On Rank-Ordered Nested Multinomial Logit Model and D-Optimal Design for this Model

Habib Jafari

Razi University

Abstract. In contrast to the classical discrete choice experiment, the respondent in a rank-order discrete choice experiment, is asked to rank a number of alternatives instead of the preferred one. In this paper, we study the information matrix of a rank-order nested multinomial logit model (RO.NMNL) and introduce local D-optimality criterion, then we obtain Locally D-optimal design for RO.NMNL models in the discrete choice experiment.

Keywords. Discrete choice; choice set; information matrix; D-Optimal criterion; locally D-Optimal design.

MSC 2010: 62K05.

1 Introduction

In classical logit model the respondent chooses just one alternative in the choice set with the highest utility (Jafari, 2010). There is another model in logit family called Rank-Order logit model. In this model the respondent can choose more than one alternative instead of one.

A rank-order conjoint experiment measures the importance of the features of goods or services by asking the respondent to rank a certain number of alternatives within the choice sets. Data from a rank-order experiment can be analyzed by the rank-ordered exploded Logit like Multinmial Logit

(MNL) and Nested Multinomial Logit (NMNL) models (Beggs, et al., 1981 and Hausman and Ruud, 1987).

In theory, when individuals are asked to rank the alternatives instead of only choosing the most preferred option, the parameters of the choice model and hence, the preferences can be estimated more efficiently. However, in practice, the respondents may be unable to perform (part of) the ranking task. This may be due to several reasons; First of all, respondents may not be able to perform the task itself. Secondly, the respondent may not be able to distinguish between his less-preferred alternatives. In any case, straightforwardly, using the reported rankings may lead to a substantial bias in the parameter estimates in the rank-order logit model (Chapman and Staelin, 1982). To solve this problem, Chapman and Staelin (1982) suggested only to use the first few ranks in the estimation. They considered several rules to determine the appropriate number of ranks to use, in their words "the explosion depth". One of these rules is based on a pooling test for the equality of parameter estimates based on different rank information. Chapman and Staelin (1982) proposed an alternative method to test the number of ranks to be used in estimations. However, in both approaches this number is assumed to be the same for all respondents. If ranking capabilities differ across individuals, it may lead to an efficiency loss.

In this paper, first, I have studied about the Rank-Order Multinomial Logit (RO.MNL) model and the Rank-Order Nested Multinomial Logit (RO.NMNL) model. Then, I have used the D-optimality criterion to obtain locally D-optimal design for RO.NMNL model. So, I have organized my paper as follows:

In Section 2, I review the rank-ordered multinomial logit model (Vermeulen, et al.), then in Section 3, the information matrix related to the rank-ordered NMNL model is calculated and I define a special class of design and the method of obtaining locally D-optimal design for that.

2   Rank-Order MNL (RO.MNL) Model

The Rank-Order Logit Model was introduced to the literature by Beggs, et al. (1981). In this approach, each ranking of a choice set is converted into a number of independent pseudo-choices. In this way, each ranking of alternatives in a choice set is considered as a sequential and conditional choice task. The alternative with the first rank is imagined as the preferred alternative

(with the highest utility in the classical method) of the entire choice set. The next ranked alternatives are viewed as the preferred alternatives of the choice sets consisting of all alternatives except the ones with a better ranking. In the resulting rank-ordered multinomial logit model, a ranking of a set of J_c alternatives is thus seen as a series of $J_c - 1$ choices. In this situation as in the classical MNL model there are $\mathcal{J}$ alternatives, $\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_{\mathcal{J}}$. Then we consider $\mathcal{C}$ choice sets each with $J_c > 1; \forall c$ alternatives, where there are $2^{\mathcal{J}} - (\mathcal{J} + 1)$ choice sets each with at least two alternatives.

- $\mathbb{C}_c = \{a_{1c}, a_{2c}, \ldots, a_{J_c c}\}$ denotes a choice set with J_c alternatives, where a_{jc} is the alternative j of choice set $\mathbb{C}_c$.

The rank of an alternative is determined by its utility. The utility of the alternative a_j in choice set $\mathbb{C}_c$ experienced by respondent i is modeled as (by effects-type coding),

$$U_{jc} = \mathbf{f}^T(a_{jc})\boldsymbol{\beta} + \varepsilon_{jc} = \sum_{k=1}^{K} \sum_{\ell=1}^{L_k - 1} f_{k\ell}(a_{jc})\beta_{k,\ell} + \varepsilon_{jc}; \quad j = 1, 2, \ldots, J_c, \quad (1)$$

where:

- $\mathbf{f}(a_{jc}) = (\mathbf{f}_1(a_{jc}), \ldots, \mathbf{f}_k(a_{jc}), \ldots, \mathbf{f}_K(a_{js}))^T$;
  $\mathbf{f}_k(a_{jc}) = (f_{k1}(a_{jc}), \ldots, f_{k\ell}(a_{jc}), \ldots, f_{kL_k-1}(a_{jc}))^T$

is the characteristics of attributes (there are K attributes each with $L_k; \forall k = 1, 2, \ldots, K$ levels) related to alternative a_j (Main-effects model), which is chosen by the individual i and:

- $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k, \ldots, \boldsymbol{\beta}_K)^T; \quad \boldsymbol{\beta}_k = (\beta_{k,1}, \ldots, \beta_{k,\ell}, \ldots, \beta_{k,L_k-1})^T$

is p-dimensional vector of parameters ($p = \sum_{k=1}^{K}(L_k - 1)$) where $\sum_{\ell=1}^{L_k} \beta_{k\ell} = 0$ and:

- ε_{js} are error terms which have $i.i.d$ extreme value distribution (type II).

Now, suppose that $Y_{(1)}, \ldots, Y_{(j)}, \ldots, Y_{(J_c)}$ denote the rank-alternative variables of a choice set with J_c alternatives. For example, $Y_{(1)} = r_{(1)}$ means that alternative $a_{r_{(1)}}; (r_{(1)} \in \{1, 2, \ldots, J_c\})$ has the first rank with the highest utility ($U_{r_{(1)}c} = \max_{j \in \mathbb{C}_c} U_{jc}$) and $Y_{(2)} = r_{(2)}$ means that alternative $a_{r_{(2)}}$ has the second rank, its utility is less than the utility of $a_{r_{(1)}}$ and greater than

www.SID.ir

the remaining alternatives ($U_{r_{(2)}c} = \max_{j \in \mathbb{C}_{c(r_{(1)})}} U_{jc}$ and $U_{r_{(2)}c} < U_{r_{(1)}c}$) or we can say, $Y_{(2)}$ denotes an alternative with the second rank in the original choice set. However, after removing the alternative with the first rank, $Y_{(2)}$ will be denoted an alternative with the first rank in the new choice set (a choice set without first rank alternative). In this state $\mathbb{C}_{c(j)}$ denotes a choice set (choice set c), which excludes alternative a_j. In this situation and to obtain choice probabilities, we can also define the observation variables as follows:

$$Y_{r_{(1)}c} = \begin{cases} 1, & U_{r_{(1)}c} = \max_{j \in \mathbb{C}_c} U_{jc} \\ 0, & \text{Otherwise} \end{cases},$$

$$Y_{r_{(2)}s} = \begin{cases} 1, & U_{r_{(2)}c} = \max_{j \in \mathbb{C}_{c(r_{(1)})}} U_{jc} \\ 0, & \text{Otherwise} \end{cases}$$

and so on. Now, the probabilities of rank-order alternatives can be defined as below:

$$P(Y_{(1)} = r_{(1)}) = P_{r_{(1)}c} = \frac{\exp\left\{\mathbf{f}^T(a_{r_{(1)}c})\boldsymbol{\beta}\right\}}{\sum_{\ell \in \mathbb{C}_s} \exp\left\{\mathbf{f}^T(a_{\ell c})\boldsymbol{\beta}\right\}}; \quad a_{r_{(1)}} \in \mathbb{C}_c,$$

$$P(Y_{(2)} = r_{(2)}) = P_{r_{(2)}c} = \frac{\exp\left\{\mathbf{f}^T(a_{r_{(2)}c})\boldsymbol{\beta}\right\}}{\sum_{\ell \in \mathbb{C}_{c(r_{(1)})}} \exp\left\{(\mathbf{f}^T(a_{\ell c})\boldsymbol{\beta}\right\}}; \quad a_{r_{(2)}} \in \mathbb{C}_{c(r_{(1)})},$$

$$P(Y_{(3)} = r_{(3)}) = P_{r_{(3)}c} = \frac{\exp\left\{\mathbf{f}^T(a_{r_{(3)}c})\boldsymbol{\beta}\right\}}{\sum_{\ell \in \mathbb{C}_{c(r_{(1)},r_{(2)})}} \exp\left\{\mathbf{f}^T(a_{\ell c})\boldsymbol{\beta}\right\}}; \quad a_{r_{(3)}} \in \mathbb{C}_{c(r_{(1)},r_{(2)})},$$

$$\vdots$$

$$P(Y_{(J_c)} = r_{(J_c)}) = P_{r_{(J_c)}c} = 1; \quad a_{r_{(J_c)}} \in \mathbb{C}_{c(r_{(1)},r_{(2)},...,r_{(J_c-1)})},$$

where $P(A)$ denotes probability of an event A. In this model, we expect to obtain more information about the preferences of respondents than in the classical conjoint choice experiment. Also, if we use the same number choice sets (to compare with classical conjoint experiment), the parameters of model can be more accurately estimated. Chapman and Staelin (1982) have also attempted to achieve a desired degree of precision of the estimates, less choice sets are required in a rank-order conjoint experiment.

This model may be better than the classical model but there are some problems in its application, for example, the major disadvantage of using

a rank-order conjoint experiment is the weak link with reality: in real life, respondents choose the alternative which they like most and hardly ever select a second best item, and the other is the alternatives with lower ranking.

According to Chapman and Staelin (1982), lower rankings are less reliable if the number of alternatives to rank is high. Of course, we try to solve this problem by considering the number of alternatives less than the total number of alternatives (in a choice set).

Now, considering a combination of alternatives like $(a_{r_{(1)}}, a_{r_{(2)}}, \ldots, a_{r_{(J_c)}})$; $r_{(1)}, r_{(2)}, \ldots, r_{(J_c)} \in \{1, 2, \ldots, J_c\}$ to analyze data, means that the alternative $a_{r_{(1)}}$ has the first rank, alternative $a_{r_{(2)}}$ has the second rank and so forth $(r_{(j)} \neq r_{(j')}, j \neq j' = 1, 2, \ldots, J_c)$. Since there are J_c cases for $r_{(1)}$ and $J_c - 1$ cases for $r_{(2)}$ and at last there is just one case for $r_{(J_c)}$, we define the following variable to introduce observation variable in RO.MNL model:

$$\mathcal{Y}_{(r_{(1)}r_{(2)}\ldots r_{(J_c)})} = \begin{cases} 1, & \text{if } Y_{(1)} = r_{(1)}, Y_{(2)} = r_{(2)}, \ldots, Y_{(J_c)} = r_{(J_c)} \\ 0, & \text{Otherwise,} \end{cases} \tag{2}$$

where $E(\mathcal{Y}_{(r_{(1)}r_{(2)}\ldots r_{(J_c)})}) = P_{r_{(1)}c} \cdot P_{r_{(2)}c} \ldots P_{r_{(J_c)}c}$.

