SELECTION DISTRIBUTIONS AND DETERMINISTIC REDUCTION OF GENERALIZED RANDOM UTILITY MODELS

N. Kalouptsidis, University of Athens
V. Psaraki, National Technical University of Athens

Conference paper
Session III

Moving through nets:
The physical and social dimensions of travel
10th International Conference on Travel Behaviour Research
Lucerne, 10-15. August 2003

SELECTION DISTRIBUTIONS AND DETERMINISTIC REDUCTION OF GENERALIZED RANDOM UTILITY MODELS

N. Kalouptsidis
Department of Informatics and Telecommunications, Division of Communications and Signal Processing
University of Athens
Athens

Phone: +30-2107275304
Fax: +30-2107275601
eMail: kalou@di.uoa.gr

V. Psaraki
Faculty of Civil Engineering
National Technical University of Athens
Athens

Phone: +30-2107721204
Fax:
eMail: kaloup@otenet.gr

Abstract

A generalized random utility model including latent variables, flexible disturbances and combined revealed and stated preferences is considered. It is shown that the selection distributions governing the choice probabilities in multiple choices are outputs of linear filters and hence can be evaluated by the fast Fourier transform. In the binary case and under the most probable choice rule, maximization of correct classifications is tightly related to maximum score maximization. Parameter estimation is accomplished by algorithms that seek the solution of a maximum number of a given set of linear inequalities. This is extended to multiple responses under proper block iid assumptions on the disturbances, and provided the random parameters are associated solely to choice invariant factors such as individual characteristics.

Keywords

Random Utility Model, Choice Behaviour, Mixed Logic, Latent Variables, Maximum Score Estimation

Preferred citation

N. Kalouptsidis, V. Psaraki (2003) ’SELECTION DISTRIBUTIONS AND DETERMINISTIC REDUCTION OF GENERALIZED RANDOM UTILITY MODELS’, paper presented at the 10 th International Conference on Travel Behaviour Research, Lucerne, August 2003.

1. Introduction

Many popular discrete choice models including multinomial logit, multinomial probit and nested logit models are built upon the random utility maximization (RUM) model (Domencich and Mc Fadden, 1975), (Ben-Akiva and Lerman, 1985), (Ortuzar and Williumsen, 1996), (Greene, 1997), (Bolduc and Ben-Akiva, 1991),(Koppelman and Sethi, 2000), (Mc Fadden, 2000). RUM as a behavioural model follows the economists' theory of consumer behaviour where preferences are represented by utilities and choices are made by utility maximizing calculations. Furthermore, utilities are treated as random rather than deterministic. This is so because features of the taste template are heterogeneous across individuals and hence unknown to the analyst. Likewise unobserved aspects of experience and information on the attributes of alternatives are interpreted as random factors (Mc Fadden, 2000).

The RUM model postulates that an individual n choosing among $1 \leq i \leq k+1$ alternatives, bases his/her decision on the random utilities U_{ni} associated with these alternatives. In particular the alternative with the maximum utility is chosen. The most popular utility representation is the linear in the parameters form

$$U_{ni} = \mathbf{z}_{ni}^T \beta_n + \epsilon_{ni} \tag{1}$$

or in matrix form

$$U_n = Z_n \beta_n + \epsilon_n \tag{2}$$

T denotes matrix transpose. The rows of matrix Z_n, $\mathbf{z}_{ni}^T$ represent the factors that affect utility of individual n for alternative i. β_n designates the vector of weights individual n attaches to these factors. The disturbance term ϵ_n incorporates unobserved characteristics of attributes and of the individual that render utilities unobservable, latent variables. The decision maker n chooses alternative i if

$$U_{ni} \geq U_{nj} \tag{3}$$

for all j in the set of alternatives faced by n. The above choice is not observable by the analyst,who is then content to assess the probabilities of choosing each alternative i by the rule

$$P_n(i; \beta_n) = P[U_{ni} > U_{nj}, \quad j \neq i] \tag{4}$$

Suppose $i(n)$ is the chosen alternative by the individual n. A set of measurements $(Z_n, i(n))$ over a set of N individuals provides a window through which we observe eq.(2)and eq.(3). Parameter estimation is then performed using the above set of measurements and an optimization criterion. The most popular practice rests upon maximization of the log likelihood function

$$L(\beta) = \sum_{n=1}^{N} \log P_n(i(n); \beta_n) \tag{5}$$

An alternative to maximum likelihood optimisation rule considered in this paper, relies on maximization of correct classifications through the most probable alternative rule. Since the analyst has no access to the latent utilities due to the unobserved disturbances he/she cannot deduce the winning alternative $i(n)$ from utility maximization (3). A reasonable guess for the analyst is the most probable choice $\hat{i}(n)$ satisfying

$$P_n(\hat{i}(n); \beta_n) = \max_{1 \leq j \leq k+1} P_n(j; \beta_n) \tag{6}$$

On the basis of the most probable choice rule, a parameter vector β classifies correctly an individual n in the data set if

$$P_n(\hat{i}(n); \beta) = P_n(i(n); \beta) \tag{7}$$

that is

$$P_n(\hat{i}(n); \beta) > P_n(j; \beta), \qquad 1 \leq j \leq k+1, \qquad j \neq i(n) \tag{8}$$

Parameter estimation is then performed by maximizing the number of correct classifications in a given data set. This approach is tightly related to maximum score estimation studied and made operative by (Manski, 1975; 1986). Both maximum likelihood and maximization of correct classifications based on the most probable alternative are discussed in subsequent sections.

The model capability to approximate the real choice process and to improve reliability of the resulting forecasts is enhanced if the role of the disturbances, the factors and the measurements are further clarified. This is done in the next section along the lines of the integrated choice model developed in (Walker and Ben Akiva, 2002). Section 3 defines the selection distributions. Section 4 deals with the binary choice case. Section 5 discusses multiple choices. Conclusions are summarized in section 6.

