OPTIMAL DESIGN FOR A THREE –LEVEL NESTED MULTINOMIAL LOGIT MODEL IN DISCRETE CHOICE EXPERIMENTS

Habib Jafari

Institute for Mathematical Stochastics (IMST),
OvG University, Universitätsplatz 2, 39106 Magdeburg-Germany

ABSTRACT

In this paper we calculate the information matrix for a three-level nested Multinomial logit model and derive the locally D-optimal design to estimate the parameter vector.

Keywords: *Conjoint Analysis, Three-level Nested MNL Model, Discrete Choice Experiment, D-Optimal Criterion..*

1. INTRODUCTION

The multinomial logit (MNL) model is most widely used in discrete choice models due to its closed-form choice probabilities and its consistency with the random utility maximization (RUM). However, the MNL model suffers from restrictive independence from irrelevant alternatives (IIA) property, which states that the ratio of two choice probabilities is independent of the other alternatives in the model. This implies that a change in an attribute of one alternative will have the same proportional impact on the probability of each of the other alternatives being chosen. The NMNL model relaxes the IIA property by dividing the alternatives into subsets or nests, allowing the IIA assumption to hold within each nest but not for alternatives in different nests. Notwithstanding that there is the same IIA property for the nests that it is the IIN (Independent from Irrelevant Nest). As opposed to the more flexible Multinomial Probit and Mixed Logit models, the NMNL model has closed-form choice probabilities, which can be estimated without resorting to simulation methods. Due to its simplicity and allowing for a variety of substitution patterns, the NMNL model remains the most common extension of the MNL model in applied work. *Daly and Zachary (1978)* and *McFadden (1978a)* have shown that the two-level NMNL model is consistent with RUM under the condition that the dissimilarity parameters are constrained within the unit interval. In many practical applications, however, this condition has not been met. *Börch-Supan (1990)* argues that the DZM condition is unnecessarily strong given that the NMNL model should be viewed as a local approximation. Based on the work of *Börch-Supan*, *Herriges and Kling (1996)* who derive the necessary conditions for local consistency with random utility maximization for two-level NMNL models; the two-level NMNL model is consistent with RUM when dissimilarity parameters vary in interval [0,1) and when the dissimilarity parameters are greater than one. Therefore, the two-level NMNL model is consistent for some range of the characteristics of attributes with RUM. A two-level NMNL model is not consistent with RUM when there is a dissimilarity parameter less than zero.

In some cases of two-level NMNL models, the IIA property may not hold within some or all of the nests. In this situation, we can divide the alternatives of these nests into several sub-sets, called sub-nests. This kind of NL model is termed the three-level NMNL model, since within it there are three kinds of choice probabilities that will be discussed in the section 2.

The rest of this chapter is structured as follows. Section 2 discusses the model specifications of three-level NMNL models. Section 3 presents the information matrix for a three-level NMNL models (with two nests). We will introduce the D-optimal criterion by subsection 3.1 .

2. MODEL SPECIFICATIONS

Following *Gil-Molton and Hole, (2004)*, let us consider a sample of I individuals with J discrete possible alternatives (in choice set C), which are produced by K attributes, each with L_k levels. In this paper, for three-level NMNL model, the total number of alternatives is showed by $\sum_{m=1}^{M}\sum_{h=1}^{H_m} J_{hm}$, where J_{hm} is the number of alternatives in the sub-nest h of nest m. In this case, there are S choice sets each containing J_s alternatives to fit model, where if C_s is a choice set with J_s alternatives then $C=\bigcup_{s=1}^{S} C_s$. Certainly, in such a model, the total number of alternatives in choice set C is denoted by $\prod_{k=1}^{K} L_k$ with regard to the attributes and their levels. This model was obtained based on selection of an alternative with the highest utility.

The utility related to the three-level NMNL model (i.e. choice set, C_s), where the individual i is derived when choosing alternative j as denoted by U_{ijhms}. This utility is partitioned into a systematic component, v_{ijhms}, and a random component, ε_{ijhms} (s denotes the choice set C_s), to produce (Because, the conditions are the same for all individuals then we ignore index i):

$$U_{jhms} = U_{j|hms} + U_{h|ms} + U_{ms}, \qquad (1)$$

So that:

$$U_{j|hms} = v_{j|hms} + \varepsilon_{j|hms}, \qquad U_{h|ms} = v_{h|ms} + \varepsilon_{h|ms}, \qquad U_{ms} = v_{ms} + \varepsilon_{ms},$$

where $\varepsilon_{j|hms}$ have EVD (Extreme Value Distribution type (II)) with variance σ^2_{hm} (They are correlated in the same sub-nest, $\rho_{hm} = corr(\varepsilon_{j|hms}, \varepsilon_{j'|hms})$), the distributions of $\varepsilon_{h|ms}$ is such that variable $\max_{j\in C_{hms}} U_{ij|hms}$ with variance σ^2_m (C_{hms} denotes a choice set (s) which includes alternatives in sub-nest h of nest m and $\rho_m = corr(\varepsilon_{h|ms}, \varepsilon_{h'|ms})$) and the distribution of ε_{ms} is such that variable $\max_{h\in H_m} U_{ih|ms}$ will have EVD (Type II) with variance σ^2 so that $corr(\varepsilon_{ims}, \varepsilon_{im's}) = 0$, where H_m denotes the number of sub-nest in nest m. Naturally, these three error terms are independent ($\varepsilon_{ij|hms}$, $\varepsilon_{ih|ms}$ and ε_{ims}). Now, with consideration to utility (1) $v_{j|hms}$ can be written by a regression function as:

$$v_{j|hms} = \mathbf{x}^{\mathrm{T}}_{j|hms}\boldsymbol{\beta} = \sum_{k=1}^{K} \mathbf{x}^{\mathrm{T}}_{j|hmks}\boldsymbol{\beta}_k; \quad \boldsymbol{\beta}_k^T = (\beta_{k,1}, \ldots, \beta_{k,\ell}, \ldots \beta_{k,L_k}).$$

According to effect-type coding:

$$\sum_{\ell=1}^{L_k} \beta_{k,\ell} = 0 \Rightarrow \beta_{k,L_k} = -\sum_{\ell=1}^{L_k-1} \beta_{k,\ell}.$$

In this situation, $\boldsymbol{\beta}_k^T$ will be rewritten by $\boldsymbol{\beta}_k^T = (\beta_{k,1}, \ldots, \beta_{k,\ell}, \ldots \beta_{k,L_k-1})$. Similarly, we will have:

$\mathbf{x}^T_{j|hms} = (\mathbf{x}^T_{j|hm1s}, \ldots, \mathbf{x}^T_{j|hmks}, \ldots, \mathbf{x}^T_{j|hmKs})$; $\mathbf{x}^T_{j|hmks} = (x_{j|hmk,1,s}, x_{j|hmk,2,s}, \ldots, x_{j|hmk,L_k-1,s})$, where $\boldsymbol{\beta}_k^T$ and $\mathbf{x}^T_{j|hmks}$ denote the characteristics of the attributes k related to choosing alternative j by individual i (ignored) in the sub-nest h of the nest m according to choice set C_s. Now, according to previous assumptions we will have:

$$\Sigma_{U_s} = Cov(U_{ms}, U_{m's}) = \begin{cases} \Sigma_m, & m = m'; \\ 0, & m \neq m', \end{cases} \text{ where } \Sigma_m = \begin{cases} \Sigma_{hm}, & h = h'; \\ \Sigma_{hm,h'm}, & m \neq m' \end{cases},$$

$$\Sigma_{hm} = Cov(U_{jhm}, U_{j'hm}) = \sigma^2_{hm}(1-\rho_{hm})\mathbf{I}_{J_{hms}} + (\rho_{hm}\sigma^2_{hm} + \sigma^2_m + \sigma^2)\mathbf{J}_{J_{hm}},$$

$$\Sigma_{hm,h'm} = Cov(U_{jhm}, U_{j'h'm}) = (\rho_m\sigma^2_m + \sigma^2)\mathbf{J}_{J_{hm}},$$

where $i = 1,2,\ldots,I$, $j = 1,2,\ldots,J_{hms}$, $h = 1,2,\ldots,H_m$; $m = 1,2,\ldots,M$ and $\mathbf{I}_r$ is identity matrix of size r and $\mathbf{J}_r = \mathbf{1}_r\mathbf{1}_r^T$.

