Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/072L50

On Locally Optimal Criterion for a Logit Model with Random Parameters

Habib Jafari

Department of Statistics-Razi University-Kermanshah-Iran

Received: 4 Oct. 2012, Revised: 25 Jan. 2013, Accepted: 27 Jan. 2013
Published online: 1 Jun. 2013

Abstract: Mixed logit is a highly flexible model that can approximate any random utility model. In this paper, two kinds of mixed multinomial logit models were considered. The main aim was to introduce a locally *D*-optimal criterion to obtain an optimal combination of the levels of attributes for producing alternatives and an optimal combination of alternatives in choice sets. Thus, a design including choice sets as the support points was designed.

Keywords: Mixed Logit models; Conjoint choice experiments; Multinomial Logit models; D-optimal criterion

1 Introduction

Like probit logit models, the mixed logit model has been known for many years; but, it has just become fully applicable since the advent of simulation. The first application of mixed logit was apparently in automobile demand models created jointly by [4] and [6]. In these studies, the explanatory variables did not vary over decision makers, and the observed dependent variable was market shares rather than individual customer's choices. As a result, the computationally intensive integration that is inherent in the mixed logit was needed to be performed only once for the market as a whole rather than for each decision maker in a sample. Early applications on customer-level data, such as [15] and [1], included only one or two dimensions of integration, which could be calculated by quadrature. Improvements in computer speed and in the understanding of simulation methods have allowed the full power of mixed logit to be utilized.

Among the studies to evidence this power are those conducted by [2] and [5] on cross-sectional data, and [7], [12] and [3] on panel data.

Mixed logit models can be derived under a variety of different behavioral specifications and each derivation provides a particular interpretation. The mixed logit model is defined on the basis of the functional form for its choice probabilities. Any behavioral specification, derived choice probability take this particular form, is called a mixed logit model. Mixed logit probabilities are the integrals of standard logit probabilities over a density of parameters. In this paper, the information matrix was required to be known which was calculated by the choice probabilities related to both random and fix attributes. Then, it a design based on choice sets was introduced in order to obtain the best choice of the alternatives produced by combining the levels of the attributes.

Therefore, this paper is composed of three sections. First of all, model specifications are described. The optimal criterion is introduced in the second section. At last, conclusion is presented.

2 Model Specifications

Considering the random utility of the standard MNL model [9], if heterogeneity is taken into account in the parameter across consumers, then the following utility function can be defined:

$$U_{ijs} = \mathbf{f}_i^T(a_{js})\beta_i + \varepsilon_{ijs}; \begin{cases} i = 1,2,\dots,N, & \text{the number of Individuals;} \\ j = 1,2,\dots,J, & \text{the number of Alternatives;} \\ s = 1,2,\dots,S, & \text{the number of Choice sets.} \end{cases} \tag{1}$$

where,

$$\mathbf{f}_i(a_{js}) = (\mathbf{f}_{i1}^T(a_{js}), \mathbf{f}_{i2}^T(a_{j2}), \dots, \mathbf{f}_{iK}^T(a_{js}))^T,$$

$$\mathbf{f}_{ik}(a_{js}) = (f_{ik1}(a_{js}), f_{ik2}(a_{js}), \dots, f_{ikL_k}(a_{js}))^T$$

characterize the levels of attribute k and $C_{is} = [\mathbf{f}_{ik}(a_{js})]_{j=1,\dots,J}^{k=1,\dots,K}$ denotes the choice set with J

* Corresponding author e-mail: jafari_habib@yahoo.com

alternatives. Thus, it is assumed that the parameter vector $\beta_i = (\beta_{i1}^T, \beta_{i2}^T, \ldots, \beta_{iK}^T)^T$ is different from all individuals and has multivariate normal distribution with mean $\mu_i = (\mu_{i1}^T, \mu_{i2}^T, \ldots, \mu_{iK}^T)^T$ and variance Σ_i, where there are K attributes, each with $\mathscr{L}_k; k = 1, \ldots, K$ levels. Since β_i is said to be different from all individuals, then β_i is a random variable. Each individual selects a part of levels of attributes, not all of their levels. Thus, the following can be considered:

$$\beta_{ik} = (\beta_{ik1}, \beta_{ik2}, \ldots, \beta_{ikL_k})^T,$$

where $L_k \le \mathscr{L}_k; k = 1, \ldots, K$. Considering the effects-type coding, the following can be given:

$$\sum_{\ell=1}^{\mathscr{L}_k} \beta_{ik\ell} = 0, \quad \sum_{\ell=1}^{L_k} \beta_{ik\ell} \neq 0.$$

Now, it is supposed that $\beta_{ik\ell}$ and $\beta_{ik\ell'}$; $\forall \ell \neq \ell'$ $(\ell = 1, 2, \ldots, L_k)$ are not correlate; thus:

$$cov(\beta_{ik\ell}, \beta_{ik\ell'}) = \begin{cases} \sigma_{ik\ell}^2, & \ell = \ell'; \\ 0, & \ell \neq \ell'. \end{cases}$$

Also:

$$Cov(\beta_{ik}, \beta_{ik'}) = \begin{cases} \Sigma_{ik}, & k = k'; \\ \mathbf{0}, & k \neq k'. \end{cases}$$

where $\Sigma_{ik} = diag\left(\sigma_{ik1}^2, \sigma_{ik2}^2, \ldots, \sigma_{ikL_k}^2\right)$. Thus, $\Sigma_i = \bigoplus_{k=1}^K \Sigma_{ik}$.