2. Model specification

Model specification results from the split of roles of the major actors involved in the choice equations. The linear form is maintained throughout. Measurements combine revealed preferences (RP) and stated preferences (SP). Disturbances are split into flexible and core disturbances. Flexible disturbances are attributed to random parameter variations. Factors are divided into observable and latent variables.

2.1 Combining revealed preferences and stated preferences

The generalized random utility model combines revealed measurement data $(Z_n^R, i^R(n))$ and a set of stated preference measurements $(Z_n^{S_j}, i^{S_j}(n))$, $n = 1, 2, \ldots, N$, $j = 1, 2, \ldots, S$. Revealed preferences indicate market behaviour. Additional surveys in the form of stated preferences significantly enhance the estimation process and provide information on the nature of the unobservable elements of the choice process. Using the approach suggested by (Morikawa, 1994), eq.(2) is decomposed as

$$U_n^R = Z_n^R \beta_n + \psi^R e^{RS} + \mathbf{e}_n^R \tag{9}$$
$$U_n^{S_j} = Z_n^{S_j} \beta_n + \psi^S e^{RS} + \mathbf{e}_n^{S_j} \qquad j = 1, 2, \ldots, S \tag{10}$$

The single RP survey is complemented by S stated preference surveys. e_n^{RS} is a scalar random variable that is constant across responses for a single individual. It accounts for the correlation across multiple responses for a given individual. ψ^R and ψ^S are deterministic vectors and account for the magnitude of the correlation effect. The main parameters β_n remain the same in both utility expressions.

2.2 Factors

Factors are split into observable explanatory variables and latent explanatory variables. The latent variables are in turn divided into two groups. The first group of latent variables incorporates psychological factors affecting utilities, such as attitudes and perception. Comfort and convenience are typical psychological factors encountered in transportation demand modelling. Following (Walker and Ben Akiva, 2002) we assume that factors in this first group are partially explained by observable attributes of the alternatives and characteristics of the individual through the linear additive model

$$x_{nil}^* = \mathbf{x}_{nil}^{lT} \lambda_l + e_{nil}^l \qquad l = 1, 2, \ldots, L \tag{11}$$

The latter equation expresses each of the L latent variables x_{nil}^* as a function of observable explanatory variables $\mathbf{x}_{nil}^l$, an individual and choice invariant parameter vector λ_l and a disturbance term e_{nil}^l. The superscript l emphasizes the latent nature of the variable. Collecting all latent variables in a single vector we obtain

$$\mathbf{x}_{ni}^{*T} = \mathbf{x}_{ni}^{lT} \mathbf{\Lambda} + e_{ni}^{lT} \tag{12}$$

where Λ is the block diagonal matrix

$$\Lambda = \mathrm{diag}\,(\,\lambda_1, \lambda_2, \ldots, \lambda_L\,)$$

The second group of latent variables includes attributes of alternatives not actually chosen by the individual and for which information is gathered through possibly separate surveys. Travel times for example exhibit random fluctuations due to varying traffic conditions. For each individual n and alternative i, the set of choice specific latent attributes is given by

$$\tilde{\mathbf{x}}_{ni} = \bar{\mathbf{x}}_{ni} + \mathbf{e}_{ni}^x \tag{13}$$

$\bar{\mathbf{x}}_{ni}$ is the mean attribute vector for alternative i. The random vector $\mathbf{e}^x_{ni}$ represents the random fluctuation of the factor $\tilde{\mathbf{x}}_{ni}$ from its mean. The superscript x signifies the difference of the attribute error $\mathbf{e}^x_{ni}$ from the error $\mathbf{e}_{ni}$ in the utility specification. It may occur that for certain alternative j some of the components of the factor vector $\tilde{\mathbf{x}}_{ni}$ are deterministic. The travel cost of a transit mode is an example. This is accommodated by assuming that the variance of the corresponding error component is zero. $\bar{\mathbf{x}}_{ni}$ can be estimated by averaging over SP data. In the absence of adequate data, it can be approximated by means of subpopulation classes. Collecting equations (12) for all alternatives we obtain the matrix form

$$X^*_n = X^l_n \Lambda + E^l_n \tag{14}$$

X^*_n and E^l_n are $(k+1) \times L$ matrices. Likewise eq.(13) is written in matrix form as

$$\tilde{X}_n = \bar{X}_n + E^x_n \tag{15}$$

On the basis of the above discussion the factors are decomposed as follows

$$Z^R_n = (\, X^{Ro}_n, \tilde{X}^R_n, X^*_n \,) \tag{16}$$

where the matrix X^{Ro}_n consists of the observable explanatory variables. We assume that the same latent variables enter both the SP and RP measurements. Therefore

$$Z^{S_j}_n = \left(X^{S_j o}_n, \tilde{X}^{S_j}_n, X^*_n \right) \tag{17}$$

In correspondence with the above equations we partition the parameter vector β_n as

$$\beta^T_n = (\, \beta^{oT}_n, \tilde{\beta}^T, \beta^{*T} \,) \tag{18}$$

where

- β^o_n corresponds to the observable explanatory variables.
- $\tilde{\beta}$ weighs the random and choice specific latent factors. It is choice invariant and individual invariant deterministic vector.
- β^* corresponds to the latent variables. It is choice invariant and individual invariant deterministic vector.