To define a design, $\mathcal{S}$ $(\mathcal{S} \subset \mathcal{C})$ choice sets each with $J_s = J; \forall s$ alternatives will be considered. Thus, according to the choice rank probabilities and Equation (2), the Likelihood function can be defined as follows:

$$L(\mathbb{C}_s, \boldsymbol{\beta}) = \underbrace{\prod_{r_{(1)}=1}^{J} \prod_{r_{(2)}=1}^{J} \cdots \prod_{r_{(J_s)}=1}^{J}}_{r_{(1)} \neq r_{(2)} \neq \ldots \neq r_{(J)}} \left(P_{r_{(1)}s} \cdot P_{r_{(2)}s} \ldots P_{r_{(J)}s} \right)^{\mathcal{Y}_{(r_{(1)}r_{(2)}\ldots r_{(J)})}}$$

and the log-likelihood function as:

$$\begin{aligned}
\ell(\mathbb{C}_s, \boldsymbol{\beta}) &= \underbrace{\sum_{r_{(1)}=1}^{J} \sum_{r_{(2)}=1}^{J} \cdots \sum_{r_{(J)}=1}^{J}}_{r_{(1)} \neq r_{(2)} \ldots \neq r_{(J)}} \mathcal{Y}_{(r_{(1)}r_{(2)}\ldots r_{(J)})} \cdot \ln \left(P_{r_{(1)}s} \cdot P_{r_{(2)}s} \ldots P_{r_{(J)}s} \right) \\
&= \underbrace{\sum_{r_{(1)}=1}^{J} \sum_{r_{(2)}=1}^{J} \cdots \sum_{r_{(J)}=1}^{J}}_{r_{(1)} \neq r_{(2)} \ldots \neq r_{(J)}} \mathcal{Y}_{(r_{(1)}r_{(2)}\ldots r_{(J)})} \\
&\quad \times \left(\ln P_{r_{(1)}s} + \ln P_{r_{(2)}s} + \ldots + \ln P_{r_{(J)}s} \right).
\end{aligned} \tag{3}$$

Lemma 1. *The information matrix based on Log-Likelihood (3) is obtained by:*

$$\mathbf{I}_{\mathrm{R(MNL)}}(\mathbb{C}_s, \boldsymbol{\beta}) = \mathbf{I}_{\mathrm{MNL}}(\mathbb{C}_s, \boldsymbol{\beta}) + \sum_{r_{(1)}=1}^{J} P_{r_{(1)}s} \cdot \mathbf{I}_{\mathrm{MNL}}(\mathbb{C}_{s(r_{(1)})}, \boldsymbol{\beta})$$

$$+ \sum_{r_{(1)} \neq r_{(2)}}^{J} P_{r_{(1)}s} P_{r_{(2)}} \cdot \mathbf{I}_{\mathrm{MNL}}(\mathbb{C}_{s(r_{(1)}, r_{(2)})}, \boldsymbol{\beta}) +$$

$$+ \ldots + \sum_{r_{(1)} \neq r_{(2)} \neq \ldots \neq r_{(J-2)}}^{J} P_{r_{(1)}} P_{r_{(2)}} \ldots P_{r_{(J-2)}}$$

$$\times \mathbf{I}_{\mathrm{MNL}}(\mathbb{C}_{s(r_{(1)}, r_{(2)}, \ldots, r_{(J-2)})}, \boldsymbol{\beta}),$$

where:

- $\mathbf{I}_{\mathrm{MNL}}(\mathbb{C}_s, \boldsymbol{\beta}) = \mathbf{F}_s^T (\mathbf{P}_s - \mathbf{p}_s \mathbf{p}_s^T) \mathbf{F}_s$: *The information matrix of a discrete choice experiment (classic) by choice set with size J (Sandor and Wedel, 2001), where:*
$$\mathbf{F}_s = (\mathbf{f}(a_{1s}), \mathbf{f}(a_{2s}), \ldots, \mathbf{f}(a_{Js}))^T.$$

- $\mathbf{I}_{\mathrm{MNL}}(\mathbb{C}_{s(j)}, \boldsymbol{\beta}) = \mathbf{F}_{s(j)}^T (\mathbf{P}_{s(j)} - \mathbf{p}_{s(j)} \mathbf{p}_{s(j)}^T) \mathbf{F}_{s(j)}$.

- $\mathbb{C}_{s(j)}$ *denotes a choice set without alternative* a_j.

- $\mathbf{p}_{s(j)} = (p_{1s}, \ldots, p_{j-1s}, p_{j+1s}, \ldots, p_{Js})^T$ *is a $(J-1)$-dimensional vector containing the probabilities.*

- $\mathbf{P}_{s(j)}$ *is a diagonal matrix with the elements of* $\mathbf{p}_{s(j)}$ *on its diagonal.*

- $\mathbf{F}_{s(j)} = (\mathbf{f}(a_{1s}), \ldots, \mathbf{f}(a_{j-1s}), \mathbf{f}(a_{j+1s}), \ldots, \mathbf{f}(a_{Js}))^T$ *is the $((J-1) \times p)$ design matrix containing all attribute levels of the profiles in choice set, except profile* a_j.

The expression for the information matrix of a rank-order experiment proves that asking the respondents to rank the alternatives in a choice set provides extra information. Since this difference $\mathbf{I}_{\mathrm{R(MNL)}}(\mathbb{C}_s, \boldsymbol{\beta}) - \mathbf{I}_{\mathrm{MNL}}(\mathbb{C}_s, \boldsymbol{\beta})$ is a non-negative definite matrix, it ensures that the amount of information in a ranking experiment is more extensive than in classical experiments in a choice experiment. In other words, ranking is always better than its classical counterpart.

Similar to the discrete choice experiment which used the criterion (Atkinson, et al., 2007):

$$\Psi'(\xi, \boldsymbol{\beta}) = \ln\left[\det\{\mathbf{I}_{\mathrm{MNL}}(\xi, \boldsymbol{\beta})\}\right]^{-1}$$

to obtain the local D-optimal criterion, for RO.MNL, also, a similar situation was used, it means that:

$$\Psi'_R(\xi, \boldsymbol{\beta}) = \ln\left[\det\{\mathbf{I}_{\mathrm{R(MNL)}}(\xi, \boldsymbol{\beta})\}\right]^{-1}$$

to obtain locally D-optimal design, where ξ is a design which includes $\mathcal{S}$ choice sets, $\mathbb{C}_1, \mathbb{C}_2, \ldots, \mathbb{C}_{\mathcal{S}}$.

3 Rank-Order Nested MNL (RO.NMNL) Models

Suppose that there are J alternatives (in choice set $\mathbb{C}_s$), which have an upper ranking in comparison to others and which have been divided into M nests each with $\mathcal{J}_m; \forall m$ alternatives, $\tilde{a}_{1m}, \tilde{a}_{2m}, \ldots, \tilde{a}_{\mathcal{J}_m m}$. To analyze data, some choice sets each with the same number of alternatives will be considered, where:

$$\mathbb{C}_m = \{a_{1m}, a_{2m}, \ldots, a_{J_m m}\}$$

denotes a choice set with J_m alternatives and:

$$\mathbb{C} = \bigcup_{m=1}^{M} \mathbb{C}_m; \ \mathbb{C}_m \bigcap \mathbb{C}_{m'} = \phi, \quad \forall m \neq m',$$

where $\mathbb{C}$ includes $J = \sum_{m=1}^{M} J_m$ alternatives. In this case, the utility of choosing an alternative a_{jm} by individual i is calculated as follows (w.r.t choice set $\mathbb{C}_s$):

$$U_{jms} = U_{j(m)s} + U_{ms},$$

where $U_{j(m)s} = v_{j(m)s} + \varepsilon_{j(m)s}$ and $U_{ms} = v_{ms} + \varepsilon_{ms}$ (Train, 2003):

- $v_{j(m)s} = \mathbf{f}^T(a_{jms})\boldsymbol{\beta}; \quad \mathbf{f}(a_{jms}) = (\mathbf{f}_1^T(a_{jms}), \mathbf{f}_2^T(a_{jms}), \ldots, \mathbf{f}_K^T(a_{jms}))^T;$

$$\mathbf{f}_k(a_{jms}) = (f_{k1}(a_{jms}), f_{k2}(a_{jms}), \ldots, f_{kL_k-1}(a_{jms}))^T.$$

- $v_{ms} = E(\max_{j \in \mathbb{C}_{ms}} U_{j(m)s}),$

www.SID.ir

- $\varepsilon_{j(m)s}$ have EVD with $\text{corr}(\varepsilon_{j(m)s}, \varepsilon_{j'(m)s}) = \rho_m; \forall j \neq j'$, but ε_{ms} have $i.i.d$ EVD.

Now, let that $(Y_{(1)}, Z_{(1)})$, $(Y_{(2)}, Z_{(2)})$, ..., $(Y_{(J)}, Z_{(J)})$ denote the joint rank alternative variables of a choice set with J alternatives, where variables $Y_{(j)}$ and $Z_{(j)}; j \in \{1, 2, \ldots, J\}$ denote alternatives and a nest with rank j. For example, $(Y_{(1)} = r_{(1)}, Z_{(1)} = m_{(1)})$ means that the alternative $a_{r_{(1)}} \in \mathbb{C}_{m_{(1)}s}$ from nest $m_{(1)} \in M$ have the first rank, where $U_{r_{(1)}(m_{(1)})s} = \max_{j \in \mathbb{C}_{m_{(1)}s}} U_{j(m_{(1)})s}$ and $U_{m_{(1)}s} = \max_{m \in M} U_{ms}$, also $(Y_{(2)} = r_{(2)}, Z_{(2)} = m_{(2)})$ means that alternative $a_{r_{(2)}}$ from nest $m_{(2)}$ have the second rank. Certainly it is the case in the original choice set, but in the new choice set (it is denoted by $\mathbb{C}_{s(r_{(1)})}$ after removing the alternative $a_{r_{(1)}}$) alternative $a_{r_{(2)}}$ and nest $m_{(2)}$ will have the first rank, where $m_{(2)}$ may be the same as nest $m_{(1)}$ or not. In this state, if there is $m_{(1)} = m_{(2)}$ then $a_{r_{(1)}}$ and $a_{r_{(2)}}$ selected from the same nest, then $U_{r_{(2)}(m_{(2)})s} = \max_{j \in \mathbb{C}_{m_{(2)}s(r_{(1)})}} U_{j(m_{(2)})s}$, where $\mathbb{C}_{ms(j)}$ denotes the choice set of nest m, which excludes the alternative a_j.

We know that if $\text{corr}(\varepsilon_{j(m_{(1)})s}, \varepsilon_{j'(m_{(1)})s}) = \rho_{m_{(1)}}$ and $\text{corr}(\varepsilon_{j(m_{(2)})s}, \varepsilon_{j'(m_{(2)})s}) = \rho_{m_{(2)}}$; $j \neq j'$ then $\rho_{m_{(1)}} = \rho_{m_{(2)}}$ if $m_{(1)} = m_{(2)}$. Now, keeping to the relation between ρ and λ ($\lambda = \sqrt{1 - \rho}$), it can be written that $\lambda_{m_{(1)}} = \lambda_{m_{(2)}}$, where $\lambda_{m_{(1)}}$ and $\lambda_{m_{(2)}}$ are dissimilarity parameters related to nests, which include the first rank alternative and the second rank alternative, respectively. Now, we denote the choice probability related to choosing an alternative with the first rank as follows (w.r.t the choice set $\mathbb{C}_s$):

$$
\begin{aligned}
P_{r_{(1)}m_{(1)}s} &= P(Y_{(1)} = r_{(1)}, Z_{(1)} = m_{(1)}) \\
&= P(Y_{(1)} = r_{(1)}|Z_{(1)} = m_{(1)}) \cdot P(Z_{(1)} = m_{(1)}) \\
&= P_{r_{(1)}|m_{(1)}s} \cdot P_{m_{(1)}s}.
\end{aligned}
$$

Thus:

$$
P_{r_{(1)}m_{(1)}s} = \frac{\exp\left\{ \frac{\mathbf{f}^T(a_{r_{(1)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}} \right\}}{\sum_{j=1}^{J_{m_{(1)}s}} \exp\left\{ \frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}} \right\}}
$$

$$
\times \frac{\left[\sum_{j \in \mathbb{C}_{m_{(1)}s}} \exp\left\{ \frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}} \right\} \right]^{\lambda_{m_{(1)}}}}{\left[\sum_{j \in \mathbb{C}_{m_{(1)}s}} \exp\left\{ \frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}} \right\} \right]^{\lambda_{m_{(1)}}} + \sum_{m \neq m_{(1)}}^{M} \left[\sum_{j \in \mathbb{C}_{ms}} \exp\left\{ \frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_m} \right\} \right]^{\lambda_m}},
$$

where $\mathbb{C}_{ms}$ denotes a choice set, which includes all of the alternatives in the nest m and $\mathbb{C}_{m_{(1)}s}$ denotes a set of all of alternatives in nest $m_{(1)}$ which has the first rank with respect to choice set $\mathbb{C}_s$. Also, vector $\mathbf{f}(a_{r_{(j)}s})$ denotes the characteristics of the attributes related to alternative $a_{r_{(j)}}$, which has the rank $r_{(j)}$ (w.r.t choice set $\mathbb{C}_s$)

Similarly, the choice probabilities of the second rank alternative and nest are obtained by:

$$P_{r_{(2)}m_{(2)}s} = P(Y_{(2)} = r_{(2)}, Z_{(2)} = m_{(2)})$$
$$= P(Y_{(2)} = r_{(2)}|Z_{(2)} = m_{(2)}) \cdot P(Z_{(2)} = m_{(2)})$$
$$= P_{r_{(2)}|m_{(2)}s} \cdot P_{m_{(2)}s},$$

where:

$$P_{r_{(2)}|m_{(2)}s} = \frac{\exp\left\{\frac{\mathbf{f}^T(a_{r_{(2)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right\}}{\sum_{j=1}^{J^{r_{(1)}}_{m_{(1)}s}} \exp\left\{\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right\}}; \quad m_{(1)} = m_{(2)},$$

$$= \frac{\exp\left\{\frac{\mathbf{f}^T(a_{r_{(2)}s})\boldsymbol{\beta}}{\lambda_{m_{(2)}}}\right\}}{\sum_{j=1}^{J_{m_{(2)}s}} \exp\left\{\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(2)}}}\right\}}; \quad m_{(1)} \neq m_{(2)}$$

and

$$P_{m_{(2)}s} = \frac{\left[\sum_{j=1}^{J^{r_{(1)}}_{m_{(1)}s}} \exp\left\{\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right\}\right]^{\lambda_{m_{(1)}}}}{\left[\sum_{j=1}^{J^{r_{(1)}}_{m_{(1)}s}} \exp\left\{\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right\}\right]^{\lambda_{m_{(1)}}} + \sum_{\substack{m=1 \\ (m \neq m_{(1)})}}^{M} \left[\sum_{j=1}^{J_{ms}} \exp\left\{\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_m}\right\}\right]^{\lambda_m}}; m_{(1)} = m_{(2)},$$

$$= \frac{\left[\sum_{j=1}^{J_{m_{(2)}s}} \exp\left\{\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(2)}}}\right\}\right]^{\lambda_{m_{(2)}}}}{\left[\sum_{j=1}^{J^{r_{(1)}}_{m_{(1)}s}} \exp\left\{\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right\}\right]^{\lambda_{m_{(1)}}} + \left[\sum_{j=1}^{J_{m_{(2)}s}} \exp\left\{\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(2)}}}\right\}\right]^{\lambda_{m_{(2)}}} + \sum_{\substack{m=1 \\ (m \neq m_{(1)}, m_{(2)})}}^{M}\left[\sum_{j=1}^{J_{ms}} \exp\left\{\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_m}\right\}\right]^{\lambda_m}}$$

$$m_{(1)} \neq m_{(2)},$$

where:

- J_{ms} denotes the number of alternatives in nest m of choice set $\mathbb{C}_s$.