In this kind of situation, the following case should be concentrated on:

$$\beta_{ik} \sim N_{L_k}(\mu_{ik}, \Sigma_{ik}) \quad \text{and} \quad \beta_i \sim N_p(\mu_i, \Sigma_i), \quad (2)$$

where $p = \sum_{k=1}^K L_k$. Now, with respect to (2):

$$\beta_i = \mu_i + \mathbf{V}\sigma_i, \quad (3)$$

where $\mathbf{V} = \bigoplus_{k=1}^K \mathbf{V}_k$; $\mathbf{V}_k = diag(v_{k1}, v_{k2}, \ldots, v_{kL_k})$ is a diagonal matrix with pair wise with independent standard normal elements on the diagonal and

$\sigma_i = (\sigma_{i1}^T, \sigma_{i2}^T, \ldots, \sigma_{iK}^T)^T; \sigma_{ik} = (\sigma_{ik1}, \sigma_{ik2}, \ldots, \sigma_{ikL_k})^T$.

According to the above description, the obtained model is called the mixed or heterogeneous *logit* model [5]. Since β_i is random variable, then it can be written as follows:

$$p_i(a_{js}|\beta_i) = \frac{\exp\left(\mathbf{f}_i^T(a_{js})\beta_i\right)}{\sum_{j'=1}^J \exp\left(\mathbf{f}_i^T(a_{j's})\beta_i\right)}, \quad (4)$$

where $p_i(a_{js}|\beta_i) = P(Y_{ijs} = 1|\beta_i)$ is the conditional probability in which individual i selects alternative a_j in the choice set s and (with respect to (1));

$$Y_{ijs} = \begin{cases} 1, & if\ U_{ijs} = \max_{a_{j'} \in C_s} U_{ij's}; \\ 0, & \text{otherwise}. \end{cases}$$

But in the MMNL models,

$$\mathscr{P}_i(a_{js}|\theta_i) = P(Y_{ijs} = 1|\theta_i); \theta_i = (\mu_i^T, \sigma_i^T)^T$$

the unconditional probability that alternative a_j is selected by individual i (4) so that:

$$\mathscr{P}_i(a_{js}|\theta_i) =$$

$$= \int_{\Re^p} p_i(a_{js}|\beta_i) g(\beta_i|\theta_i) d\beta_i \quad (5)$$

$$= \int_{\Re^p} p_i(a_{js}|(\mu_i + \mathbf{V}\sigma_i))\phi(\mathbf{V}_1)\ldots\phi(\mathbf{V}_K)d\mathbf{V} \quad (6)$$

$$= \int_{\Re^p} \frac{\exp\left(\mathbf{f}_i^T(a_{js})(\mu_i + \mathbf{V}\sigma_i)\right)}{\sum_{j'=1}^J \exp\left(\mathbf{f}_i^T(a_{j's})(\mu_i + \mathbf{V}\sigma_i)\right)}\phi(\mathbf{V}_1)\ldots\phi(\mathbf{V}_K)d\mathbf{V} \quad (7)$$

where $\phi(.)$ is the normal standard density function. There is not an analytical method for calculating integral (5). In these kinds of situations, Quadrature (Gauss-Hrmite quadrature) technique can be used; but, it must be noted that if the dimension of integration is greater than two, Quadrature technique can not compute the integrals in sufficient speed and precision for maximum likelihood estimation. Thus, simulation techniques are usually applied for estimating *Mixed Logit* models. According to the previous description, the integrals in the choice probabilities are approximated using a Mote-Carlo technique; then, the resulting simulated log-likelihood function is maximized. For a given θ_i, a vector of values for β_i is drawn from $g(\beta_i|\theta_i)$ for individual. the values of this draw can then be used to calculate $\hat{p}_i(a_{js}|\beta_i)$. This process is repeated R times, meaning that:

$$\beta_i^{(1)}, \ldots, \beta_i^{(R)} \sim^{i.i.d} N_p(\mu_i, \sigma_i);$$

$$\beta_i^{(r)} = \mu_i + \mathbf{V}^{(r)}\sigma_i\ ; \ \forall i = 1, 2, \ldots, N, \ r = 1, 2, \ldots, R,$$

where integration over $g(\beta_i|\theta_i)$ is approximated by averaging the R draws. If it is supposed that

$$\hat{p}_i(a_{js}|\beta_i^{(r)}) = \frac{\exp\left(\mathbf{f}_i^T(a_{js})\beta_i^{(r)}\right)}{\sum_{j'=1}^J \exp\left(\mathbf{f}_i^T(a_{j's})\beta_i^{(r)}\right)} \quad (8)$$

is the realization of the choice probability related to the choosing alternative a_j by individual i for the r^{th} draw of β_i, then this can be written;

$$\hat{\mathscr{P}}_i(a_{js}|\theta_i) = \frac{1}{R}\sum_{r=1}^R \hat{p}_i\left(a_{js}|\beta_i^{(r)}\right), \quad (9)$$

where $\hat{\mathscr{P}}_i(a_{js}|\theta_i)$ is the simulated choice probability of individual i choosing alternative a_j given θ_i. In this situation, the *Simulated Log-Likelihood* function can be defined as follows;

$$\mathscr{S}\ell_s(\theta_i) = \sum_{i=1}^N \sum_{j=1}^J Y_{ijs} \ln\left(\hat{\mathscr{P}}_i(a_{js}|\theta_i)\right). \quad (10)$$

The estimated parameter vector $\hat{\theta}$ is the vector that maximizes the *Simulated Log-Likelihood*.

