Contents

List of Tables

List of Figures

1 INTRODUCTION

"*Conjoint analysis*" is concerned with understanding how people make choice between products or services or a combination of product and service, so that businesses can design new products or services that better meet customers underlying needs. A key benefit of *"conjoint analysis"* is the ability to produce dynamic market models that enable companies to test out what steps they would need to take to improve their market share, or how competitors behavior will affect their customers. *"Conjoint analysis"*, also called multi-attribute compositional models, is a statistical technique that originated in mathematical psychology. Today it is used in many of the social sciences and applied sciences including marketing, product management, and operations research. The objective of *"Conjoint analysis"* is to determine what combination of a limited number of attributes is most preferred by respondents.

Discrete choice experiments play an important role in psychology and market research for measuring the consumer's preferences. Usually, the choice behavior is modeled by a *multinomial* response, where the probabilities of preferences are given by a logistic model.

(Thurston 1927) has introduced his law of comparative judgment, a model based on utility function $U_{ij} = v_{ij} + \varepsilon_{ij}$ with a normal error where i and j denote individual and alternative. He has also showed that the probability $P_{i(j,j')}$ that alternative j is chosen over alternative j' has a form that now it is called binomial probit. *(Marschak 1960)* generalized Thurston' law of comparative judgment to stochastic utility maximization in multinomial choice sets which is called the Random Utility Maximization (RUM) model. *(Luce 1959)* introduced an axiomatic treatment of choice behavior that the ratio of choice probabilities for j and j' not depends on the other alternatives (in every choice set) which is called Independence from Irrelevant Alternatives (IIA). In continuation, the strict utilities in Luce model has been parameterized in a form suitable for econometric applications which is called conditional logit, now known as multinomial logit (MNL)*(McFadden 1974)*. Also, *(Williams 1977)*, *(Daly and Zachary 1978)* and *(Ben-Akiva and Lerman 1985)* have developed RUM justifications for the Nested MNL (NMNL) model. The decade has seen extensive development and use of open (Multinomial Probit (MNP)) and closed form choice models consistent with RUM, including General Extreme Value (GEV) models and mixing in the parameters of MNL and NMNL models. These models belong to a family of models which are called discrete choice models. From these models are used to analyze data in classical discrete choice experiment, where the respondent is asked to choose an alternative with the highest utility among alternatives in a choice set. But there is the other experiments from discrete choice experiments that the respondent is asked to rank a number of alternatives instead of preferred one which are called Rank-Order experiments. Data from a rank-order experiment can be analyzed by the rank-ordered exploded Logit models *(Beggs, et al. 1981)*, *(Hausman and Ruud 1987)*.

Discrete choice models have been applied in divers fields and are increasingly employed in the social sciences. Of the existing discrete choice model, the Probit, Logit and the Nested logit models are the most commonly used in practice. The logit and the Nested logit models belong to the family of generalized extreme value (GEV) models. The Probit model, although allowing for flexible covariance structures for the disturbance terms, is computationally burdensome for problems with more than a few alternatives because it requires the evaluation of multiple integrals. Such that until recently, the model was not feasible for choice problems with more than three alternatives. The GEV class, unlike the Probit model, is computationally manageable for large choice sets because it involves at most one dimensional integration regardless of the number of choice alternatives in the model. However, it suffers from the restriction of Homoscedastic disturbance. Homoscedasticity is a troubling assumption, since choice models are most often used with micro-level data that are frequently Heteroscedastic.

The error distribution function of the GEV class in particular is less familiar than the normal distribution for the Probit models, and the fact that the distribution imposes Homoscedasticity is not readily noticed. The effect of Heteroscedasticity in choice models is for more serious than in linear models. In a linear model, if Heteroscedasticity is ignored, the least squares estimate is still unbiased and consistent, although inefficient. In a choice model, the maximum likelihood estimators are not only biased and inefficient but even inconsistent *(Yatchew and Griliches 1985), (Greene 1997).*

(Zeng 2000) has explained a technique to relax the restriction of homoscedasticity in the entire GEV class of models to allow for heteroscedasticity across alternatives as well as across decision makers.

Optimal design for conjoint analysis is the topic of this thesis. The design of a choice experiment comprises a select number of choice sets administered to each respondent. The aim of a choice experiment is to estimate the importance of each attribute and their levels based on the respondents preferences. The estimates are then used to mimic real marketplace choices by making predictions about consumer future purchases. At present, two design approaches are prevalent; (i) The Linear design approach and (ii) The Baysian design approach. Bayesian choice designs have so far been constructed for the Logit models. Since the Logit models are nonlinear in the parameters, the quality of a given design depends on the unknown parameter vector. The Bayesian design approach deals with this problem by assuming a prior distribution of likely parameters. To date, most of the Bayesian research focus has been on designs for main-effects models. *(Sandor and Wedel 2001)* were the first to introduce the Bayesian design procedure in the choice design literature. They generated Bayesian designs using the D-optimality criterion for the MNL model. This design criterion seeks to minimize the determinate of the variance-covariance matrix of the parameter estimators. In the Bayesian framework, it is referred to as the D_b-optimality criterion. Of course, four optimality criteria are used in the Bayesian context which are labeled the D_b- ,A_b-,G_b- and V_b-optimality criteria. In this thesis, we use D-optimality criterion, since, *(Yu, et al. Preprint)* have written that D-optimality criterion is invariance to the scale or coding of the attributes. Also, the relative efficiency of the designs does not change when different codings of the attributes are used *(Goos 2002)*. Also, *(Kessels, et al. 2006b)* have denoted that D-and A-optimal designs are nearly as good as the

G-and V-optimal designs in terms of prediction quality but much faster to compute compared to G- and V -optimal designs.

The thesis is organized as follows.

In chapter 2, two kinds of logit models will be discussed, which are called Multinomial logit (MNL) and Nested MNL model. According to similarity between alternatives in a choice set, the Nested MNL model may be Two-level, Three-level,... . Of course, in this thesis we have just discussed about Two and Three-level NMNL model.

Optimal design and some of optimal criteria like D- and A-criterion are explored in chapter 3. In this chapter, we concentrate of D-criterion which is a function of the determinant of the information matrix. Bayesian criterion, which is one of the suitable criteria to obtain optimal design in nonlinear model (specially), is introduced in this chapter. In addition, in Subsection 3.7.1 of chapter three have been discussed about optimal design in the MNL model.

The three remaining chapters are the principle chapters of the thesis. In chapter 4 we calculate the information matrix for the two-level NMNL model with M nests. Afterwards, we illustrate two examples based on the local D-optimality criterion. In Chapter 5 the information matrix for a three-level nested MNL model will be calculated. With respect to Random Utility Maximization (RUM) conditions and D-optimality criterion we obtain the locally D-optimal design based on experiments $2^3/5/6$ (there are three attributes each with two levels, where six choice sets each with five alternatives has been selected from population). In chapter 6, we introduce a model of the logit family, which also includes the probabilities of choosing alternatives with lower utility, in a choice set. In this situation, the alternatives with upper utility (Ranking) are removed when we want to obtain the choice probabilities of the alternatives with lower utility. This model is called Rank-Order logit model. In this chapter 6 we have introduced the likelihood function of the Rank-Order Nested MNL model, then we have calculated the information matrix for this model. Also, we have obtained Locally D-optimal design for this model.

2 MODEL SPECIFICATION

Modeling the individual behavior of consumers is one of the main topics in marketing research. This individual behavior is influenced by socio-economic characteristics, marketing instruments or latent variables. The connection between these influencing variables and the choice of a product is typically studied by using a statistical choice model for disaggregated data.

A classic choice model is the conditional logit model of (McFadden 1974). It is widely discussed and a standard in marketing (Guadagni and Little 1983). This model however has some disadvantages in particular the IIA (Independence of Irrelevant Alternatives) and a very restrictive assumption about the errors. This led to many approaches for relaxing these assumptions. For overviews see *(Ben-Akiva, 1973)* and *(Horowitz, et al. 1994)*.

All these approaches present alternative ways for modeling consumer purchase and obtain results which adapt better to the data than the classic approach. How ever, to our knowledge no general statistical test to check adequateness of the logit model was applied to marketing data until now.

(Bartelts, et al. 1999) have introduced a test procedure which will help in finding an appropriate consumer purchase model. The test is based on a nonparametric test statistic which makes it a very flexible and general tool.

The simplest model in logit family is Multinomial logit model. But this model has a retraction that all of alternatives are independent in choice set and it is called IIA. In logit model family there are other models that this property dose not hold between all of alternatives like Nested MNL model which hold just between alternatives in each of nests. And there are some models in this family that IIA dose not hold between all of alternatives like Probit logit model. But, we concentrate on two models of logit model, multinomial and nested logit models.

2.1 Multinomial Logit Model

Many researchers use data on individuals to analyze postsecondary attendance behavior. With these data the enrollment choices of the individuals are made over a limited number of "discrete" alternatives that constitute the exhaustive set of available education options. It is now well-known that using ordinary least squares (OLS) to analyze relationships in which the dependent variable is discrete or qualitative is not appropriate. If there are just two alternatives in the choice set, logit or probit analyzes are often used to estimate the relationship between the option selected and the characteristics of the alternatives and of the individuals in the data sample. In analysis of enrollment choices these methods have been used to explain the choice between attendance and nonattendance or between attendance at a particular institution and not attending that institution.

It is clearly of interest to extend the analysis to choice among several types of attendance options or even to a limited number of individual institutions. When there are more than two alternatives in the choice set, the most widely used approach is multinomial or conditional logit (MNL). This method is computationally simple with today's software and computers and also has another desirable property: It is easy to use the estimated results to forecast choices when a new alternative is introduced or when an existing one is eliminated so long as no parameters are added or deleted as a result. This is a useful property, for example, in estimating the effects of either opening or closing a postsecondary educational institution. However, MNL also has a distinct limitation, a property known as "independence of irrelevant alternatives" (IIA), which implies that the odds of choosing alternative j relative to alternative j' are independent of the characteristics of or the availability of alternatives other than j and j'. This is clearly a very restrictive property when the alternatives being studied have different degrees of "nearness" or similarity.

Now, here, we consider choice models based on the assumption of stochastic utility maximization. Under this assumption, a decision maker chooses the alternative that maximizes his or her utility function, which has both a deterministic component and a stochastic component. In formal notations, assume a sample of $\mathcal{I}$ decision makers, each choosing among $\mathcal{C}$ choice sets which each of them include J_c discrete alternatives, where $J_c > 1; \forall c \in \mathcal{C}$. Now, if we suppose that $\mathbb{C}$ is choice set, which consists all of alternatives,$\mathcal{J}$, ($\mathbb{C} = \{\tilde{a}_1, \ldots, \tilde{a}_j, \ldots, \tilde{a}_{\mathcal{J}}\}$) and $\mathbb{C}_c$ is a choice set which includes J_c alternatives then we can write that $\bigcup_{c=1}^{\mathcal{C}} \mathbb{C}_c = \mathbb{C}: \ \exists c, c';\ \mathbb{C}_c \bigcap \mathbb{C}_{c'} \neq \phi$, where $2^{\mathcal{J}} - (\mathcal{J}+1)$ is the number of choice sets in set $\mathcal{C}$ ($\mathcal{C}$ is the set of all of choice sets each with $J_c > 1; \forall c$ alternatives). Here, it has been used notation $\mathbb{C}_c$ to denote a choice set with J_c alternatives, where $\mathbb{C}_c = \{a_{1c}, \ldots, a_{jc}, \ldots, a_{J_c c}\}$ (a_{jc} denotes the j^{th} alternative of choice set c). In this situation, it has been supposed that there are K attributes each with $L_k; k = 1, 2, \ldots, K$ levels. Consequently, it has been considered S choice sets each with J_s alternatives to fit model.