Now, with consideration to the utility (1) observation variables as follows can be introduced:

$$Y_{j|hms}=\begin{cases}1, & U_{j|hms}=\max_{j'\in C_{hms}} U_{j'|hms};\\ 0, & Otherwise,\end{cases} \quad , Y_{h|ms}=\begin{cases}1, & U_{h|ms}=\max_{h'\in H_m} U_{h'|ms};\\ 0, & Otherwise,\end{cases},$$

$$Y_{ms}=\begin{cases}1, & U_{ms}=\max_{m'} U_{m's};\\ 0, & Otherwise.\end{cases}$$

Thus, when the variables $Y_{j|hms}, Y_{h|ms}$ and Y_{ms} are independent $\left(Y_{jhms}=Y_{j|hms}\times Y_{h|ms}\times Y_{ms}\right)$:

$$p_{jhms}=p_{j|hms}\times p_{h|ms}\times p_{ms}, \tag{2}$$

where $p_{j|hms}=\Pr\left(Y_{j|hms}=1\right)$ is the conditional probability of choosing alternative j, given that sub-nest h and nest m have been chosen, $p_{h|ms}=\Pr\left(Y_{h|ms}=1\right)$ is the conditional probability of choosing sub-nest h when nest m is chosen and $p_{ms}=\Pr\left(Y_{ms}=1\right)$ is the marginal probability of choosing nest m (with respect to choice set s) and $\Pr$ denotes the probability of an event. Based on the distribution of the error terms of the utility, these probabilities can be calculated by (*McFadden, (1978))*:

$$p_{j|hms}=\frac{e^{\frac{v_{j|hms}}{\lambda_{hm}}}}{\sum_{j'=1}^{J_{hms}} e^{\frac{v_{j'|hms}}{\lambda_{hm}}}}, \qquad p_{h|ms}=\frac{e^{\frac{\lambda_{hm}}{\mu_m}IV_{hms}}}{\sum_{h'=1}^{H_m} e^{\frac{\lambda_{h'm}}{\mu_m}IV_{h'ms}}}, \qquad p_{ms}=\frac{e^{\mu_m IV_{ms}}}{\sum_{m'=1}^{M} e^{\mu_{m'} IV_{m's}}},$$

where $IV_{ms}=E\left(\max_{h\in H_m} U_{ih|ms}\right)=\ln\left(\sum_{h=1}^{H_m} e^{\frac{\lambda_{hm}}{\mu_m}IV_{hms}}\right), IV_{hms}=E\left(\max_{j\in C_{hms}} U_{j|hms}\right)=\ln\left(\sum_{j=1}^{J_{hms}} e^{\frac{v_{j|hms}}{\lambda_{hm}}}\right).$

3. INFORMATION MATRIX

There are criteria like D-, A-, G-criterion, etc, for obtaining optimal design. In this chapter, we use D-optimal criterion (a function of the determinant of the information matrix) in order to obtain an optimal design. Thus, first we must obtain the information matrix for the three-level nested logit model. In this situation, the log-likelihood function is required, defined for the choice set C_s and one individual as follows:

$$\ell(\boldsymbol{\theta};C_s)=\sum_{m=1}^{M}\sum_{h=1}^{H_m}\sum_{j=1}^{J_{hms}} y_{jhms}\times\ln(p_{jhms}),$$

where J_{hms} denotes the number of alternatives in sub-set hm corresponding to choice set C_s.

Based on the definition of the information matrix (w.r.t choice set C_s):

$$-E\left(\frac{\partial^2\ell(\boldsymbol{\theta};C_s)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right)=\sum_{m,h,j} p_{j|hms}p_{h|ms}p_{ms}\times\left(\frac{-\partial^2\ln\left(p_{j|hms}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}+\frac{-\partial^2\ln\left(p_{h|ms}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}+\frac{-\partial^2\ln\left(p_{ms}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right), \tag{3}$$

where:

$$\mathbf{I}\left(\boldsymbol{\theta};C_s\right)=\begin{pmatrix}\mathbf{I}_{\beta} & \mathbf{I}_{\beta\mu} & \mathbf{I}_{\beta\lambda}\\ \mathbf{I}_{\beta\mu}^T & \mathbf{I}_{\mu} & \mathbf{I}_{\mu\lambda}\\ \mathbf{I}_{\beta\lambda}^T & \mathbf{I}_{\mu\lambda}^T & \mathbf{I}_{\lambda}\end{pmatrix} \tag{4}$$

is the information matrix of the choice set C_s and $\boldsymbol{\theta}$ is the full parameters vector, so that $\boldsymbol{\theta}^T=\left(\boldsymbol{\beta}^T,\boldsymbol{\mu}^T,\boldsymbol{\lambda}^T\right)$ and $\boldsymbol{\beta}^T=\left(\boldsymbol{\beta}_1^T,\boldsymbol{\beta}_2^T,...,\boldsymbol{\beta}_K^T\right); \boldsymbol{\beta}_k^T=\left(\beta_{k,1},\ldots,\beta_{k,\ell},...,\beta_{k,L_k-1}\right)$ and $\boldsymbol{\mu}^T=\left(\mu_1,\mu_2,...,\mu_M\right)$ and

$\boldsymbol{\lambda}^T = \left(\boldsymbol{\lambda}_1^T, \boldsymbol{\lambda}_2^T, \ldots, \boldsymbol{\lambda}_M^T\right)$; $\boldsymbol{\lambda}_m^T = \left(\lambda_{1m}, \lambda_{2m}, \ldots, \lambda_{H_m m}\right)$; $\forall m$. This means that parameter $\beta_{k,\ell}$ is related to the ℓ^{th} level of attribute k, μ_m, the dissimilarity parameter of the nest m and $\boldsymbol{\lambda}_m^T$ is the dissimilarity parameters vector of the nest m, where λ_{hm} denotes the dissimilarity parameter of the sub-nest hm in nest m. Thus, the number of parameters in the three-level NMNL model there are as follows:

$$q = \underbrace{\sum_{k=1}^{K}(L_k - 1)}_{part-worth} + \underbrace{\sum_{m=1}^{M} H(m)}_{sub-nests} + \underbrace{M}_{nests} = q_1 + q_2 + M,$$

where q_1 is the number of part-worth parameters, q_2 is the number of the dissimilarity parameters of the sub-nests and M is the number of dissimilarity parameters of the nests, hence, the information matrix (4) is a symmetric positive semi definite $q \times q -$ matrix.

In order to fit the three-level NMNL model, let us consider the following experiments:

$$\sum_{m=1}^{M}\sum_{h=1}^{H_m} J_{hm} / \mathrm{J_s} / \mathrm{S}; \qquad \mathrm{S} \geq q_1 + q_2 + M, \tag{5}$$

where $\sum_{m=1}^{M}\sum_{h=1}^{H_m} J_{hm}$ denotes the total number of alternatives in population and $\mathrm{J_s} = \sum_{m=1}^{M}\sum_{h=1}^{H_m} \mathrm{J_{hms}}$ denotes the number of alternatives in choice set s, selected from population, randomly. In particular, suppose that $\mathrm{J_s} = \mathrm{J}; \forall \mathrm{s}$ then in this case, there will be S choice sets each with J alternatives. However, based on (5), the $S \leq \mathsf{S}$ choice sets can be considered instead of S (in reality the number of choice sets S increases dramatically when the number of attributes and their levels increase, then S must be often reduced to S, by employing a suitable technique (See *Grasshof, et al. (2004)*)). Also, the number of alternatives, which will be selected from sub-nests, may vary. Thus, there are different classes can be used in order to obtain a sample with size J from the population by:

$$\mathsf{S}_n = \binom{J_{11}}{\mathrm{J_{n11}}} \cdots \binom{J_{H_1 1}}{\mathrm{J_{nH_1 1}}} \cdots \binom{J_{hm}}{\mathrm{J_{nhm}}} \cdots \binom{J_{1M}}{\mathrm{J_{n1M}}} \cdots \binom{J_{H_M M}}{\mathrm{J_{nH_M M}}} \tag{6}$$

where $\sum_{m=1}^{M}\sum_{h=1}^{H_m} \mathrm{J_{nhm}} = \mathrm{J}; \forall \mathrm{n} \in N$ and $\mathsf{S_n}$ is the number of choice sets, each including J alternatives.