Additionally, it is supposed that there are two kinds of attributes. One part has fixed parameters which include K_1 attributes, each with $\mathscr{L}_k; k = 1,...,K_1$ levels. Another group has random parameters which consist K_2 attributes, each with $L_k; k = K_1+1,...,K$ ($L_k \leq \mathscr{L}_k$, $K_1+K_2 = K$) levels. Therefore, the utility function is introduced as follow,

$$U_{ijs} = \mathbf{f}_i^T(a_{js})\beta + \mathbf{h}_i^T(a_{js})\mathbf{b}_i + \varepsilon_{ijs}. \tag{11}$$

where $\beta = (\beta_1^T, \beta_2^T, \ldots, \beta_{K_1}^T)^T$ (according to effects-type coding $\sum_{\ell=1}^{\mathscr{L}_k} \beta_{k\ell} = 0$) is fixed for all individuals; but $\mathbf{b}_i = (\mathbf{b}_{i1}^T, \mathbf{b}_{i2}^T, \ldots, \mathbf{b}_{iK_2}^T)^T$ (according to effects-type coding $\sum_{\ell=1}^{L_k} b_{ik\ell} \neq 0$) is not the same for all the individuals and is a random variable, for example, with the multivariate normal distribution;

$$\mathbf{b}_i \sim N_{p_2}(\mu_{b_i}, \sigma_{b_i}) : \forall i = 1, 2, \ldots, N$$

where $p_1 = \sum_{k=1}^{K_1}(\mathscr{L}_k - 1)$ and $p_2 = \sum_{k=K_1+1}^{K} L_k$ are the dimension of the fixed and random parameters, respectively.

Similar to (8) and (9), this can be written:

$$p_i(a_{js}|\mathbf{b}_i) = \frac{\exp\left(\mathbf{f}_i^T(a_{js})\beta + \mathbf{h}_i^T(a_{js})\mathbf{b}_i\right)}{\sum_{j'=1}^{J} \exp\left(\mathbf{f}_i^T(a_{j's})\beta + \mathbf{h}_i^T(a_{j's})\mathbf{b}_i\right)}.$$

If η_i is assumed to be the parameter vector of the distribution function of $\mathbf{b}_i$, meaning that $g(\mathbf{b}_i|\eta_i)$; thus, two type parameters, β and η_i, will be present so that:

$$\mathscr{P}_i(a_{js}|\eta_i) = \int_{\Re^{p_2}} P_i(a_{js}|\mathbf{V})\, \phi(\mathbf{V}_1)\ldots\phi(\mathbf{V}_K)d\mathbf{V},$$

where

$$P_i(a_{js}|\mathbf{V}) = \frac{\exp\left(\mathbf{f}_i^T(a_{js})\beta + \mathbf{h}_i^T(a_{js})(\mu_{b_i} + \mathbf{V}\sigma_{(b)i})\right)}{\sum_{j'=1}^{J} \exp\left(\mathbf{f}_i^T(a_{j's})\beta + \mathbf{h}_i^T(a_{j's})(\mu_{bi} + \mathbf{V}\sigma_{(b)i})\right)}.$$

Similar to (8) and (9), a vector of values for $\mathbf{b}_i$ is drawn from $g(\mathbf{b}_i|\eta_i)$ for individual. The values of this draw can then be used for calculating $\hat{p}_i(a_{js}|\mathbf{b}_i)$. This process repeated R times and integration over $g(\mathbf{b}_i|\eta_i)$ is approximated by averaging the R draws. If it is supposed that $\hat{p}_i(a_{js}|\mathbf{b}_i^{(r)})$ is the realization of the choice probability of individual i for selecting alternative a_j for the r^{th} draw "**b**", then the following can be given:

$$\hat{\mathscr{P}}_i(a_{js}|\beta, \eta_i) = \frac{1}{R}\sum_{r=1}^{R} \hat{p}_i(a_{js}|\mathbf{b}_i^{(r)}),$$

where $\hat{\mathscr{P}}_i(a_{js}|\beta,\eta_i)$ is the simulated choice probability of individual i choosing alternative a_j given η_i. Thus, the *Simulated Log-Likelihood* function is defined as follows;

$$\mathscr{S}\ell_s(\beta, \eta_i) = \sum_{i=1}^{N}\sum_{j=1}^{J} Y_{ijs} \ln\left(\hat{\mathscr{P}}_i(a_{js}|\beta, \eta_i)\right). \tag{12}$$

The estimated parameter vectors $\hat{\beta}$ and $\hat{\eta}$ are the vectors that maximize the *Simulated Log-Likelihood* function.