Obtaining the probability related to choosing an alternative with the highest utility is principle aim in the MNL models. In this thesis, for each individual i, each alternative j provides utility (ignoring index i):

$$U_{jc} = v_{jc} + \varepsilon_{jc}. \tag{2.1}$$

In above function, v_{jc} is a deterministic component, usually specified as a linear function of observed independent variables, such as:

$$v_{jc} = \mathbf{f}^T(a_{jc})\boldsymbol{\beta} = \sum_{k=1}^{K} \mathbf{f}_k^T(a_{jc})\boldsymbol{\beta}_k^T,$$

where:

- $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_k^T, \ldots, \boldsymbol{\beta}_K^T); \quad \boldsymbol{\beta}_k^T = (\beta_{k,1}, \ldots, \beta_{k,\ell}, \ldots, \beta_{k,L_k}),$
- $\mathbf{f}(a_{jc}) = (\mathbf{f}_1^T(a_{jc}), \ldots, \mathbf{f}_k^T(a_{jc}), \ldots, \mathbf{f}_K^T(a_{jc}))^T;\ \mathbf{f}_k(a_{jc}) = (f_{k1}(a_{jc}), \ldots, f_{k\ell}(a_{jc}), \ldots, f_{kL_k}(a_{jc}))^T.$

In this case, $\boldsymbol{\beta}_k^T$ and $\mathbf{f}_k^T(a_{jc})$ denote the part-worth parameter and the characterizes of attribute k in choice set $\mathbb{C}_c$. The ε_{jc} is an unobserved, stochastic disturbance ($i.i.d$) with mean zero and variance σ^2 (for each choice set), thus:

$$cov(U_{jc}, U_{j'c}) = \begin{cases} \sigma^2, & j = j'; \\ 0, & j \neq j'. \end{cases}$$

Thus the variance-covariance matrix for the vector $U_c = (U_{1c}, \ldots, U_{jc}, \ldots, U_{J_c c})^T$ is as below:

$$V(U_s) = \Sigma_{U_s} = \begin{pmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{I}_{J_c}.$$

Under stochastic utility maximization, an individual i chooses alternative j (in choice set $\mathbb{C}_c$) if and only if $U_{jc} > U_{j'c}, \ \forall j \neq j'$:

$$U_{jc} = \max_{j' \in \mathbb{C}_c} U_{j'c}, \tag{2.2}$$

where $\mathbb{C}_c$ denotes a choice set with J_c alternatives.

Here, we are going to calculate the probability of choosing alternative j which has the highest utility among the other (in the choice set). Therefore, the observation variables to calculate the choice probabilities are defined as follows:

$$Y_{jc} = \begin{cases} 1, & U_{jc} = \max_{j' \in \mathbb{C}_c} U_{j'c}; \\ 0, & \text{otherwise.} \end{cases},$$

where:

$$p_{jc} = P(Y_{jc} = 1) = P(U_{jc} = \max_{j' \in \mathbb{C}_c} U_{j'c}) = P(U_{jc} > U_{j'c}; \forall j' \neq j \in \mathbb{C}_c), \tag{2.3}$$

where $P(A)$ denotes the probability related to occur event A. In this situation, $E(Y_{jc}) = p_{jc}$ and

$$cov(Y_{jc}, Y_{j'c}) = \begin{cases} p_{jc} \cdot (1 - p_{jc}), & j = j'; \\ -p_{jc} \cdot p_{j'c}, & j \neq j'. \end{cases}$$

Thus the variance-covariance matrix for the vector $Y_c = (Y_{1c}, \ldots, Y_{jc}, \ldots, Y_{J_c c})^T$ is calculated as follow:

$$\begin{aligned} V(Y_c) = \Sigma_{Y_c} &= \begin{pmatrix} p_{1c} \cdot (1 - p_{1c}) & \cdots & -p_{1c} \cdot p_{jc} & \cdots & -p_{1c} \cdot p_{J_c c} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ -p_{jc} \cdot p_{1c} & \cdots & p_{jc} \cdot (1 - p_{jc}) & \cdots & -p_{jc} \cdot p_{J_c c} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ -p_{J_c c} \cdot p_{1c} & \cdots & -p_{J_c c} \cdot p_{jc} & \cdots & p_{J_c c} \cdot (1 - p_{J_c c}) \end{pmatrix} \\ &= \mathbf{P}_c - \mathbf{p}_c \mathbf{p}_c^T, \end{aligned}$$

where:

- $\mathbf{P}_c = diag[p_{1c}, \ldots, p_{jc}, \ldots, p_{J_c c}]$,
- $\mathbf{p}_c = [p_{1c}, \ldots, p_{jc}, \ldots, p_{J_c c}]^T$.

According to (2.1) we rewrite (2.3) as follow:

$$p_{jc} = P(\varepsilon_{j'c} < (v_{jc} - v_{j'c} + \varepsilon_{jc}); \forall j' \neq j \in \mathbb{C}_c). \tag{2.4}$$

For calculating this probability, we need to know the distribution of "ε"s. There are two important criteria for choosing the error distribution function are functional flexibility and computational efficiency. The *Probit* model is obtained by assuming the multivariate normal distribution for the disturbances. Because, there are no prior restrictions on the form of the covariance matrix of a multivariate normal distribution, hence, the *probit* model is functionally flexible *(Hausman and Wise 1978)*. However, under the multivariate normal distribution there is no closed form solution to (2.4) when there are more than three alternatives. With the recent development of simulation methods *(McFadden 1989)*, has made it theoretically possible to apply the model to a large number of choice. Due to problems of fragile identification, however, in practice it is still rarely used for more than three or four alternatives. Because many theoretically important decisions involve more than a few alternatives, researchers must often turn to alternatives models, usually the Generalized Extreme Value (GEV) class, which includes the familiar logit and nested logit model. Now, we first introduce the GEV classes then the MNL model will be obtained.

2.1.1 The Generalized Extreme Value(GEV)

In this section, it is introduced a family of choice models derived from Stochastic Utility Maximization (SUM) or Random Utility Maximization (RUM), which includes *multinomial* and nested *logit* model.

GEV distributions have application in the study of discrete choice behavior, and were initially studied by *(McFadden 1978a)*, *(McFadden 1981)*, *(McFadden 1984)*, *(McFadden 2001)* with the following result, which characterizes this family (ignoring index i for simplicity and considering a choice set ($\mathbb{C}_c$) with J_c alternatives).

Theorem 2.1. (McFadden 1981) *Suppose that $G_{J_c}(z_{1c}, \ldots, z_{jc}, \ldots, z_{J_c c})$ is non-negative, homogeneous of degree one function of $(z_{1c}, \ldots, z_{jc}, \ldots, z_{J_c c}) \geq 0$ such that;*

$$\lim_{z_{jc} \to +\infty} G_{J_c}(z_{1c}, \ldots, z_{jc}, \ldots, z_{J_c c}) = +\infty; \forall j = 1, 2, \ldots, J_c,$$

where $G_{J_c} : \Re^{J_c} \to \Re$. Also for any distinct $(j_1, \ldots, j_n)$ from $\{1, \ldots, J_c\}$ have that:

$$\frac{\partial^n G_{J_c}(z_{1c}, \ldots, z_{jc}, \ldots, z_{J_c c})}{\partial z_{j_1}, \ldots, \partial z_{j_n}} \begin{cases} \geq 0, & \textit{if } n \textit{ be odd;} \\ \leq 0, & \textit{if } n \textit{ be even.} \end{cases}$$

In this case: $p_{js} = \frac{z_{js} \cdot G_{J_c}^{jc}(z_{1c},\ldots,z_{jc},\ldots,z_{J_c c})}{G(z_{1c},\ldots,z_{jc},\ldots,z_{J_c c})}$, *where* $G_{J_c}^{jc} = \frac{\partial G_{J_c}(z_{1c},\ldots,z_{jc},\ldots,z_{J_c c})}{\partial z_{jc}}$ *and with assumption* $z_{jc} = \exp(v_{jc})$ *will be:*

$$p_{jc} = \frac{\exp(v_{jc}) \cdot G_{J_c}^{jc}(\exp(v_{1c}),\ldots,\exp(v_{jc}),\ldots,\exp(v_{J_c c}))}{G_{J_c}(\exp(v_{1c}),\ldots,\exp(v_{jc}),\ldots,\exp(v_{J_c c}))}. \tag{2.5}$$

Example 2.1. *suppose that:*

$$G(z) = G_{J_c}(z_{1c},\ldots,z_{jc},\ldots,z_{J_c c}) = \sum_{j=1}^{J_c} z_{jc},$$

where $G(\alpha z) = \alpha G(z)$ *is a generating function with the homogeneous of degree one. According to assumption* $z_{jc} = \exp(v_{jc})$, *we will have:*

$$\frac{\partial G_{J_c}(\exp(v_{1c}),\ldots,\exp(v_{jc}),\ldots,\exp(v_{J_c c}))}{\partial \exp(v_{jc})} = 1 \quad \Rightarrow \quad p_{jc} = \frac{\exp(v_{jc})}{\sum_{j'=1}^{J_c} \exp(v_{jc})}. \tag{2.6}$$

Consequently, the model (2.6) is called the MNL model. Under the assumptions of Theorem 2.1, the expectation of maximum utility is calculated as follow *(McFadden 1984)*:

$$E(\max_{j\in\mathbb{C}_c} U_{jc}) = \log G_{J_c}(\exp(v_{1c}),\ldots,\exp(v_{jc}),\ldots,\exp(v_{J_c c})) + \gamma, \tag{2.7}$$

where γ is *Euler*'s constant and $\log G_{J_c}(\exp(v_{1c}),\ldots,\exp(v_{jc}),\ldots,\exp(v_{J_c c}))$ is called inclusive value, IV_c (the expected utility for the choice of alternative within choice set $\mathbb{C}_c$). In this situation and by noting to Equation (2.7) and Theorem 2.1 we will have:

$$\begin{aligned}
\frac{\partial E(\max_{j\in\mathbb{C}_c} U_{jc})}{\partial v_{jc}} &= \frac{\frac{\partial G_{J_c}(\exp(v_{1c}),\ldots,\exp(v_{jc}),\ldots,\exp(v_{J_c c}))}{\partial v_{jc}}}{G_{J_c}(\exp(v_{1c}),\ldots,\exp(v_{jc}),\ldots,\exp(v_{J_c c}))} \\
&= \frac{\exp(v_{jc}) \cdot G_{J_c}^{jc}(\exp(v_{1c}),\ldots,\exp(v_{jc}),\ldots,\exp(v_{J_c c}))}{G_{J_c}(\exp(v_{1c}),\ldots,\exp(v_{jc}),\ldots,\exp(v_{J_c c}))} \\
&= p_{jc}
\end{aligned}$$

Now, since (based on *Euler*'s Law):

$$\sum_{j=1}^{J_c} \exp(v_{jc}) \cdot G_{J_c}^{jc}(\exp(v_{1c}),\ldots,\exp(v_{jc}),\ldots,\exp(v_{J_c c})) = G_{J_c}(\exp(v_{1c}),\ldots,\exp(v_{jc}),\ldots,\exp(v_{J_c c}))$$

thus:

$$\sum_{j=1}^{J_c} \frac{\partial E(\max_{j\in\mathbb{C}_c} U_{jc})}{\partial v_{jc}} = \sum_{j=1}^{J_c} p_{jc} = 1.$$

A Special Case: According to the utility (2.1), it has been assumed that ε_j's have *i.i.d* extreme value distribution type(II) ($\varepsilon_{1c}, \ldots, \varepsilon_{J_c c}$ have *i.i.d* EVD type II), so that:

$$F_{\varepsilon_j}(t) = \exp(-\exp(-t)); t \in \Re. \tag{2.8}$$

With regards to the definition of the choice probability related to choosing alternative j (in a choice set, $\mathbb{C}_c$) will be:

$$\begin{aligned} p_{jc} &= P(U_{jc} > U_{j'c}; \forall j' \neq j \in \mathbb{C}_c) = P(v_{jc} + \varepsilon_{jc} > v_{j'c} + \varepsilon_{j'c}; \forall j' \neq j \in \mathbb{C}_c) \\ &= \int_{-\infty}^{+\infty} P(\underbrace{\varepsilon_{1c} < v_{jc} - v_{1c} + \varepsilon_{jc}, \ldots, \varepsilon_{J_c c} < v_{jc} - v_{J_c c} + \varepsilon_{jc}}_{\text{without } j^{th} \text{ component}}) f_{\varepsilon_{jc}}(\epsilon_{jc}) d\epsilon_{jc} \\ &= \int_{-\infty}^{+\infty} \underbrace{F_{\varepsilon_{1c}}(v_{jc} - v_{1c} + \epsilon_{jc}) \ldots F_{\varepsilon_{J_c c}}(v_{jc} - v_{J_c c} + \epsilon_{jc})}_{\text{without } j^{th} \text{ distribution function}}) f_{\varepsilon_{jc}}(\epsilon_{jc}) d\epsilon_{jc}. \end{aligned}$$

By noting Equation (2.8) the solution of the above integral is obtained as follow:

$$p_{jc} = \frac{\exp(v_{jc})}{\sum_{j'=1}^{J_c} \exp(v_{j'c})}; \quad j = 1, 2, \ldots, J_c. \tag{2.9}$$

Model (2.9) is called standard multinomial logit model. Since, the variance of this model for all of decision makers are as the same, then this model is also called Homoscedastic MNL model.