- $J_{m_{(1)}s}$ is the number of alternatives in nest $m_{(1)}$ which has the first rank, with respect to choice set $\mathbb{C}_s$.

- $J_{m_{(2)}s}$ denotes the number of alternatives in nest $m_{(2)}$ which has the second rank based on the original choice set and has the first rank with respect to a new choice set (after removing alternative $a_{r_{(1)}}$) so that if $m_{(1)} = m_{(2)}$ then $J_{m_{(2)}s} = J_{m_{(1)}s} - 1$.

- $J_{m_{(1)}s}^{r_{(1)}}$ is the number of alternatives in nest $m_{(1)}$ after choosing alternative $a_{r_{(1)}}$ and removing it.

Also, to obtain the other choice probabilities related to the alternatives with lower ranking like $P_{r_{(3)}m_{(3)}}, \ldots, P_{r_{(J)}m_{(J)}}$, we act in a similar way, where $J = \sum_{m=1}^{M} J_{ms}$ for each choice set, $\mathbb{C}_s$, and $J_{ms}^{j,j'}$ denote the number of alternatives in nest m after removing the alternatives a_j and $a_{j'}$.

In this situation, the observation variables to analyze the RO.NMNL model is defined as follows:

$$\mathcal{Y}_{((r_{(1)},m_{(1)}),(r_{(2)},m_{(2)}),\ldots,(r_{(J)},m_{(J)}))} = \begin{cases} 1, & (Y_{(1)} = r_{(1)}, Z_{(1)} = m_{(1)}),\ldots,(Y_{(J)} = r_{(J)}, \\ & \qquad\qquad Z_{(J)} = m_{(J)}); \\ 0, & \text{Otherwise}, \end{cases}$$

where:

$$E(\mathcal{Y}_{((r_{(1)},m_{(1)}),(r_{(2)},m_{(2)}),\ldots,(r_{(J)},m_{(J)}))}) = P_{r_{(1)}m_{(1)}s} \cdot P_{r_{(2)}m_{(2)}s} \ldots P_{r_{(J)}m_{(J)}s}$$

and $J = \sum_{m=1}^{M} J_{ms}$ are the number of alternatives in choice set $\mathbb{C}_s$ so that:

$$\underbrace{\sum_{m_{(1)}=1}^{M} \sum_{m_{(2)}=1}^{M} \ldots \sum_{m_{(J)}=1}^{M}}_{\text{some nests maybe equal}} \underbrace{\sum_{r_{(1)}=1}^{J_{m_{(1)}s}} \sum_{r_{(2)}=1}^{J_{m_{(2)}s}} \ldots \sum_{r_{(J)}=1}^{J_{m_{(J)}s}}}_{r_{(1)} \neq r_{(2)} \cdots \neq r_{(J)}} P_{r_{(1)}m_{(1)}s} \cdot P_{r_{(2)}m_{(2)}s} \ldots P_{r_{(J)}m_{(J)}s} = 1.$$

3.1   Information Matrix for RO.NMNL Model

In this section, the information matrix for the two-level RO.NMNL model will be obtained. Then, we define the local D-optimality criterion, which is a function based on the determinate of the information matrix. As we know, the information matrix is calculated by log-likelihood function, where

Likelihood function for RO.NMNL model is calculated as follows (w.r.t choice set $\mathbb{C}_s$):

$$L(\mathbb{C}_s, \boldsymbol{\theta}) = \prod_{m_{(1)}=1}^{M} \prod_{m_{(2)}=1}^{M} \cdots \prod_{m_{(J)}=1}^{M} \underbrace{\prod_{r_{(1)}=1}^{J_{m_{(1)}s}} \prod_{r_{(2)}=1}^{J_{m_{(2)}s}} \cdots \prod_{r_{(J)}=1}^{J_{m_{(J)}s}}}_{r_{(1)} \neq r_{(2)} \neq \ldots \neq r_{(J)}} \left(P_{r_{(1)}m_{(1)}s} \cdot P_{r_{(2)}m_{(2)}s} \right. \\ \left. \ldots P_{r_{(J)}m_{(J)}s} \right)^{\mathcal{Y}'},$$

where $\mathcal{Y}' = \mathcal{Y}_{((r_{(1)},m_{(1)}),(r_{(2)},m_{(2)}),\ldots,(r_{(J)},m_{(J)}))}$.

The log-likelihood function for the two-level RO.NMNL model and based on the above likelihood function is calculated by:

$$\ell(\mathbb{C}_s, \boldsymbol{\theta}) = \underbrace{\sum_{m_{(1)}=1}^{M} \sum_{m_{(2)}=1}^{M} \cdots \sum_{m_{(J)}=1}^{M}}_{\text{some nests maybe equal}} \underbrace{\sum_{r_{(1)}=1}^{J_{m_{(1)}s}} \sum_{r_{(2)}=1}^{J_{m_{(2)}s}} \cdots \sum_{r_{(J)}=1}^{J_{m_{(J)}s}}}_{r_{(1)} \neq r_{(2)} \ldots \neq r_{(J)}} \mathcal{Y}' \\ \times G_s\big((r_{(1)}, m_{(1)}), (r_{(2)}, m_{(2)}), \ldots, (r_{(J)}, m_{(J)})\big), \qquad (4)$$

where $G_s : \mathbb{C}_s \to \Re^-$ and:

$$G_s\big((r_{(1)}, m_{(1)}), (r_{(2)}, m_{(2)}), \ldots, (r_{(J)}, m_{(J)}))\big) = \ln P_{r_{(1)}m_{(1)}s} + \ln P_{r_{(2)}m_{(2)}s} \\ + \ldots + \ln P_{r_{(J)}m_{(J)}s}.$$

According to the number of attributes, K, each with $L_k; k = 1, 2, \ldots, K$ levels, we define parameters vector and the characteristics of attributes as follows:

- $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\lambda})^T$; $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k, \ldots, \boldsymbol{\beta}_K)^T$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m, \ldots, \lambda_M)^T$,

- $\boldsymbol{\beta}_k = (\beta_{k,1}, \ldots, \beta_{k,\ell}, \ldots, \beta_{k,L_k-1})^T$ (w.r.t effect type codding, $\sum_{\ell=1}^{L_k} \beta_{k\ell} = 0$),

where λ_m have been introduced by symbol $\lambda_{m_{(r)}}$, which is the dissimilarity parameter related to a nest with rank r (or it had rank r). In reality, parameters vector $\boldsymbol{\theta}$ includes $p + M$ parameters, where $p = \sum_{k=1}^{K}(L_k - 1)$. Also:

- $v_{r_{(j)}(m_{(r)})s} = \mathbf{f}^T(a_{r_{(j)}m_{(r)}s})\boldsymbol{\beta}; \mathbf{f}(a_{r_{(j)}m_{(r)}s}) = (\mathbf{f}_1(a_{r_{(j)}m_{(r)}s}), \ldots,$
  $\mathbf{f}_k(a_{r_{(j)}m_{(r)}s}), \ldots, \mathbf{f}_K(a_{j_{(r)}m_{(r)}s}))^T;$

- $\mathbf{f}_k(a_{r_{(j)}m_{(r)}s}) = (f_{k1}(a_{r_{(j)}m_{(r)}s}), \ldots, f_{k\ell}(a_{r_{(j)}m_{(r)}s}), \ldots, f_{kL_k-1}(a_{r_{(j)}m_{(r)}s}))^T,$

where $f_{k\ell}(a_{r_{(j)}m_{(r)}s})$ denotes the characteristics of the ℓth level of attribute k for alternative $a_{r_{(j)}m_{(r)}}$ (has the rank r) in nest $m_{(r)}$ (with rank r).

Corresponding to Equation (4) and definition $-E\left(\frac{\partial^2 \ell(\mathbb{C}_s, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right)$ for the local information matrix, we obtain:

$$
\mathbf{I}_{R(NMNL)}(\mathbb{C}_s, \boldsymbol{\theta}) = -E\left(\frac{\partial^2 \ell(\mathbb{C}_s, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right)
$$

$$
= \sum_{m_{(1)}=1}^{M} \sum_{m_{(2)}=1}^{M} \cdots \sum_{m_{(J)}=1}^{M} \underbrace{\sum_{r_{(1)}=1}^{J_{m_{(1)}s}} \sum_{r_{(2)}=1}^{J_{m_{(2)}s}} \cdots \sum_{r_{(J)}=1}^{J_{m_{(J)}s}}}_{r_{(1)} \neq r_{(2)} \ldots \neq r_{(J)}} E(\mathcal{Y}')
$$

$$
\times \frac{-\partial^2 G_s(r_{(1)}, r_{(2)} \ldots r_{(J)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T},
$$

which is the information matrix of the two-level RO.NMNL model, where:

- $E(\mathcal{Y}') = P_{r_{(1)}m_{(1)}s} \cdot P_{r_{(2)}m_{(2)}s} \ldots P_{r_{(J)}m_{(J)}s}$

- $\frac{-\partial^2 G_s(r_{(1)}, r_{(2)} \ldots r_{(J)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = -\frac{\partial^2 \ln P_{r_{(1)}m_{(1)}s}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} - \frac{\partial^2 \ln P_{r_{(2)}m_{(2)}s}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} - \cdots - \frac{\partial^2 \ln P_{r_{(J-1)}m_{(J-1)}s}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}.$

Lemma 2. *According to the above descriptions, the information matrix of RO.NMNL model is calculated as follows:*

$$
\mathbf{I}_{R(NMNL)}(\mathbb{C}_s, \boldsymbol{\theta}) = \mathbf{I}_{NMNL}(\mathbb{C}_s, \boldsymbol{\theta}) + \sum_{m_{(1)}=1}^{M} \sum_{r_{(1)}=1}^{J_{m_{(1)}s}} P_{r_{(1)}m_{(1)}s} \cdot \mathbf{I}_{NMNL}(\mathbb{C}_{s(r_{(1)})}, \boldsymbol{\theta})
$$

$$
+ \underbrace{\sum_{m_{(1)}, m_{(2)}=1}^{M}}_{maybe\ equal} \underbrace{\sum_{r_{(1)}, r_{(2)}=1}^{J_{m_{(1)}s}, J_{m_{(2)}s}}}_{r_{(1)} \neq r_{(2)}} P_{r_{(1)}m_{(1)}s} \cdot P_{r_{(2)}m_{(2)}s} \cdot \mathbf{I}_{NMNL}(\mathbb{C}_{s(r_{(1)}, r_{(2)})}, \boldsymbol{\theta}) + \ldots +
$$

$$
+ \underbrace{\sum_{m_{(1)}, \ldots, m_{(J-2)}=1}^{M}}_{maybe\ equal} \underbrace{\sum_{r_{(1)}, \ldots, r_{(J-2)}=1}^{J_{m_{(1)}s}, \ldots, J_{m_{(J-2)}s}}}_{r_{(1)} \neq \ldots \neq r_{(J-2)}} P_{r_{(1)}m_{(1)}s} \ldots P_{r_{(J-2)}m_{(J-2)}s} \cdot \mathbf{I}_{NMNL}(\mathbb{C}_{s(r_{(1)}, \ldots, r_{(J-2)})}, \boldsymbol{\theta}),
$$

Proof.

$$
\mathbf{I}_{R(NMNL)}(\mathbb{C}_s, \boldsymbol{\theta}) = -E\left(\frac{\partial^2 \ell(\mathbb{C}_s, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right) = \sum_{m_{(1)}=1}^{M} \sum_{r_{(1)}=1}^{J_{m_{(1)}s}} P_{r_{(1)}m_{(1)}s} \cdot \left(\frac{-\partial^2 \ln P_{r_{(1)}m_{(1)}s}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right)
$$

$$
+ \sum_{m_{(1)}=1}^{M} \sum_{r_{(1)}=1}^{J_{m_{(1)}s}} P_{r_{(1)}m_{(1)}s} \sum_{m_{(2)}=1}^{M} \sum_{r_{(2)}=1}^{J_{m_{(2)}s}} P_{r_{(2)}m_{(2)}s} \cdot \left(\frac{-\partial^2 \ln P_{r_{(2)}m_{(2)}s}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right) + \cdots +
$$

$$
+ \sum_{m_{(1)}=1}^{M} \sum_{m_{(2)}=1}^{M} \cdots \sum_{m_{(J-2)}=1}^{M} \underbrace{\sum_{r_{(1)}=1}^{J_{m_{(1)}s}} \sum_{r_{(2)}=1}^{J_{m_{(2)}s}} \cdots \sum_{r_{(J-2)}=1}^{J_{m_{(J-2)}s}}}_{r_{(1)} \neq r_{(2)} \cdots \neq r_{(J-2)}} P_{r_{(1)}m_{(1)}s} \cdot P_{r_{(2)}m_{(2)}s} \cdots P_{r_{(J-2)}m_{(J-2)}s}
$$

$$
\times \left\{ \sum_{m_{(J-1)}=1}^{M} \sum_{r_{(J-1)}=1}^{J_{m_{(J-1)}s}} P_{r_{(J-1)}m_{(J-1)}s} \cdot \left(\frac{-\partial^2 \ln P_{r_{(J-1)}m_{(J-1)}s}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right) \right\},
$$

where:

