frac{f_i(z_i - z_1)}{1 + \sum_{j=2}^{n} f_j(z_j - z_1)}, \qquad \boldsymbol{z} \in \mathbb{R}^n. \tag{13}$$

Using also the condition (C5) for $i > 1$, $j > 1$, $i \neq j$, and putting $z_1 = 0$, we have that $f_i'(z_i)/f_i(z_i) = f_j'(z_j)/f_j(z_j)$ for all z_i, z_j (where $'$ denotes differentiation with respect to a single argument), so that there exists some constant λ such that, for all $i > 1$, $f_i'(z)/f_i(z) = \lambda$. Using (13) again, it follows that p_i is as given by (12) for all $i > 1$, and so also for $i = 1$ by the condition (C2). The condition (C3) ensures that $\lambda > 0$.

As remarked above, the SMLM has the property that, in the representation (1), the distribution of $X_i - X_j$, and in particular the variance of this distribution, is the same for all pairs of distinct options i and j. This property is generally accepted to be unduly restrictive. For example, two options may be perceived as being very similar to each other in comparison with a third. Thus the use of this model may even pose difficulties in the the definition of options. These problems are very considerably eased in the ARUM known as the *nested multinomial logit model* (NMLM), which we now discuss. (See also Daly and Zachary (1978) and McFadden (1978, 1981).) This model is so called because its mathematical expression imposes a formal hierarchical structure on the set of options $\{1, \ldots, n\}$. However we emphasize that it continues to model a simultaneous maximization of utility over this set.

For simplicity we define the model first in the case of three options. Given parameters $\lambda > 0$, $\nu > 0$, define the function $u : \mathbb{R}^3 \to \mathbb{R}$ by

$$u(\boldsymbol{z}) = \theta_{2,\nu}(z_1, \theta_{2,\lambda}(z_2, z_3)). \tag{14}$$

For $\nu = \lambda$ this is the surplus function of the SMLM with parameter λ. For general λ, ν, we may use Theorem 2 to determine whether u remains the surplus function of some ARUM. The condition (S1) of that theorem is easily seen to be satisfied.

Since $\lambda > 0$, $\nu > 0$, the condition (S2) is also readily seen to be satisfied. Finally, from the second of the two forms given for $D_1 u$, $D_2 u$, $D_3 u$ above and from (10), it is easily checked that if $\lambda \geq \nu$ then $D_{123} u$ is non-negative everywhere. An explicit calculation gives

$$\begin{aligned} D_{123} u(\boldsymbol{z}) = \nu \exp\{&-2\nu u(\boldsymbol{z}) + (\nu - 2\lambda)\theta_{2,\lambda}(z_2, z_3) + \nu z_1 + \lambda z_2 + \lambda z_3\} \\ &\times [\lambda - \nu + 2\nu \exp\{-\nu u(\boldsymbol{z}) + \nu \theta_{2,\lambda}(z_2, z_3)\}], \end{aligned}$$

which, for $\lambda < \nu$ and fixed z_2, z_3, becomes negative for all sufficiently large z_1.

It follows that u is the surplus function of a ARUM (the NMLM for three options) if and only if

$$\lambda \geq \nu. \tag{15}$$

When this condition holds the joint distribution of the random variables X_1, X_2, X_3 in the representation (1) is such that, for $j = 2$, 3, $X_j - X_1$ has a logistic distribution with parameter ν (variance $\pi^2/3\nu^2$), while $X_2 - X_3$ has a logistic distribution with parameter λ.

Thus, in the interpretation of the model, for λ strictly greater than ν, the variation in the relative perception of options 2 and 3 is less than that for either of the other two pairs. This model is therefore natural in those applications in which the options 2 and 3 are more closely related to each other than is either to option 1. In extreme situations it may not be clear that the options 2 and 3 should be regarded as distinct. There is no problem with the model here: for fixed ν, as $\lambda \to \infty$, $\theta_{2,\lambda}(z_2, z_3) \to \max(z_2, z_3)$, and the model approaches in the obvious sense a two-option simple logit model with parameter ν in which the choice is between the option 1 and the pair $\{2, 3\}$, the further choice between the options 2 and 3 being made deterministically.