Based on class n to create an experiment, $\mathrm{J_s}$ can be rewrite as $\sum_{m=1}^{M}\sum_{h=1}^{H_m} \mathrm{J_{nhms}} = \mathrm{J}_{ns}$, where $\mathrm{J_{ns}} = \mathrm{J}; \forall \mathrm{s} \in \mathsf{S_n}, \forall n \in N$ and $\mathrm{J_{nhms}} = \mathrm{J_{nhms'}}; \forall \mathrm{s} \neq \mathrm{s'} \in \mathsf{S_n}$ but $\mathrm{J_{nhms}}$ and $\mathrm{J_{n'hms'}}$ (for different class and different choice set) may be equal or not equal. According to reduce the total number of choice sets (S) to a reasonable number (S), we reduce $\mathsf{S_n}$ to S_n in each class, where $S_n \leq q$ (avoiding singular information matrix) then consider $\sum_{m=1}^{M}\sum_{h=1}^{H_m} J_{hm} / \mathrm{J} / \mathrm{S_n}; \quad \mathrm{S_n} \geq q; \forall n \in N$ instead of (5). This involves choosing S_n choice sets each of them with J alternatives in each class. According to the type of model, it is possible that $\mathsf{S_n} < q; \forall n \in N$. In such a case and in order to avoid a singularity information matrix, we can combine them ($\mathsf{S_n}$) together to create a new design. To obtain information matrix related to each choice set this is used (4).

Table 1: Three-level NMNL model with two nests

First Nest (1)		Second Nest (2)
Sub-nest (1)	Sub-nest (2)	
J_{11}	J_{21}	J_2

Lemma 1: The information matrix is related to a three-level nested logit model (choice set C_s) with two nests, the first nest has two sub-nests with J_{11s} and J_{21s} alternatives and the second, J_{2s} alternatives (Table 1 denotes a population with $J_{11}+J_{21}+J_2$ alternatives) is calculated as follows:

$$\mathbf{I}(\boldsymbol{\theta};C_s)=\begin{pmatrix}\mathbf{I}_{11s} & \mathbf{I}_{12s} & \mathbf{I}_{13s} & \mathbf{I}_{14s} & \mathbf{I}_{15s}\\ \mathbf{I}_{12s}^T & I_{22s} & I_{23s} & I_{24s} & I_{25s}\\ \mathbf{I}_{13s}^T & I_{23s} & I_{33s} & I_{34s} & I_{35s}\\ \mathbf{I}_{14s}^T & I_{24s} & I_{34s} & I_{44s} & I_{45s}\\ \mathbf{I}_{15s}^T & I_{25s} & I_{35s} & I_{45s} & I_{55s}\end{pmatrix},$$

where $\boldsymbol{\theta}^T=\left(\boldsymbol{\beta}^T,\mu_1,\mu_2,\lambda_{11},\lambda_{21}\right)$. For simplicity, we suppose that $\beta_{1,1}=\beta_1\ldots\ \beta_{K,L_K-1}=\beta_{q_1}$ and $\lambda_1=\lambda_{11}$, $\lambda_2=\lambda_{21}$ then $\boldsymbol{\theta}^T=\left(\beta_1,\ldots,\beta_{q_1},\mu_1,\mu_2,\lambda_1,\lambda_2\right)$ (See Appendix-**AI**).

To fit this model (Table 1), examine the following experiments:

$$\left(J_{11}+J_{21}+J_2\right)/\mathrm{J}/S_n\,;q_1+\leq S_n\leq \mathsf{S}_\mathrm{n}\,,$$

where $\mathrm{J}=\mathrm{J}_{11n}+\mathrm{J}_{21n}+\mathrm{J}_{2n}$ and:

$$\mathsf{S}_n=\begin{pmatrix}J_{11}\\ \mathrm{J}_{n11}\end{pmatrix}\times\begin{pmatrix}J_{21}\\ \mathrm{J}_{n21}\end{pmatrix}\times\begin{pmatrix}J_{2}\\ \mathrm{J}_{n2}\end{pmatrix};\quad \mathrm{n}=1,2,\ldots,\mathrm{N}.$$

Corollary 1: When obtaining a locally optimal design when $\boldsymbol{\beta}=\mathbf{0}$, the above information matrix (Lemma 1) should be rewritten by the following Table (w.r.t **Appendix AI**):

$\mathbf{B}_{1\|1s}=\frac{1}{\mathrm{J}_{11s}}\mathbf{X}_{1\|1s}^T\mathbf{I}_{\mathrm{J}_{11s}}\mathbf{X}_{1\|1s}$	$\mathbf{A}_{1\|1s}=\frac{1}{\mathrm{J}_{11s}}\mathbf{X}_{1\|1s}^T\mathbf{1}_{\mathrm{J}_{11s}}$
$\mathbf{B}_{2\|1s}=\frac{1}{\mathrm{J}_{21s}}\mathbf{X}_{2\|1s}^T\mathbf{I}_{\mathrm{J}_{21s}}\mathbf{X}_{2\|1s}$	$\mathbf{A}_{2\|1s}=\frac{1}{\mathrm{J}_{21s}}\mathbf{X}_{2\|1s}^T\mathbf{1}_{\mathrm{J}_{21s}}$
$\mathbf{B}_{2s}=\frac{1}{\mathrm{J}_{2s}}\mathbf{X}_{2s}^T\mathbf{I}_{\mathrm{J}_{2s}}\mathbf{X}_{2s}$	$\mathbf{A}_{2s}=\frac{1}{\mathrm{J}_{2s}}\mathbf{X}_{2s}^T\mathbf{1}_{\mathrm{J}_{2s}}$
$a_{1\|1s}=\ln(\mathrm{J}_{11s}),\ a_{2\|1s}=\ln(\mathrm{J}_{21s}),\ a_{2s}=\ln(\mathrm{J}_{2s})$	$p_{j\|11s}=\frac{1}{\mathrm{J}_{11s}},\ p_{j\|21s}=\frac{1}{\mathrm{J}_{21s}},\ p_{j\|2s}=\frac{1}{\mathrm{J}_{2s}}$
$a_{21\|1s}=\ln\left((\mathrm{J}_{11s})^{\frac{\lambda_1}{\mu_1}}+(\mathrm{J}_{21s})^{\frac{\lambda_2}{\mu_1}}\right)$	$\mathbf{X}_{\ell\|1s}^T=\left(\mathbf{x}_{1\|\ell 1s},\ldots,\mathbf{x}_{\mathrm{J}_{\ell 1}\|\ell 1s}\right)$
$\mathbf{x}_{j\|\ell 1s}^T=\left(x_{j1\|\ell 1s},\ldots,x_{jq_1\|\ell 1s}\right);\ \ell=1,2$	$\mathbf{X}_{2s}^T=\left(\mathbf{x}_{1\|2s},\ldots,\mathbf{x}_{\mathrm{J}_{2s}\|2s}\right);\ \mathbf{x}_{j\|2s}^T=\left(x_{j1\|2s},\ldots,x_{jq_1\|2s}\right)$
$p_{1s}=\frac{\left(\sum_{h=1}^{2}(\mathrm{J}_{h1s})^{\frac{\lambda_h}{\mu_1}}\right)^{\mu_1}}{\left(\sum_{h=1}^{2}(\mathrm{J}_{h1s})^{\frac{\lambda_h}{\mu_1}}\right)^{\mu_1}+(\mathrm{J}_{2s})^{\mu_2}};\ p_{2s}=1-p_{1s}$	$p_{1\|1s}=\frac{(\mathrm{J}_{11s})^{\frac{\lambda_1}{\mu_1}}}{\sum_{h=1}^{2}(\mathrm{J}_{h1s})^{\frac{\lambda_h}{\mu_1}}};\ p_{2\|1s}=1-p_{1\|1s},$

where $\mathbf{I}_r$ denotes an $r\times r$ -identity matrix and $\mathbf{1}_r$ is a r dimensional vector which all of its elements are one.