- $\sum_{m_{(1)}=1}^{M} \sum_{r_{(1)}=1}^{J_{m_{(1)}}} P_{r_{(1)}m_{(1)}s} \cdot \left(\frac{-\partial^2 \ln P_{r_{(1)}m_{(1)}s}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right) = \mathbf{I}_{NMNL}(\mathbb{C}_s, \boldsymbol{\theta}),$

- $\sum_{m_{(2)}=1}^{M} \sum_{r_{(2)}=1}^{J_{m_{(2)}s}} P_{r_{(2)}m_{(2)}s} \cdot \left(\frac{-\partial^2 \ln P_{r_{(2)}m_{(2)}s}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right) = \mathbf{I}_{NMNL}(\mathbb{C}_{s(r_{(1)})}, \boldsymbol{\theta}),$

- $\sum_{m_{(J-1)}=1}^{M} \sum_{r_{(J-1)}=1}^{J_{m_{(J-1)}s}} P_{r_{(J-1)}m_{(J-1)}s} \cdot \left(\frac{-\partial^2 \ln P_{r_{(J-1)}m_{(J-1)}s}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right) =$
  $\mathbf{I}_{NMNL}(\mathbb{C}_{s(r_{(1)},\ldots,r_{(J-2)})}, \boldsymbol{\theta}).$

$\square$

where $\mathbb{C}_{s(j)}$ denotes a choice set without considering alternatives a_j. In this case, since $P_{r_{(J)}m_{(J)}} = 1$ then $\ln P_{r_{(J)}m_{(J)}} = 0$ and:

- $\mathbf{I}_{\text{NMNL}}(\mathbb{C}_s, \boldsymbol{\theta})$ (full alternatives),

- $\mathbf{I}_{\text{NMNL}}(\mathbb{C}_{s(r_{(1)})}, \boldsymbol{\theta})$ (without alternative $a_{r_{(1)}}$),

- $\mathbf{I}_{\text{NMNL}}(\mathbb{C}_{s(r_{(1)},r_{(2)})}, \boldsymbol{\theta})$ (without alternatives $(a_{r_{(1)}}, a_{r_{(2)}})$),

- $\mathbf{I}_{\text{NMNL}}(\mathbb{C}_{s(r_{(1)},\ldots,r_{(J-2)})}, \boldsymbol{\theta})$ (without alternatives $(a_{r_{(1)}}, \ldots, a_{r_{(J-2)}})$).

will be calculated by the following lemma (Lemma 3).

www.SID.ir

Lemma 3. *According to a NMNL model with two nests each with J_{1s} and J_{2s} alternatives, the information matrix $\mathbf{I}_{\mathrm{NMNL}}(\mathbb{C}_{s(j)}, \boldsymbol{\theta})$ will be calculated as follows:*

$$\mathbf{I}_{\mathrm{NMNL}}(\mathbb{C}_{s(j)}, \boldsymbol{\theta}) = -E\left(\frac{\partial^2 \ell(\mathbb{C}_{s(j)}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right) = \begin{pmatrix} \mathbf{I}_{11s(j)} & \mathbf{I}_{12s(j)} & \mathbf{I}_{13s(j)} \\ \mathbf{I}_{12s(j)}^T & I_{22s(j)} & I_{23s(j)} \\ \mathbf{I}_{13s(j)}^T & I_{23s(j)} & I_{33s(j)} \end{pmatrix},$$

where $\mathbf{I}_{..s(j)}$ and $I_{..s(j)}$ denote the elements of above information matrix without considering alternative a_j.

Proof. According to the following assumptions:

- $\mathbf{A}_{ms(j)} = \mathbf{F}_{ms(j)}^T \mathbf{p}_{.|ms(j)}$,

- $\mathbf{B}_{ms(j)} = \mathbf{F}_{ms(j)}^T \mathbf{P}_{.|ms(j)} \mathbf{F}_{ms(j)}$,

- $\mathbf{p}_{.|ms(j)} = (p_{1|ms(j)}, \ldots, p_{j-1|ms(j)}, p_{j+1|ms(j)}, \ldots, p_{J_{ms(j)}|ms(j)})^T$,

- $\mathbf{P}_{.|ms(j)} = \mathrm{diag}[p_{1|ms(j)}, \ldots, p_{j-1|ms(j)}, p_{j+1|ms(j)}, \ldots, p_{J_{ms(j)}|ms(j)}]$,

- $\mathbf{F}_{ms(j)} = (\mathbf{f}(a_{1ms}), \ldots, \mathbf{f}(a_{j-1,ms}), \mathbf{f}(a_{j+1,ms}), \ldots, \mathbf{f}(a_{J_{ms}^j ms}))^T$,

- $\mathbf{f}(a_{j'ms}) = (f_1(a_{j'ms}), \ldots, f_h(a_{j'ms}), \ldots, f_p(a_{j'ms}))^T$; $j' \neq j$,

- $v_{ms(j)} = \ln\left[\sum_{j'=1}^{J_{ms}^j} \exp\left\{\frac{\mathbf{f}^T(a_{j'ms})\boldsymbol{\beta}}{\lambda_m}\right\}\right]$,

- $p_{j'|ms(j)} = \dfrac{\exp\left\{\frac{\mathbf{f}^T(a_{j'ms})\boldsymbol{\beta}}{\lambda_m}\right\}}{\sum_{l=1}^{J_{ms}^j} \exp\left\{\frac{\mathbf{f}^T(a_{lms})\boldsymbol{\beta}}{\lambda_1}\right\}}$,

- $p_{ms(j)} = \dfrac{\left[\sum_{l=1}^{J_{ms}^j} \exp\left\{\frac{\mathbf{f}^T(a_{lms})\boldsymbol{\beta}}{\lambda_m}\right\}\right]^{\lambda_m}}{\left[\sum_{l=1}^{J_{1s}^j} \exp\left\{\frac{\mathbf{f}^T(a_{l1s})\boldsymbol{\beta}}{\lambda_1}\right\}\right]^{\lambda_1} + \left[\sum_{l=1}^{J_{2s}^j} \exp\left\{\frac{\mathbf{f}^T(a_{l2s})\boldsymbol{\beta}}{\lambda_2}\right\}\right]^{\lambda_2}}$.

We will have:

$$[\mathbf{I}_{11s(j)}]_{p \times p} = \frac{p_{1s(j)}}{\lambda_1^2}(\mathbf{B}_{1s(j)} - \mathbf{A}_{1s(j)}\mathbf{A}_{1s(j)}^T) + \frac{p_{2s(j)}}{\lambda_2^2}(\mathbf{B}_{2s(j)} - \mathbf{A}_{2s(j)}\mathbf{A}_{2s(j)}^T)$$
$$+ p_{1s(j)}p_{2s(j)}(\mathbf{A}_{1s(j)}\mathbf{A}_{1s(j)}^T + \mathbf{A}_{2s(j)}\mathbf{A}_{2s(j)}^T - \mathbf{A}_{1s(j)}\mathbf{A}_{2s(j)}^T - \mathbf{A}_{2s(j)}\mathbf{A}_{1s(j)}^T)$$

$$[\mathbf{I}_{12s(j)}]_{p \times 1} = -\frac{p_{1s(j)}}{\lambda_1^3}(\mathbf{B}_{1s(j)} - \mathbf{A}_{1s(j)}\mathbf{A}_{1s(j)}^T)\boldsymbol{\beta} + \frac{p_{1s(j)}p_{2s(j)}}{\lambda_1}(\mathbf{A}_{1s(j)} - \mathbf{A}_{2s(j)})(\lambda_1.v_{1s(j)}$$
$$- \mathbf{A}_{1s(j)}^T\boldsymbol{\beta})$$

$$[\mathbf{I}_{13s(j)}]_{p\times 1} = -\frac{p_{2s(j)}}{\lambda_2^3}(\mathbf{B}_{2s(j)} - \mathbf{A}_{2s(j)}\mathbf{A}_{2s(j)}^T)\boldsymbol{\beta} + \frac{p_{1s(j)}p_{2s(j)}}{\lambda_2}(\mathbf{A}_{2s(j)} - \mathbf{A}_{1s(j)})(\lambda_2.v_{2s(j)}$$
$$-\mathbf{A}_{2s(j)}^T\boldsymbol{\beta})$$

$$[I_{22s(j)}]_{1\times 1} = \frac{p_{1s(j)}}{\lambda_1^4}\boldsymbol{\beta}^T(\mathbf{B}_{1s(j)} - \mathbf{A}_{1s(j)}\mathbf{A}_{1s(j)}^T)\boldsymbol{\beta} + \frac{p_{1s(j)}p_{2s(j)}}{\lambda_1^2}(\lambda_1.v_{1s(j)} - \boldsymbol{\beta}^T\mathbf{A}_{1s(j)})(\lambda_1.v_{1s(j)}$$
$$-\mathbf{A}_{1s(j)}^T\boldsymbol{\beta})$$

$$[I_{33s(j)}]_{1\times 1} = \frac{p_{2s(j)}}{\lambda_2^4}\boldsymbol{\beta}^T(\mathbf{B}_{2s(j)} - \mathbf{A}_{2s(j)}\mathbf{A}_{2s(j)}^T)\boldsymbol{\beta} + \frac{p_{1s(j)}p_{2s(j)}}{\lambda_2^2}(\lambda_2.v_{2s(j)} - \boldsymbol{\beta}^T\mathbf{A}_{2s(j)})(\lambda_2.v_{2s(j)}$$
$$-\mathbf{A}_{2s(j)}^T\boldsymbol{\beta})$$

$$[I_{23s(j)}]_{1\times 1} = -\frac{1}{\lambda_1.\lambda_2}p_{1s(j)}p_{2s(j)}(\lambda_1.v_{1s(j)} - \boldsymbol{\beta}^T\mathbf{A}_{1s(j)})(\lambda_2.v_{2s(j)} - \mathbf{A}_{2s(j)}^T\boldsymbol{\beta})$$

$\square$

In above information matrix, all of the notations with (j) will be calculated without alternative a_j and J_{ms}^j denotes the number of alternatives in nest m of choice set s without alternative a_j, such that $J_{ms}^j = J_{ms} - 1$.

Corollary 1. For $\boldsymbol{\beta} = \mathbf{0}$, the information matrix (Lemma 3) was calculated as follows (let $J_{1s}^j = a$, $J_{2s}^j = b$):

$$[\mathbf{I}_{11s(j)}] = \left\{\frac{a^{\lambda_1}}{a\lambda_1^2(a^{\lambda_1} + b^{\lambda_2})}\right\} \cdot \left[\mathbf{F}_{1s(j)}^T\left(\mathbf{I}_a - \frac{1}{a}\cdot\mathbf{1}_a\mathbf{1}_a^T\right)\mathbf{F}_{1s(j)}\right] + \left\{\frac{b^{\lambda_2}}{b\lambda_2^2(a^{\lambda_1} + b^{\lambda_2})}\right\}$$
$$\left[\mathbf{F}_{2s(j)}^T\left(\mathbf{I}_b - \frac{1}{b}\cdot\mathbf{1}_b\mathbf{1}_b^T\right)\mathbf{F}_{2s(j)}\right] - \left\{\frac{b^{\lambda_2}a^{\lambda_1}}{(a^{\lambda_1} + b^{\lambda_2})^2}\right\}\cdot\left(\frac{1}{b^2}\cdot\mathbf{F}_{1s(j)}^T\mathbf{1}_a\mathbf{1}_a^T\mathbf{F}_{1s(j)}\right.$$
$$\left. + \frac{1}{b^2}\cdot\mathbf{F}_{2s(j)}^T\mathbf{1}_b\mathbf{1}_b^T\mathbf{F}_{2s(j)} - \frac{1}{ab}\cdot\mathbf{F}_{1s(j)}^T\mathbf{1}_a\mathbf{1}_b^T\mathbf{F}_{2s(j)} - \frac{1}{ab}\cdot\mathbf{F}_{2s(j)}^T\mathbf{1}_b\mathbf{1}_a^T\mathbf{F}_{1s(j)}\right)$$

$$[\mathbf{I}_{12s(j)}] = \left(\frac{b^{\lambda_2}a^{\lambda_1}\ln(a)}{(a^{\lambda_1} + b^{\lambda_2})^2}\right)\cdot\left[\frac{1}{a}\cdot\mathbf{F}_{1s(j)}^T\mathbf{1}_a - \frac{1}{b}\cdot\mathbf{F}_{2s(j)}^T\mathbf{1}_b\right],$$

$$[I_{13s(j)}] = \left\{\frac{b^{\lambda_2}a^{\lambda_1}\ln(b)}{(a^{\lambda_1} + b^{\lambda_2})^2}\right\}\cdot\left[\frac{1}{b}\cdot\mathbf{F}_{2s(j)}^T\mathbf{1}_b - \frac{1}{a}\cdot\mathbf{F}_{1s(j)}^T\mathbf{1}_a\right]$$

$$[I_{22s(j)}] = \frac{a^{\lambda_1}b^{\lambda_2}}{(a^{\lambda_1} + b^{\lambda_2})^2}\cdot\{\ln(a)\}^2$$

$$[I_{23s(j)}] = -\frac{a^{\lambda_1}b^{\lambda_2}}{(a^{\lambda_1} + b^{\lambda_2})^2}\cdot\ln(a)\ln(b),$$

$$[I_{33s(j)}] = \frac{a^{\lambda_1}b^{\lambda_2}}{(a^{\lambda_1} + b^{\lambda_2})^2}\cdot\{\ln(b)\}^2.$$