3 Optimal Criterion

In this section, the D-optimal criterion is introduced for the Mixed MNL (MMNL) model. For this purpose, the information matrix of the MMNL model should be evaluated. Since it is assumed that the choices in different choice sets are independent, the information matrix based on (12) for each choice set, s, is as follow;

$$\mathbf{I}(\beta, \eta_i|C_{is}) = -E\begin{pmatrix} \frac{\partial^2 \mathscr{S}\ell_s(\beta,\eta_i)}{\partial\beta^T\partial\beta} & \frac{\partial^2 \mathscr{S}\ell_s(\beta,\eta_i)}{\partial\beta^T\partial\mu_i} & \frac{\partial^2 \mathscr{S}\ell_s(\beta,\eta_i)}{\partial\beta^T\partial\sigma_i} \\ \frac{\partial^2 \mathscr{S}\ell_s(\beta,\eta_i)}{\partial\mu_i^T\partial\beta} & \frac{\partial^2 \mathscr{S}\ell_s(\beta,\eta_i)}{\partial\mu_i^T\partial\mu_i} & \frac{\partial^2 \mathscr{S}\ell_s(\beta,\eta_i)}{\partial\mu_i^T\partial\sigma_i} \\ \frac{\partial^2 \mathscr{S}\ell_s(\beta,\eta_i)}{\partial\sigma_i^T\partial\beta} & \frac{\partial^2 \mathscr{S}\ell_s(\beta,\eta_i)}{\partial\sigma_i^T\partial\mu_i} & \frac{\partial^2 \mathscr{S}\ell_s(\beta,\eta_i)}{\partial\sigma_i^T\partial\sigma_i} \end{pmatrix}.$$

where $E(.)$ denotes the expectation of the random variable.

Now, for simplicity, the above matrix is rewritten as follow;

$$\mathbf{I}(\beta, \eta_i|C_{is}) = \begin{pmatrix} \mathscr{D}_{is}^T\mathbb{P}_{is}^{-1}\mathscr{D}_{is} & \mathscr{D}_{is}^T\mathbb{P}_{is}^{-1}\mathscr{M}_{is} & \mathscr{D}_{is}^T\mathbb{P}_{is}^{-1}\mathscr{Q}_{is} \\ \mathscr{M}_{is}^T\mathbb{P}_{is}^{-1}\mathscr{D}_{is} & \mathscr{M}_{is}^T\mathbb{P}_{is}^{-1}\mathscr{M}_{is} & \mathscr{M}_{is}^T\mathbb{P}_{is}^{-1}\mathscr{Q}_{is} \\ \mathscr{Q}_{is}^T\mathbb{P}_{is}^{-1}\mathscr{D}_{is} & \mathscr{Q}_{is}^T\mathbb{P}_{is}^{-1}\mathscr{M}_{is} & \mathscr{Q}_{is}^T\mathbb{P}_{is}^{-1}\mathscr{Q}_{is} \end{pmatrix} \tag{13}$$

along with

$$\begin{aligned} \mathscr{D}_{is} &= \int_{\Re^{p_2}} \left[\mathbf{P}_{is}(\mathbf{V}) - \mathbf{p}_{is}(\mathbf{V})\mathbf{p}_{is}^T(\mathbf{V})\right]\mathbf{F}_{is}\phi(\mathbf{V}_1)\ldots\phi(\mathbf{V}_K)d\mathbf{V} \\ \mathscr{M}_{is} &= \int_{\Re^{p_2}} \left[\mathbf{P}_{is}(\mathbf{V}) - \mathbf{p}_{is}(\mathbf{V})\mathbf{p}_{is}^T(\mathbf{V})\right]\mathbf{H}_{is}\phi(\mathbf{V}_1)\ldots\phi(\mathbf{V}_K)d\mathbf{V} \\ \mathscr{Q}_{is} &= \int_{\Re^{p_2}} \left[\mathbf{P}_{is}(\mathbf{V}) - \mathbf{p}_{is}(\mathbf{V})\mathbf{p}_{is}^T(\mathbf{V})\right]\mathbf{H}_{is}\mathbf{V}\phi(\mathbf{V}_1)\ldots\phi(\mathbf{V}_K)d\mathbf{V} \end{aligned} \tag{14}$$

where, $\eta_i = (\mu_i^T, \sigma_i^T)^T$,

$\mathbf{p}_{is}(\mathbf{V}) = (p_i(a_{1s}|\eta_i), p_i(a_{2s}|\eta_i), \ldots, p_i(a_{Js}|\eta_i))^T$
$\mathbf{P}_{is}(\mathbf{V}) = diag\,(p_i(a_{1s}|\eta_i), p_i(a_{2s}|\eta_i), \ldots, p_i(a_{Js}|\eta_i))$
$\mathbb{P}_{is} = diag(\mathscr{P}_i(a_{1s}|\eta_i), \mathscr{P}_i(a_{2s}|\eta_i), \ldots, \mathscr{P}_i(a_{Js}|\eta_i))$

$\mathbf{F}_{is} = [\mathbf{f}_{ik}(a_{js})]_{j=1,2,\ldots,J}^{k=1,\ldots,K_1}$ and $\mathbf{H}_{is} = [\mathbf{h}_{ik}(a_{js})]_{j=1,2,\ldots,J}^{k=K_1+1,\ldots,K}$ are design matrix. **See appendix A**.