3.1 D-Optimal Criterion

Taking into account (5) and (6), consider the following designs to fit the model, which was introduced in Table 1:

$$\xi_n = \begin{Bmatrix} C_{1n} & C_{2n} & \dots & C_{S_n n} \\ w_{1n} & w_{2n} & \dots & w_{S_n n} \end{Bmatrix} \in \Xi_n \; ; S_n \geq q_1 + 4, \forall n \in N \qquad (7)$$

where $C_{sn}; s = 1,2,\dots,S_n$ denotes a choice set in n[th] class, which includes J alternatives. As the number of attributes (K) and their levels $(L_k; k = 1,2,\dots,K)$ increase (design (7)), the total number of possible classes (S_n) increases dramatically. In this situation, there is a need to search for techniques to reduce the number of support points or sample size $(S_n; \forall n \in N)$ such that we can obtain a reasonable number of choice sets (see *Graßhoff, et al. (2004)*). As have been told the D-optimality criterion in linear models typically leads to an optimal number of support points which is the same number of unknown parameters and the design takes an equal number of observations at each point (*Silvey, 1980, pp.42*). The bound also applies to most local optimality criteria and Bayesian criteria for linear models (*Chernoff, (1972)*). In contrast for non-linear models there is no such bound available on the number of support points. Then we consider condition $S_n \geq q; \forall n = 1,2,..., N$ (reduced) to obtain the D-optimal criterion for design (7).

The information matrix of design (7) is calculated as follow:

$$\mathbf{I}(\boldsymbol{\theta};\xi_n) = \sum_{s=1}^{S_n} w_{sn} \mathbf{I}(\boldsymbol{\theta}; C_{sn}), \qquad (8)$$

where w_{sn} is the weight (frequency) of the choice sets C_{sn}, $\mathbf{I}(\boldsymbol{\theta}; C_{sn})$ is the information matrix of choice set s in n[th] class, which is calculated by Lemma 1 and the local D-optimality criterion at $\boldsymbol{\theta}_0$ is $(\det(\mathbf{I}(\boldsymbol{\theta}_0;\xi_n)))^{-1}$ (θ_0 is true value of full parameters vector).

The D_b-optimal criterion in relation to the prior distribution $\pi(\boldsymbol{\theta})$ on the parameters can be defined as follows (*Atkinson, et al. (2007)*);

$$D_b(\xi_n) = E_\theta\left(\det(\mathbf{I}(\boldsymbol{\theta};\xi_n))^{-\frac{1}{q}}\right) = \int_{\Re^{q_1}} \int_{\Omega^M} \int_{\Lambda^{q2}} \left(\det(\mathbf{I}(\boldsymbol{\theta};\xi_n))^{-\frac{1}{q}}\right) \pi(\boldsymbol{\theta}) d\boldsymbol{\beta} d\boldsymbol{\mu} d\boldsymbol{\lambda}, \qquad (9)$$

where $\Re, \Omega$ and Λ are the spaces of $\boldsymbol{\beta}, \boldsymbol{\mu}$ and $\boldsymbol{\lambda}$, respectively and $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\lambda})$. Specifically, suppose that $\boldsymbol{\beta}$ is independent of $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ such that $\pi(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\mu}, \boldsymbol{\lambda})$. Consider, even, the independence between $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ to be not complex; means that $\pi(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\mu})\pi(\boldsymbol{\lambda})$. For example, uniform distribution for $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, and multivariate normal distribution for $\boldsymbol{\beta}\left(N_{q_1}(\boldsymbol{\beta}_0, \Sigma_\beta)\right)$.

Since usually there is not analytical expression for quantity (9), therefore, it may be approximated by the Monte-Carlo technique that takes a large number, R, of independent draws, $\boldsymbol{\theta}_r$, of $\boldsymbol{\theta}$ from the prior distribution, π, and the average of local criterion $\left(\det(\mathbf{I}(\boldsymbol{\theta};\xi_n))^{-\frac{1}{q}}\right)$ over all draws. Thus, the weighted D-criterion is approximated by:

$$D_b(\xi_n) \cong \frac{1}{R} \sum_{r=1}^{R} \left(\det(\mathbf{I}(\boldsymbol{\theta}_r;\xi_n))^{-\frac{1}{q}}\right) \qquad (10)$$

if

$$\Pr\left(\lim_{R\to\infty} \frac{1}{R} \sum_{r=1}^{R} (\det(\mathbf{I}(\boldsymbol{\theta}_r;\xi_n)))^{-\frac{1}{q}} = E_{\boldsymbol{\theta}}(\det(\mathbf{I}(\boldsymbol{\theta}_r;\xi_n)))^{-\frac{1}{q}}\right) = 1.$$

This case, in which ξ_n^* minimizes the D_b-approximate criterion (10), is called the D_b-optimal design and will be approximated by the solution:

$$\xi_n^* = \arg\min_{\xi_n \in \Xi_n} D_b(\xi_n) \tag{11}$$

then:

$$\xi_n^* = \begin{Bmatrix} C_{n1} & C_{n2} & \dots & C_{nS_n} \\ w_{n1}^* & w_{n2}^* & \dots & w_{nS_n}^* \end{Bmatrix}; \forall n \in N$$

is D-optimal design in $\Xi_n = \left\{ w_{sn} \middle| \sum_{s=1}^{S_n} w_{sn} = 1, 0 \le w_{sn} \le 1; s = 1,2,\dots,S_n \right\}$. Finally, according to (10) and (11) it can be theorized that $\xi_{n'}^*$ is the most suitable design for estimating parameters, if $n' \in N$ exists with the result that $\xi_{n'}^* = \arg\min_n D_b(\xi_n^*)$.

Explained previously, $(\det(\mathbf{I}(\boldsymbol{\theta}_0;\xi))^{-1})$ will be the local D-optimality criterion, where $\boldsymbol{\theta}_0$ is the true value of the parameters. Thus, we say that ξ^* is locally D-optimal design in Ξ if $\xi^* = \arg\min_{\xi \in \Xi} (\det \mathbf{I}(\boldsymbol{\theta}_0;\xi))^{-1}$ and it has been used here to obtain a locally D-optimal design.

Table 2: Three-level NMNL model with with six alternatives

First Nest (1)		Second Nest (2)
Sub-nest 1(1)	Sub-nest 2(1)	
a_1 , a_2	a_3 , a_4	a_5 , a_6

Illustration:

Here is a population with three attributes, each comprised of two levels. In this situation, consider a three-level NMNL model, which includes six possible alternatives in two nests (Table 2; $\sum_{m=1}^{M}\sum_{h=1}^{H_m} J_{hm} = 2+2+2$, where (1,1,1), (1,1,-1), (1,-1,1), (1,-1,-1), (-1,1,1), (-1,1,-1) characterize alternatives a_1, a_2, a_3, a_4, a_5 and a_6. Fitting this model (Table 2), consider experiment $(2+2+2)/5/\mathrm{S}$, based on Equation (5):

$$\mathrm{S} = \underbrace{\binom{2}{2}\binom{2}{2}\binom{2}{1}}_{\mathrm{S}_1} + \underbrace{\binom{2}{2}\binom{2}{1}\binom{2}{2}}_{\mathrm{S}_2} + \underbrace{\binom{2}{1}\binom{2}{2}\binom{2}{2}}_{\mathrm{S}_3} = 6 \cdot$$

In this case, three classes ($N = 3$) we found to define the design and because of $\mathrm{S}_n < 7; \forall n \in N$, we can combine them in order to define a suitable design. Thus, there are six choice sets (Table 3) with their design matrixes, as shown by Table 4. In this situation, suppose that $\beta_{11} = \beta_1, \beta_{21} = -\beta_1$, $\beta_{31} = \beta_3, \lambda_{11} = \lambda_1$ and $\lambda_{21} = \lambda_2$, for clarity, then $\boldsymbol{\theta}^T = (\beta_1, \beta_3, \mu_1, \mu_2, \lambda_1, \lambda_2)$ is full parameters vector. In this case and keeping to RUM conditions (*Gil-Molton and Hole, (2004)*), we will encounter the two conditions as follows:

1) $\mu_m \le \dfrac{1}{1-p_{ms}}; m = 1,2$ and 2) $\lambda_{hm} \le \dfrac{\mu_m}{(1+\mu_m p_{h|ms})(1-p_{ms})}; h = 1,2$, $\forall s \in \mathrm{S}$.