2.1.2 Consistency with Random Utility Maximization (RUM)

(McFadden 1981) has showed that any set of choice probabilities that satisfies a set of compatibility conditions defines a stochastic utility maximization model (based on the maximum utility of alternatives) with an implied joint distribution of the stochastic utility components. These compatibility conditions are as follow (ignoring index i and with respect to a choice set with J_c alternatives):

$$p_{jc} = p_j(\mathbf{v}_c) \geq 0, \sum_{j=1}^{J_c} p_j(\mathbf{v}_c) = 1, \quad p_j(\mathbf{v}_c) = p_j(\mathbf{v}_c + r\mathbf{1}), \forall r \in \Re, \tag{2.10}$$

$$\frac{\partial p_j(\mathbf{v}_c)}{\partial v_{j'c}} = \frac{\partial p_{j'}(\mathbf{v}_c)}{\partial v_{jc}}, \tag{2.11}$$

$$\underbrace{\frac{\partial^{(n)} p_j(\mathbf{v}_c)}{\partial v_{1c} \partial v_{2c} \ldots [\partial v_{jc}] \ldots \partial v_{nc}}}_{\textit{is not with respect to } v_{jc}} \begin{cases} \geq 0 & n \textit{ is even} \\ \leq 0 & n \textit{ is odd} \end{cases} ; \quad n \leq J_c, \tag{2.12}$$

where $p_{jc} : \Re^{J_c} \to [0,1]$. Here, v_{jc} is the mean utility of the alternative a_{jc} and $\mathbf{v}_c = (v_{1c}, \ldots, v_{jc}, \ldots, v_{J_c c})^T$.

Condition (2.10) represents the basic requirements of non-negativity and adding-up of the J_c choice probabilities as well as the dependence of the comparison only on the difference in utilities (*translation invariance*).

Condition (2.11) guarantees the integrability of the p_{jc} and is a straightforward analogue to the *(Slutsky 1952)* condition in continuous demand analysis (*integrability*).

Condition (2.12) is the essential requirement for the implied distribution function to be property defined, to have a non-negative density function, mean that p_{jc} must have non-negative even and non-positive odd mixed partial derivatives with respect to components of $\mathbf{v}_c$ other than v_{jc} (*nonnegative density function*).

Three above conditions have been proved by *(Koning and Ridder 2002)*, completely.

Example 2.2. *According to the MNL model* (2.9)*, three conditions* (2.10)*,* (2.11) *and* (2.12) *are held out, so that:*

- $p_j(\mathbf{v}_c) \geq 0$; $\sum_{j=1}^{J_c} p_j(\mathbf{v}_c) = \frac{\sum_{i=1}^{J_c} \exp(v_{jc})}{\sum_{j'=1}^{J_c} \exp(v_{j'c})} = 1$;

$$\begin{aligned} p_j(\mathbf{v}_c + r\mathbf{1}) &= \frac{\exp(v_{jc} + r)}{\sum_{j'=1}^{J_c} \exp(v_{j'c} + r)} \\ &= \frac{\exp(v_{jc})}{\sum_{j'=1}^{J_c} \exp(v_{j'c})} = p_j(\mathbf{v}_c), \forall r \in \Re. \end{aligned}$$

- $\frac{\partial p_j(\mathbf{v}_c)}{\partial v_{j'c}} = -\frac{\exp(v_{jc})\cdot\exp(v_{j'c})}{(\sum_{l=1}^{J_c} \exp(v_{lc}))^2} = -\frac{\exp(v_{j'c})\cdot\exp(v_{jc})}{(\sum_{l=1}^{J_c} \exp(v_{lc}))^2} = \frac{\partial p_{j'}(\mathbf{v}_c)}{\partial v_{jc}}$

- $\frac{\partial^{J_c-1} p_j(\mathbf{v}_c)}{\underbrace{\partial v_{1c}, \ldots, [\partial v_{jc}], \ldots, \partial v_{J_c c}}_{without\ j^{th}}} = \frac{(-1)^{J_c-1}(J_c-1)!}{(\sum_{l=1}^{J_c} \exp(v_{lc}))^{J_c-1}} \begin{cases} \geq 0, & J_c - 1 \textit{ is even;} \\ \leq 0, & J_c - 1 \textit{ is odd.} \end{cases}$

Hence, the MNL model (2.9) *is consistency with RUM.*

2.1.3 Independence from Irrelevant Alternative (IIA)

The MNL model (2.9) is the most widely used discrete choice model due to its closed-form choice probabilities and consistency with random utility maximization. But there exist a problem in this model. The problem arise with the standard MNL because it is derived from random utility maximization, based on the assumption that the error terms are independent across alternatives, choice set, and subjects. This leads to the property of Independence from Irrelevant Alternatives

(IIA). This property (at really restriction), which arises from the assumption of independent random errors and equal variance for the choice alternative mean that; for any two alternatives, the ratio of probabilities is independent of the attributes or existence of all other alternative:

$$\frac{p_{jc}}{p_{j'c}} = \frac{\frac{\exp(v_{jc})}{\sum_{l=1}^{J_c}\exp(v_{lc})}}{\frac{\exp(v_{j'c})}{\sum_{l=1}^{J_c}\exp(v_{lc})}} = \frac{\exp(v_{jc})}{\exp(v_{j'c})}; \forall j, j' \in \mathbb{C}_c;\ \forall c \in \mathcal{C},$$

where $v_{jc} = v_c(a_j)$; $\mathbb{C}_c = \{a_{1c}, \ldots, a_{jc}, \ldots, a_{J_c c}\}$. According to the IIA property we consider the two following corollary.

Corollary 2.1. *If A be a subset of $\mathbb{C}_c$, ($A \subset \mathbb{C}_c$), In this case:*

$$\frac{p_{jA}}{p_{j'A}} = \frac{p_{j\mathbb{C}_s}}{p_{j'\mathbb{C}_s}}; \quad \forall j, j' \in A,$$

where $p_{jc} = p_{j\mathbb{C}_c}$.

Corollary 2.2. *If $\mathbb{C}_c = \{a_{1c}, \ldots, a_{jc}, \ldots, a_{J_c c}\}$ is a choice set and A is a subset of $\mathbb{C}_c$ with element exclusive mutually then $p_{jc} = p_{j\mathbb{C}_c} = p_{jA} \cdot p_{A\mathbb{C}_c} = p_{jA} \cdot p_{Ac}$ and with respect to the MNL model $p_{jc} = p_{jA} \cdot p_{Ac}$, where $p_{Ac} = \sum_{a\in A} p_{ac}$. In this situation, we will have:*

$$\begin{aligned} p_{jc} &= p_{jA} \cdot \sum_{a\in A} p_{ac} \\ &= p_{jA} \cdot \frac{\sum_{a\in A}\exp(v_{ac})}{\sum_{j\in\mathbb{C}_c}\exp(v_{jc})}, \end{aligned}$$

thus $p_{jA} = \frac{\exp(v_{jc})}{\sum_{a\in A}\exp(v_{ac})}$.

Then it can be told that the probability of choosing alternative j by individual i in choice set A is independent of the other alternatives in choice set $\mathbb{C}_c$.

2.1.4 Likelihood Function and Parameters Estimator

It was defined that the observation y_{ijc} of the variables Y_{ijc} equals 1 if the i^{th} individual selects the j^{th} alternative (in choice set $\mathbb{C}_c$) and 0, otherwise, where:

$$Y_{ijs} = \begin{cases} 1, & \text{if } U_{ijc} > U_{ij'c}, \forall j' \neq j \in \mathbb{C}_c; \\ 0, & \text{otherwise} \end{cases}$$

Now, for a random sample size $\mathcal{I}$ of the population of individual and with regards to p_{ijc}, which denotes the probability of choosing alternative j by individual i, the log-likelihood function based on choice set $\mathbb{C}_c$ is defined as follow:

$$\ell(\mathbb{C}_c, \boldsymbol{\beta}) = \ln(L(\mathbb{C}_c, \boldsymbol{\beta})) = \sum_{i=1}^{\mathcal{I}}\sum_{j=1}^{J_c} y_{ijc}\ln(p_{ij}(\mathbf{v}_c)), \tag{2.13}$$

where $\mathbf{v}_c = (v_{1c}, \ldots, v_{jc}, \ldots, v_{J_c c})^T$; $v_{jc} = \mathbf{f}^T(a_{jc})\boldsymbol{\beta}$ (see Section 2.1). The parameters of a MNL model can be estimated by standard maximum likelihood techniques. Substituting the choice probabilities of expression (2.9) into the log-likelihood function gives an explicit function of the parameters of this model. The values of the parameters that maximize this function are, under fairly general conditions, consistent and efficient *(Brownstone and Small 1989)*.

(Manski and McFadden 1981) and *(Cosslett 1981)* have described estimation methods under a variety of sampling procedures. *(Train 2003)* has discussed estimation under the most prominent of these sampling schemes. He has first described estimation when the sample is exogenous and all alternatives are used in estimation. He has then discussed estimation on a subset of alternatives and with certain types of choice-based (i.e., non-exogenous) samples.

Now, based on Equation (2.13) the estimator is the value of $\boldsymbol{\beta}$, which maximizes this function. *(McFadden 1974)* has showed that ($\mathcal{C}$ denotes the number of choice sets each with J_c alternatives, see Section 2.1):

$$\ell(\boldsymbol{\beta}) = \ell(\mathbb{C}_1, \ldots, \mathbb{C}_c, \ldots, \mathbb{C}_{\mathcal{C}}; \boldsymbol{\beta}) = \sum_{c=1}^{\mathcal{C}} \ell(\mathbb{C}_c, \boldsymbol{\beta})$$

is globally concave for linear-in-parameters utility, and many statistical packages are available for estimation of these models. When parameters enter the representative utility nonlinearly, the researcher may need to write her own estimation code using the procedures which were described by *(Train 2003), Chapter 8*.

Maximum likelihood estimation in this situation can be reexpressed and reinterpreted in a way that assists in understanding the nature of the estimates. At the maximum of the likelihood function, its derivative with respect to each of the parameters is zero (if maximum exists):

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}, \tag{2.14}$$

the maximum likelihood estimates are therefore the values of $\boldsymbol{\beta}$, which satisfy this first-order condition. For convenience, let the representative utility be linear in parameters as Equation (2.1). This specification is not required, but makes the notation and discussion more succinct. Using the log-likelihood function (2.13) and the formula for the logit probabilities, we show at the end of this subsection that the first-order condition (2.14) becomes:

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{c=1}^{\mathcal{C}} \sum_{i=1}^{\mathcal{I}} \sum_{j=1}^{J_c} \mathbf{f}^T(a_{ijc})(y_{ijc} - p_{ijc}) = 0. \tag{2.15}$$

2.2 Nested Multinomial Logit Model (NMNL)

The standard logit model (MNL) exhibits IIA, which implies proportional substitution across alternatives. As already was discussed, this property can be seen either as a restriction imposed by the model or as the natural outcome of a well specified model that captures all source of correlation over alternatives into representative utility, so that only white noise remains. Often the researcher is unable to capture all source of correlation explicitly, so that the unobserved options of utility are correlated and IIA does not hold. In this cases, a more general model than standard logit is needed.