To calculate integrals (14), there are no analytical techniques; thus, they should be calculated by the Monte-Carlo technique. Now, the distribution of the matrix $\mathbf{V}$ and the random sample with size n from its distribution mean that;

$$\mathbf{V}^{(1)}, \mathbf{V}^{(2)}, \ldots, \mathbf{V}^{(n)} \sim^{i.i.d} N_{p_2}(\mathbf{0}, \mathbf{I}_{p_2}),$$

where, $\mathbf{I}_{p_2}$ is a p_2-dimensional identically matrix; thus, (14) can be rewritten as;

$$\begin{aligned} \mathscr{D}_{is}^{(n)} &= \tfrac{1}{n}\textstyle\sum_{m=1}^{n}[\mathbf{P}_{is}(\mathbf{V}^{(m)}) - \mathbf{p}_{is}(\mathbf{V}^{(m)})\mathbf{p}_{is}^T(\mathbf{V}^{(m)})]\mathbf{F}_{is} \\ \mathscr{M}_{is}^{(n)} &= \tfrac{1}{n}\textstyle\sum_{m=1}^{n}[\mathbf{P}_{is}(\mathbf{V}^{(m)}) - \mathbf{p}_{is}(\mathbf{V}^{(m)})\mathbf{p}_{is}^T(\mathbf{V}^{(m)})]\mathbf{H}_{is} \\ \mathscr{Q}_{is}^{(n)} &= \tfrac{1}{n}\textstyle\sum_{m=1}^{n}[\mathbf{P}_{is}(\mathbf{V}^{(m)}) - \mathbf{p}_{is}(\mathbf{V}^{(m)})\mathbf{p}_{is}^T(\mathbf{V}^{(m)})]\mathbf{H}_{is}\mathbf{V}^{(m)} \end{aligned}$$

$$\mathscr{P}_i^{(n)}(a_{js}|\eta_i) = \tfrac{1}{n}\textstyle\sum_{m=1}^{n} p_i^{(m)}(a_{js}|\eta_i)$$

$$p_i^{(m)}(a_{js}|\eta_i) = \frac{\exp\left(\mathbf{f}_i^T(a_{js})\beta + \mathbf{h}_i^T(a_{js})(\mu_i + \mathbf{V}^{(m)}\sigma_i)\right)}{\sum_{j'=1}^{J}\exp\left(\mathbf{f}_i^T(a_{j's})\beta + \mathbf{h}_i^T(a_{j's})(\mu_i + \mathbf{V}^{(m)}\sigma_i)\right)}. \tag{15}$$

Now, based on Equations (15), the following can be presented:

$$\mathbf{I}^{(n)}(\beta,\eta_i|C_{is}) = \begin{pmatrix} \mathscr{D}_{is}^{(n)T}\mathbb{P}_{is}^{(n)-1}\mathscr{D}_{is}^{(n)} & \mathscr{D}_{is}^{(n)T}\mathbb{P}_{is}^{(n)-1}\mathscr{M}_{is}^{(n)} & \mathscr{D}_{is}^{(n)T}\mathbb{P}_{is}^{(n)-1}\mathscr{Q}_{is}^{(n)} \\ \mathscr{M}_{is}^{(n)T}\mathbb{P}_{is}^{(n)-1}\mathscr{D}_{is}^{(n)} & \mathscr{M}_{is}^{(n)T}\mathbb{P}_{is}^{(n)-1}\mathscr{M}_{is}^{(n)} & \mathscr{M}_{is}^{(n)T}\mathbb{P}_{is}^{(n)-1}\mathscr{Q}_{is}^{(n)} \\ \mathscr{Q}_{is}^{(n)T}\mathbb{P}_{is}^{(n)-1}\mathscr{D}_{is}^{(n)} & \mathscr{Q}_{is}^{(n)T}\mathbb{P}_{is}^{(n)-1}\mathscr{M}_{is}^{(n)} & \mathscr{Q}_{is}^{(n)T}\mathbb{P}_{is}^{(n)-1}\mathscr{Q}_{is}^{(n)} \end{pmatrix}. \quad (16)$$

In this situation and based on the number of the levels of the fixed and random attributes: $\mathscr{L}_1,\mathscr{L}_2,\ldots,\mathscr{L}_{K_1},L_{K_1+1},\ldots,L_K$. The total number of alternatives can be obtained as follows;

$$\mathscr{J} = \prod_{k=1}^{K_1}\mathscr{L}_k \prod_{k=K_1+1}^{K} L_k.$$

In this paper, since the designs with the same number of choice sets are considered for each individual; thus these designs have S choice sets as support points. Consequently, since $\mathscr{S} = \binom{\mathscr{J}}{J}$ denotes the total number of choice sets, each with J alternatives, then just a part of them is considered where $\binom{\mathscr{S}}{S}$ denotes the total number of choice sets, each with J alternatives.