Table 3: Three-level NMNL model with six choice sets

Choice Set	First Nest (1)		Second Nest (2)
	Sub-nest (1)	Sub-nest (2)	
C_1	a_1 , a_2	a_3 , a_4	a_5
C_2	a_1 , a_2	a_3 , a_4	a_6
C_3	a_1 , a_2	a_3	a_5 , a_6
C_4	a_1 , a_2	a_4	a_5 , a_6
C_5	a_1	a_3 , a_4	a_5 , a_6
C_6	a_2	a_3 , a_4	a_5 , a_6

Table 4: The design matrix of three-level NMNL model with six choice sets

Choice Set	First Nest (1)		Second Nest (2)
	Sub-nest (1)	Sub-nest (2)	
C_1	$\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & -1\end{bmatrix}$	$\begin{bmatrix}1 & -1 & 1\\ 1 & -1 & -1\end{bmatrix}$	$\begin{bmatrix}-1 & 1 & 1\end{bmatrix}$
C_2	$\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & -1\end{bmatrix}$	$\begin{bmatrix}1 & -1 & 1\\ 1 & -1 & -1\end{bmatrix}$	$\begin{bmatrix}-1 & 1 & -1\end{bmatrix}$
C_3	$\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & -1\end{bmatrix}$	$\begin{bmatrix}1 & -1 & 1\end{bmatrix}$	$\begin{bmatrix}-1 & 1 & 1\\ -1 & 1 & -1\end{bmatrix}$
C_4	$\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & -1\end{bmatrix}$	$\begin{bmatrix}1 & -1 & 1\end{bmatrix}$	$\begin{bmatrix}-1 & 1 & 1\\ -1 & 1 & -1\end{bmatrix}$
C_5	$\begin{bmatrix}1 & 1 & 1\end{bmatrix}$	$\begin{bmatrix}1 & -1 & 1\\ 1 & -1 & -1\end{bmatrix}$	$\begin{bmatrix}-1 & 1 & 1\\ -1 & 1 & -1\end{bmatrix}$
C_6	$\begin{bmatrix}1 & 1 & -1\end{bmatrix}$	$\begin{bmatrix}1 & -1 & 1\\ 1 & -1 & -1\end{bmatrix}$	$\begin{bmatrix}-1 & 1 & 1\\ -1 & 1 & -1\end{bmatrix}$

For estimating the parameters of the model, which have been described on Table 2 and based on experiments $(2+2+2)/5/6$ and Equation (7), consider the following design:

$$\xi = \begin{Bmatrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 \\ w_1 & w_2 & w_3 & w_4 & w_5 & w_6 \end{Bmatrix} \in \Xi. \qquad (12)$$

The information matrix of design (12) is calculated by $\mathbf{I}(\boldsymbol{\theta};\xi)=\sum_{s=1}^{6} w_s \mathbf{I}(\boldsymbol{\theta};C_s)$. Specifically, let $\boldsymbol{\beta}=\mathbf{0}$. Now, according to Lemma 1 and Corollary 1, the elements of the information matrix $\mathbf{I}(\boldsymbol{\theta};C_s)$ can be calculated.

According to the rule of permutation, the levels of third attribute in choice sets C_1 and C_2 (second nest) will acquire permutation between these two choice sets. Also, permutation between the two choice sets C_3 and C_4 will exist with respect to permutation and the levels of the third attribute in the second sub-nest of the first nest. By permutation, the levels of the third attribute in the first sub-nest of the first nest, we will encounter permutation between the two choice sets C_5 and C_6. Thus, we can define a new design to fit the model, already introduced by

Table 2 and according to Table 3, as follows: $\xi' = \begin{Bmatrix} C_2 & C_1 & C_4 & C_3 & C_6 & C_5 \\ w'_1 & w'_2 & w'_3 & w'_4 & w'_5 & w'_6 \end{Bmatrix} \in \Xi.$ **(13)**

In this situation, in order to have equation between the two designs ξ (12) and ξ' (13), the following design can be considered:

$$\xi'' = \begin{Bmatrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 \\ \alpha_1 & \alpha_1 & \alpha_2 & \alpha_2 & \alpha_3 & \alpha_3 \end{Bmatrix} \in \Xi\,, \qquad (14)$$

where $\alpha_1 + \alpha_2 + \alpha_3 = \dfrac{1}{2}$.

Now, suppose that $\lambda_1 = \lambda_2 = \lambda$. Moreover, we know that $\lambda \le \mu_1$ and according Table 2, it is to be expected that $\mu_2 \ge \mu_1$. Then we can assume that $\mu_1 = 2\lambda$ and $\mu_2 = 4\lambda$, thus $\det(\mathbf{I}(\theta;\xi''))$ will be changed to a more function of λ, α_1 and α_2 where $\alpha_3 = \dfrac{1}{2} - (\alpha_1 + \alpha_2)$. In this situation, the RUM conditions will be upheld when $0 \le \lambda \le 0.25$. According to this condition for λ, some locally optimal design has been calculated in Table 5.

Table 5: $\mu_1 = 2\lambda$ and $\mu_2 = 4\lambda$, locally optimal design when $0 \le \lambda \le 0.25$.

λ	0.01	0.05	0.10	0.15	0.17	0.20	0.25
α_1^*	0.3092	0.3120	0.3150	0.3180	0.3195	0.3210	0.3270
α_2^*	0.0954	0.0940	0.0925	0.0910	0.0905	0.0899	0.0895
α_3^*	0.0954	0.0940	0.0925	0.0910	0.0900	0.0891	0.0877
$D(\xi'')$	0.00368	0.08985	0.35081	0.77476	0.98940	1.36021	2.11334

Table 5 shows that α_1^* increases as λ increases but α_2^* and α_3^* decrease when λ increases because of the combination of alternatives (and attributes) in two choice sets C_1 and C_2 are less similar than in the other choice sets. According to Table 4 we can observe that two sub-nests of the first nest in the choice sets C_1 and C_2 are equal but there are two different alternatives in second nest. In this situation, because of equation between λ_1 and λ_2, it is observed that α_1^* increases as λ increases. In choice sets C_3 and C_4, there are two different alternatives in the second sub-nest of the first nest. We can see a similar situation for choice sets C_5 and C_6, naturally, there are two different alternatives in the first sub-nest of the first nest (there is not change in the second nest for choice sets C_3 to C_6). With respect to the combination of the alternatives in the four choice sets C_3 to C_6, then a similar result for α_2^* and α_3^* will be obtained, so that these two weights (α_2^* and α_3^*) are almost equal and decrease as λ increases ($0 \le \lambda \le 0.15$). But, the decreasing trend of α_3^* is faster than α_2^* when $\lambda > 0.17$, then the combination of these attributes and their levels in the two choice sets C_5 and C_6 are more similar than the choice sets C_3 and C_4 (Table 4).

Now, suppose that $\lambda_1 = 0.1$, $\mu_1 = 0.15$ and $\mu_2 = 0.25$, then the RUM conditions are hold if $0 \le \lambda_2 \le 0.15$. In Table 6 several locally D-optimal designs based on Table 4 were obtained. In this situation, α_2^* increases as λ_2 increases but α_3^* decreases (Table 6). That means the alternatives in the second sub-nest (first nest) of choice sets

C_5 and C_6 are much similar, but the alternatives in choice sets C_3 and C_4 (second sub-nest) are much more dissimilar (Table 6).

Table 6: $\mu_1 = 0.15$, $\mu_2 = 0.25$ and $\lambda_1 = 0.1$, locally optimal design when $0 \le \lambda_2 \le 0.15$.