If we substitute equations (16),(17) and (18) into the main utility equations (9), (10) we obtain

$$\begin{aligned} U^R_n &= X^{Ro}_n \beta^o_n + \bar{X}^R_n \tilde{\beta} + E^{Rx}_n \tilde{\beta} + X^l_n \Lambda \beta^* + E^l_n \beta^* + \psi^R e^{RS}_n + \mathbf{e}^R_n \\ U^{S_j}_n &= X^{S_j o}_n \beta^o_n + \bar{X}^{S_j}_n \tilde{\beta} + E^{S_j x}_n \tilde{\beta} + X^l_n \Lambda \beta^* + E^l_n \beta^* + \psi^{S_j} e^{RS}_n + \mathbf{e}^{S_j}_n \end{aligned}$$

2.3 Disturbances

Disturbances are split into flexible disturbances and core disturbances with distinct role assignments each. In this paper flexible disturbances are generated by the random nature of some of the parameters (Brownstone and Train, 1999), (Mc Fadden and Train, 2000), (Walker and Ben-Akiva, 2002). By the previous discussion, the random parameters are embodied in the β^o_n part. Hence we consider the split

$$\beta^o_n = \begin{pmatrix} \beta^d \\ \beta^r_n \end{pmatrix} \tag{19}$$

The block β^d comprises of individual and choice invariant deterministic parameters. The remaining block is random and individual specific with mean $\bar{\beta}$. The variation from the mean

$$\mathbf{e}^\beta_n = \beta^r_n - \bar{\beta}$$

is characterized by the probability density function $f^\beta(x)$. In accordance with the above parameter split, we further partition the observable explanatory variables as

$$\begin{aligned} X^{Ro}_n &= (\, X^{Rd}_n, X^{Rr}_n \,) \\ X^{S_j o}_n &= \left(X^{S_j d}_n, X^{S_j r}_n \right) \end{aligned}$$

where X^{Rd}_n corresponds to the deterministic parameters and X^{Rr}_n corresponds to the random parameters. Similar interpretations hold for the SP attributes.

2.4 Generalized utility model

Taking into account the combination of the RP and SP measurements, the latent variables and the random parameters, the main utility equation is reformulated more compactly as

$$U_n^R = X_n^R \gamma + Y_n^R(\gamma)\mathbf{w}_n^R = X_n^R \gamma + \epsilon_n^R \tag{20}$$

where

$$\begin{aligned} X_n^R &= (\, X_n^{Rd}, X_n^{Rr}, \bar{X}_n^R, X_n^l \,) && (21) \\ \gamma^T &= (\, \beta^{dT}, \bar{\beta}^T, \tilde{\beta}^T, (X_n^l \Lambda \beta^*)^T \,) && (22) \\ Y_n^R(\gamma) &= (\, \psi^R, X_n^{Rr}, \tilde{\beta}^T, \beta^{*T}, I \,) && (23) \\ \mathbf{w}_n^{RT} &= (\, e_n^{RS}, \mathbf{e}_n^{\beta T}, \mathbf{e}_n^{x_1 RT}, \ldots, \mathbf{e}_n^{x_M RT}, \mathbf{e}_n^{l_1 T}, \ldots, \mathbf{e}_n^{l_L T}, \mathbf{e}_n^{RT} \,) && (24) \end{aligned}$$

The matrix X_n^R consists of observable covariates. The parameter γ is directly obtained from β. Identifiability requires additional constraints to ensure that β^* is recovered from $X_n^l \Lambda \beta^*$. The generalized disturbance is factored as

$$\epsilon_n^R = Y_n^R(\gamma)\mathbf{w}_n^R \tag{25}$$

The matrix $Y_n^R(\gamma)$ is deterministic and depends on the parameters as well as on explanatory variables. I denotes the identity matrix. The stochastic part of the generalized error is described by the random vector $\mathbf{w}_n^R$ of dimension $n(\beta) + M + L + k + 2$ where

- $n(\beta)$ is the number of stochastic parameters
- $M + L$ is the number of the two groups of latent variables
- $k + 1$ is the number of alternatives

SP utilities are similar with R replaced by S.Thus

$$U_n^{S_j} = X_n^{S_j}\gamma + Y_n^{S_j}(\gamma)\mathbf{w}_n^{S_j} \tag{26}$$

where

$$\begin{aligned} X_n^{S_j} &= \left(\, X_n^{S_j d}, X_n^{S_j r}, \bar{X}_n^{S_j}, X_n^l \,\right) \\ Y_n^{S_j}(\gamma) &= \left(\, \psi^{S_j}, X_n^{S_j r}, \tilde{\beta}^T, \beta^{*T}, I \,\right) \\ \mathbf{w}_n^{S_j} &= \left(\, e_n^{RS}, \mathbf{e}_n^{\beta T}, \mathbf{e}_n^{x_1 S_j T}, \ldots, \mathbf{e}_n^{x_M S_j T}, \mathbf{e}_n^{l_1 T}, \ldots, \mathbf{e}_n^{l_L T}, \mathbf{e}_n^{S_j T} \,\right) \end{aligned}$$

We shall assume throughout that all blocks comprising $\mathbf{w}_n^R$ and $\mathbf{w}_n^{S_j}$ are jointly independent. Furthermore each subblock vector is independent but not necessarily identically distributed. Thus the generalized disturbance is decomposed into independent blocks of different significance. The error $\mathbf{e}_n^R$ characterizes unobserved attributes affecting choice, measurement error and functional misspecification. Unobserved tastes of the individual, intra-individual variations including different states of mind, different consumption occasions and intra-personal dynamics such as state dependence and short run preference inertia are to some extent captured by the remaining blocks. The same error blocks cope with inter-individual differences in utility. e_n^{RS} describes correlation between hypothetical and actual behavior.

Using the above generalized linear choice model we derive next expressions for the choice probabilities.