Generalized extreme value (GEV) models constitute a large class of models that exhibit a variety of substitution patterns. The unifying attribute of these models is that the unobserved portions of utility for all alternatives are jointly distributed as a generalized extreme value. This distribution allows for correlations over alternatives and, as its name implies, is a generalization of the univariate extreme value distribution, which is used for standard logit models. When all correlations are zero, the GEV distribution becomes the product of independent extreme value distribution and the GEV model becomes standard logit. The class therefore includes logit but also includes a variety of other models. Hypothesis tests on the correlation within a GEV model can be used to examine whether the correlations are zero, which is equivalent to testing whether standard logit provides an accurate representation of the substitution patterns.

The most widely used member of the GEV family is called the nested logit model. This model has been applied by many researchers in a variety of situations, including energy, transformation, housing, telecommunications, and a host of other fields; see, for example, *(Ben-Akiva, 1973)*, *(Train 1986)* and *(Train, et al. 1987)*, *(Forinash and Koppelman 1993)* and *(Lee 1999)*.

The nested logit model has become an important tool for the empirical analysis of discrete outcomes. It is attractive since it relaxes the strong assumptions of the multinomial (or conditional) logit model. At the same time, it is computationally straightforward and fast compared to the multinomial probit, mixed logit, or other even more flexible models due to the existence of a closed-form expression for the likelihood function. There is some confusion about the specification of the outcome probabilities in the nested logit models. Two substantially different formulas and many minor variations of them are presented and used in the empirical literature and in textbooks. Many researchers are neither aware of this issue nor of which version is actually implemented by the software they use. This obscures the interpretation of their results. This problem has been previously discussed by *(Hensher and Greene 2002)*, *Hunt, (2000)*, *(Koppelman and Wen 1998)* and *(Louviere, et al. 2000)*.

(McFadden 1978a), (McFadden 1981) has described a useful generalization of the MNL model and a way to relax the restrictive IIA assumption, namely the Nested Multinomial Logit model (NMNL) that uses a nested structure to estimate the probability of choosing a specific alternative. Another more general type of MNL models that also relaxes IIA assumption is the Mixed Multinomial Logit model (MMNL) introduced by *(Boyed and Mellman 1980)*.

2.2.1 Two-Level Nested MNL Model

In this section we consider a two-level nested MNL model, which the J_c alternatives (in a choice set $\mathbb{C}_c$) are grouped into M subsets (nests), each consisting of $J_{mc}; m = 1, 2, \dots, M$ alternatives:

$$\mathbb{C}_{mc} = \{a_{1mc}, a_{2mc}, \dots, a_{J_{mc}mc}\},$$

so that:

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

where $\mathbb{C}_{mc}$ denotes the choice set, which includes the alternatives of nest m based on choice set $\mathbb{C}_c$ of size J_{mc} ($J_c = \sum_{m=1}^{M} J_{mc}$). In this situation, we face to two steps for choosing. First step, choosing a nest with the highest utility and choosing an alternative with the highest utility, in second step. For this propose we consider the utility related to choice alternative j and nest m by individual i as follow (ignoring index i):

$$U_{jmc} = U_{mc} + U_{j|mc}; \begin{cases} j = 1, 2, \ldots, J_{mc}, & \text{Alternatives;} \\ m = 1, 2, \ldots, M & \text{Nests,} \end{cases} \tag{2.16}$$

where:

1. $U_{mc} = v_{mc} + \varepsilon_{mc}; \quad v_{mc} = E(\max_{j \in \mathbb{C}_{mc}} U_{j|mc})$,

2. $U_{j|mc} = v_{j|mc} + \varepsilon_{j|mc}; \ v_{j|mc} = \mathbf{f}^T(a_{jmc})\boldsymbol{\beta}$ (Section 2.1)
   - $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_k^T, \ldots, \boldsymbol{\beta}_K^T)^T; \quad \boldsymbol{\beta}_k = (\beta_{k,1}, \ldots, \beta_{k,\ell}, \ldots, \beta_{k,L_k})$,
   - $\mathbf{f}(a_{jmc}) = (\mathbf{f}_1^T(a_{jmc}), \ldots, \mathbf{f}_k^T(a_{jmc}), \ldots, \mathbf{f}_K^T(a_{jmc}))^T$;

$$\mathbf{f}_k(a_{jmc}) = (f_{k1}(a_{jmc}), \ldots, f_{k\ell}(a_{jmc}), \ldots, f_{kL_k}(a_{jmc}))^T,$$

so that ε_{mc} have the same distribution ($i.i.d$) such as $\max_{j \in \mathbb{C}_{mc}} U_{j|mc}$ and $\varepsilon_{j|mc}$ have EVD *(Ben-Akiva, 1973)*. In this situation we consider the variances of ε_{mc} and $\varepsilon_{j|mc}$ with symbols σ^2 and σ_m^2, respectively. In this model $\varepsilon_{j|mc}$'s are correlated in the same nest, $corr(\varepsilon_{j|mc}, \varepsilon_{j'|mc}) = \rho_m; \forall j \neq j' \in \mathbb{C}_{mc}$, but $corr(\varepsilon_{mc}, \varepsilon_{m'c}) = 0; \forall m \neq m'$, also, ε_{mc} and $\varepsilon_{j|mc}$ are independence then we will have:

$$cov(U_{jmc}, U_{j'm'c}) = \begin{cases} \sigma_m^2 + \sigma^2, & j = j', m = m'; \\ (1 - \lambda_m^2)\sigma_m^2 + \sigma^2, & j \neq j', m = m'; \\ 0, & m \neq m', \end{cases}$$

where $\lambda_m = \sqrt{1 - \rho_m}$. Also, based on vector $U_c = (U_{1c}, \ldots, U_{mc}, \ldots, U_{Mc})^T$, the variance-covariance matrix of U_c is calculated as below:

$$V(U_c) = \Sigma_{U_c} = \begin{pmatrix} \Sigma_{U_{1c}} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & \cdots & \Sigma_{U_{mc}} & \cdots & 0 \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \Sigma_{U_{Mc}} \end{pmatrix},$$

where:

$$\Sigma_{U_{mc}} = \begin{pmatrix} \sigma_m^2+\sigma^2 & \cdots & \rho_m\sigma_m^2+\sigma^2 & \cdots & \rho_m\sigma_m^2+\sigma^2 \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ \rho_m\sigma_m^2+\sigma^2 & \cdots & \sigma_m^2+\sigma^2 & \cdots & \rho_m\sigma_m^2+\sigma^2 \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ \rho_m\sigma_m^2+\sigma^2 & \cdots & \rho_m\sigma_m^2+\sigma^2 & \cdots & \sigma_m^2+\sigma^2 \end{pmatrix}.$$

Thus,

$$\Sigma_{U_{mc}} = \sigma_m^2(1-\rho_m)\mathbf{I}_{J_m} + (\sigma_m^2\rho_m+\sigma^2)\mathbf{J}_{J_m}; \quad \forall c \in \mathcal{C},\ m=1,2,\ldots,M,$$

where $\mathbf{I}_r$ is $r \times r$ identity matrix and $\mathbf{J}_r$ denotes a $r \times r$ matrix, which all of its elements are one.

In this case, the observation variables to obtain choice probabilities in two-level NMNL models are introduced as follow:

$$Y_{j|mc} = \begin{cases} 1, & U_{j|mc} = \max_{j' \in \mathbb{C}_{mc}} U_{j'|mc}; \\ 0, & \text{otherwise.} \end{cases} \qquad Y_{mc} = \begin{cases} 1, & U_{mc} = \max_{m'} U_{m'c}; \\ 0, & \text{otherwise.} \end{cases}$$

In this situation:

$$Y_{jmc} = Y_{j|mc} \times Y_{mc} \quad \Rightarrow \ p_{jmc} = p_{j|mc} \times p_{mc},$$

where $p_{jmc} = P(Y_{jmc} = 1)$ and $p_{mc} = P(Y_{mc} = 1)$ and $p_{j|mc} = P(Y_{j|mc} = 1)$ have the familiar functional form of simple marginal and conditional *logit* choice probabilities, respectively. According to the definition of variables Y_c we will have, $E(Y_{jmc}) = p_{jmc}$ and:

$$cov(Y_{jmc}, Y_{j'm'c}) = \begin{cases} p_{jmc}\cdot(1-p_{jmc}), & j=j', m=m'; \\ -p_{jmc}\cdot p_{j'mc}, & j \neq j', m=m'; \\ -p_{jmc}\cdot p_{j'm'c}, & m \neq m'. \end{cases}$$

According to the vector $Y_c = (Y_{1c}, \ldots, Y_{mc}, \ldots, Y_{Mc})^T$ we will have:

$$V(Y_c) = \Sigma_{Y_c} = \begin{pmatrix} \Sigma_{1,c} & \cdots & \Sigma_{1m,c} & \cdots & \Sigma_{1m',c} & \cdots & \Sigma_{1M,c} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \Sigma_{m1,c} & \cdots & \Sigma_{m,c} & \cdots & \Sigma_{mm',c} & \cdots & \Sigma_{mM,c} \\ \vdots & \cdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ \Sigma_{m'1,c} & \cdots & \Sigma_{m'm,c} & \cdots & \Sigma_{m',c} & \cdots & \Sigma_{m'M,c} \\ \vdots & \cdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \Sigma_{M1,c} & \cdots & \Sigma_{Mm,c} & \cdots & \Sigma_{Mm',c} & \cdots & \Sigma_{M,c} \end{pmatrix},$$

where:

$$\Sigma_{m,c} = \begin{pmatrix} p_{1mc}\cdot(1-p_{1mc}) & \cdots & -p_{1mc}\cdot p_{jmc} & \cdots & -p_{1mc}\cdot p_{J_{mc}mc} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ -p_{jmc}\cdot p_{1mc} & \cdots & p_{jmc}\cdot(1-p_{jmc}) & \cdots & -p_{jmc}\cdot p_{J_{mc}mc} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ -p_{J_{mc}mc}\cdot p_{1mc} & \cdots & -p_{J_{mc}mc}\cdot p_{jmc} & \cdots & p_{J_{mc}mc}\cdot(1-p_{J_{mc}mc}) \end{pmatrix}; \ m=1,2,\ldots,M,$$

with:

$$\Sigma_{mm',c} = \begin{pmatrix} -p_{1mc}\cdot p_{1m'c} & \cdots & -p_{1mc}\cdot p_{jm'c} & \cdots & -p_{1mc}\cdot p_{J_{m'c}m'c} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ -p_{jmc}\cdot p_{1m'c} & \cdots & -p_{jmc}\cdot p_{jm'c} & \cdots & -p_{jmc}\cdot p_{J_{m'c}m'c} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ -p_{J_{mc}mc}\cdot p_{1m'c} & \cdots & -p_{J_{mc}mc}\cdot p_{jm'c} & \cdots & -p_{J_{mc}mc}\cdot p_{J_{m'c}m'c} \end{pmatrix} ;\ m \neq m' = 1, 2, \ldots, M.$$

With regards to the generating function $G_c(z)$ (Subsection 2.1.1), which is defined by the dissimilarity parameter λ_m of nest m as follow:

$$G_c(z_{11}, \ldots, z_{jm}, \ldots, z_{1M}, \ldots, z_{J_M M}) = \sum_{m=1}^{M} \left(\sum_{j=1}^{J_{mc}} z_{jmc} \right)^{\lambda_m},$$

where $z_{jmc} = \exp(\frac{-\varepsilon_{jmc}}{\lambda_m})$. Thus, according to $F_Z(z) = \exp(-G_c(z_{11}, \ldots, z_{jm}, \ldots, z_{1M}, \ldots, z_{J_M M}))$, we will have:

$$F_{NMNL}(\boldsymbol{\varepsilon}; \boldsymbol{\lambda}) = \exp\left(-\sum_{m=1}^{M} \left(\sum_{j=1}^{J_m} \exp(\frac{-\varepsilon_{jm}}{\lambda_m}) \right)^{\lambda_m} \right). \tag{2.17}$$