λ_2	0.01	0.05	0.06	0.08	0.10	0.12	0.15
α_1^*	0.3310	0.3338	0.3245	0.3133	0.3095	0.3103	0.3165
α_2^*	0.0000	0.0001	0.0285	0.0657	0.0953	0.1183	0.1458
α_3^*	0.1690	0.1661	0.1497	0.1210	0.0952	0.0713	0.0377
$D\left(\xi''\right)$	0.05879	0.15900	0.17522	0.19926	0.21502	0.22534	0.23478

Another Table 7, which includes was calculated some locally D-optimal designs based on $\mu_1 = \mu_2 = \lambda_1 = \lambda_2 = \lambda$. In this case, the RUM conditions hold if $0 < \lambda \le 1$. Table 7 denotes: α_1^* increases as λ increases, but α_2^* and α_3^* decrease. Noting the decreasing trend of α_2^* and α_3^*, we can observe that the decreasing trend of α_3^* is faster than α_2^*, because of more similarity (alternatives) in the choice sets C_5 and C_6 in contrast of that between C_3 and C_4.

Table 7: $\mu_1 = \mu_2 = \lambda_1 = \lambda_2 = \lambda$, locally optimal design when $0 \le \lambda \le 1$.

λ	0.05	0.10	0.15	0.20	0.30	0.40	0.50
α_1^*	0.2889	0.2908	0.2926	0.2943	0.2979	0.3011	0.3040
α_2^*	0.1055	0.1046	0.1037	0.1029	0.1012	0.0998	0.0988
α_3^*	0.1055	0.1046	0.1037	0.1028	0.1009	0.0991	0.0971
$D_b\left(\xi''\right)$	0.02689	0.10466	0.22959	0.39868	0.86051	1.47810	2.24747

With respect to fixed values for $\mu_1 = 0.1$ and $\lambda_1 = \lambda_2 = 0.08$ (Table 8), α_2^* and α_3^* are equal and they decrease as μ_2 increases, but α_1^* increases. Then, the alternatives in the second nest (choice sets C_3 to C_6) are more similar than the alternatives in the second nest of the choice sets C_1 and C_2.

Table 8: $\mu_1 = 0.1$ and $\lambda_1 = \lambda_2 = 0.08$, locally optimal design when $0 \le \mu_2 \le 1$.

μ_2	0.10	0.15	0.20	0.25	0.30	0.40	0.50
α_1^*	0.2964	0.3038	0.3074	0.3095	0.3110	0.3130	0.3150
α_2^*	0.1018	0.0981	0.0963	0.0952	0.0945	0.0933	0.0924
α_3^*	0.1018	0.0981	0.0963	0.0952	0.0945	0.0933	0.0924
$D\left(\xi''\right)$	0.09941	0.10191	0.10319	0.10409	0.10484	0.10619	0.10750

Suppose that $\mu_2 = 0.5$ and $\lambda_1 = 0.1,\ \lambda_2 = 0.2$. In this situation, RUM conditions hold if $0.2 \le \mu_1 \le 0.5$. Table 9 showed that α_1^* increases (almost as always, with a decreasing trend) as μ_1 increases. The third row of Table 9 denotes, α_2^* decreases (with a very weak decreasing trend) and α_3^* is equal zero as μ_1 increases. That means that the alternatives in the choice sets C_5 and C_6 are much more similar than are the others. And we can say, if $\mu_2 = 0.5$ and $\lambda_1 = 0.1,\ \lambda_2 = 0.2$ and $0.45 \le \mu_1$, then:

$$\xi''^* = \begin{Bmatrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 \\ 0.3364 & 0.3364 & 0.1636 & 0.1636 & 0.0000 & 0.0000 \end{Bmatrix}$$

is locally D-optimal design in $\Xi = \bigcup_{n=1}^{3} \Xi_n$.

According to the results which were obtained in the different classes in Table 5 to Table 9, we can say that the alternatives in the two choice sets C_1 and C_2 are less dissimilar than the others and the alternatives in the choice sets C_5 and C_6 are more similar than the others.

Table 9: $\mu_2 = 0.5$ and $\lambda_1 = 0.1,\ \lambda_2 = 0.2$, locally optimal design when $0.2 \le \mu_1 \le 0.5$.

μ_1	0.20	0.25	0.30	0.35	0.40	0.45	0.50
α_1^*	0.3333	0.3347	0.3355	0.3360	0.3363	0.3364	0.3364
α_2^*	0.1667	0.1653	0.1645	0.1640	0.1637	0.1636	0.1636
α_3^*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$D(\xi'')$	0.39644	0.56598	0.75789	0.97082	1.20394	1.45688	1.72959

Note: To obtain locally D-optimal design, $D(\xi'') = \left(\det\left(\mathbf{I}(\xi''; \theta_{\mathbf{0}})\right)\right)^{-1}$ (Table 5 to Table 9), Maple has been used with initial values $\alpha_2 = \alpha_3 = 0.2,\ \alpha_1 = 0.1$. The Sequential Quadratic Programming (SQP) method was also used and naturally the number 1000 was considered for the iteration limit.

4. CONCULSION

We know that in two-level NMNL model, all of alternatives are divided into several nests. According to IIA property which holds in each nest, it may be necessary that the alternatives of some nests or all of those to divide into several sub-nests.

In this paper, to fit three-level NMNL model it has been used D-optimal design, which is a function of the determinant of the information matrix. Also, we have calculated the information matrix of a three-level NMNL model for local D-optimality criterion when $\boldsymbol{\beta} = \mathbf{0}$. Based on example, we have discussed about different classes. We have observed, the optimal weights (choice sets) have an increasing or decreasing trend; of course, this trend depends on similarity between the alternatives in the choice sets. For example, there isn't much similarity between alternatives when the optimal weights (choice sets) have an increasing trend (dissimilarity parameter increased) and vice versa.

APPENDIX

Appendix AI:

The elements of the information matrix related to Lemma 1 are as follows

$$\mathbf{I}_{11s}=\frac{p_{1s}\cdot p_{1|1s}}{\lambda_1^2}\left(\mathbf{B}_{1|1s}-\mathbf{A}_{1|1s}\mathbf{A}_{1|1s}^T\right)+\frac{p_{1s}\cdot p_{2|1s}}{\lambda_2^2}\left(\mathbf{B}_{2|1s}-\mathbf{A}_{2|1s}\mathbf{A}_{2|1s}^T\right)+\frac{p_{2s}}{\mu_2^2}\left(\mathbf{B}_{2s}-\mathbf{A}_{2s}\mathbf{A}_{2s}^T\right)+$$
$$\frac{p_{1s}\cdot p_{1|1s}\cdot p_{2|1s}}{\mu_1^2}\left(\mathbf{A}_{1|1s}\mathbf{A}_{1|1s}^T+\mathbf{A}_{2|1s}\mathbf{A}_{2|1s}^T-\mathbf{A}_{1|1s}\mathbf{A}_{2|1s}^T-\mathbf{A}_{2|1s}\mathbf{A}_{1|1s}^T\right)+p_{1s}.p_{2s}\left(p_{1|1s}^2.\mathbf{A}_{1|1s}\mathbf{A}_{1|1s}^T+p_{2|1s}^2.\mathbf{A}_{2|1s}\mathbf{A}_{2|1s}^T+\mathbf{A}_{2s}\mathbf{A}_{2s}^T\right)+$$
$$p_{1s}.p_{2s}.p_{1|1s}.p_{2|1s}\left(\mathbf{A}_{1|1s}\mathbf{A}_{2|1s}^T+\mathbf{A}_{2|1s}\mathbf{A}_{1|1s}^T\right)-p_{1s}.p_{2s}\left[\left(p_{1|1s}.\mathbf{A}_{1|1s}+p_{2|1s}.\mathbf{A}_{2|1s}\right)\mathbf{A}_{2s}^T+\mathbf{A}_{2s}\left(p_{1|1s}.\mathbf{A}_{1|1s}^T+p_{2|1s}.\mathbf{A}_{2|1s}^T\right)\right]$$

$$\mathbf{I}_{12s}=\frac{p_{1s}\cdot p_{1|1s}\cdot p_{2|1s}}{\mu_1^3}\left(\lambda_2.a_{2|1s}-\lambda_1.a_{1|1s}\right)\left(\mathbf{A}_{1|1s}-\mathbf{A}_{2|1s}\right)$$
$$-\frac{p_{1s}\cdot p_{2s}}{\mu_1}\left(\lambda_1.p_{1|1s}.a_{1|1s}+\lambda_2.p_{2|1s}.a_{2|1s}-\mu_1.a_{21|1s}\right)\times\left(p_{1|1s}.\mathbf{A}_{1|1s}+p_{2|1s}.\mathbf{A}_{2|1s}-\mathbf{A}_{2s}\right)$$