Similar to classical MNL models [9], a criterion can be defined in order to obtain a locally D-optimality criterion as follow (for individual i);

$$\Psi_{(Mixed)}(\xi_i|\beta,\eta_i) = -\ln\left(\det\left(\mathbf{I}^{(n)}(\beta,\eta_i|\xi_i)\right)\right) \quad (17)$$

where, $\mathbf{I}^{(n)}(\beta,\eta_i|\xi_i) = \sum_{s=1}^{S} w_{is}\cdot\mathbf{I}^{(n)}(\beta,\eta_i|C_{is})$ for design ξ_i which includes S support points as follows;

$$\xi_i = \begin{Bmatrix} C_{i1} & C_{i2} & \ldots & C_{iS} \\ w_{i1} & w_{i2} & \ldots & C_{iS} \end{Bmatrix} \in \Xi_i;\ \ \forall i = 1,2,\ldots,N. \quad (18)$$

Now, based on (18) and to obtain the locally D-optimal design, a new design can be defined in the following way (because they are identifiable for the parameters of the model);

$$\xi = \sum_{i=1}^{N}\alpha_i\cdot\xi_i, \quad (19)$$

where (Convex combination property);

$$\Psi_{(Mixed)}(\xi|\beta,\eta) \le \sum_{i=1}^{N}\alpha_i\cdot\Psi_{(Mixed)}(\xi_i|\beta,\eta_i), \quad (20)$$

such that:

$$\xi = \begin{Bmatrix} C_1 & C_2 & \ldots & C_{\mathbb{S}} \\ w_1 & w_2 & \ldots & C_{\mathbb{S}} \end{Bmatrix} \in \Xi, \quad (21)$$

where

$C_1 = C_{11},\ldots,C_{\mathbb{S}} = C_{NS}$,

$w_1 = \alpha_1 w_{11}, w_2 = \alpha_1 w_{12}\ldots, w_{\mathbb{S}} = \alpha_N w_{NS}$.

Also, it can be written as;

$$\Psi_{(Mixed)}(\xi|\beta,\eta) = -\ln\left(\det\left(\mathbf{I}^{(n)}(\beta,\eta|\xi)\right)\right),$$

where

$$\mathbf{I}^{(n)}(\beta,\eta|\xi) = \sum_{s=1}^{\mathbb{S}} w_s\cdot\mathbf{I}^{(n)}(\beta,\eta|C_s);$$

$$\eta = (\eta_1^T,\eta_2^T,\ldots,\eta_N^T)^T$$

and

$$\eta_i = (\mu_{i11},\ldots,\mu_{iKL_K},\sigma_{i11},\ldots,\sigma_{iKL_K})^T; i = 1,2,\ldots,N.$$

With respect to (11), the number of parameters is $p = p_1 + 2p_2$, where p_1 denotes the number of parameters for fix attributes and $2p_2$ is the number of parameters for random attributes. Thus, to obtain the optimal combination of the levels of attributes for producing alternatives and the optimal combination of alternatives in choice sets, design (21) can be considered, where $\mathbb{S}$ can be hold in interval $[p,\frac{p(p+1)}{2}]$ (*Caratheodory* ' theorem [8]). In this case, ξ^* is the locally D-optimal design if:

$$\xi^* = \arg\min_{\xi\in\Xi}\Psi_{(Mixed)}(\xi|\beta_0,\eta_0), \quad (22)$$

where β_0,η_0 denote the true values of parameters and $\Xi = \bigcup_{i=1}^{N}\Xi_i; \forall i = 1,2,\ldots,N$, where;

$$\Xi_i = \left\{(C_{is},w_{is})|0\le w_{is}\le 1;\sum_{s=1}^{S}w_{is} = 1,\ \ C_{is} = \left[\mathbf{F}_{is}\ \mathbf{H}_{is}\right]\right\}.$$

4 Conclusion

In marketing and business, there are many important things like the quality on which the combination of attributes (products or services) depends. But, there exist an important thing, without which all these products or services are meaningless; that is the customer. If customers needs are satisfied, then, success will be achieved; otherwise; not.

It is known that a customer tends to choose an alternative (product or service) which has the highest utility; thus, a model is required which can show this situation. One of the most important introduced model is the MNL model which belongs to the logit family. Of course, this model (MNL) can be used when all alternatives are independent; in other words, when IIA [14] is hold; otherwise, other models of the logit family should be used which include two-level or three-level NMNL models [10].

In this kind of models, the focus should be on the combination of the levels of attributes (alternatives) and the combination of alternatives in choice sets.

To obtain the best combination, the levels of attributes and the most suitable combination of choice sets, the optimal design can be a nice idea.

There are many criteria to obtain optimal design for do this idea; but, in this paper, only the D-optimal criterion was introduced because this criterion was not affected by reparameterizations of the model [13].

The D-optimal criterion for the MMNL model is a function of the determinate of information matrix, depending on unknown parameters as the standard MNL model [9]. But, there is a different situation from the standard MNL model. With respect to the MMNL model and dependence of the information matrix on unknown parameters, a difficult task was encountered. To obtain the information matrix, some integrals were solved which did not have a closed form. In this situation, the Monte-Carlo technique was used for obtaining the information matrix and D-criterion based on a sample from a vector with multiple standard normal distribution. Of course, in most cases, a locally optimal design should be obtained which acts as a linear model instead of an optimal design. Sometimes, the D-criterion can not be obtained because there is not any suitable software which can calculate the determinant of the information matrix. Although, it could be done for several examples, considering some conditions and limitations on the parameters. For example, in this paper, the optimal design was discussed for the standard MNL model [9]. Also, different designs were introduced and the one which is locally more suitable was given.