3.2 D-Optimal Design

Similar to the classical choice experiments in which the formula:

$$\Psi(\xi, \boldsymbol{\theta}) = \ln\det\left(\mathbf{I}_{\text{NMNL}}^{-1}(\xi, \boldsymbol{\theta})\right)$$

www.SID.ir

Table 1. NMNL Model: There are two Nests each with $\mathcal{J}_1$ and $\mathcal{J}_2$ alternatives

nest(I)	nest(II)
$\tilde{a}_{11}, \ldots, \tilde{a}_{j1}, \ldots, \tilde{a}_{\mathcal{J}_1 1}$	$\tilde{a}_{12}, \ldots, \tilde{a}_{j2}, \ldots, \tilde{a}_{\mathcal{J}_2 2}$

has been used to obtain the local D-optimality criterion, it is also used in the RO.NMNL:

$$\Psi_R(\xi, \boldsymbol{\theta}) = \ln \det \left(\mathbf{I}^{-1}_{R(\text{NMNL})}(\xi, \boldsymbol{\theta}) \right)$$

to calculate the local D-optimality criterion, where ξ is a design with choice sets $\mathbb{C}_1, \ldots, \mathbb{C}_s, \ldots, \mathbb{C}_\mathcal{S}$. In this position, ξ^*, which minimizes $\Psi_R(\xi, \boldsymbol{\theta})$ for true value of $\boldsymbol{\theta}$ is called locally D-optimal design, where:

$$\xi^* = \arg \min_{\xi \in \Xi} \Psi_R(\xi, \boldsymbol{\theta})$$

for true value of parameters.

3.3   Example

Imagine that there is a two-level NMNL model with two nests in a way that one of them has $\mathcal{J}_1$ alternatives and the other includes $\mathcal{J}_2$ alternatives (Table 1), where $\mathcal{J}_1 + \mathcal{J}_2 = \mathcal{J}$.

In this situation, we select three alternatives, $J_s = 3; \forall s \in \mathcal{S}$, from Table 1 (It is assumed that just three alternatives have suitable ranking). In this case, we will encounter two classes ($N = 2$), where:

$$\mathcal{S}_1 = \left(\begin{array}{c} \mathcal{J}_1 \\ 1 \end{array} \right) \times \left(\begin{array}{c} \mathcal{J}_2 \\ 2 \end{array} \right), \qquad \mathcal{S}_2 = \left(\begin{array}{c} \mathcal{J}_1 \\ 2 \end{array} \right) \times \left(\begin{array}{c} \mathcal{J}_2 \\ 1 \end{array} \right).$$

According to the dimensions of parameters, $\boldsymbol{\beta}$ (p-dimensional) and $\boldsymbol{\lambda}$ (2-dimensional), there is a $(p + 2)$-dimensional parameters vector. In most of non-Bayesian linear problems, an upper bound on the number of support points in an optimal design is available (Pukelsheim, 1993). The D-optimality criterion in linear models typically leads to an optimal number of support points that is the same number of unknown parameters and the design takes an equal number of observations at each point (Silvey, 1980). The bound also applies to most local optimality criteria and Bayesian criteria for linear models (Chernoff, 1972). In contrast for nonlinear models, there is no

such bound available on the number of support points. Thus we define the following design based on two classes:

$$\xi_n = \left\{ \begin{array}{cccc} \mathbb{C}_{n1} & \mathbb{C}_{n2} & \cdots & \mathbb{C}_{n\mathcal{S}_n} \\ w_{n1} & w_{n2} & \cdots & w_{n\mathcal{S}_n} \end{array} \right\} \in \Xi_n; \quad n = 1, 2. \tag{5}$$

The information matrix corresponding to ξ_n (5) is calculated by:

$$\mathbf{I}_{\mathrm{R(NMNL)}}(\xi_n, \boldsymbol{\theta}) = \sum_{s=1}^{\mathcal{S}_n} w_{ns} \cdot \mathbf{I}_{\mathrm{R(NMNL)}}(\mathbb{C}_{ns}, \boldsymbol{\theta}),$$

where the local D-optimality criterion is defined as follows:

$$\Psi_R(\xi_n, \boldsymbol{\theta}) = \ln \det \left(\mathbf{I}_{\mathrm{R(NMNL)}}^{-1}(\xi_n, \boldsymbol{\theta}) \right).$$

Now, with respect to $\xi_n; n = 1, 2$ and the combination of them, $\xi = \sum_{n=1}^{2} \alpha_n \cdot \xi_n$, we will have:

$$\xi^* = \arg \min_{\xi \in \Xi} \Psi_R(\xi, \boldsymbol{\theta}) \tag{6}$$

is locally D-optimal design in Ξ ($\Xi = \bigcup_{n=1}^{2} \Xi_n$), where ($\sum_{n=1}^{2} \alpha_n = 1$, $\alpha_n \geqslant 0; n = 1, 2$):

$$\Psi_R(\xi, \boldsymbol{\theta}) \leqslant \sum_{n=1}^{2} \alpha_n \cdot \Psi_R(\xi_n, \boldsymbol{\theta}).$$

Lemma 4. *The information matrix of a choice set which includes two nests, so that one of them includes two alternatives and the other has one (Figure 1), is calculated by:*

$$\mathbf{I}_{\mathrm{R(NMNL)}}(\mathbb{C}_s, \boldsymbol{\theta}) = \mathbf{I}_{\mathrm{NMNL}}(\mathbb{C}_s, \boldsymbol{\theta}) + \sum_{m_{(1)}=1}^{2} \sum_{r_{(1)}=1}^{J_{m_{(1)}s}} P_{r_{(1)}m_{(1)}s} \cdot \mathbf{I}_{\mathrm{NMNL}}(\mathbb{C}_{s(r_{(1)})}, \boldsymbol{\theta}),$$

where $P_{r_{(1)}m_{(1)}s} = P_{r_{(1)}|m_{(1)}s} \cdot P_{m_{(1)}s}$ and

$$P_{r_{(1)}|m_{(1)}s} = \left\{ \begin{array}{ll} \dfrac{\exp\left\{ \frac{\mathbf{f}^T(a_{r_{(1)}m_{(1)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}} \right\}}{\exp\left\{ \frac{\mathbf{f}^T(a_{r_{(1)}m_{(1)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}} \right\} + \exp\left\{ \frac{\mathbf{f}^T(a_{jm_{(1)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}} \right\}}, & \mathbb{C}_{m_{(1)}s} = \{a_{r_{(1)}m_{(1)}s}, a_{jm_{(1)}s}\}; \\ 1, & \mathbb{C}_{m_{(1)}s} = \{a_{r_{(1)}m_{(1)}s}\}. \end{array} \right. \tag{7}$$

$$P_{m_{(1)}s} = \begin{cases} \dfrac{\left[\exp\left\{\dfrac{\mathbf{f}^T(a_{r_{(1)}m_{(1)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right\}+\exp\left\{\dfrac{\mathbf{f}^T(a_{jm_{(1)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right\}\right]^{\lambda_{m_{(1)}}}}{\left[\exp\left\{\dfrac{\mathbf{f}^T(a_{r_{(1)}m_{(1)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right\}+\exp\left\{\dfrac{\mathbf{f}^T(a_{jm_{(1)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right\}\right]^{\lambda_{m_{(1)}}}+\exp\left\{\mathbf{f}^T(a_{j'm_{(1)}s})\boldsymbol{\beta}\right\}}, \\[4pt] \qquad\qquad\qquad \mathbb{C}_{m_{(1)}s}=\{a_{r_{(1)}m_{(1)}s},a_{jm_{(1)}s}\}; \\[10pt] \dfrac{\exp\left\{\mathbf{f}^T(a_{r_{(1)}m_{(1)}s})\boldsymbol{\beta}\right\}}{\left[\exp\left\{\dfrac{\mathbf{f}^T(a_{jm_{(1)}s})\boldsymbol{\beta}}{\lambda_m}\right\}+\exp\left\{\dfrac{\mathbf{f}^T(a_{j'm_{(1)}s})\boldsymbol{\beta}}{\lambda_m}\right\}\right]^{\lambda_m}+\exp\left\{\mathbf{f}^T(a_{r_{(1)}m_{(1)}s})\boldsymbol{\beta}\right\}}, \\[4pt] \qquad\qquad\qquad \mathbb{C}_{m_{(1)}s}=\{a_{r_{(1)}m_{(1)}s}\}. \end{cases} \tag{8}$$

with $r_{(1)}, j, j' \in \mathbb{C}_s$; $r_{(1)} \neq j \neq j'$ and $m_{(1)}, m = 1, 2$ (two nests) and $\lambda_{m_{(1)}}$ is the dissimilarity parameter of a nest, which includes an alternative with the highest utility (first rank). In Equations (7) and (8), the notation $\mathbb{C}_{m_{(1)}s}$ denotes the choice set with the first rank nest.

Corollary 2. Now, consider a special case (Figure 1, Case 1). In this situation, $\mathbf{I}_{\mathrm{R(NMNL)}}(\mathbb{C}_s, \boldsymbol{\theta})$ is calculated as follows:

$$\mathbf{I}_{\mathrm{R(NMNL)}}(\mathbb{C}_s, \boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I}_{11s(0)} & \mathbf{I}_{12s(0)} & \mathbf{0} \\ \mathbf{I}_{12s(0)}^T & I_{22s(0)} & 0 \\ \mathbf{0}^T & 0 & 0 \end{bmatrix} + P_{a_{11s}1s} \cdot \begin{bmatrix} \mathbf{I}_{11s(a_{11s})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}^T & 0 & 0 \\ \mathbf{0}^T & 0 & 0 \end{bmatrix}$$

Figure 1. NMNL Model: There are two nests (for choice set $\mathbb{C}_s$), one of one of them with two alternatives and one for another (a_{jms} denotes the jth alternative of the mth nest from choice set s)

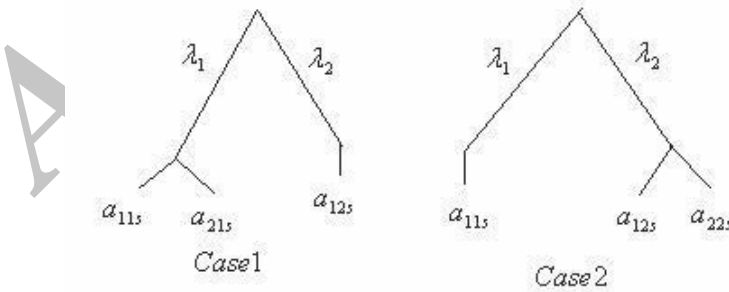$$+P_{a_{21s}1s} \cdot \begin{bmatrix} \mathbf{I}_{11s(a_{21s})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}^T & 0 & 0 \\ \mathbf{0}^T & 0 & 0 \end{bmatrix} + P_{a_{12s}2s} \cdot \begin{bmatrix} \mathbf{I}_{11s(a_{12s})} & \mathbf{I}_{12s(a_{12s})} & \mathbf{0} \\ \mathbf{I}_{12s(a_{12s})}^T & I_{22s(a_{12s})} & 0 \\ \mathbf{0}^T & 0 & 0 \end{bmatrix}$$

where:

- $P_{a_{11s}1s} = P(Y_{(1)} = a_{11s}, Z_{(1)} = 1) = P_{a_{11s}|1s} \cdot P_{1s}$,

- $P_{a_{21s}1s} = P(Y_{(1)} = a_{21s}, Z_{(1)} = 1) = P_{a_{21s}|1s} \cdot P_{1s}$,

- $P_{a_{12s}2s} = P(Y_{(1)} = a_{12s}, Z_{(1)} = 2) = P_{a_{12s}|2s} \cdot P_{2s}$

are calculated by (7) and (8), where $P_{r_{(j)}m_{(r)}s}$ has already been defined (See Section 3), $\mathbf{I}_{\ell\ell's(0)}$ ($I_{\ell\ell'(0)}$) denotes the ℓth row and ℓ'th column of the information matrix $\mathbf{I}$ (based on choice set $\mathbb{C}_s$) with respect to all of the alternatives in the choice set $\mathbb{C}_s$. But, $\mathbf{I}_{\ell\ell'(j)}$ ($I_{\ell\ell'(j)}$) denotes the ℓth row and ℓ'th column of the information matrix $\mathbf{I}$ (based on choice set $\mathbb{C}_s$) without alternative a_j. Moreover,