3. Selection Distributions

Let us first consider the error term (25). For each alternative $i = 1, 2, \ldots, k + 1$ we define the random vector

$$\xi_{ni}^{RT} = (\, \epsilon_{n1}^R - \epsilon_{ni}^R, \epsilon_{n2}^R - \epsilon_{ni}^R, \ldots, \epsilon_{ni-1}^R - \epsilon_{ni}^R, \epsilon_{ni+1}^R - \epsilon_{ni}^R, \ldots, \epsilon_{nk+1}^R - \epsilon_{ni}^R \,) \tag{27}$$

Let $f_{ni}^{\xi R}(\mathbf{x};\gamma)$ be the corresponding probability density function and $F_{ni}^{\xi R}(\mathbf{x};\gamma)$ the probability distribution. Note that

$$\sum_{i=1}^{k+1} F_{ni}^{\xi R}(\mathbf{x};\gamma) = 1 \tag{28}$$

Consider the 'systematic' part of the generalized utility expression (20)

$$V_n^R = X_n^R \gamma \tag{29}$$

with components $V_{n,i}^R$, $i = 1, 2, \ldots, k+1$. For each alternative i define the vector of utility differences

$$\mathbf{v}_{ni}^{RT} = (\, V_{ni}^R - V_{n1}^R, \ldots, V_{ni}^R - V_{ni-1}^R, V_{ni}^R - V_{ni+1}^R, \ldots, V_{ni}^R - V_{nk+1}^R \,)$$

Each vector $\mathbf{v}_{ni}^R$ is readily determined once the parameter vector γ given by eq.(22) is computed. In fact,

$$\mathbf{v}_{ni}^R = -A_i V_n^R = -A_i X_n^R \gamma \tag{30}$$

where

$$A_i = \begin{pmatrix} I_{i-1} & -\mathbf{1_{i-1}} & 0_{i-1,k+1-i} \\ 0_{k+1-i,i-1} & -\mathbf{1}_{k+1-i} & I_{k+1-i} \end{pmatrix}, \qquad i = 1, \ldots, k+1 \tag{31}$$

I_q is the $q \times q$ identity matrix, $0_{q\times r}$ is the $q \times r$ zero matrix and $\mathbf{1}_q$ a vector of length q consisting of 1's.

With the above definitions, equation (4) takes the form

$$P_n^R(i;\beta_n) = F_{ni}^{\xi R}(\mathbf{v}_{ni}^R;\gamma) = F_{ni}^{\xi R}(-A_i X_n^R \gamma;\gamma) \tag{32}$$

Similar expressions are valid for the SP parameters.

The probability distributions $F_{ni}^{\xi R}$ and $F_{ni}^{\xi S_j}$ are referred to as *selection distributions*. They are not available as they depend on unknown parameters as well as on the distribution of the disturbances. Additional information on the structure of the selection distributions is extracted in the sequel. The analysis is simpler in the binary case of two alternatives. This is presented in the next section.

4. Binary choice

In the binary case, the two selection distributions are related by $F_{n2}^{\xi R} = 1 - F_{n1}^{\xi R}$. Moreover the random vector ξ_{n1}^R becomes the scalar random variable

$$\begin{aligned} \xi_n^R \quad = \quad & (\psi_1^R - \psi_2^R) e_n^{RS} + (\mathbf{x}_{n1}^{RrT} - \mathbf{x}_{n2}^{RrT})\mathbf{e}_n^\beta + \tilde{\beta}^T(\mathbf{e}_{n1}^x - \mathbf{e}_{n2}^x) + \\ & \beta^{*T}(\mathbf{e}_{n1}^l - \mathbf{e}_{n2}^l) + e_{n1}^R - e_{n2}^R \end{aligned} \tag{33}$$

The difference $\mathbf{e}_{n1}^x - \mathbf{e}_{n2}^x$ is an M dimensional vector. It consists of the differences of the two entries of each 2×1 dimensional vector $\mathbf{e}^{x_j}$, $j = 1, 2, \ldots, M$. $\mathbf{e}_{n1}^l - \mathbf{e}_{n2}^l$ is similarly defined.

The following table summarizes the notation we shall employ. Given a function f, let $\hat{f}$ denote the Fourier transform of f. The first column contains the disturbance blocks of $\mathbf{w}_n^R$. The second column contains corresponding probability density functions. The third column includes the terms of ξ_n^R in eq.(33). The last column indicates the characteristic function of the random terms in column 3.

The computations make use of the following well-known facts from probability theory.

1. Suppose Y is a linear combination of random variables X_j:

$$Y = \sum_j \beta_j X_j = \beta^T X$$

The characteristic function of Y, $\hat{f}_Y(\omega)$, namely the Fourier transform of the probability density function $f_y(-x)$ is given by

$$\hat{f}_Y(\omega) = \prod_j \hat{f}_{X_j}(\beta_j \omega)$$

Table 1: Disturbance characteristics and notation

Disturbance	*pdf*	*Terms in* ξ	*Characteristic function*
e_n^{RS}	f^{RS}	$\psi_1^R - \psi_2^R)e_n^{RS}$	$\hat{f}^{RS}((\psi_1^R - \psi_2^R)\omega)$
$\mathbf{e}_{n,j}^{\beta}$, j-entry of $\mathbf{e}_n^{\beta}$	f_j^{β}	$(\mathbf{x}_{n1}^{RrT} - \mathbf{x}_{n2}^{RrT})\mathbf{e}_n^{\beta}$	$\prod_j \hat{f}_j^{\beta}((\mathbf{x}_{n1,j}^{Rr} - \mathbf{x}_{n2,j}^{Rr})\omega)$
$\mathbf{e}_{n1j}^{x}$, j-entry of $\mathbf{e}_{n1}^{x}$	f_{1j}^x, f_{2j}^x	$\tilde{\beta}^T(\mathbf{e}_{n1}^x - \mathbf{e}_{n2}^x)$	$\prod_j \hat{f}_{2j}^x(-\tilde{\beta}_j\omega)\hat{f}_{1j}^x(\tilde{\beta}_j\omega)$
e_{nj}^l	f_{1j}^l, f_{2j}^l	$\beta^{*T}(\mathbf{e}_{n1}^l - \mathbf{e}_{n2}^l)$	$\prod_j \hat{f}_{2j}^l(-\beta_j^*\omega)\hat{f}_{1j}^l(\beta_j^*\omega)$
e_{nj}^R	f_{1j}^R, f_{2j}^R	$e_{n1}^R - e_{n2}^R$	$\hat{f}_2^R(-\omega)f_1^R(\omega)$