Acknowledgements

This work is partially supported by the Cooperative 4407 knowledge Base New Ideas.

References

[1] Ben-Akiva, M., Bolduc, D. and Bradley, M. (1993). Estimation of travel model choice models with randomly distributed values of time, *Transportation Research Record, 1413, 88-97*.

[2] Bhat, C. (1998a). Accomodating variations in responsivness to level-of-service variables in travel mode choice models, *Transportation Research A 32, 455-507*.

[3] Bhat, C. (2000). Incorporating observed and unobserved heterogeneity in urban work mode choice modeling, *Transportation Science 34, 228-238*.

[4] Boyed, J. and J. Mellman (1980). The effect of fuel economy standards on the U.S. automotive market: A hedonic demand analysis, *Transportation Reseach A, 14, 367-378*.

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

[6] Cardell, S. and Dunbar, F. (1980). Measuring societal impacts of automobile downsizing, *Transportation Research A , 14, 423-434*.

[7] Erdem, T. (1996). A dynamic analysis of market structure based on panel data, *Marketing Science, 15, 359-378*.

[8] Fedorov, V.V. (1972). *Theory of optimal experiments*, Academic Press, New York.

[9] Jafari, H. (2011). Locally *D*-optimal design for a logit model in disceret choice experiment, *Comunication Statistics Thorey and Method, Accepted , April 2011*.

[10] Jafari, H. (2010). *Optimal Design in Conjoint Analysis* , PhD Thsis, IMST, OvG University Magdeburg.

[11] McFadden, D. and Train, K. (2000). Mixed *MNL* models of discrete response, *Journal of Applied Econometrics, 15, 447-470*.

[12] Revelt, D. and Train, K.(1998). Mixed logit with repeated choices, *Review of Economics and Statistics 80, 647-657*.

[13] Schwabe, R. (1996). *Optimum designs for multi-factor models*, Springer, New York.

[14] Train, K.(1987). *Qualitative choice analysis*, MIT Press, Cambridge, MA.

[15] Train, K., McFadden, D. and Ben-Akiva, M. (1987a). The demand for local telephone service: A fully discrete model of residential calling patterns and service choice, *Rand Journal of Econometrics 18, 109-123*

Appendix A:

According to (14):

$$\frac{\partial S\ell(\beta,\eta_i)}{\partial\beta} =$$

$$= \sum_{j=1}^{J} Y_{ij}\cdot\frac{\partial \ln(\mathscr{P}_i(a_j|\eta_i))}{\partial\beta}$$

$$= \sum_{j=1}^{J} Y_{ij}.\mathscr{P}_i^{-1}(a_j|\eta_i)\left(\mathbf{f}_i(a_j)\mathscr{P}_i(a_j|\eta_i) - \int_{\Re^{2p_2}} p_i(a_j|\eta_i)\mathbf{p}_i^T(\mathbf{V})\mathbf{F}_i\, d\Phi(\mathbf{V})\right)$$

$$= \sum_{j=1}^{J} Y_{ij}\mathbf{f}_i(a_j) - \int_{\Re^{2p_2}} \sum_{j=1}^{J} Y_{ij}p_i(a_j|\eta_i)\mathscr{P}_i^{-1}(a_j|\eta_i)\mathbf{p}_i^T(\mathbf{V})\mathbf{F}_i\, d\Phi(\mathbf{V})$$

$$= \mathbf{Y}_i^T\mathbf{F}_i - \int_{\Re^{2p_2}} \mathbf{Y}_i^T\mathbb{P}_i^{-1}\mathbf{p}_i(\mathbf{V})\mathbf{p}_i^T(\mathbf{V})\mathbf{F}_i\, d\Phi(\mathbf{V})$$

$$= \mathbf{Y}_i^T\mathbb{P}_i^{-1}\left(\mathbb{P}_i\mathbf{F}_i - \int_{\Re^{2p_2}} \mathbf{p}_i(\mathbf{V})\mathbf{p}_i^T(\mathbf{V})\mathbf{F}_i\, d\Phi(\mathbf{V})\right)$$

$$= \mathbf{Y}_i^T\mathbb{P}_i^{-1}\left(\int_{\Re^{2p_2}} \mathbf{P}_i(\mathbf{V})\mathbf{F}_i\, d\Phi(\mathbf{V}) - \int_{\Re^{2p_2}} \mathbf{p}_i(\mathbf{V})\mathbf{p}_i^T(\mathbf{V})\mathbf{F}_i\, d\Phi(\mathbf{V})\right)$$

$$= \mathbf{Y}_i^T\mathbb{P}_i^{-1}\underbrace{\left(\int_{\Re^{2p_2}} \left(\mathbf{P}_i(\mathbf{V}) - \mathbf{p}_i(\mathbf{V})\mathbf{p}_i^T(\mathbf{V})\right)\mathbf{F}_i\, d\Phi(\mathbf{V})\right)}_{\mathscr{D}_i}$$