Distribution (2.17) is a type of GEV distribution . This distribution for unobserved components of utility gives rise to following the choice probability for alternative j and nest m based on choice set $\mathbb{C}_c$ *(McFadden 1978a)*:

$$p_{j|mc} = \frac{\exp(\frac{v_{j|mc}}{\lambda_m})}{\sum_{j'=1}^{J_{mc}} \exp(\frac{v_{j'|mc}}{\lambda_m})}, \quad p_{mc} = \frac{\exp(\lambda_m v_{mc})}{\sum_{m'=1}^{M} \exp(\lambda_{m'} v_{m'c})},$$

where $v_{j|mc} = \mathbf{f}^T(a_{jmc})\boldsymbol{\beta}$ (Equation (2.16)) and according to:

$$\begin{aligned} E(U_{mc}) &= E(v_{mc} + \varepsilon_{mc}) \\ &= v_{mc} + v'_{mc} - \ln\left(\sum_{j=1}^{J_{mc}} \exp(\frac{v_{j|mc}}{\lambda_m}) \right) \end{aligned}$$

then $v_{mc} = \mu'_{mc} + \ln\left(\sum_{j=1}^{J_{mc}} \exp(\frac{v_{j|mc}}{\lambda_m})\right)$, where $\mu'_{ms} = E(U_{mc}) - v'_{mc}$. Now, suppose that $\mu'_{mc} = 0$ thus $v_{mc} = \ln\left(\sum_{j=1}^{J_{mc}} e^{\frac{v_{j|mc}}{\lambda_m}}\right)$. In hence, p_{jmc} can be rewritten as follows:

$$p_{jmc} = \frac{\exp(\frac{v_{j|mc}}{\lambda_m})}{\sum_{l=1}^{J_{mc}} \exp((\frac{v_{l|mc}}{\lambda_m}))} \cdot \frac{\left(\sum_{l=1}^{J_{mc}} \exp((\frac{v_{l|m}}{\lambda_m}))\right)^{\lambda_m}}{\sum_{m'=1}^{M} \left(\sum_{l=1}^{J_{m'c}} \exp((\frac{v_{l|m'c}}{\lambda_{m'}}))\right)^{\lambda_{m'}}}, \tag{2.18}$$

where λ_m is the so-called dissimilarity parameter of subset m. The parameter λ_m is a measure of the degree of independence in unobserved utility among the alternatives in nest m. A higher value of λ_m means greater independence and less correlation. The statistic $1-\lambda_m$ is a measure of correlation, in the sense that as λ_m rises, indicating less correlation, this statistic drops. As *(McFadden 1978a), (McFadden 1978b)* points out, the correlation is actually more complex than $1-\lambda_m$, but $1-\lambda_m$ can be used as an indication of correlation. A value of $\lambda_m = 1$ indicates complete independence within nest m, that is, no correlation. When λ_m for all m representing independence among all the alternatives in all nests, the GEV distribution becomes the produce of independent extreme value term, whose distribution is given in (2.8). In this case, the nested logit model reduces to the standard logit model.

IIA and IIN Properties

We can use Equation (2.18) to show that IIA holds within each nest of alternatives but not across nest. Considering alternatives $j \in \mathbb{C}_{mc}$ and $j' \in \mathbb{C}_{m'c}$ (based on choice set $\mathbb{C}_c$). Since the denominator of (2.18) is the same for all alternatives, the ratio of probabilities is the ratio of numerators:

$$\frac{p_{jmc}}{p_{j'm'c}} = \frac{\exp(\frac{v_{j|mc}}{\lambda_m})\left(\sum_{l=1}^{J_{mc}} \exp((\frac{v_{l|mc}}{\lambda_m}))\right)^{\lambda_m}}{\exp(\frac{v_{j'|m'c}}{\lambda_{m'}})\left(\sum_{l=1}^{J_{m'c}} \exp((\frac{v_{l|m'c}}{\lambda_{m'}}))\right)^{\lambda_{m'}}}. \tag{2.19}$$

In this situation, if $m = m'$ (i.e., j and j' are in the same nest) then we will have:

$$\frac{p_{jmc}}{p_{j'mc}} = \frac{\exp(\frac{v_{j|mc}}{\lambda_m})}{\exp(\frac{v_{j'|mc}}{\lambda_m})}.$$

This ratio just depends on two alternatives j and j' and is independent of all other alternatives. This means that the IIA property is hold in nest, each. But, for $m \neq m'$ (i.e., j and j' are in the different nests), Equation (2.19) denotes that this ratio dependents on all of alternatives in nests m and m' and is independent of all other alternatives. A form of IIA holds, therefore, even for alternatives in different nests. This form of IIA can be loosely described as *Independent from Irrelevant Nests* or IIN. Thus, with a nested logit model, IIA holds over alternatives in each nest and IIN holds over alternatives in different nests.

Consistency with RUM

(McFadden 1981) has shown that any set of the choice probabilities that satisfy the compatibility conditions (2.10), (2.11), (2.12) are consistent with RUM. In the case of the Nested logit model only (2.12) compatibility condition is restrictive.

(Daly and Zachary 1978), (McFadden 1978a), (McFadden 1978b) and *(Williams 1977)* showed, independently and using different proofs, that the two-level nested logit model is consistent with random utility maximization (RUM). Such that this the dissimilarity parameter for the nested are restricted to lie within the unit interval:

$$\lambda_m \leq 1, \forall m = 1, 2, \ldots, M,$$

where m is index relative to m^{th} nest and λ_m is dissimilarity parameter associated to m^{th} nest.

(Börsch-Supan 1990) showed that the *Daly, Zachary, McFadden*(DZM) conditions of the validity of stochastic utility maximization in nested MNL model is unnecessarily strong. Therefore it may be too often rejected NMNL model because of their large dissimilarity parameter, since by noting to these conditions (DZM), λ_m should be in unit interval, means, $0 < \lambda_m \leq 1, m = 1, 2, \ldots, M$, at really, it reveals that nested MNL models with large dissimilarity parameters are only compatible with stochastic utility maximization are sufficiently large (Theorem 2.2).

Theorem 2.2. *(Börsch-Supan 1990): In two-level NMNL models, a necessary criterion for the nonnegativity condition are sufficiently large marginal choice probabilities of the subsets of similar alternatives (w.r.t Choice set $\mathbb{C}_c$):*

$$p_{mc} \geq \frac{\lambda_m - 1}{\lambda_m}; \forall m = 1, 2, \ldots, M.$$

(Herriges and Kling 1996) extended the results of *Börsch-Supan*'s theorem and examining the extent to which it is likely to expand the set of consistent NMNL models.

Theorem 2.3. (Herriges and Kling 1996)*: In two-level NMNL models, the following are necessary conditions for consistency with stochastic utility maximization (based on choice set $\mathbb{C}_c$). Let $\tau_m = \frac{\lambda_m - 1}{\lambda_m}$ then:*

$$p_{mc} \geq \tau_m; m = 1, 2, \ldots, M \tag{2.20}$$

$$\left(2(\tau_m - p_{mc})^2 + \tau_m p_{mc}\right) \geq \tau_m; \forall m \in \mathcal{M}_3 \equiv \{m | J_{mc} \geq 3\} \tag{2.21}$$

$$\left[6(p_{mc} - \tau_m)^3 + \tau_m \left[2(p_{mc} - 1) - \tau_m\right](1 - p_{mc})\right] \geq 0; \forall m \in \mathcal{M}_4 \equiv \{m | J_{mc} \geq 4\}, \tag{2.22}$$

where J_{mc} is the number of alternatives in the m^{th} nest and $\mathcal{M}_3$, $\mathcal{M}_4$ denote the sets of nests, which have at least three and four alternatives. To proof Theorem 2.3 from the first, second and third mixed partial derivations of Equation (2.18) have been used.

Corollary 2.3. *In two-level NMNL models, consistency with stochastic utility maximization places the following necessary restrictions on dissimilarity coefficients:*

$$\lambda_m \leq \frac{1}{1 - p_{mc}}, m = 1, 2, \ldots, M, \tag{2.23}$$

$$\lambda_m \leq \frac{4}{3(1 - p_{mc}) + \left[(1 + 7\lambda_m)(1 - \lambda_m)\right]^{\frac{1}{2}}}, \forall m \in \mathcal{M}_3 \equiv \{m | J_{mc} \geq 3\}, \tag{2.24}$$

where the condition (2.12) *is to be valid for all deterministic utility components* $\mathbf{v}_c \in \Re^{J_c}$.

Example 2.3. *Suppose that there is a choice set with three alternatives,* $\mathbb{C} = \{\tilde{a}_{1c}, \tilde{a}_{2c}, \tilde{a}_{3c}\}$, *and due to alternatives* $\tilde{a}_{1c}$ *and* $\tilde{a}_{2c}$ *have common properties, thus, we divide all of alternatives into two nests,* $\mathbb{C}_1 = \{a_{11c}(\tilde{a}_{1c}), a_{21c}(\tilde{a}_{2c})\}$ *and* $\mathbb{C}_2 = \{a_{12c}(\tilde{a}_{3c})\}$. *In this situation, generating function is denoted by:*

$$\begin{aligned} G(\exp(-\varepsilon_1), \exp(-\varepsilon_2), \exp(-\varepsilon_3)) &= \sum_{m=1}^{2} \left(\sum_{j=1}^{J_m} \exp((-\varepsilon_{j|m})^{\frac{1}{\lambda_m}}) \right)^{\lambda_m} \\ &= \left(\exp(-\frac{\varepsilon_1}{\lambda_1}) + \exp(-\frac{\varepsilon_2}{\lambda_1}) \right)^{\lambda_1} + \left(\exp(-\frac{\varepsilon_3}{\lambda_2}) \right)^{\lambda_2} \end{aligned}$$

and we will have:

- $p_{1c} = \frac{\left(\exp\left(\frac{v_{a_{11c}|1c}}{\lambda_1}\right) + \exp\left(\frac{v_{a_{21c}|1c}}{\lambda_1}\right)\right)^{\lambda_1}}{\exp(v_{a_{12c}|2c}) + \left(\exp\left(\frac{v_{a_{11c}|1c}}{\lambda_1}\right) + \exp\left(\frac{v_{a_{21c}|1c}}{\lambda_1}\right)\right)^{\lambda_1}}$,
- $p_{2c} = \frac{\exp(v_{a_{12c}|2c})}{\exp\left(v_{a_{12c}|2c}\right) + \left(\exp(\frac{v_{a_{11c}|1c}}{\lambda_1}) + \exp\left(\frac{v_{a_{12c}|1c}}{\lambda_1}\right)\right)^{\lambda_1}}$,
- $p_{j|1c} = \frac{\exp(\frac{v_{j|1c}}{\lambda_1})}{\exp\left(\frac{v_{a_{11c}|1c}}{\lambda_1}\right) + \exp\left(\frac{v_{a_{21c}|1c}}{\lambda_1}\right)}; j \in \{a_{11c}, a_{12c}\}, \quad p_{a_{12c}|2c} = 1.$