2. Pick any one of X_j, say X_k. Let F_{X_k} denote the pdf of X_k and $\hat{F}_{X_k}$ its Fourier transform. Then
$$\hat{F}_Y(\omega) = \prod_{j=1}^{k-1} \hat{f}_{X_j}(\beta_j\omega)\hat{F}_{\beta_k X_k}(\omega)$$

3. Consider the difference $Y = X_1 - X_2$. Then
$$\hat{f}_Y(\omega) = \hat{f}_{X_1}(\omega)\hat{f}_{X_2}(-\omega) \qquad \hat{F}_Y(\omega) = \hat{f}_{X_2}(-\omega)\hat{F}_{X_1}(\omega)$$

The selection distribution is determined from the product of the elements of the last column of the above table and property 3. The independence assumptions imposed on the error $\mathbf{w}_n^R$ imply

$$\begin{aligned}\hat{F}_{n1}^{\xi R}(\omega) \quad = \quad & \hat{f}^{RS}((\psi_1^R - \psi_2^R)\omega)\prod_j \hat{f}_j^{\beta}((\mathbf{x}_{n1,j}^{Rr} - \mathbf{x}_{n2,j}^{Rr})\omega) \times \\ & \prod_j \hat{f}_{2j}^x(-\tilde{\beta}_j\omega)\hat{f}_{1j}^x(\tilde{\beta}_j\omega) \times \\ & \prod_j \hat{f}_{2j}^l(-\beta_j^*\omega)\hat{f}_{1j}^l(\beta_j^*\omega)\hat{f}_2^R(-\omega)f_1^R(\omega)\end{aligned} \tag{34}$$

If we take the inverse Fourier transform we express the desired selection distribution as a convolution product. Similar expressions hold for the SP measurements.

Equations (34) and (32) together with their SP counterparts can be employed in the maximization of the log likelihood as an alternative to maximum simulated likelihood (MSL) algorithms. It is well known (Ben-Akiva and Lerman, 1985) that descent maximization algorithms require at each step the valuation of the choice probabilities and the derivatives of the choice probabilities with respect to parameters. For the generalized choice model under consideration, choice probabilities are calculated via successive conditioning over the probability density of the latent variables, the flexible disturbances and the combined RP/SP disturbance (Walker and Ben-Akiva, 2002). These calculations involve high dimensional integrals. As a consequence when the number of parameters plus the number of latent variables is not small, considerable computational burden is faced. For this reason these integrals are simulated with sample averages where draws are selected from the underlying densities (Hajivassiliou et al, 1996), (Hensher and Greeene, 2003). The production of draws is based on Monte Carlo pseudo random sequences or quasi random sequences (for example Halton draws); for a discussion see (Bhat, 2003). A binary choice study is undertaken in (Walker and Ben Akiva, 2002).

Alternatively, taking into account the special structure of the utilities (20) and the block independence of the disturbances, valuation of the choice probabilities can be effected via the selection distributions $F_{n1}^{\xi R}$, $1 - F_{n1}^{\xi R}$, $F_{n1}^{\xi S_j}$, $1 - F_{n1}^{\xi S_j}$ and eq.(34). The passage to and from the frequency domain is accomplished by a fast Fourier transform (FFT), see for instance, (Kalouptsidis, 1997). If the densities of the elements of the disturbance vector $\mathbf{w}_n^R$ have a known parametrized functional form, the characteristic functions can be precomputed. Then the likelihood maximization algorithm computes only the arguments of these functions at each step, in accordance with (34). Alternatively, non parametric estimators for the characteristic functions can by used.

Let us next consider the most probable alternative rule and maximization of correct classifications. Application of the MPA rule on the revealed preference data set yields

$$P_n^R(1;\beta_n) > P_n^R(2;\beta_n) = 1 - P_n^R(1;\beta_n)$$

if and only if

$$P_n^R(1;\beta_n) = F_{ni}^{\xi R}(-A_1 X_n^R \gamma;\gamma) > 1/2 \tag{35}$$

Let $\delta_n^R(\gamma_n)$ denote the median of the selection distribution $F_{n1}^{\xi R}$:

$$F_{n1}^{\xi R}(\delta_n^R(\gamma_n)) = 1/2$$

Typically, $F_{n1}^{\xi R}$ is strictly increasing. We shall further assume that the median is constant over the sample population. Thus (35) holds if and only if

$$\mathbf{v}_{n1}^R > \delta_n^R(\gamma)$$

or

$$\begin{gathered}(X_{n1}^{Rd} - X_{n2}^{Rd})^T \beta^d + (X_{n1}^{Rr} - X_{n2}^{Rr})^T \bar{\beta} + \\ (\bar{X}_{n1}^R - \bar{X}_{n2}^R)^T \tilde{\beta} + (X_{n1}^l - X_{n2}^l)^T (X_n^l \Lambda \beta^*) > \delta^R(\gamma)\end{gathered} \tag{36}$$

In much the same way the stated responses of individual n in conjunction with the most probable alternative rule lead to the inequality

$$\begin{gathered}(X_{n1}^{S_j d} - X_{n2}^{S_j d})^T \beta^d + (X_{n1}^{S_j r} - X_{n2}^{S_j r})^T \bar{\beta} + \\ (\bar{X}_{n1}^{S_j} - \bar{X}_{n2}^{S_j})^T \tilde{\beta} + (X_{n1}^l - X_{n2}^l)^T (X_n^l \Lambda \beta^*) > \delta^{S_j}(\gamma)\end{gathered} \tag{37}$$

The collection of the above inequalities for all individuals in the given data set must be solved with respect to β^d, $\bar{\beta}$ $\tilde{\beta}$, β^* and $\delta^R(\gamma)$, $\delta^{S_j}(\gamma)$.