$$\mathbf{I}_{13s}=-\frac{p_{2s}}{\mu_2^3}\left(\mathbf{B}_{2s}-\mathbf{A}_{2s}\mathbf{A}_{2s}^T\right)\boldsymbol{\beta}+p_{1s}.p_{2s}\left(p_{1|1s}.\mathbf{A}_{1|1s}+p_{2|1s}.\mathbf{A}_{2|1s}-\mathbf{A}_{2s}\right)\left(\frac{1}{\mu_2}\mathbf{A}_{2s}^T\boldsymbol{\beta}-a_{2s}\right)$$

$$\mathbf{I}_{14s}=-\frac{p_{1s}\cdot p_{1|1s}}{\lambda_1^3}\left(\mathbf{B}_{1|1s}-\mathbf{A}_{1|1s}\mathbf{A}_{1|1s}^T\right)\boldsymbol{\beta}+\frac{p_{1s}\cdot p_{1|1s}\cdot p_{2|1s}}{\lambda_1.\mu_1^2}\left(\mathbf{A}_{1|1s}-\mathbf{A}_{2|1s}\right)\left(\lambda_1.a_{1|1s}-\mathbf{A}_{1|1s}^T\boldsymbol{\beta}\right)$$
$$+\frac{p_{1s}\cdot p_{2s}\cdot p_{1|1s}}{\lambda_1}\left(p_{1|1s}.\mathbf{A}_{1|1s}+p_{2|1s}.\mathbf{A}_{2|1s}-\mathbf{A}_{2s}\right)\left(\lambda_1.a_{1|1s}-\mathbf{A}_{1|1s}^T\boldsymbol{\beta}\right)$$

$$\mathbf{I}_{15s}=-\frac{p_{1s}\cdot p_{2|1s}}{\lambda_2^3}\left(\mathbf{B}_{2|1s}-\mathbf{A}_{2|1s}\mathbf{A}_{2|1s}^T\right)\boldsymbol{\beta}+\frac{p_{1s}\cdot p_{1|1s}\cdot p_{2|1s}}{\lambda_2.\mu_1^2}\left(\mathbf{A}_{2|1s}-\mathbf{A}_{1|1s}\right)\left(\lambda_2.a_{2|1s}-\mathbf{A}_{2|1s}^T\boldsymbol{\beta}\right)$$
$$+\frac{p_{1s}\cdot p_{2s}\cdot p_{2|1s}}{\lambda_2}\left(p_{2|1s}.\mathbf{A}_{1|1s}+p_{1|1s}.\mathbf{A}_{1|1s}-\mathbf{A}_{2s}\right)\left(\lambda_2.a_{2|1s}-\mathbf{A}_{2|1s}^T\boldsymbol{\beta}\right)$$

$$I_{22s}=\frac{p_{1s}\cdot p_{1|1s}\cdot p_{2|1s}}{\mu_1^4}\left(\lambda_1.a_{1|1s}-\lambda_2.a_{2|1s}\right)^2+\frac{p_{1s}\cdot p_{2s}}{\mu_1^2}\left(\lambda_1.p_{1|1s}.a_{1|1s}+\lambda_2.p_{2|1s}.a_{2|1s}\right)^2$$
$$-\frac{2p_{1s}.p_{2s}.a_{21|1s}}{\mu_1}\left(\lambda_1.p_{1|1s}.a_{1|1s}+\lambda_2.p_{2|1s}.a_{2|1s}\right)+p_{1s}.p_{2s}.a_{21|1s}^2$$

$$I_{23s}=\frac{p_{1s}\cdot p_{2s}}{\mu_1.\mu_2}\left(\mu_1.a_{21|1s}-\lambda_1.p_{1|1s}.a_{1|1s}-\lambda_2.p_{2|1s}.a_{2|1s}\right)\left(\mathbf{A}_{2s}^T\boldsymbol{\beta}-\mu_2.a_{2s}\right)$$

$$I_{24s}=\frac{p_{1s}\cdot p_{1|1s}\cdot p_{2|1s}}{\lambda_1.\mu_1^3}\left(\lambda_1.a_{1|1s}-\lambda_2.a_{2|1s}\right)\left(\mathbf{A}_{1|1s}^T\boldsymbol{\beta}-\lambda_1.a_{1|1s}\right)$$
$$+\frac{p_{1s}\cdot p_{2s}\cdot p_{1|1s}}{\lambda_1.\mu_1}\left(\lambda_1.p_{1|1s}.a_{1|1s}+\lambda_2.p_{2|1s}.a_{2|1s}-\mu_1.a_{21|1s}\right)\left(\mathbf{A}_{1|1s}^T\boldsymbol{\beta}-\lambda_1.a_{1|1s}\right)$$

$$I_{25s}=\frac{p_{1s}\cdot p_{1|1s}\cdot p_{2|1s}}{\lambda_2.\mu_1^3}\left(\lambda_2.a_{2|1s}-\lambda_1.a_{1|1s}\right)\left(\mathbf{A}_{2|1s}^T\boldsymbol{\beta}-\lambda_2.a_{2|1s}\right)+\frac{p_{1s}\cdot p_{2s}\cdot p_{2|1s}}{\lambda_2.\mu_1}\left(\lambda_1.p_{1|1s}.a_{1|1s}+\lambda_2.p_{2|1s}.a_{2|1s}-\mu_1.a_{21|1s}\right)\left(\mathbf{A}_{2|1s}^T\boldsymbol{\beta}-\lambda_2.a_{2|1s}\right)$$

$$I_{33s}=\frac{p_{2s}}{\mu_2^4}\boldsymbol{\beta}^T\left(\mathbf{B}_{2s}-\mathbf{A}_{2s}\mathbf{A}_{2s}^T\right)\boldsymbol{\beta}+\frac{p_{1s}\cdot p_{2s}}{\mu_2^2}\left(\boldsymbol{\beta}^T\mathbf{A}_{2s}-\mu_2.a_{2s}\right)\left(\mathbf{A}_{2s}^T\boldsymbol{\beta}-\mu_2.a_{2s}\right)$$

$$I_{34s}=-\frac{p_{1s}\cdot p_{2s}\cdot p_{1|1s}}{\lambda_1.\mu_2}\left(\boldsymbol{\beta}^T\mathbf{A}_{2s}-\mu_2.a_{2s}\right)\left(\mathbf{A}_{1|1s}^T\boldsymbol{\beta}-\lambda_1.a_{1|1s}\right)$$

$$I_{35s}=-\frac{p_{1s}\cdot p_{2s}\cdot p_{2|1s}}{\lambda_2.\mu_2}\left(\boldsymbol{\beta}^T\mathbf{A}_{2s}-\mu_2.a_{2s}\right)\left(\mathbf{A}_{2|1s}^T\boldsymbol{\beta}-\lambda_2.a_{2|1s}\right)$$

$$I_{44s}=\frac{p_{1s}\cdot p_{1|1s}}{\lambda_1^4}\boldsymbol{\beta}^T\left(\mathbf{B}_{1|1s}-\mathbf{A}_{1|1s}\mathbf{A}_{1|1s}^T\right)\boldsymbol{\beta}+\frac{p_{1s}\cdot p_{1|1s}}{\lambda_1^2.\mu_1^2}\left(p_{2|1s}+\mu_1^2.p_{2s}.p_{1|1s}\right)\left(\boldsymbol{\beta}^T\mathbf{A}_{1|1s}-\lambda_1.a_{1|1s}\right)\left(\mathbf{A}_{1|1s}^T\boldsymbol{\beta}-\lambda_1.a_{1|1s}\right)$$

$$I_{45s}=-\frac{p_{1s}\cdot p_{1|1s}\cdot p_{2|1s}}{\lambda_1.\lambda_2.\mu_1^2}\left(1-p_{2s}.\mu_1^2\right)\left(\boldsymbol{\beta}^T\mathbf{A}_{1|1s}-\lambda_1.a_{1|1s}\right)\left(\mathbf{A}_{2|1s}^T\boldsymbol{\beta}-\lambda_2.a_{2|1s}\right)$$