Let, $\lambda_1 = \lambda$ thus:

$$p_{a_{11c}1c} = p_{1c} \cdot p_{a_{11c}|1c} = \frac{\left(\exp\left(\frac{v_{a_{11c}|1c}}{\lambda}\right) + \exp\left(\frac{v_{a_{21c}|1c}}{\lambda}\right)\right)^{\lambda-1} \cdot \exp\left(\frac{v_{a_{11c}|1c}}{\lambda}\right)}{\left(\exp\left(v_{a_{12c}|2c}\right) + \left(\exp\left(\frac{v_{a_{11c}|1c}}{\lambda}\right) + \exp\left(\frac{v_{a_{21c}|1c}}{\lambda}\right)\right)^{\lambda}\right)}, \tag{2.25}$$

$$p_{a_{21c}1c} = p_{1c} \cdot p_{a_{21c}|1c} = \frac{\left(\exp\left(\frac{v_{a_{11c}|1c}}{\lambda}\right) + \exp\left(\frac{v_{a_{21c}|1c}}{\lambda}\right)\right)^{\lambda-1} \cdot \exp\left(\frac{v_{a_{21c}|1c}}{\lambda}\right)}{\left(\exp\left(v_{a_{12c}|2c}\right) + \left(\exp\left(\frac{v_{a_{11c}|1c}}{\lambda}\right) + \exp\left(\frac{v_{a_{21c}|1c}}{\lambda}\right)\right)^{\lambda}\right)}, \tag{2.26}$$

$$p_{a_{12c}2c} = p_{2c} \cdot p_{a_{12c}|2c} = \frac{\exp\left(v_{a_{12c}|2c}\right)}{\exp\left(v_{a_{12c}|2c}\right) + \left(\exp\left(\frac{v_{a_{11c}|1c}}{\lambda}\right) + \exp\left(\frac{v_{a_{21c}|1c}}{\lambda}\right)\right)^{\lambda}}, \tag{2.27}$$

where p_{jmc}, $p_{j|mc}$ and p_{mc} denote the joint, conditional and marginal choice probabilities related to choosing alternative j and nest m in choice set c. Since, $J_{mc} \leq 2; m = 1, 2$ then Condition (2.20) must be just considered. In this case, the NMNL model , which has been denoted by Equations (2.25), (2.26) and (2.27) is consistent with RUM when $\lambda \leq \frac{1}{1-p_{1c}}$. In this situation, if $\lambda = 1$ then the NMNL model reduce to the MNL model, thus testing $H_0 : \lambda = 1$ via $H_1 : \lambda \neq 1$

can be interesting. According to the normalized log-likelihood function of the NMNL model, which is defined as follow:

$$\ell(\mathbb{C}_c, \beta, \lambda) = \sum_{i=1}^{\mathcal{I}} \sum_{m=1}^{M} \sum_{j=1}^{J_{mc}} y_{ijmc} \ln(p_{ijmc}). \tag{2.28}$$

Let us, suppose that $\mathbf{f}(a_{11c}) = 1$ and $\mathbf{f}(a_{21c}) = \mathbf{f}(a_{12c}) = 0$, where $v_{j|mc} = \mathbf{f}^T(a_{jmc})\beta; j \in \{a_{11c}, a_{21c}, a_{12c}\}$, $m = 1, 2$. According to Equations (2.25), (2.26) and (2.27) we will have:

$$p_{a_{11c}1c} = \frac{\left(1+\exp\left(\frac{\beta}{\lambda}\right)\right)^{\lambda-1} \cdot \exp\left(\frac{\beta}{\lambda}\right)}{\left(1+(1+\exp\left(\frac{\beta}{\lambda}\right))^{\lambda}\right)} \quad , \quad p_{a_{21c}1c} = \frac{\left(1+\exp\left(\frac{\beta}{\lambda}\right)\right)^{\lambda-1}}{\left(1+(1+\exp\left(\frac{\beta}{\lambda}\right))^{\lambda}\right)} \quad , \quad p_{a_{12c}2c} = \frac{1}{1+(1+\exp\left(\frac{\beta}{\lambda}\right))^{\lambda}}.$$

Let, $\delta = \frac{\beta}{\lambda}$ and $\mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 = \mathcal{I}$ (See Example 2.3) thus:

$$\ell(\delta, \lambda) = \mathcal{I}_1\delta + (\mathcal{I}_1 + \mathcal{I}_2)(\lambda - 1)\ln(1 + \exp(\delta)) - \mathcal{I} \cdot \ln(1 + (1 + \exp(\delta))^{\lambda}). \tag{2.29}$$

Then, the maximum likelihood estimators for δ and λ are calculated as below:

$$\hat{\delta} = \ln(\frac{\mathcal{I}_1}{\mathcal{I}_2}), \qquad \hat{\lambda} = \frac{\ln(\frac{\mathcal{I}_1+\mathcal{I}_2}{\mathcal{I}_3})}{\ln(\frac{\mathcal{I}_1+\mathcal{I}_2}{\mathcal{I}_2})}. \tag{2.30}$$

Based on the definition of the information matrix, we will have:

$$\mathbf{I}(\delta, \lambda) = -E(\frac{\partial^2 \ell(\delta, \lambda)}{\partial\delta\partial\lambda}) = -E\begin{pmatrix} \frac{\partial^2 \ell(\delta,\lambda)}{\partial\delta^2} & \frac{\partial^2 \ell(\delta,\lambda)}{\partial\delta\partial\lambda} \\ \frac{\partial^2 \ell(\delta,\lambda)}{\partial\delta\partial\lambda} & \frac{\partial^2 \ell(\delta,\lambda)}{\partial\lambda^2} \end{pmatrix}.$$

Suppose that:

$$\mathbf{I}(\delta, \lambda) = \begin{pmatrix} I_{\delta\delta} & I_{\delta\lambda} \\ I_{\delta\lambda} & I_{\lambda\lambda} \end{pmatrix} \quad \Rightarrow \quad \mathbf{I}^{-1}(\delta, \lambda) = \frac{1}{I_{\delta\delta}I_{\lambda\lambda} - I_{\delta\lambda}^2} \begin{pmatrix} I_{\lambda\lambda} & -I_{\delta\lambda} \\ -I_{\delta\lambda} & I_{\delta\delta} \end{pmatrix}.$$

With regards to the properties of the maximum likelihood estimator (MLE), we know:

$$\hat{\lambda} \sim^a N(\lambda, I^{-1}(\lambda)),$$

where $\sim^a$ denotes the asymptotically distribution and $\Sigma_\lambda = I^{-1}(\lambda)$. According to asymptotically distribution of $\hat{\lambda}$, the **rejection region** for above hypothesis is obtained as follow:

$$r.r = \left\{\hat{\lambda}|\hat{\lambda} > \left(1 + z_{\frac{\alpha}{2}} I^{-\frac{1}{2}}(\lambda)\right) \; or \; \hat{\lambda} < \left(1 - z_{\frac{\alpha}{2}} I^{-\frac{1}{2}}(\lambda)\right)\right\}$$

Since, $\mathbf{I}^{-1}(\delta, \lambda)$ is a function of unknown parameters (λ, δ) then we use their estimators $(\hat{\lambda}, \hat{\delta})$ to obtain the inverse of the information matrix $(\mathbf{I}^{-1}(\delta, \lambda))$, so that:

$$r.r = \left\{\hat{\lambda}|\hat{\lambda} > \left(1 + z_{\frac{\alpha}{2}} \hat{\Sigma}_\lambda\right) \; or \; \hat{\lambda} < \left(1 - z_{\frac{\alpha}{2}} \hat{\Sigma}_\lambda\right)\right\},$$

where $\Sigma_\lambda = \sigma^2_\lambda = \frac{a}{b}$ with:

$$\begin{aligned} a &= \Big[(\mathcal{I}\lambda - (\mathcal{I}_1+\mathcal{I}_2)(\lambda-1))\,(1+\exp{(\delta)})^{2\lambda} + \big(\mathcal{I}\lambda^2\exp{(\delta)} + \mathcal{I}\lambda - 2(\mathcal{I}_1+\mathcal{I}_2)(\lambda-1)\big)\,(1+\exp{(\delta)})^{\lambda} - (\mathcal{I}_1+\mathcal{I}_2)(\lambda-1)\Big] \\ &\quad \times \Big[1+(1+\exp{(\delta)})^{\lambda}\Big] \end{aligned}$$

$$\begin{aligned} b &= \exp{(\delta)}\,(-\mathcal{I}_1-\mathcal{I}_2+\mathcal{I})^2\,(1+\exp{(\delta)})^{3\lambda} + (-(\mathcal{I}\lambda-(\mathcal{I}_1+\mathcal{I}_2)(\lambda-1))\mathcal{I}(\ln(1+\exp{(\delta)}))^2 \\ &\quad +2\mathcal{I}\exp{(\delta)}\,\lambda(-\mathcal{I}_1-\mathcal{I}_2+\mathcal{I})\ln(1+\exp{(\delta)}) + \exp{(\delta)}\,(-3\mathcal{I}_1-3\mathcal{I}_2+\mathcal{I})(-\mathcal{I}_1-\mathcal{I}_2+\mathcal{I}))((1+\exp{(\delta)})^{\lambda})^2 \\ &\quad -2\left[-\frac{1}{2}\mathcal{I}(\lambda-1)(\ln(1+\exp{(\delta)}))^2 + \exp{(\delta)}\,\mathcal{I}\ln(1+\exp{(\delta)})\lambda + \exp{(\delta)}\,(-\frac{3}{2}\mathcal{I}_1 - \frac{3}{2}\mathcal{I}_2+\mathcal{I})\right] \\ &\quad \times\Big[(\mathcal{I}_1+\mathcal{I}_2)(1+\exp{(\delta)})^{\lambda} + \exp{(\delta)}\,(\mathcal{I}_1+\mathcal{I}_2)^2\Big]. \end{aligned}$$

According to (2.25) $\Lambda = (0, \frac{1}{1-p_1}]$ is parameter space of λ then $\hat{\lambda}$ must vary in Λ, thus;

- if $(1 - z_{\frac{\alpha}{2}}\hat{\Sigma}_\lambda) < 0 < \frac{1}{1-p_I} < (1 + z_{\frac{\alpha}{2}}\hat{\Sigma}_\lambda) \ \Rightarrow\ r.r = \left\{\hat{\lambda} | \hat{\lambda} < \frac{1}{1-p_I}\right\}$,
- if $0 < (1 + z_{\frac{\alpha}{2}}\hat{\Sigma}_\lambda) < \frac{1}{1-p_I} \ \Rightarrow\ r.r = \left\{\hat{\lambda} | (1 + z_{\frac{\alpha}{2}}\hat{\Sigma}_\lambda) < \hat{\lambda} < \frac{1}{1-p_I}\right\}$; $p_I > \frac{z_{\frac{\alpha}{2}}\hat{\Sigma}_\lambda}{(1+z_{\frac{\alpha}{2}}\hat{\Sigma}_\lambda)}$,
- if $0 < (1 - z_{\frac{\alpha}{2}}\hat{\Sigma}_\lambda) < (1 + z_{\frac{\alpha}{2}}\hat{\Sigma}_\lambda) < \frac{1}{1-p_I} \ \Rightarrow\ r.r = \left\{\hat{\lambda} | \hat{\lambda} > \left(1 + z_{\frac{\alpha}{2}}\hat{\Sigma}_\lambda\right) or \hat{\lambda} < \left(1 - z_{\frac{\alpha}{2}}\hat{\Sigma}_\lambda\right)\right\}$,
- if $0 < (1 - z_{\frac{\alpha}{2}}\hat{\Sigma}_\lambda) < \frac{1}{1-p_I} < (1 + z_{\frac{\alpha}{2}}\hat{\Sigma}_\lambda) \ \Rightarrow\ r.r = \left\{\hat{\lambda} | (1 - z_{\frac{\alpha}{2}}\hat{\Sigma}_\lambda) > \hat{\lambda}\right\}$.

WALD's Test-Statistic : $\hat{\lambda}$ has asymptotically normal distribution, thus;

$$W = (\hat{\lambda} - \lambda)^T \left(\mathbf{I}^{-1}(\lambda)\right)^{-1} (\hat{\lambda} - \lambda) \tag{2.31}$$

has Chi-Square distribution asymptotically under null hypothesis, $H_0 : \lambda = \lambda_0$, means that:

$$W = (\hat{\lambda} - \lambda_0)^T \left(\mathbf{I}_{\lambda\lambda} - \frac{\mathbf{I}^2_{\delta\lambda}}{\mathbf{I}_{\delta\delta}}\right) (\hat{\lambda} - \lambda_0) \sim^a \chi^2_1, \tag{2.32}$$

where $\left(\mathbf{I}^{-1}(\lambda)\right)^{-1} = \mathbf{I}_{\lambda\lambda} - \frac{\mathbf{I}^2_{\alpha\lambda}}{\mathbf{I}_{\delta\delta}}$.

Test statistic (2.32) is called $WALD$'s test statistic and if $W > \chi^2_{\alpha,1}$, $H_0 : \lambda = \lambda_0 = 1$ will be rejected, else, H_0 will be accepted, means that the MNL model is true.