If the random blocks e_n^{RS}, e_n^{β}, $\mathbf{e}_{n1}^x - \mathbf{e}_{n2}^x$, $\mathbf{e}_{n1}^l - \mathbf{e}_{n2}^l$, $e_{n1}^R - e_{n2}^R$ are independent and each has a symmetric probability density function, it holds $\delta^R(\gamma) = 0$ and likewise for $\delta^{S_j}(\gamma)$ and the SP data. In this case the disturbance statistics do not influence the most probable rule. Instead, random utility maximization reduces to a deterministic utility maximization in the sense that $P_n(1) > P_n(2)$ if and only if $V_{n1}^R > V_{n2}^R$ and $V_{n1}^{S_j} > V_{n2}^{S_j}$ where

$$\begin{aligned} V_{n1}^R &= X_{n1}^{Rd}\beta^d + X_{n1}^{Rr}\bar{\beta} + \bar{X}_{n1}^R\tilde{\beta} + X_{n1}^l \Lambda\beta^* \\ V_{n2}^R &= X_{n2}^{Rd}\beta^d + X_{n2}^{Rr}\bar{\beta} + \bar{X}_{n2}^R\tilde{\beta} + X_{n2}^l \Lambda\beta^* \end{aligned}$$

$V_{n1}^{S_j}$, $V_{n2}^{S_j}$ are similarly defined. The above expressions are linear inequalities with unknowns β^d, $\bar{\beta}$, $\tilde{\beta}$ and $\Lambda\beta^*$, comprising the parameter γ. If the above inequalities have a solution, variants of the simplex method or the perceptron algorithm can be applied. Such solution rarely exists. Then we seek for algorithms that find parameters satisfying the largest number of the given set of inequalities. Such algorithms are described in (Manski, 1975; 1986) and (Kalouptsidis et al, 2002).

5. Multiple Choice

If more than two choices occur, the selection distributions are functions of several variables and their relationship is more complicated. This is the subject of this section. It is established that the selection distributions are obtained as outputs of linear filters whose input is one of them. Then the advantages of this interrelationship in maximum likelihood as well as maximum score computations are demonstrated.

The selection distribution $F_{ni}^{\xi R}$ is the probability distribution function of the random vector ξ_{ni}^R. ξ_{ni}^R is obtained from ξ_{n1}^R via the linear transformation

$$\xi_{ni}^R = L_i \xi_{n1}^R \tag{38}$$

where L_i is a $k \times k$ invertible triangular matrix.

Let $\hat{f}^R_{ni}(\omega)$ be the characteristic function of ξ^R_{ni}. Then

$$\hat{f}^R_{ni}(\omega) = E[e^{j\omega^T \xi^R_{ni}}] = E[e^{j\omega^T L_i \xi^R_{n1}}] = \hat{f}^R_{n1}(L_i^T \omega) \tag{39}$$

Let $u_s(t)$ denote the unit step function: $u_s(t)$ is 1 for $t \geq 0$ and 0 for $t < 0$. The multidimensional unit step is defined as

$$u_s(t_1, t_2, \ldots, t_k) = u_s(t_1)u_s(t_2)\cdots u_s(t_k).$$

The above function is separable and its Fourier transform is

$$\hat{u}_s(\omega_1, \omega_2, \ldots, \omega_k) = \hat{u}_s(\omega_1)\hat{u}_s(\omega_2)\cdots\hat{u}_s(\omega_k).$$

It is well known that

$$\hat{u}_s(\omega) = \frac{1}{j\omega} + \pi\delta(\omega),$$

where $\delta(\omega)$ is the Dirac distribution.

Using $u_s(t_1, t_2, \ldots, t_k)$, the distribution can be obtained from the density via the convolution

$$F(t_1, t_2 \ldots, t_k) =$$

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(t_1, t_2, \ldots, t_k) u_s(t_1 - s_1, t_2 - s_2, \ldots, t_k - s_k) ds_1 ds_2 \cdots ds_k$$

Passing to the frequency domain, convolution converts to pointwise multiplication. Therefore,

$$\hat{F}(\omega_1, \omega_2, \ldots, \omega_k) = \hat{f}(\omega_1, \omega_2, \ldots, \omega_k)\hat{u}_s(\omega_1, \omega_2, \ldots, \omega_k) \tag{40}$$

Hence

$$\frac{\hat{F}^{\xi R}_{nj}(\omega)}{\hat{F}^{\xi R}_{n1}(\omega)} = \frac{\hat{f}^{\xi R}_{nj}(\omega)}{\hat{f}^{\xi R}_{n1}(\omega)} = \frac{\hat{f}^{\xi R}_{n1}(L_j^T \omega)}{\hat{f}^{\xi R}_{n1}(\omega)} \tag{41}$$

Therefore

$$\hat{F}^{\xi R}_{nj}(\omega) = \frac{\hat{f}^{\xi R}_{n1}(L_j^T \omega)}{\hat{f}^{\xi R}_{n1}(\omega)} \hat{F}^{\xi R}_{n1}(\omega) \tag{42}$$

The latter equation states that all *selection* distributions are obtained if one of them, say $\hat{F}^{\xi R}_{n1}(\omega)$, is filtered by a multidimensional linear filter with frequency response $\hat{f}^{\xi R}_{n1}(L_j^T \omega)/\hat{f}^{\xi R}_{n1}(\omega)$. The characteristics of this frequency response are solely determined by the selection distribution $F^{\xi R}_{n1}$.