$$I_{55s}=\frac{p_{1s}\cdot p_{2|1s}}{\lambda_2^4}\boldsymbol{\beta}^T\left(\mathbf{B}_{2|1s}-\mathbf{A}_{2|1s}\mathbf{A}_{2|1s}^T\right)\boldsymbol{\beta}+\frac{p_{1s}\cdot p_{2|1s}}{\lambda_2^2.\mu_1^2}\left(p_{1|1s}+\mu_1^2.p_{2s}.p_{2|1s}\right)\left(\boldsymbol{\beta}^T\mathbf{A}_{2|1s}-\lambda_2.a_{2|1s}\right)\left(\mathbf{A}_{2|1s}^T\boldsymbol{\beta}-\lambda_2.a_{2|1s}\right)$$

where

$\mathbf{B}_{1|1s}=\mathbf{X}_{1|1s}^T\mathbf{P}_{.|11s}\mathbf{X}_{1|1s}$, $\mathbf{B}_{2|1s}=\mathbf{X}_{2|1s}^T\mathbf{P}_{.|21s}\mathbf{X}_{2|1s}$, $\mathbf{A}_{1|1s}=\mathbf{X}_{1|1s}^T\mathbf{p}_{.|11s}$, $\mathbf{A}_{2|1s}=\mathbf{X}_{2|1s}^T\mathbf{p}_{.|21s}$

$\mathbf{B}_{2s}=\mathbf{X}_{2s}^T\mathbf{P}_{.|2s}\mathbf{X}_{2s}$, $\mathbf{A}_{2s}=\mathbf{X}_{2s}^T\mathbf{p}_{.|2s}$, $\mathbf{X}_{\ell|1s}^T=\left(\mathbf{x}_{1|\ell 1s},\dots,\mathbf{x}_{\mathrm{J}_{\ell 1s}|\ell 1s}\right)$,

$\mathbf{x}_{j|\ell 1s}^T=\left(x_{j1|\ell 1s},\dots,x_{jq_1|\ell 1s}\right);\ \ell=1,2$, $\mathbf{P}_{.|\ell 1s}=diag\left(p_{1|\ell 1s},\dots,p_{\mathrm{J}_{\ell 1s}|\ell 1s}\right)$

$\mathbf{p}_{.|\ell 1s}=diag\left(p_{1|\ell 1s},\dots,p_{\mathrm{J}_{\ell 1s}|\ell 1s}\right)$, $\mathbf{X}_{2s}^T=\left(\mathbf{x}_{1|2s},\dots,\mathbf{x}_{\mathrm{J}_{2s}|2s}\right)$, $\mathbf{x}_{j|2s}^T=\left(x_{j1|2s},\dots,x_{jq_1|2s}\right)$

, $\mathbf{P}_{.|2s}=diag\left(p_{1|2s},\dots,p_{\mathrm{J}_{2s}|2s}\right)$, $\mathbf{p}_{.|2s}=diag\left(p_{1|2s},\dots,p_{\mathrm{J}_{2s}|2s}\right)$ $a_{1|1s}=\ln\left(\sum_{j=1}^{\mathrm{J}_{11s}}e^{\left(\frac{\mathbf{X}_{j|11s}^T\boldsymbol{\beta}}{\lambda_1}\right)}\right)$,

$a_{2|1s}=\ln\left(\sum_{j=1}^{\mathrm{J}_{21s}}e^{\left(\frac{\mathbf{X}_{j|21s}^T\boldsymbol{\beta}}{\lambda_2}\right)}\right)$, $a_{2s}=\ln\left(\sum_{j=1}^{\mathrm{J}_{2s}}e^{\left(\frac{\mathbf{X}_{j|2s}^T\boldsymbol{\beta}}{\mu_2}\right)}\right)$ $p_{j|11s}=\dfrac{e^{\left(\frac{\mathbf{X}_{j|11s}^T\boldsymbol{\beta}}{\lambda_1}\right)}}{\sum_{j=1}^{\mathrm{J}_{11s}}e^{\left(\frac{\mathbf{X}_{j|11s}^T\boldsymbol{\beta}}{\lambda_1}\right)}}$,

$p_{j|21s}=\dfrac{e^{\left(\frac{\mathbf{X}_{j|21s}^T\boldsymbol{\beta}}{\lambda_2}\right)}}{\sum_{j=1}^{\mathrm{J}_{21s}}e^{\left(\frac{\mathbf{X}_{j|21s}^T\boldsymbol{\beta}}{\lambda_2}\right)}}$, $p_{j|2s}=\dfrac{e^{\left(\frac{\mathbf{X}_{j|2s}^T\boldsymbol{\beta}}{\mu_2}\right)}}{\sum_{j=1}^{\mathrm{J}_{2s}}e^{\left(\frac{\mathbf{X}_{j|2s}^T\boldsymbol{\beta}}{\mu_2}\right)}}$,

$$a_{21|1s} = \ln\left(\left(\sum_{j=1}^{J_{11s}} e^{\left(\frac{\mathbf{X}_{j|11s}^T\boldsymbol{\beta}}{\lambda_1}\right)}\right)^{\frac{\lambda_1}{\mu_1}} + \left(\sum_{j=1}^{J_{21s}} e^{\left(\frac{\mathbf{X}_{j|21s}^T\boldsymbol{\beta}}{\lambda_2}\right)}\right)^{\frac{\lambda_2}{\mu_1}}\right),$$

$$p_{1s} = \frac{\left(\sum_{h=1}^{2}\left(\sum_{j=1}^{J_{11s}} e^{\left(\frac{\mathbf{X}_{j|11s}^T\boldsymbol{\beta}}{\lambda_1}\right)}\right)^{\frac{\lambda_h}{\mu_1}}\right)^{\mu_1}}{\left(\sum_{h=1}^{2}\left(\sum_{j=1}^{J_{11s}} e^{\left(\frac{\mathbf{X}_{j|11s}^T\boldsymbol{\beta}}{\lambda_1}\right)}\right)^{\frac{\lambda_h}{\mu_1}}\right)^{\mu_1} + \left(\sum_{j=1}^{J_{2s}} e^{\left(\frac{\mathbf{X}_{j|2s}^T\boldsymbol{\beta}}{\mu_2}\right)}\right)^{\mu_2}}, \quad p_{1|1s} = \frac{\left(\sum_{j=1}^{J_{11s}} e^{\left(\frac{\mathbf{X}_{j|11s}^T\boldsymbol{\beta}}{\lambda_1}\right)}\right)^{\frac{\lambda_1}{\mu_1}}}{\sum_{h=1}^{2}\left(\sum_{j=1}^{J_{h1s}} e^{\left(\frac{\mathbf{X}_{j|h1s}^T\boldsymbol{\beta}}{\lambda_h}\right)}\right)^{\frac{\lambda_h}{\mu_1}}}$$

5. REFRENCES:

[1] Atkinson. A.C. and A.N. Donev and R.D. Tobias (2007). Optimum experimental designs, with SAS, *Oxford Univ. Press*

[2] Ben-Akiva, M. (1973). The structure of travel demand models, *PhD Thesis, MIT.*

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

[4] Chernoff, H. (1972). Sequential analysis and optimal design. *Society for Industrial and Applied Mathematics, Philadelphia PA.*

[5] Daly, A. and S. Zachary (1978). Improved multiple choice models, in D.Hensher and M.Dalvi, eds., *Determinates of Travel Choice, Saxon House, Sussex.*

[6] Gil-Molton,M. and Hole, A. (2004). Tests for the consistency of three-level tested logit models with utility maximization, *Economics Letters,85, 133-137.*

[7] Graßhoff U., Großmann H., Holling H. and Schwabe R., (2004). Optimal Designs for Main-Effects in Linear Paired Comparison Models, *Journal of Statistical Planning and Inference, 126:361-376.*

[8] Herriges, JA. And CL. Kling (1996). Testing the consistency of nested logit models with utility maximization, *Econometrics Letters, Vol.50, No.1, 33-39.*

[9] McFadden, D., (1978). Modeling the choice of residential location, in: A. Karlquist et. al., eds., Spatial interaction theory and residential location, *North-Holland, Amsterdam 75-96.*

[10] Silvey, S.D.(1980). Optimal design. *Chapman and Hall, London.*