- $\mathbf{I}_{11s(j)} = p_{1s(j)} \cdot p_{2s(j)} \cdot \left(\mathbf{F}_{1s(j)}^T \mathbf{F}_{1s(j)} + \mathbf{F}_{2s(j)}^T \mathbf{F}_{2s(j)} - \mathbf{F}_{1s(j)}^T \mathbf{F}_{2s(j)} - \mathbf{F}_{2s(j)}^T \mathbf{F}_{1s(j)} \right); \ \forall j = a_{11s}, a_{21s}$,

- $\mathbf{I}_{11s(a_{12s})} = \frac{1}{\lambda_1^2} \cdot \left(\mathbf{B}_{1s(a_{12s})} - \mathbf{A}_{1s(a_{12s})} \mathbf{A}_{1s(a_{12s})}^T \right)$,

- $\mathbf{I}_{12s(a_{12s})} = -\frac{1}{\lambda_1^3} \cdot \left(\mathbf{B}_{1s(a_{12s})} - \mathbf{A}_{1s(a_{12s})} \mathbf{A}_{1s(a_{12s})}^T \right) \boldsymbol{\beta}$,

- $I_{22s(a_{12s})} = \frac{1}{\lambda_1^4} \cdot \boldsymbol{\beta}^T \left(\mathbf{B}_{1s(a_{12s})} - \mathbf{A}_{1s(a_{12s})} \mathbf{A}_{1s(a_{12s})}^T \right) \boldsymbol{\beta}$,

- $p_{1s(j)} = \frac{\exp\left\{\mathbf{f}^T(a_{j's})\boldsymbol{\beta}\right\}}{\exp\left\{\mathbf{f}^T(a_{j's})\boldsymbol{\beta}\right\} + \exp\left\{\mathbf{f}^T(a_{12s})\boldsymbol{\beta}\right\}}; \quad j, j' \in \{a_{11s}, a_{21s}\}; \quad j \neq j'$ (see Section 3.1)

where $\mathbf{F}_{ms(j)}, p_{ms(j)}$ denote the design matrix and the probability of choosing nest (in choice set $\mathbb{C}_s$) when the alternative a_j is removed.

Example 1. For a two-level RO.NMNL model, we have two attributes each with two levels, then there are four alternatives $\tilde{a}_{11}, \tilde{a}_{21}, \tilde{a}_{12}, \tilde{a}_{22}$, where:

$\tilde{a}_{jm}$	$At.1(f_1(\tilde{a}_{jm}))$	$At.2(f_2(\tilde{a}_{jm}))$
$\tilde{a}_{11}$	$+1$	$+1$
$\tilde{a}_{21}$	-1	$+1$
$\tilde{a}_{12}$	$+1$	-1
$\tilde{a}_{22}$	-1	-1

Table 2. Two-level Nested MNL model: Four choice sets,$\mathbb{C}_s$; $s = 1, 2, 3, 4$, each with three alternatives (a_{jms} denote the jth alternative of the nest m in choice set s)

Choice set ($\mathbb{C}_{ns}$)	First nest (I)	Second nest (II)
$\mathbb{C}_{11} = \mathbb{C}_1$	$a_{111} = \tilde{a}_{11}$, $a_{211} = \tilde{a}_{21}$	$a_{121} = \tilde{a}_{12}$
$\mathbb{C}_{12} = \mathbb{C}_2$	$a_{112} = \tilde{a}_{11}$, $a_{212} = \tilde{a}_{21}$	$a_{122} = \tilde{a}_{22}$
$\mathbb{C}_{21} = \mathbb{C}_3$	$a_{113} = \tilde{a}_{11}$	$a_{123} = \tilde{a}_{12}$, $a_{223} = \tilde{a}_{22}$
$\mathbb{C}_{22} = \mathbb{C}_4$	$a_{114} = \tilde{a}_{21}$	$a_{124} = \tilde{a}_{12}$, $a_{224} = \tilde{a}_{22}$

Let's consider experiments which include four choice sets each with three alternatives (Table 2, Figure 1), where $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$ thus we can write $\boldsymbol{\theta} = (\beta_1, \beta_2, \lambda_1, \lambda_2)^T$. In this position, there are two classes in which their information matrices will be calculated as follows:

1. $\xi_1 = \left\{ \begin{array}{cc} \mathbb{C}_{11} & \mathbb{C}_{12} \\ w_{11} & w_{12} \end{array} \right\} \in \Xi_1$, where

$$\mathbf{I}_{\mathrm{R(NMNL)}}(\xi_1, \boldsymbol{\theta}) = \sum_{s=1}^{2} w_{1s} \cdot \mathbf{I}_{\mathrm{R(NMNL)}}(\mathbb{C}_{1s}, \boldsymbol{\theta}),$$

2. $\xi_2 = \left\{ \begin{array}{cc} \mathbb{C}_{21} & \mathbb{C}_{22} \\ w_{21} & w_{22} \end{array} \right\} \in \Xi_2$, where

$$\mathbf{I}_{\mathrm{R(NMNL)}}(\xi_2, \boldsymbol{\theta}) = \sum_{s=1}^{2} w_{2s} \cdot \mathbf{I}_{\mathrm{R(NMNL)}}(\mathbb{C}_{2s}, \boldsymbol{\theta}).$$

According to $\xi = \alpha_1 \cdot \xi_1 + \alpha_2 \cdot \xi_2$ we will have;

$$\xi = \left\{ \begin{array}{cccc} \mathbb{C}_1 & \mathbb{C}_2 & \mathbb{C}_3 & \mathbb{C}_4 \\ w_1 & w_2 & w_3 & w_4 \end{array} \right\} \in \Xi. \tag{9}$$

where $w_1 = \alpha_1 \cdot w_{11}$, $w_2 = \alpha_1 \cdot w_{12}$, $w_3 = \alpha_2 \cdot w_{21}$ and $w_4 = \alpha_2 \cdot w_{22}$. The information matrix of the design (9) is calculated by:

$$\mathbf{I}_{\mathrm{R(NMNL)}}(\xi, \boldsymbol{\theta}) = \sum_{s=1}^{4} w_s \cdot \mathbf{I}_{R_{(NMNL)}}(\mathbb{C}_s, \boldsymbol{\theta}), \tag{10}$$

where $\mathbf{I}_{\mathrm{R(NMNL)}}(\mathbb{C}_s, \boldsymbol{\theta})$ for each of choice sets are calculated by (Lemma 4 and Section 3):

For $\mathbb{C}_1$:

$$\mathbf{I}_{\mathrm{R(NMNL)}}(\mathbb{C}_1, \boldsymbol{\theta}) = \left[\begin{array}{ccc} \mathbf{I}_{111(0)} & \mathbf{I}_{121(0)} & \mathbf{0} \\ \mathbf{I}_{121(0)}^T & I_{221(0)} & 0 \\ \mathbf{0} & 0 & 0 \end{array} \right] + P_{a_{111}11} \cdot \left[\begin{array}{ccc} \mathbf{I}_{111(a_{111})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{array} \right]$$

$$+ P_{a_{211}11} \cdot \left[\begin{array}{ccc} \mathbf{I}_{111(a_{211})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{array} \right] + P_{a_{121}21} \cdot \left[\begin{array}{ccc} \mathbf{I}_{111(a_{121})} & \mathbf{I}_{121(a_{121})} & \mathbf{0} \\ \mathbf{I}_{121(a_{121})} & I_{221(a_{121})} & 0 \\ \mathbf{0} & 0 & 0 \end{array} \right],$$

where

$$\mathbf{I}_{111(0)} = 4p_{11(0)} \cdot \left[\begin{array}{cc} (1-p_{1|11})\left\{ \frac{p_{1|11}}{\lambda_1^2} + (1-p_{11(0)})(1-p_{1|11}) \right\} & -(1-p_{11(0)})(1-p_{1|11}) \\ -(1-p_{11(0)})(1-p_{1|11}) & (1-p_{11(0)}) \end{array} \right]$$

$$\mathbf{I}_{121(0)} = 2p_{11(0)} \cdot \left[\begin{array}{c} (1-p_{1|11})\left\{ -\frac{2\beta_1 p_{1|11}}{\lambda_1^3} + \frac{(1-p_{11(0)})(\lambda_1 \ln(1-p_{1|11})+2\beta_1 p_{1|11})}{\lambda_1} \right\} \\ -\frac{(1-p_{11(0)})(\lambda_1 \ln(1-p_{1|11})+2\beta_1 p_{1|11})}{\lambda_1} \end{array} \right]$$

$$I_{221(0)} = \frac{4p_{1|11}p_{11(0)}(1-p_{1|11})\beta_1^2}{\lambda_1^4} + \frac{p_{11(0)}(1-p_{11(0)})(\lambda_1 \ln(1-p_{1|11})+2\beta_1 p_{1|11})^2}{\lambda_1^2}$$

$$\mathbf{I}_{111(a_{111})} = 4p_{11(a_{111})} \cdot (1-p_{11(a_{111})}) \cdot \left[\begin{array}{cc} +1 & -1 \\ -1 & +1 \end{array} \right],$$

$$\mathbf{I}_{111(a_{211})} = 4p_{11(a_{211})} \cdot (1-p_{11(a_{211})}) \cdot \left[\begin{array}{cc} 0 & 0 \\ 0 & +1 \end{array} \right]$$

$$\mathbf{I}_{111(a_{121})} = \frac{4p_{1|11}\cdot(1-p_{1|11})}{\lambda_1^2} \cdot \left[\begin{array}{cc} +1 & 0 \\ 0 & 0 \end{array} \right], \quad \mathbf{I}_{121(a_{121})} = -\frac{4p_{1|11}(1-P_{1|11})}{\lambda_1^3} \left[\begin{array}{c} \beta_1 \\ 0 \end{array} \right]$$

$$I_{221(a_{121})} = \frac{4p_{1|11}(1-p_{1|11})}{\lambda_1^4}\beta_1^2, \quad p_{11(a_{111})} = \frac{\exp(-\beta_1+\beta_2)}{\exp(-\beta_1+\beta_2)+\exp(\beta_1-\beta_2)},$$

$$p_{11(a_{211})} = \frac{\exp(\beta_1+\beta_2)}{\exp(\beta_1+\beta_2)+\exp(\beta_1-\beta_2)}$$

$$p_{11(0)} = \frac{\left\{ \exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)+\exp\left(\frac{-\beta_1+\beta_2}{\lambda_1}\right) \right\}^{\lambda_1}}{\left\{ \exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)+\exp\left(\frac{-\beta_1+\beta_2}{\lambda_1}\right) \right\}^{\lambda_1}+\exp(\beta_1-\beta_2)},$$

$$p_{1|11} = \frac{\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)}{\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)+\exp\left(\frac{-\beta_1+\beta_2}{\lambda_1}\right)}.$$

www.SID.ir

Table 3. RO.NMNL model $\lambda_1 = .6, \lambda_2 = .4$ (Two nests): Locally D-optimal design for Design (9), where there are four choice sets each with three alternatives; w.r.t local D -optimal criterion, $\Psi_R(\xi, \boldsymbol{\theta}_0) = (\det(\mathbf{I}_{R(NMNL)}(\xi, \boldsymbol{\theta}_0)))^{-1}$

β_1	β_2	w_1^*	w_2^*	w_3^*	w_4^*	$\Psi_R(\xi^*, \boldsymbol{\theta}_0)$
	-.7	.489	.000	.297	.214	1.370
	-.4	.534	.000	.466	.000	1.233
-.8	0.0	.500	.000	.500	.000	1.164
	.3	.473	.000	.527	.000	1.183
	.6	.459	.000	.541	.000	1.269
	-.7	.487	.019	.494	.000	1.383
	-.4	.501	.000	.499	.000	1.170
-.3	0.0	.504	.000	.496	.000	1.061
	.3	.505	.000	.495	.000	1.096
	.6	.507	.000	.493	.000	1.230
	-.7	.235	.242	.260	.263	3.680
	-.4	.249	.248	.252	.251	2.974
0.0	0.0	.273	.260	.235	.232	2.702
	.3	.293	.268	.223	.217	2.943
	.6	.309	.273	.212	.206	3.619
	-.7	.000	.528	.232	.240	1.183
	-.4	.000	.533	.060	.407	1.102
.6	0.0	.000	.509	.000	.491	1.069
	.3	.256	.262	.000	.482	1.088
	.6	.467	.079	.000	.454	1.146