Again, for $\lambda \geq \nu$, it may be helpful to exhibit a particular member of the equivalence class of ARUMs (1) with surplus function u given by (14). One such is the *generalized extreme value model* in which the joint distribution function F of $X_1, \ldots, X_n$ is given by

$$F(\boldsymbol{x}) = \exp[-\exp(-\nu x_1) - \{\exp(-\lambda x_2) + \exp(-\lambda x_3)\}^{\nu/\lambda}]$$

(see McFadden, 1978, 1981, Strauss, 1981). (It may also be verified directly that this is a distribution function if and only if $\lambda \geq \nu$.)

Concluding remarks

Choice models based on the maximization of (random) utilities have considerable logical coherence and are readily interpretable. The preceding discussion has illustrated this in the case of ARUMs. In particular, the nested multinomial logit models considered in the previous section constitute a fairly flexible family of ARUMs, with the same good statistical properties as are possessed by simple logit models for the analysis of binary data (Cox and Snell, 1989). In many situations the most appropriate member of this family will not be (or even be close to) a simple multinomial logit model.

Nevertheless NMLMs continue to place some constraints on the variance structure of the random variables $X_1, ..., X_n$ in the representation (1). For example, in the three-option model discussed in the previous section, the variance of $X_2 - X_1$ is equal to that of $X_3 - X_1$, and is greater than or equal to that of $X_2 - X_3$. These conditions, and in particular the latter, are to some extent natural in many circumstances. (Recall that the square roots of these three variances must at least satisfy the obvious triangle inequality.) However these constraints are not always entirely appropriate. Börsch-Supan (1990) has shown that it is possible to construct ARUMs which have the property that, *for values of their arguments within a restricted range* (which may be sufficient for a particular application), their choice probability functions agree with those of NMLMs, but which require weaker constraints on their parameters than necessarily exist on those of the latter.

It should be remembered, however, that ARUMs form a relatively simple class of random utility models. In particular, for every such model, the utility of each option depends on its observable component only through its location parameter. In applications ARUMs are perhaps best regarded as good first approximations.

Acknowledgement. The author grateful to Tony Marley for help with the background research for this note.

References

[1] Apostol, T. M. (1957). Mathematical Analysis (Addison Wesley, Reading, M.A).

[2] Börsch-Supan, A. (1990). On the compatibility of nested logit models with utility maximization, Journal of Econometrics 43, 373-388.

[3] Bunch, D. S. (1987). Maximum likelihood estimation of probabilistic choice models, SIAM Journal of Scientific and Statistical Computing 8, 56-70.

[4] Cox, D. R. & E. J. Snell. (1989). Analysis of binary data (Chapman and Hall, London).

[5] Daly, A. J. & S. Zachary. (1978). Improved multiple choice models, in: D. A. Henscher & Q. Dalvi, eds., Determinants of travel choice (Teakfield, Hampshire) 335-357.

[6] Harris, A. J. & J. C. Tanner. (1974). Transport demand models based on personal characteristics, Transport and Road Research Laboratory Supplementary Report 65 UC (Berkshire).

[7] Luce, R. D. (1959). Individual choice behavior (Wiley, New York).

[8] Luce, R. D. & P. Suppes. (1965). Preference, utility, and subjective probability, in: R. D. Luce, R. R. Bush & E. Galanter, eds., Handbook of Mathematical Psychology, Vol. III (Wiley, New York).

[9] Mcfadden, D. (1978). Modelling the choice of residential location, in: A. Karlqvist et al., eds., Spatial interaction theory and planning models, (North Holland, Amsterdam) 75-96.

[10] Mcfadden, D. (1981). Econometric models of probabilistic choice, in: D. McFadden & C. Mansky, eds., Structure and analysis of discrete data (MIT Press, Cambridge, M.A.) 198-272.

[11] Stern, H. (1990). A continuum of paired comparisons models, Biometrika 77, 265-273.

[12] Stern, H. (1990). Models for distributions on permutations, Journal of the American Statistical Association 85, 558-564.

[13] Strauss, D. J. (1979). Some results on random utility models, Journal of Mathematical Psychology 20, 35-52.

[14] Strauss, D. J. (1981). Choice by features: an extension of Luce's choice model to account for similarities, British Journal of Mathematical and Statistical Psychology 34, 50-61.

[15] 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.

[16] Yellott, J. I. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution, Journal of Mathematical Psychology 15, 109-144.

[17] Yellott, J. I. (1980). Generalized Thurstone models for ranking: equivalence and reversibility, Journal of Mathematical Psychology 22, 48-69.