LRT's Test-Statistic: The above hypothesis can be done by the LRT test statistic, also. In this case, we assume that: $L(\hat{\delta}, \hat{\lambda}) = L\left(\ln\left(\frac{\mathcal{I}_1}{\mathcal{I}_2}\right) \frac{\ln\left(\frac{\mathcal{I}_3}{\mathcal{I}_1+\mathcal{I}_2}\right)}{\ln\left(\frac{\mathcal{I}_2}{\mathcal{I}_1+\mathcal{I}_2}\right)}\right)$ is maximum likelihood function without restriction and $L(\tilde{\delta}, \tilde{\lambda}) = L\left(\ln\left(\frac{2\mathcal{I}_1}{\mathcal{I}-\mathcal{I}_1}\right), 1\right)$ is likelihood restricted, thus:

$$LRT = -2\ln\left(\frac{L(\tilde{\delta}, \tilde{\lambda})}{L(\hat{\delta}, \hat{\lambda})}\right) \sim^a \chi^2_{(2-1)}. \tag{2.33}$$

Test statistic (2.33) also has Chi-square distribution asymptotically with 1 degree of freedom, then:

$$LRT = 2\mathcal{I}_2 \ln\left(\frac{2\mathcal{I}_2}{\mathcal{I}_2 + \mathcal{I}_3}\right) + 2\mathcal{I}_3 \ln\left(\frac{2\mathcal{I}_3}{\mathcal{I}_2 + \mathcal{I}_3}\right) \sim^a \chi^2_1.$$

Similarity, if $LRT > \chi^2_{\alpha,1}$, $H_0 : \lambda = \lambda_0 = 1$ will be rejected, else, the MNL model is true.

Noting to (2.25), since $\frac{1}{1-p_{ms}} \geq 1$ then it is possible that $\exists m; \lambda_m > 1$. However, we can construct a set like $D(\lambda)$ so that for all of points in this set the NMNL choice probabilities do indeed represent a choice system compatible with stochastic utility maximization (See (Börsch-Supan 1990)).

Totaly, suppose that there is a NMNL model by M nests. In this case if $\lambda_m = 1; \forall m$ then NMNL model reduces to MNL model. for this purpose we define the following hypothesis test:

$$\begin{cases} H_0 : \Lambda = \mathbf{1}_M \\ H_1 : \Lambda \neq \mathbf{1}_M \end{cases},$$

where $\Lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_M \end{pmatrix}$ and $\mathbf{1}_M = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$. To do this test, the Wald test can be considered. Thus we suppose that $\hat{\Lambda}$ be the MLE for Λ, which has the properties of the MLE (asymptotically). Now if Σ_Λ is variance-covariance matrix related to all of the dissimilarity parameters then the Wald statistic is as follow:

$$W = (\hat{\Lambda} - \Lambda_0)\Sigma_\Lambda^{-1}(\hat{\Lambda} - \Lambda_0)^T.$$

This statistic under H_0 has chi-square distribution with M degree of freedom. Now, according to $\Lambda_0 = \mathbf{1}_M$ we will have:

$$W = (\hat{\Lambda} - \mathbf{1}_M)\Sigma_\Lambda^{-1}(\hat{\Lambda} - \mathbf{1}_M)^T \sim \chi^2_M.$$

In this situation, if $W > \chi^2_{\alpha,M}$ then null hypothesis will be rejected else it will be accepted.

2.2.2 Three-Level Nested MNL Model

The nested logit model *(McFadden 1978a)*, *(McFadden 1981)* allows partial relaxation of the assumption of independence of the stochastic components of utility of alternatives. In some choice situations, the IIA property holds for some pairs of alternatives but not all. In these situations, the nested logit model can be used if the set of alternatives faced by an individual can be partitioned into subsets such that the IIA property holds within subsets but not across subsets. In the nested logit model, the joint distribution of the errors is generalized extreme value (GEV). This is a generalization of the type I extreme-value distribution that gives rise to the conditional logit model. Note that all within each subset are correlated with each other. Nested logit models can be described analytically following the notation of *(McFadden 1981)*.

In subsection 2.2.1 was told that if there are some alternatives, which are more similar than the others, the alternatives of the choice set are divided into several sets which are called nests. In this situation, we can divide alternatives in some nests to sub-nests if it is necessary (if IIA be not hold in at least one nest). By dividing the alternatives of some (all) nests into sub-nests, we will confront with a new model of logit family which is called three-level nested *logit* models, due to there are three kind of choice probabilities.

We first introduce utility and observation variables to obtain the choice probabilities in the three-level NMNL models. In this case, the utility related to choosing alternatives include three component as follow (based on choice set $\mathbb{C}_c$):

$$U_{jhmc} = U_{mc} + U_{h|mc} + U_{j|hmc}; \begin{cases} j = 1, 2, \ldots, J_{hmc}, & \text{(alternative in sub-nest } hm \text{);} \\ h = 1, 2, \ldots, H_m, & \text{(sub-nest);} \\ m = 1, 2, \ldots, M, & \text{(nest).} \end{cases} \tag{2.34}$$

The alternatives are divided into M nests, with H_m sub-nests in nest m and J_{hmc} alternatives in sub-nest hm. The choice can be considered as first choosing among the M nests, then among the H_m sub-nests in the chosen nest m, and finally among the J_{hmc} alternatives in the chosen sub-nest hm so that:

$$\mathbb{C}_{hmc} = \{a_{1hmc}, a_{2hmc}, \ldots, a_{J_{hmc}mc}\}.$$

In this case, we define $\mathbb{C}_c$ a choice set, which includes J_c alternatives, where $J_c = \sum_{m=1}^{M}\sum_{h=1}^{H_m} J_{hmc}$. The alternatives of the choice set $\mathbb{C}_c$ will be divided into M nests each with J_{mc} alternatives while the choice set of nest m is denoted by $\mathbb{C}_{mc}$. According to the definition of three-level NMNL model, the alternatives of each nest(at least one of them) are divided to several sub-nests each with J_{mhc} alternatives (choice set $\mathbb{C}_{hmc}$), where:

$$\bigcup_{m=1}^{M} \mathbb{C}_{mc} = \mathbb{C}_c;\ \mathbb{C}_{mc} \bigcap \mathbb{C}_{m'c} = \phi \ \ and \ \ \bigcup_{h=1}^{H_m} \mathbb{C}_{hmc} = \mathbb{C}_{mc};\ \mathbb{C}_{hmc} \bigcap \mathbb{C}_{h'mc} = \phi.$$

Consequently, $J_c = \sum_{m=1}^{M} J_{mc}$ and $J_{mc} = \sum_{h=1}^{H_m} J_{hmc}$.

Also, Equation (2.34) can be rewritten by (ignoring index i):

1. $U_{mc} = v_{mc} + \varepsilon_{mc}$
   - $v_{mc} = E(\max_{h\in H_m} U_{h|mc})$.
2. $U_{h|mc} = v_{h|mc} + \varepsilon_{h|mc}$
   - $v_{h|mc} = E(\max_{j\in \mathbb{C}_c} U_{j|hmc})$.
3. $U_{j|hmc} = v_{j|hmc} + \varepsilon_{j|hmc}$
   - $v_{j|hmc} = \mathbf{f}^T(a_{jhmc})\boldsymbol{\beta},\ \ \mathbf{f}(a_{jhmc}) = (\mathbf{f}_1^T(a_{jhmc}), \ldots, \mathbf{f}_k^T(a_{jhmc}), \ldots, \mathbf{f}_K^T(a_{jhmc}))^T;$

$$\mathbf{f}_k(a_{jhmc}) = (f_{k1}(a_{jhmc}), \ldots, f_{k\ell}(a_{jhmc}), \ldots, f_{kL_k}(a_{jhmc}))^T,$$

where suppose that $\varepsilon_{j|hmc}$ have EVD with variance σ^2_{hm} and they are correlated in the same sub-nest ($\rho_{hm} = corr(\varepsilon_{j|hmc}, \varepsilon_{j'|hmc})$), $\varepsilon_{h|mc}$ has distribution such as $\max_{j\in\mathbb{C}_{hmc}} U_{j|hmc}$ ($\mathbb{C}_{hmc}$ denotes a set of alternatives in sub-nest hm) and we use of symbol σ^2_m to denote for its variance. $\varepsilon_{h|mc}$ are correlated in the same nest ($\rho_m = corr(\varepsilon_{h|mc}, \varepsilon_{h'|mc})$). ε_{mc} have ($i.i.d$) distribution such as $\max_{h\in H_m} U_{h|mc}$, with variance σ^2, where $corr(\varepsilon_{mc}, \varepsilon_{m'c}) = 0$; $m \neq m'$ (for simplicity it has been used three symbols $\sigma^2_{h_m}, \sigma^2_m$ and σ^2 instead of their variances). In this situation, $\varepsilon_{j|hmc}$, $\varepsilon_{h|mc}$ and ε_{mc} has been assumed that are independence. In this case we will have:

$$cov(U_{jhmc}, U_{j'h'm'c}) = \begin{cases} \sigma^2_{hm} + \sigma^2_m + \sigma^2, & j = j', h = h', m = m'; \\ \rho_{hm}\sigma^2_{hm} + \sigma^2_m + \sigma^2, & j \neq j', h = h', m = m'; \\ \rho_m\sigma^2_m + \sigma^2, & h \neq h', m = m'; \\ 0, & m \neq m'. \end{cases}$$

Thus with respect to vector $U_c = (U_{1c}, \ldots, U_{mc}, \ldots, U_{Mc})^T$ we will have:

$$\Sigma_{U_c} = Cov(U_{mc}, U_{m'c}) \quad = \begin{pmatrix} \Sigma_1 & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ \mathbf{0} & \cdots & \Sigma_m & \cdots & \mathbf{0} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \cdots & \Sigma_M \end{pmatrix}.$$

Also, the variance-covariance matrix of the vector $U_{mc} = (U_{1mc}, \ldots, U_{hmc}, \ldots, U_{H_m mc})^T$ is calculated as follow:

$$\Sigma_m = Cov(U_{hmc}, U_{h'mc}) \quad = \begin{pmatrix} \Sigma_{1m} & \cdots & \Sigma_{1m,hm} & \cdots & \Sigma_{1m,h'm} & \cdots & \Sigma_{1m,Hm} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \Sigma^T_{1m,hm} & \cdots & \Sigma_{hm} & \cdots & \Sigma_{hm,h'm} & \cdots & \Sigma_{hm,H_m m} \\ \vdots & \cdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ \Sigma^T_{1m,h'm} & \cdots & \Sigma^T_{hm,h'm} & \cdots & \Sigma_{h'm} & \cdots & \Sigma_{h'm,H_m m} \\ \vdots & \cdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \Sigma^T_{1m,H_m m} & \cdots & \Sigma^T_{hm,H_m m} & \cdots & \Sigma^T_{h'm,H_m m} & \cdots & \Sigma_{H_m m} \end{pmatrix},$$

where (Let $r = J_{hmc}$):

$$\Sigma_{hm} = Cov(U_{jhmc}, U_{j'hmc}) \; = \sigma^2_{hm}(1 - \rho_{h_m})\mathbf{I}_r + (\rho_{hm}\sigma^2_{hm} + \sigma^2_m + \sigma^2)\mathbf{J}_r,$$

$$\Sigma_{hm,h'm} = Cov(U_{jhmc}, U_{j'h'mc}) \; = (\rho_m\sigma^2_m + \sigma^2)\mathbf{J}_r$$

where $\mathbf{I}_r$ and $\mathbf{J}_r$ have been defined in subsection 2.2.1.