Next we express $F^{\xi R}_{n1}$ in terms of the disturbance statistics. Recall that F^R_{n1} is the distribution of the random vector ξ^R_{n1}. Taking into account eqs (23)-(25), ξ^R_{n1} is written as the sum of 5 main terms:

$$\xi_1 = A_1 \psi e^{RS} + A_1 X^r \mathbf{e}^{\beta} + \sum_j \tilde{\beta}_j A_1 \mathbf{e}^{x_j} + \sum_j \beta_j^* A_1 \mathbf{e}^{l_j} + A_1 \mathbf{e} \tag{43}$$

where, the indices R, γ and n are omitted for simplicity. The five random vectors appearing in the right hand side of the above expression are independent because they are functions of the respective independent disturbances. Therefore the characteristic function of ξ^R_{n1} is the product of the characteristic functions of these five random vectors. Furthermore each one of them is obtained by a linear transformation of the respective disturbance. Hence the following factorisation results.

$$\begin{aligned} \hat{f}^{\xi}_1(\omega) \quad = \quad & \hat{f}^{RS}(\psi^T A_1^T \omega)\hat{f}^{\beta}(X^{rT} A_1^T \omega) \prod_{j=1}^{M} \hat{f}^{x_j}(\tilde{\beta}_j A_1^T \omega) \times \\ & \prod_{j=1}^{L} \hat{f}^{l_j}(\beta_j^* A_1^T \omega) \hat{f}^{e}(A_1^T \omega) \end{aligned}$$

If each of the contributing disturbance vectors is independent, the following more refined factorisation obtains

$$\begin{aligned} \hat{f}^{\xi R}_{n1}(\omega) &= \hat{f}^{RS}(\psi^T A_1^T \omega) \prod_j \hat{f}^{\beta}_j (\Delta x^{RrT}_{nj} A_1^T \omega) \times \\ & \prod_{j=1}^{M} \hat{f}^{x_j}_1 (\tilde{\beta}_j (-\sum_i \omega_i)) \prod_{m=1}^{M} \hat{f}^{x_j}_{m+1} (\tilde{\beta}_j \omega_m) \times \\ & \prod_{j=1}^{M} \hat{f}^{l_j}_1 (\beta^*_j (-\sum_i \omega_i)) \prod_{m=1}^{M} \hat{f}^{l_j}_{m+1} (\beta^*_j \omega_m) \hat{f}^e_1 (-\sum \omega_i) \times \\ & \prod_m \hat{f}^e_{m+1}(\omega_m) \end{aligned} \tag{44}$$

If a disturbance is an iid vector, further simplications result since the individual factors are identical: $\hat{f}^b_j = \hat{f}^b$, $\hat{f}^{x_j}_m = \hat{f}^{x_j}$, $\hat{f}^{l_j}_m = \hat{f}^{l_j}$ and $\hat{f}^{ej} = \hat{f}^e$.

Similar expressions are valid for the SP measurements and the respective selection distributions.

It follows from the above analysis that in analogy with the binary case, only one dimensional Fourier transforms are involved in the calculation of the choice probabilities. The various correlation and heterogeneity effects are immersed into the frequency variables solely. This fact facilitates computations in the frequency domain.

5.1 Equality of the selection distributions

In this paragraph we explore conditions under which the selection distributions coincide. As in the previous subsection we omit the indices R, γ and n. We deduce from eq. (42) that $F^{\xi}_j = F^{\xi}_1$ for all j if

$$\hat{f}^{\xi}_1 (L_j^T \omega) = \hat{f}^{\xi}_1 (\omega) \qquad j = 2, \dots, k+1 \tag{45}$$

According to the factorisation formula (44) the latter holds if it is satisfied for each disturbance factor. More specifically, consideration of the combined RP/SP term yields

$$\hat{f}^{RS}(\psi^T A_1^T L_j^T \omega) = \hat{f}^{RS}(\psi^T A_1^T \omega)$$

or

$$L_j A_1 \psi = A_1 \psi$$

which in effect states that the RP/SP correlation vector ψ is choice invariant. This means that with the exception of one, all components of ψ are equal. Consideration of the flexible disturbance term leads to

$$L_j A_1 X^r = A_1 X^r$$

This constraint states that all attributes included in X^r, namely attributes associated with random parameters, must be choice invariant, such as the characteristics of indidual. The remaining three factors, namely random attributes, latent variables and core disturbances conform to eq.(45) if each $\mathbf{e}^{x_j}$, $\mathbf{e}^{l_j}$ and $\mathbf{e}$ are iid vectors.

5.2 Most probable choice rule and deterministic reduction

The most probable rule (6) and maximization of correct classifications for multiple discrete responses is considered next. We shall demonstrate that under the iid assumptions that ensure equality of the selection distributions, the most probable rule reduces to the deterministic maximization rule

$$V^R_{ni^*(n)} = \max_{1 \le i \le k+1} V^{S_j}_{ni} \tag{46}$$

$$V^{S_j}_{ni^*(n)} = \max_{1 \le i \le k+1} V^{S_j}_{ni} \tag{47}$$

where V_{ni} is the systematic utility given by eq. (29).