For $\mathbb{C}_2$:

$$
\mathbf{I}_{R(NMNL)}(\mathbb{C}_2, \boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I}_{112(0)} & \mathbf{I}_{122(0)} & \mathbf{0} \\ \mathbf{I}_{122(0)}^T & I_{222(0)} & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix} + P_{a_{112}12} \begin{bmatrix} \mathbf{I}_{112(a_{112})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix}
$$

$$
+ P_{a_{212}12} \begin{bmatrix} \mathbf{I}_{112(a_{212})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix} + P_{a_{122}22} \begin{bmatrix} \mathbf{I}_{112(a_{122})} & \mathbf{I}_{122(a_{122})} & \mathbf{0} \\ \mathbf{I}_{122(a_{122})}^T & I_{222(a_{122})} & 0 \\ \mathbf{0}^T & 0 & 0 \end{bmatrix},
$$

where

www.SID.ir

$$\mathbf{I}_{112(0)} = 4p_{12(0)} \begin{bmatrix} p_{1|11}\left\{\frac{1-p_{1|11}}{\lambda_1^2} + p_{1|11}(1-p_{12(0)})\right\} & p_{1|11}(1-p_{12(0)}) \\ p_{1|11}(1-p_{12(0)}) & (1-p_{12(0)}) \end{bmatrix}$$

$$\mathbf{I}_{122(0)} = -2p_{12(0)} \begin{bmatrix} p_{1|11}\left\{\frac{2\beta_1(1-p_{1|11})}{\lambda_1^3} + \frac{(1-p_{12(0)})(\lambda_1\ln(1-p_{1|11})+2\beta_1 p_{1|11})}{\lambda_1}\right\} \\ -\frac{(1-p_{12(0)})(\lambda_1\ln(1-p_{1|11})+2\beta_1 p_{1|11})}{\lambda_1} \end{bmatrix}$$

$$I_{222(0)} = \frac{4p_{1|11}p_{12(0)}(1-p_{1|11})\beta_1^2}{\lambda_1^4} + \frac{p_{12(0)}(1-p_{12(0)})(\lambda_1\ln(1-p_{1|11})+2\beta_1 p_{1|11})^2}{\lambda_1^2}$$

$$\mathbf{I}_{112(a_{112})} = 4p_{12(a_{112})}(1-p_{12(a_{112})})\begin{bmatrix} 0 & 0 \\ 0 & +1 \end{bmatrix},$$

$$\mathbf{I}_{112(a_{212})} = 4p_{12(a_{212})}(1-p_{12(a_{212})})\begin{bmatrix} +1 & +1 \\ +1 & +1 \end{bmatrix}$$

$$\mathbf{I}_{112(a_{122})} = \frac{4p_{1|11}(1-p_{1|11})}{\lambda_1^2}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{I}_{122(a_{122})} = -\frac{4p_{1|11}(1-p_{1|11})}{\lambda_1^3}\begin{bmatrix} \beta_1 \\ 0 \end{bmatrix}$$

$$I_{222(a_{122})} = \frac{4p_{1|11}(1-p_{1|11})}{\lambda_1^4}\beta_1^2, \quad p_{12(a_{112})} = \frac{\exp(-\beta_1+\beta_2)}{\exp(-\beta_1+\beta_2)+\exp(-\beta_1-\beta_2)},$$

$$p_{12(a_{212})} = \frac{\exp(\beta_1+\beta_2)}{\exp(\beta_1+\beta_2)+\exp(-\beta_1-\beta_2)}$$

$$p_{1|11} = \frac{\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)}{\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)+\exp\left(\frac{-\beta_1+\beta_2}{\lambda_1}\right)},$$

$$p_{12(0)} = \frac{\left\{\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)+\exp\left(\frac{-\beta_1+\beta_2}{\lambda_1}\right)\right\}^{\lambda_1}}{\left\{\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)+\exp\left(\frac{-\beta_1+\beta_2}{\lambda_1}\right)\right\}^{\lambda_1}+\exp(-\beta_1-\beta_2)}.$$

For $\mathbb{C}_3$:

$$\mathbf{I}_{R(NMNL)}(\mathbb{C}_3, \boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I}_{113(0)} & \mathbf{0} & \mathbf{I}_{133(0)} \\ \mathbf{0} & 0 & 0 \\ \mathbf{I}_{133(0)}^T & 0 & I_{333(0)} \end{bmatrix} + P_{a_{113}13}\begin{bmatrix} \mathbf{I}_{113(a_{113})} & \mathbf{0} & \mathbf{I}_{133(a_{113})} \\ \mathbf{0} & 0 & 0 \\ \mathbf{I}_{133(a_{113})}^T & 0 & I_{333(a_{113})} \end{bmatrix}$$

$$+ P_{a_{223}23}\begin{bmatrix} \mathbf{I}_{113(a_{223})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix} + P_{a_{123}23}\begin{bmatrix} \mathbf{I}_{113(a_{123})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix},$$

where

www.SID.ir

$$\mathbf{I}_{113(0)} = 4(1-p_{13(0)}) \begin{bmatrix} (1-p_{3|23})\left\{\frac{p_{3|23}}{\lambda_2^2} + p_{13(0)}(1-p_{3|23})\right\} & p_{13(0)}(1-p_{3|23}) \\ p_{13(0)}(1-p_{3|23}) & p_{13(0)} \end{bmatrix}$$

$$\mathbf{I}_{133(0)} = 2(1-p_{13(0)}) \begin{bmatrix} (1-p_{3|23})\left\{-\frac{2\beta_1 p_{3|23}}{\lambda_2^3} + \frac{p_{13(0)}(\lambda_2 \ln(1-p_{3|23})+2\beta_1 p_{3|23})}{\lambda_2}\right\} \\ \frac{p_{13(0)}(\lambda_2 \ln(1-p_{3|23})+2\beta_1 p_{3|23})}{\lambda_2} \end{bmatrix}$$

$$\mathbf{I}_{333(0)} = \frac{4p_{3|23}(1-p_{13(0)})(1-p_{3|23})\beta_1^2}{\lambda_2^4} + \frac{p_{13(0)}(1-p_{13(0)})(\lambda_2 \ln(1-p_{3|23})+2\beta_1 p_{3|23})^2}{\lambda_2^2}$$

$$\mathbf{I}_{113(a_{223})} = 4p_{13(a_{223})}(1 - p_{13(a_{223})}) \begin{bmatrix} 0 & 0 \\ 0 & +1 \end{bmatrix},$$

$$\mathbf{I}_{113(a_{123})} = 4p_{13(a_{123})}(1 - p_{13(a_{123})}) \begin{bmatrix} +1 & +1 \\ +1 & +1 \end{bmatrix}$$

$$\mathbf{I}_{113(a_{213})} = \frac{4p_{3|23}(1-p_{3|23})}{\lambda_2^2} \begin{bmatrix} +1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{I}_{133(a_{213})} = -\frac{4p_{3|23}(1-p_{3|23})}{\lambda_2^3} \begin{bmatrix} \beta_1 \\ 0 \end{bmatrix}$$

$$I_{333(a_{213})} = \frac{4p_{3|23}(1-p_{3|23})}{\lambda_2^4}\beta_1^2, \quad p_{13(a_{223})} = \frac{\exp(\beta_1+\beta_2)}{\exp(\beta_1+\beta_2)+\exp(\beta_1-\beta_2)},$$

$$p_{13(a_{123})} = \frac{\exp(\beta_1+\beta_2)}{\exp(-\beta_1-\beta_2)+\exp(\beta_1+\beta_2)}$$

$$p_{13(0)} = \frac{\exp(\beta_1+\beta_2)}{\exp(\beta_1+\beta_2)+\left\{\exp\left(\frac{\beta_1-\beta_2}{\lambda_2}\right)+\exp\left(\frac{-\beta_1-\beta_2}{\lambda_2}\right)\right\}^{\lambda_2}}.$$

Table 4. C.NMNL model, $\lambda_1 = .6, \lambda_2 = .4$ (two nests): Locally D-optimal design for Design (9), where there are four choice sets each with three alternatives; w.r.t local D -optimal criterion, $\Psi(\xi, \boldsymbol{\theta}_0) = (\det(\mathbf{I}_{(NMNL)}(\xi, \boldsymbol{\theta}_0)))^{-1}$ (Comparing to Table 3)

β_1	β_2	w_1^*	w_2^*	w_3^*	w_4^*	$\Psi(\xi^*, \boldsymbol{\theta}_0)$
	-.7	.500	.000	.500	.000	2.099
	-.4	.500	.000	.500	.000	1.829
-.8	0.0	.500	.000	.500	.000	1.716
	.3	.500	.000	.500	.000	1.785
	.6	.500	.000	.500	.000	2.004
	-.7	.469	.066	.464	.000	2.123
	-.4	.486	.028	.486	.000	1.756
-.3	0.0	.500	.000	.500	.000	1.606
	.3	.500	.000	.500	.000	1.721
	.6	.500	.000	.500	.000	2.058
	-.7	.272	.272	.228	.228	4.769
	-.4	.284	.284	.216	.216	3.960
0.0	0.0	.305	.305	.195	.195	3.762
	.3	.324	.324	.176	.176	4.234
	.6	.341	.341	.159	.159	5.345
	-.7	.000	.605	.000	.395	1.691
	-.4	.000	.574	.000	.426	1.495
.6	0.0	.000	.537	.000	.463	1.445
	.3	.000	.521	.000	.479	1.550
	.6	.000	.512	.000	.488	1.807

For $\mathbb{C}_4$:

$$\mathbf{I}_{\mathrm{R(NMNL)}}(\mathbb{C}_4, \boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I}_{114(0)} & \mathbf{0} & \mathbf{I}_{134(0)} \\ \mathbf{0} & 0 & 0 \\ \mathbf{I}_{134(0)}^T & 0 & I_{334(0)} \end{bmatrix} + P_{a_{114}14} \begin{bmatrix} \mathbf{I}_{114(a_{114})} & \mathbf{0} & \mathbf{I}_{134(a_{114})} \\ \mathbf{0} & 0 & 0 \\ \mathbf{I}_{134(a_{114})}^T & 0 & I_{334(a_{114})} \end{bmatrix}$$

$$+ P_{a_{224}24} \begin{bmatrix} \mathbf{I}_{114(a_{224})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix} + P_{a_{124}24} \begin{bmatrix} \mathbf{I}_{114(a_{124})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix}$$

where

www.SID.ir

$$\mathbf{I}_{114(0)} = 4(1 - p_{14(0)}) \begin{bmatrix} p_{3|23}\left(\frac{1-p_{3|23}}{\lambda_2^2} + p_{3|23}p_{14(0)}\right) & -p_{3|23}p_{14(0)} \\ -p_{3|23}p_{14(0)} & p_{14(0)} \end{bmatrix}$$

$$\mathbf{I}_{134(0)} = -2(1 - p_{14(0)}) \begin{bmatrix} p_{3|23}\left\{\frac{2\beta_1(1-p_{3|23})}{\lambda_2^3} + \frac{p_{14(0)}(\lambda_2\ln(1-p_{3|23})+2\beta_1 p_{3|23})}{\lambda_2}\right\} \\ -\frac{p_{14(0)}(\lambda_2\ln(1-p_{3|23})+2\beta_1 p_{3|23})}{\lambda_2} \end{bmatrix}$$

$$I_{334(0)} = \frac{4p_{3|23}(1-p_{14(0)})(1-p_{3|23})\beta_1^2}{\lambda_2^4} + \frac{p_{14(0)}(1-p_{14(0)})(\lambda_2\ln(1-p_{3|23})+2\beta_1 p_{3|23})^2}{\lambda_2^2}$$

$$\mathbf{I}_{114(a_{224})} = 4p_{14(a_{224})}(1 - p_{14(a_{224})}) \begin{bmatrix} +1 & -1 \\ -1 & +1 \end{bmatrix},$$

$$\mathbf{I}_{114(a_{124})} = 4p_{14(a_{124})}(1 - p_{14(a_{124})}) \begin{bmatrix} 0 & 0 \\ 0 & +1 \end{bmatrix}$$

$$\mathbf{I}_{114(a_{114})} = \frac{4p_{3|23}(1-p_{3|23})}{\lambda_2^2} \begin{bmatrix} +1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{I}_{134(a_{114})} = -\frac{4p_{3|23}(1-p_{3|23})}{\lambda_2^3} \begin{bmatrix} \beta_1 \\ 0 \end{bmatrix}$$

$$I_{334(a_{114})} = \frac{4p_{3|23}(1-p_{3|23})}{\lambda_2^4}\beta_1^2,$$

$$p_{14(a_{224})} = \frac{\exp(-\beta_1+\beta_2)}{\exp(-\beta_1+\beta_2)+\exp(\beta_1-\beta_2)} , \quad p_{14(a_{124})} = \frac{\exp(-\beta_1+\beta_2)}{\exp(-\beta_1+\beta_2)+\exp(-\beta_1-\beta_2)} ,$$

$$p_{14(0)} = \frac{\exp(-\beta_1+\beta_2)}{\exp(-\beta_1+\beta_2)+\left\{\exp\left(\frac{\beta_1-\beta_2}{\lambda_2}\right)+\exp\left(\frac{-\beta_1-\beta_2}{\lambda_2}\right)\right\}^{\lambda_2}},$$

$$p_{3|23} = \frac{\exp\left(\frac{\beta_1-\beta_2}{\lambda_2}\right)}{\exp\left(\frac{\beta_1-\beta_2}{\lambda_2}\right)+\exp\left(\frac{-\beta_1-\beta_2}{\lambda_2}\right)}.$$

where $p_{j|ms}$ denotes the conditional probabilities and $I_{..s(j)}$, $p_{ms(j)}$ denote the elements of the information matrix and marginal probabilities without considering alternative j, respectively. We can see Table 3, which includes some locally D-optimal design based on the RO.NMNL model.