Similarly;

$$\frac{\partial S\ell(\beta,\eta_i)}{\partial\beta^T} = \underbrace{\left(\int_{\Re^{2p_2}} \left(\mathbf{P}_i(\mathbf{V}) - \mathbf{p}_i(\mathbf{V})\mathbf{p}_i^T(\mathbf{V})\right)\mathbf{F}_i\, d\Phi(\mathbf{V})\right)^T}_{\mathscr{D}_i^T}\mathbb{P}_i^{-1}\mathbf{Y}_i,$$

thus;

$$E\left(\frac{\partial S\ell(\beta,\eta_i)}{\partial\beta^T}\frac{\partial S\ell(\beta,\eta_i)}{\partial\beta}\right) = \mathscr{D}_i^T\mathbb{P}_i^{-1}\underbrace{E\left(\mathbf{Y}_i\mathbf{Y}_i^T\right)}_{\mathbb{P}_i}\mathbb{P}_i^{-1}\mathscr{D}_i = \mathscr{D}_i^T\mathbb{P}_i^{-1}\mathscr{D}_i$$

where $\int_{\Re^{2p_2}} P_i(V)\, d\Phi(V) = \mathbb{P}_i$ and;

$$\mathbf{Y}_i\mathbf{Y}_i^T = \begin{pmatrix} Y_{i1}^2 & Y_{i1}Y_{i2} & \cdots & Y_{i1}Y_{iJ} \\ Y_{i1}Y_{i2} & Y_{i2}^2 & \cdots & Y_{i2}Y_{iJ} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{i1}Y_{iJ} & Y_{i2}Y_{iJ} & \cdots & Y_{iJ}^2 \end{pmatrix} \Rightarrow E(\mathbf{Y}_i\mathbf{Y}_i^T) = \begin{pmatrix} \mathscr{P}_i(a_1|\beta_i) & 0 & \cdots & 0 \\ 0 & \mathscr{P}_i(a_2|\beta_i) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathscr{P}_i(a_J|\beta_i) \end{pmatrix} = \mathbb{P}_i$$

Similarity, the following can be given:

$$\frac{\partial S\ell(\beta,\eta_i)}{\partial \mu_i} = \mathbf{Y}_i^T\mathbb{P}_i^{-1} \underbrace{\left(\int_{\Re^{2p_2}} \left(\mathbf{P}_i(\mathbf{V}) - \mathbf{p}_i(\mathbf{V})\mathbf{p}_i^T(\mathbf{V})\right)\mathbf{H}_i\, d\Phi(\mathbf{V})\right)}_{\mathscr{M}_i} = \mathbf{Y}_i^T\mathbb{P}_i^{-1}\mathscr{M}_i$$

$$\frac{\partial S\ell(\beta,\eta_i)}{\partial \sigma_i} = \mathbf{Y}_i^T\mathbb{P}_i^{-1} \underbrace{\left(\int_{\Re^{2p_2}} \left(\mathbf{P}_i(\mathbf{V}) - \mathbf{p}_i(\mathbf{V})\mathbf{p}_i^T(\mathbf{V})\right)\mathbf{H}_i\mathbf{V}\, d\Phi(\mathbf{V})\right)}_{\mathscr{Q}_i} = \mathbf{Y}_i^T\mathbb{P}_i^{-1}\mathscr{Q}_i.$$

Hence;

$$E\left(\frac{\partial S\ell(\beta,\eta_i)}{\partial \beta^T}\frac{\partial S\ell(\beta,\eta_i)}{\partial \mu_i}\right) = \mathscr{D}_i^T\mathbb{P}_i^{-1}\underbrace{E\left(\mathbf{Y}_i\mathbf{Y}_i^T\right)}_{\mathbb{P}_i}\mathbb{P}_i^{-1}\mathscr{M}_i = \mathscr{D}_i^T\mathbb{P}_i^{-1}\mathscr{M}_i$$

$$E\left(\frac{\partial S\ell(\beta,\eta_i)}{\partial \beta^T}\frac{\partial S\ell(\beta,\eta_i)}{\partial \sigma_i}\right) = \mathscr{D}_i^T\mathbb{P}_i^{-1}\underbrace{E\left(\mathbf{Y}_i\mathbf{Y}_i^T\right)}_{\mathbb{P}_i}\mathbb{P}_i^{-1}\mathscr{Q}_i = \mathscr{D}_i^T\mathbb{P}_i^{-1}\mathscr{Q}_i$$

$$E\left(\frac{\partial S\ell(\beta,\eta_i)}{\partial \mu_i^T}\frac{\partial S\ell(\beta,\eta_i)}{\partial \sigma_i}\right) = \mathscr{M}_i^T\mathbb{P}_i^{-1}\underbrace{E\left(\mathbf{Y}_i\mathbf{Y}_i^T\right)}_{\mathbb{P}_i}\mathbb{P}_i^{-1}\mathscr{Q}_i = \mathscr{M}_i^T\mathbb{P}_i^{-1}\mathscr{Q}_i$$

Habib Jafari received the MS degree in Mathematical Statistics from Shiraz University in 2000 (Iran), and the PhD degree in Mathematical Statistics (Conjoint Analysis) from the IMST, OvG University of Magdeburg in 2010 (Germany). He is currently an assistant professor in Razi University. His research interests are in the areas of optimal design, conjoint Analysis