Suppose $F_{ni}^{\xi R} = F_{n}^{\xi R} = F$ for all alternatives i. As before, we skip R, n and γ to simplify notation. We shall show that

$$\hat{i}(n) = i^*(n)$$

Indeed, let $\mathbf{X}$ be a random vector with probability distribution function given by the selection distribution F. Such random vector always exists. Let $1 \leq j \leq k+1$. Then the event

$$[\mathbf{X} \leq \mathbf{v}_j] = \{\omega \in \Omega : X_m(\omega) \leq V_j - V_m, \quad m \neq j\}$$

is contained in the event

$$[\mathbf{X} \leq \mathbf{v}_{i^*(n)}] = \{\omega \in \Omega : X_m(\omega) \leq V_{i^*(n)} - V_m, \quad m \neq i^*(n)\}$$

Indeed, for any $m \neq i^*(n), j$,

$$V_j - V_m \leq V_{i^*(n)} - V_m$$

because of (47). For the same reason

$$V_j - V_{i^*(n)} \leq V_{i^*(n)} - V_j$$

Therefore

$$P(j) = F(\mathbf{v}_j) = P[\mathbf{X} \leq \mathbf{v}_j] \leq P[\mathbf{X} \leq \mathbf{v}_{i^*(n)}] = F(\mathbf{v}_{i^*(n)}) = P(i^*(n))$$

The latter inequality implies $\hat{i}(n) = i^*(n)$, proving the claim.

We infer from the above analysis that if the selection distributions coincide the most probable rule (6) becomes equivalent to the deterministic utility maximization rule (47). A consequence of the above equivalence is that the selection distribution F and hence the probabilistic structure of the block iid disturbances has no impact on the winning alternative. The resulting situation relates to maximum score estimation formulated in the important papers by Manski, see (Manski, 1975, 1986). As in the binary case parameter estimation can be accomplished by solving linear inequalities. If a true parameter vector exists, which is rarely the case, the most probable rule leads to the set of linear inequalities

$$[X_{ni^R(n)}^{RT} - X_{nj}^{RT}]\gamma > 0 \tag{48}$$

$$[X_{ni^{S_j}(n)}^{S_jT} - X_{nm}^{S_jT}]\gamma > 0 \tag{49}$$

$n = 1, 2, \ldots, N$, $j, m = 1, 2, \ldots, k+1$, $j \neq i^R(n)$, $m \neq i^{S_j}(n)$. Variants of the simplex method, the perceptron algorithm and several other schemes can be employed for the solution. If no true parameters exist, maximization of correct classifications is applicable. The parameter γ that satisfies the maximum number of the above inequalities is seeked. The maximum score algorithm proposed by (Manski 1975, 1986) applies. Alternative schemes are derived in (Kalouptsidis et al, 2003).

6. Conclusions

In this paper the generalized choice model including random parameters, latent variables and combined RP/SP preferences has been considered. The selection distributions governing the choice probabilities have been analysed in the frequency domain. The resulting expressions in conjunction with the fast Fourier Transform offer computational advantages in likelihood maximization algorithms. The selection distributions coincide under block iid assumptions on the disturbances. Then the most probable rule reduces to a deterministic utility maximization problem, closely related to maximum score estimation. Deterministic utility maximization can be approached by algorithms solving linear inequalities. Analysis and assessment of the proposed methods in a proper experimental setup is under way. Extensions dealing with ordered choices as well as the possibility that only a subset of the parameters is common to both RP and SP surveys are being studied.

References

[1] Ben-Akiva, M. and Lerman, R. (1985) *Discrete choice analysis*. MIT Press: Cambridge MA.

[2] Ben-Akiva, M. and Morikawa, T. (1990) Estimation of switching models from revealed preferences and stated intentions. *Transportation Research*, 24A, pp. 485-495.

[3] Bhat, C. R., Castelar S., 2002. A unified mixed logit framework for modelling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. *Transportation Research B* 36B: 593-616.

[4] Bhat, C. R., 2003. Simulation estimation of mixed discrete choice models using randomised and scrambled halton sequences. *Transportation Research*.

[5] Bolduc, D., Ben-Akiva, M., 1991. A multinomial probit formulation for large choice sets. Proceedings of the 6th international conference on Travel Behaviour 2, 243-258.

[6] Brownstone D., and K. Train (1999) Forecasting new product penetration with flexible substitution patterns, *Journal of Econometrics* vol. 89, pp. 109-129.

[7] Domencich, T. and McFadden, D. (1975) *Urban travel demand*. North Holland: Amsterdam.

[8] Greene, W. H. (1997) *Econometric Analysis, third edition*. Prentice Hall.

[9] Hajivassiliou, V., McFadden, D. and Ruud P. (1996) Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *Journal of Econometrics*, 72, 85-134.

[10] Hensher D., and W. H. Greene (2003) The Mixed Logit model: The state of practice, *Transportation* vol. 30, 133-176.

[11] Kalouptsidis, N. (1997) *Signal Processing Systems*. John Wiley.

[12] Kalouptsidis, N., K. Koutroumbas and V. Psaraki (2003) Maximizing correct classifications in random utility models, under review.

[13] Koppelman F., and Sethi V (2000) Closed form discrete choice models. In: Hensher DA and Button KJ (eds) Handbook of Transport Modelling, Volume 1 pp 211-222. Oxford: Pergamon Press.

[14] Manski C. F. (1975) Maximum Score Estimation of the Stochastic Utility Model of Choice, *Journal of Econometrics*, vol. 3, pp. 205-228.

[15] Manski C. F. (1986) Operational characteristics of maximum score estimation *Journal of Econometrics* vol. 32, pp. 85-108.

[16] McFadden, D. (1973) Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in Econometrics*, ed P. Zarembka, 105-142, Academic Press, New York.

[17] McFadden, D and Train K. (2000) Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics*, 15 (5), 447-470.

[18] Morikawa, T., 1994. Correcting state dependence and serial correlation in the RP/SP combined estimation method. *Transportation* 21, 153-165.

[19] Ortuzar, J. and Willumsen, L. G. (1996) *Modelling Transport, second edition*. John Wiley.

[20] Walker Joan and M. Ben-Akiva (2002) Generalized random utility model *Mathematical Social Sciences* vol. 43, pp.303-343.