Table 5. RO.NMNL model, $\beta_1 = \beta_2 = 0.0$ (two nests): Locally D-optimal design for Design (9), where there are four choice sets each with three alternatives; w.r.t local D -optimal criterion, $\Psi_R(\xi, \boldsymbol{\theta}_0) = (\det(\mathbf{I}_{\text{R(NMNL)}}(\xi, \boldsymbol{\theta}_0)))^{-1}$

λ_1	λ_2	w_1^*	w_2^*	w_3^*	w_4^*	$\Psi_R(\xi^*, \boldsymbol{\theta}_0)$
	0.1	0.321	0.293	0.196	0.190	1.251
	0.2	0.367	0.325	0.158	0.150	1.341
	0.3	0.376	0.331	0.151	0.143	1.358
0.1	0.4	0.378	0.332	0.149	0.140	1.366
	0.8	0.380	0.332	0.149	0.139	1.386
	0.9	0.379	0.332	0.150	0.139	1.392
	0.1	0.220	0.218	0.281	0.280	1.493
	0.2	0.247	0.239	0.258	0.255	2.018
0.4	0.3	0.286	0.268	0.226	0.220	2.306
	0.5	0.336	0.306	0.182	0.175	2.539
	0.8	0.358	0.321	0.164	0.156	2.649
	1.0	0.363	0.324	0.161	0.152	2.693
	0.1	0.214	0.213	0.286	0.286	1.526
	0.2	0.221	0.219	0.280	0.279	2.121
0.8	0.3	0.234	0.229	0.269	0.268	2.529
	0.5	0.268	0.257	0.239	0.236	3.031
	0.8	0.309	0.291	0.202	0.197	3.368

Table 4 denotes some locally D-optimal designs for C.NMNL (Classical NMNL) model. Based on Table 4 and Table 5, it is seen that:

$$\Psi_R(\xi^*, \boldsymbol{\theta}) = (\det(\mathbf{I}_{\text{R(NMNL)}}(\xi^*, \boldsymbol{\theta})))^{-1} \leqslant (\det(\mathbf{I}_{\text{NMNL}}(\xi^*, \boldsymbol{\theta})))^{-1} = \Psi(\xi^*, \boldsymbol{\theta})$$

for all of values of the parameters (true values). Then, it can be argued that the Rank-Order choice experiment is better than the classical choice experiment for estimating NMNL models.

Specially, let $\beta_2 = 0$ and $\lambda_1 = \lambda_2 = \lambda$. Now, based on the two choice sets $\mathbb{C}_1$ and $\mathbb{C}_3$, we will have $p_{1|11} = p_{1|23}$, $p_{2|11} = p_{2|23}$ (Marginal choice probabilities w.r.t all of the alternatives in choice sets $\mathbb{C}_s$; $s = 1, 3$ and $p_{11} = p_{23}$, also there exist similar considerations for two others choice sets $\mathbb{C}_2$ and $\mathbb{C}_4$ so that $p_{1|12} = p_{1|24}$, $p_{2|12} = p_{2|24}$ and $p_{12} = p_{24}$. Due to the symmetry considerations, we can derive an optimal solution with weights $w_1 = w_3$ and

www.SID.ir

$w_2 = w_4$ for the design (9), where $2w_1 + 2w_2 = 1$ or $w_2 = \frac{1}{2} - w_1$ as follows:

$$\xi = \left\{ \begin{array}{cccc} \mathbb{C}_1 & \mathbb{C}_2 & \mathbb{C}_3 & \mathbb{C}_4 \\ w_1 & \frac{1}{2} - w_1 & w_1 & \frac{1}{2} - w_1 \end{array} \right\} \in \Xi \qquad (11)$$

In this case, $w_1^* = 0.5$ when $\beta_1 < -0.05$ (Table 6) so that we can consider a locally D-optimal design, when $\lambda = 0.1$, as follows:

$$\xi_{(\beta_1 < -0.05)}^* = \left\{ \begin{array}{cccc} \mathbb{C}_1 & \mathbb{C}_2 & \mathbb{C}_3 & \mathbb{C}_4 \\ 0.5 & 0.0 & 0.5 & 0.0 \end{array} \right\}.$$

But if β_1 is positive then w_1^* first decreases (when $0 \leqslant \beta_1 < 0.01$, for too small values of β_1) then increases (when $0.05 \leqslant \beta_1$) as β_1 increases (Table 6). For instance, in comparison, when $w_{1(\beta_1=0.3)}^* = 0.000$, $w_{1(\beta_1=0.6)}^* = 0.005$, $w_{1(\beta_1=0.9)}^* = 0.068$ and $w_{1(\beta_1=1.0)}^* = 0.315$, more cases were calculated in Table 6 (the Sequential Quadratic Programming method by MAPLE).

Let's consider $\lambda = 1.0$. Table 6 denotes that w_1^* decreases as β_1 increases. So that, we will face (approximately) the following locally D-optimal design:

$$\xi^* = \left\{ \begin{array}{cccc} \mathbb{C}_1 & \mathbb{C}_2 & \mathbb{C}_3 & \mathbb{C}_4 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{array} \right\},$$

where β_1 tends to zero from both right and left.

In the other cases we suppose that $\boldsymbol{\beta} = \mathbf{0}, \lambda_2 = 0.1$. Based on the combination of alternatives in choice sets $\mathbb{C}_1$ to $\mathbb{C}_4$, there will be permutation between the two choice sets $\mathbb{C}_1$ and $\mathbb{C}_2$, by permuting the levels of the first attribute in the second nest. Furthermore, by permutation, the levels of the first attribute in the first nest will have permutation between the two nests $\mathbb{C}_3$ and $\mathbb{C}_4$. Now, we consider the following invariant design instead:

$$\xi = \left\{ \begin{array}{cccc} \mathbb{C}_1 & \mathbb{C}_2 & \mathbb{C}_3 & \mathbb{C}_4 \\ w & w & \frac{1}{2} - w & \frac{1}{2} - w \end{array} \right\} \in \Xi. \qquad (12)$$

Table 6. RO.NMNL model, $\beta_2 = 0$, $\lambda_1 = \lambda_2 = \lambda$ (two nests): Locally D-optimal design, w_1^*, with respect to Design (11)

		λ									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
	0.9	0.068	0.056	0.021	0.000	**0.000**	0.000	0.000	0.000	0.000	0.041
	0.8	0.044	0.027	0.000	**0.000**	0.000	0.000	0.000	0.000	0.000	0.047
	0.7	0.024	0.000	**0.000**	0.000	0.000	0.000	0.000	0.000	0.000	0.059
	0.6	0.005	**0.000**	0.000	0.000	0.000	0.000	0.000	**0.000**	0.019	0.078
	0.5	**0.000**	0.000	0.000	0.000	0.000	0.000	**0.000**	0.000	0.048	0.101
	0.4	0.000	0.000	0.000	0.000	0.000	**0.000**	0.000	0.025	0.084	0.129
	0.3	0.000	0.000	0.000	0.000	**0.000**	0.000	0.018	0.080	0.125	0.159
	0.2	0.000	0.000	0.000	**0.000**	0.000	0.041	0.101	0.141	0.169	0.191
	0.1	0.000	0.000	**0.000**	0.062	0.127	0.161	0.184	0.200	0.212	0.222
	0.05	0.000	0.024	0.154	0.188	0.201	0.211	0.220	0.228	0.234	0.238
β_1	0.001	0.263	0.257	0.257	0.257	0.257	0.257	0.256	0.256	0.255	0.255
	-0.001	0.281	0.265	0.263	0.261	0.260	0.259	0.258	0.257	0.257	0.256
	-0.05	0.500	0.500	0.500	0.424	0.356	0.322	0.303	0.290	0.281	0.275
	-0.1	0.500	0.500	0.500	0.500	0.492	0.408	0.360	0.330	0.310	0.296
	-0.2	0.500	0.500	0.500	0.500	0.500	0.500	0.498	0.425	0.377	0.343
	-0.3	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.454	0.398
	-0.4	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.458
	-0.5	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
	-0.6	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
	-0.7	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
	-0.8	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
	-0.9	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500

Table 7. RO.NMNL model, $\boldsymbol{\beta} = \mathbf{0}$: Locally D-optimal design, w^*, with respect to Design (12)

		λ₁									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
	0.1	**0.307**	0.243	0.225	0.219	0.217	0.215	0.214	0.214	0.213	0.213
	0.2	0.346	**0.306**	0.265	0.243	0.232	0.226	0.223	0.220	0.219	0.217
	0.3	0.353	0.335	**0.305**	0.277	0.257	0.245	0.237	0.232	0.228	0.225
	0.4	0.355	0.345	0.327	**0.305**	0.283	0.267	0.255	0.246	0.240	0.236
λ_2	0.5	0.356	0.349	0.337	0.321	**0.304**	0.287	0.273	0.263	0.254	0.248
	0.6	0.356	0.351	0.342	0.331	0.317	**0.303**	0.289	0.278	0.269	0.261
	0.7	0.356	0.352	0.345	0.336	0.325	0.313	**0.301**	0.290	0.281	0.273
	0.8	0.356	0.353	0.347	0.340	0.331	0.321	0.310	**0.300**	0.291	0.283
	0.9	0.356	0.353	0.348	0.342	0.334	0.326	0.317	0.308	**0.299**	0.291
	1.0	0.355	0.353	0.349	0.343	0.337	0.329	0.321	0.313	0.305	**0.298**

Based on design (12) and the assumptions $\boldsymbol{\beta} = \mathbf{0}, \lambda_2 = 0.1$, Table 7 denotes that w^* increases as λ_1 decreases, when $\lambda_2 = 0.1$. For different values of λ_1 and λ_2, optimal weight has been shown by Table 7 (It has used the Sequential Quadratic Programming method by MAPLE). It can be seen that the optimal weight, w^*, decreases when λ_1 and λ_2 are equal and also increase (Table 7). Also, with respect to the fixed values for the dissimilarity parameter λ_2, Table 7 denotes that the optimal weight, w^*, decreases as λ_1 increases. That means that the alternatives of the two choice sets $\mathbb{C}_1, \mathbb{C}_2$ are more similar than the combination of the alternatives in the two other choice sets; $\mathbb{C}_3, \mathbb{C}_4$. But, based on fixed amounts of λ_1 optimal weight, w^*, has an increasing trend (expect when $0 < \lambda_1 < 0.2$) as λ_2 increases. That means, the combination of the alternatives in the two choice sets $\mathbb{C}_3, \mathbb{C}_4$ are less similar than the alternatives in two choice sets $\mathbb{C}_1, \mathbb{C}_2$. Table 7 also denotes that for fixed small values of λ_1 ($0 < \lambda_1 < 0.2$), the optimal weight, first, has an increasing then a decreasing trend when λ_2 increases.

4   Discussion

This paper has dealt with a model from non-linear models. This model belongs to a very complex model from logit family. In this kind of logit model we have tried to obtain the probability related to choosing an alternative with the highest utility in the choice set. In this situation, a consumer prefers to choose an alternative with the best utility within all of the alternatives in a choice set. But, sometimes we like to know her/his preference about the other alternatives in the same choice set (except the best alternative). Therefore, we want the consumer to choose other alternatives based on his/her preference. Of course, in this step it is not necessary to rank all of the alternatives in the choice set by the consumer, instead, she/he can rank a part of them.

In this paper, in contrast to classic logit models (McFadden, 1974) the alternative with the highest utility will be removed from the choice set, then the next alternative will be chosen from the new choice set which excludes that alternative with the highest utility. This algorithm will be continued to choose suitable alternatives (not all alternatives) in the choice set.

In continuation, based on several choice sets we have defined a design (for estimating parameters), then we have obtained the information matrix of that design. Based on D-optimal criterion and with respect to the local optimality criterion (non-linear complex model) a suitable criterion was

introduced and it was determined how we can obtain a locally D-optimal design. Finally, we have described this method by an example.

References

Atkinson, A.C., Donev, A.N. and Tobias, R.D. (2007). *Optimum Experimental Design, with SAS*. Oxford University, Press.

Beggs, S., Cardell, S. and Hausman, J. (1981). Assessing thepotentail demand for for electric cars, *Journal of Econometrics,* **16**, 1-19.

Chapman, R.G. and Staelin, R. (1982). Exploiting rank ordered choice set data whitin the stochastic utility model,. *Journal of Marketing Research,* **19**, 288-301.

Chernoff, H. (1972). Sequentail analysis and optimal design. *Society for Industrial and Applied Mathematics , Philadelphia PA.*

Hausman, J.A. and Ruud, P.A. (1987). Specifing and testing econometric model for rank-order data, *Journal of Econometrics,* **34**, 83-104.

Jafari, H. (2010). Optimal Design In Conjoint Analysis, *PhD Thesis, IMST*.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka Ed. *Frontiers in Econometrics, 105-142 Academic Press.*

Pukelsheim, F., (1993). *Optimal Design of Experiments*. Wiley, New York.

Sandor, Z. and Wedel, M. (2001). Designing conjoint choice experiments using Mangaer, *Journal of Marketing Research,* **38**, 430-444.

Silvey, S.D. (1980). *Optimal Design*. Chapman and Hall, London.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Vermeulen, B., Goos, P., Vandebroek, M. Rank-Order Conjoint Experiments: Efficiency and design. *Preprint.*

Habib Jafari
Department of Statistics,
Razi University,
Kermanshah, Iran.
email: *jafari_habib@yahoo.com*
        jafari_habib@hotmail.de