According to the Utility (2.34) the observation variables can be introduced as below (ignoring index i):

- $Y_{mc} = \begin{cases} 1, & U_{mc} = \max_{m'} U_{m'c}; \\ 0, & \text{otherwise}. \end{cases}$

- $Y_{h|mc} = \begin{cases} 1, & U_{h|mc} = \max_{h' \in H_m} U_{h'|mc}; \\ 0, & \text{otherwise.} \end{cases}$

- $Y_{j|hmc} = \begin{cases} 1, & U_{j|hmc} = \max_{j' \in \mathbb{C}_{hmc}} U_{j'|hmc}; \\ 0, & \text{otherwise.} \end{cases}$

With regards to independence between above three error terms (Equation (2.34)) we will have:

$$Y_{jhmc} = Y_{j|hmc} \times Y_{h|mc} \times Y_{mc}.$$

Thus:

$$p_{jhmc} = p_{j|hmc} \times p_{h|mc} \times p_{mc}, \tag{2.35}$$

where $p_{jhmc} = P(Y_{jhmc} = 1)$, $p_{j|hmc} = P(Y_{j|hmc} = 1)$, $p_{h|mc} = P(Y_{h|mc} = 1)$ and $p_{mc} = P(Y_{mc} = 1)$ so that (McFadden 1981):

- $p_{mc} = \frac{exp(\mu_m IV_{mc})}{\sum_{m'=1}^{M} \exp(\mu_{m'} IV_{m'c})}$,

- $p_{h|mc} = \frac{exp\left(\frac{\lambda_{hm}}{\mu_m} IV_{hmc}\right)}{\sum_{h'=1}^{H_m} \exp\left(\frac{\lambda_{h'm}}{\mu_m} IV_{h'ms}\right)}$,

- $p_{j|hmc} = \frac{\exp\left(\frac{v_{j|hmc}}{\lambda_{hm}}\right)}{\sum_{j'=1}^{J_{hmc}} \exp\left(\frac{v_{j'|hmc}}{\lambda_{hm}}\right)}$.

In the above choice probabilities, IV_{mc} and IV_{hmc} are the inclusive values of nest m and sub-nest hm, respectively, where:

- $IV_{ms} = E\left(\max_{h \in H_m} U_{h|mc}\right) = \ln\left(\sum_{h=1}^{H_m} \exp\left(\frac{\lambda_{hm}}{\mu_m} IV_{hmc}\right)\right)$,

- $IV_{hmc} = E\left(\max_{j \in \mathbb{C}_{hmc}} U_{j|hmc}\right) = \ln\left(\sum_{j=1}^{J_{hmc}} \exp\left(\frac{v_{j|hmc}}{\lambda_{hm}}\right)\right)$

and the parameters μ_m and λ_{hm} are the measures of the degree of independence in unobserved utility among the sub-nests in nest, m, and the alternatives in sub-nest, hm, respectively.

Based on the definition of the observations variable (Y_c) we will have:

$$cov(Y_{jhmc}, Y_{j'h'm'c}) = \begin{cases} p_{jhmc} \cdot (1 - p_{jhmc}), & j = j', h = h', m = m'; \\ -p_{jhmc} \cdot p_{j'hmc}, & j \neq j', h = h', m = m'; \\ -p_{jhmc} \cdot p_{j'h'mc}, & h \neq h', m = m'; \\ -p_{jhmc} \cdot p_{j'h'm'c}, & m \neq m'. \end{cases}$$

Thus the variance-covariance of the vector $Y_c = (Y_{1c}, \dots, Y_{mc}, \dots, Y_{Mc})^T$:

$$\Sigma_{Y_c} = Cov(Y_{ms}, Y_{m's}) = \begin{pmatrix} \Sigma_{1c} & \cdots & \Sigma_{1m,c} & \cdots & \Sigma_{1M,c} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ \Sigma_{m1,c} & \cdots & \Sigma_{mc} & \cdots & \Sigma_{mM,c} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ \Sigma_{M1,c} & \cdots & \Sigma_{Mm,c} & \cdots & \Sigma_{Mc} \end{pmatrix},$$

where based on vector $Y_{mc} = (Y_{1mc}, \ldots, Y_{hmc}, \ldots, Y_{H_m mc})^T$ we will have:

$$\Sigma_{mc} = Cov(Y_{hmc}, Y_{h'mc}) = \begin{pmatrix} \Sigma_{1,mc} & \cdots & \Sigma_{1h,mc} & \cdots & \Sigma_{1h',mc} & \cdots & \Sigma_{1H_m,mc} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \Sigma_{h1,mc} & \cdots & \Sigma_{h,mc} & \cdots & \Sigma_{hh',mc} & \cdots & \Sigma_{hH_m,mc} \\ \vdots & \cdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ \Sigma_{h'1,mc} & \cdots & \Sigma_{h'h,mc} & \cdots & \Sigma_{h',mc} & \cdots & \Sigma_{h'H_m,mc} \\ \vdots & \cdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \Sigma_{H_m1,mc} & \cdots & \Sigma_{H_mh,mc} & \cdots & \Sigma_{H_mh',mc} & \cdots & \Sigma_{H_m,mc} \end{pmatrix},$$

$$\Sigma_{mm',c} = \begin{pmatrix} \Sigma_{1m,1m'c} & \cdots & \Sigma_{1m,h'm'c} & \cdots & \Sigma_{1m,H_{m'}m'c} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ \Sigma_{hm,1m'c} & \cdots & \Sigma_{hm,h'm'c} & \cdots & \Sigma_{hm,H_{m'}m'c} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ \Sigma_{H_mm,1m'c} & \cdots & \Sigma_{H_mm,h'm'c} & \cdots & \Sigma_{H_mm,H_{m'}m'c} \end{pmatrix},$$

$$\Sigma_{h,mc} = \begin{pmatrix} p_{1hms} \cdot (1 - p_{1hmc}) & \cdots & -p_{1hmc} \cdot p_{jhmc} & \cdots & -p_{1hmc} \cdot p_{J_{hm}hmc} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ -p_{1hmc} \cdot p_{jhmc} & \cdots & p_{jhms} \cdot (1 - p_{jhmc}) & \cdots & -p_{jhmc} \cdot p_{J_{hm}hmc} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ -p_{1hmc} \cdot p_{J_{hm}hmc} & \cdots & \cdots & \cdots & p_{J_{hm}hms} \cdot (1 - p_{J_{hm}hmc}) \end{pmatrix},$$

$$\Sigma_{hh',mc} = \begin{pmatrix} -p_{1hmc} \cdot p_{1h'mc} & \cdots & -p_{1hmc} \cdot p_{j'h'mc} & \cdots & -p_{1hmc} \cdot p_{J_{h'm}h'mc} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ -p_{jhmc} \cdot p_{1h'ms} & \cdots & -p_{jhmc} \cdot p_{j'h'mc} & \cdots & -p_{jhmc} \cdot p_{J_{h'm}h'mc} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ -p_{J_{hm}hmc} \cdot p_{1h'ms} & \cdots & -p_{J_{hm}hmc} \cdot p_{j'h'mc} & \cdots & -p_{J_{hm}hmc} \cdot p_{J_{h'm}h'mc} \end{pmatrix},$$

$$\Sigma_{hm,h'm'c} = \begin{pmatrix} -p_{1hmc} \cdot p_{1h'm'c} & \cdots & -p_{1hmc} \cdot p_{j'h'm'c} & \cdots & -p_{1hmc} \cdot p_{J_{h'm'}h'm'c} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ -p_{jhmc} \cdot p_{1h'm'c} & \cdots & -p_{jhmc} \cdot p_{j'h'm'c} & \cdots & -p_{jhmc} \cdot p_{J_{h'm'}h'm'c} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ -p_{J_{hm}hmc} \cdot p_{1h'm'c} & \cdots & -p_{J_{hm}hmc} \cdot p_{j'h'm'c} & \cdots & -p_{J_{hm}hmc} \cdot p_{J_{h'm'}h'm'c} \end{pmatrix}.$$

Consistency with RUM in the Three-Level NMNL Models

According to RUM Conditions (2.10), (2.11) and (2.12), it has been showed that the following conditions are necessary for consistency with random utility maximization.

Theorem 2.4. (Gil-Molton and Hole 2004)*: The necessary conditions for compatibility three -level nested multinomail logit models with RUM are as follow (w.r.t Choice set $\mathbb{C}_c$):*

$$\mu_m \leq \frac{1}{1-p_{mc}}, \forall m = 1, 2, \ldots, M, \tag{2.36}$$

$$\mu_m \leq \frac{4}{3(1-p_{mc}) + [(1+7p_{mc})(1-p_{mc})]^{\frac{1}{2}}}, \forall m \in \mathbb{M}_3 \equiv \{m | H_m \geq 3\}, \tag{2.37}$$

$$\lambda_{hm} \leq \frac{1}{\frac{1-p_{mc}}{\mu_m} + (1-p_m)p_{h|mc}}, \forall h, m \tag{2.38}$$

$$\lambda_{hm} \leq \frac{4}{\frac{3}{\mu_m} + 3p_{h|mc} - 3(\frac{1}{\mu_m} + p_{mc})p_{h|mc} + D^{\frac{1}{2}}}, \forall h, m \in \mathbb{M}'_3 \equiv \{h, m | J_{hmc} \geq 3\}, \tag{2.39}$$

with:

$$D = (1+7p_{mc})(1-p_{mc})p^2_{h|mc} + \frac{(1+7p_{h|mc})(1-p_{h|mc})}{\mu^2_m} - \frac{6(1-p_{mc})(1-p_{h|mc})p_{h|mc}}{\mu_m},$$

where $\mathbb{M}$ denotes a set of choice sets with at least three sub-nests in each nest and $\mathbb{M}'$ is the set of all choice sets with at least three alternatives in sub-nests.

Obtaining the above four conditions, it has been used of differentiation of equation (2.35). The conditions (2.36) and (2.37) correspond to the conditions in *(Herriges and Kling 1996)* and the conditions (2.38) and (2.39) are implied by the first and the second order mixed derivative of (2.35) according to $v_{j'|hms}$ for $j' \in \mathbb{C}_{hms}$, where $j' \neq j$. The above conditions are necessary and sufficient for a model with three alternatives per sub-nest and three sub-nest per nest.

For a model with two alternatives per sub-nest and two sub-nest per nest the conditions (2.36) and (2.38) are necessary and sufficient. The conditions are not sufficient when there are more than three alternative per sub-nest, but in practical applications testing the first and second-order conditions may be considered satisfactory.

3 OPTIMAL DESIGN

The design of experiments is an important part of scientific research. Design involves specifying all aspects of an experiment and choosing the values of variables that can be controlled before the experiment starts. Control variables might include: Choosing which treatments to study, defining the treatments precisely, choosing blocking factors, choosing how to randomize, specifying the experimental unites to be used, specifying a length of time for the experiment to be performed, choosing a sample size and choosing the proportion of observations to allocate to each treatment. These are all relevant aspects in design.

When designing an experiment, decisions must be made before data collection, and data collection is restricted by limited resources. Because information is usually available prior to experimentation and, indeed, often motivates doing the experiment, Bayesian methods are ideally suited to contributed to experimental design. Bayesian decision theory also motivates precise specification of the reason the experiment is being conducted. Like most areas of Bayesian statistics, Bayesian experimental design has gained popularity in the past two decades. But also like many areas of Bayesian statistics, applications to actual experiments still lag behind the theory. The basic idea in experimental design is that statistical inference about the quantities of interest can be improved by appropriately selecting the values of the control variables. In estimation problems, estimators with small variance are usually desirable. Control variables should therefore be selected to achieve small variability for the estimator chosen. Much depends however on what is to be estimated, and how it will be estimated. Specifying the purpose of the experiment generates various criteria for the choice of a design. *(Chaloner and Verdinelli 1995)* have addressed the fundamental principles of design by providing a general Bayesian decision theoretic framework for a coherent approach.

In this chapter is presented the definition of various optimality and collected the corresponding equivalence theorems (for linear and non-linear model) which are useful tools for checking optimality. Collecting this chapter, we have used of various paper and books like *(Fedorov 1972), (Silvey 1980), (Pukelsheim 1993), (Schwabe 1996)* and others.

Design are represented by the measure ξ over $\mathcal{X}$. If the design has trials at n distinct points in $\mathcal{X}$, we write;

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \cdots & x_n \\ w_1 & w_2 & \cdots & w_n \end{array} \right\} \in \Xi, \tag{3.1}$$

where $\Xi = \{\xi | x_i \in \mathcal{X} \ and \ \sum_{i=1}^{n} w_i = 1; 0 < w_i < 1; \forall i\}$.

In this situation, $\mathcal{X}$ is design region and the first line gives the values of the factors (explanatory variables) at the design points with the w_i the associated design weights. Since ξ is a measure then $\int_{\mathcal{X}} \xi(dx) = 1$.

Now, if we wish to stress that a measure refers to an exact design, realizable in integers for a specific N, the measure is written:

$$\xi_N = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ \frac{r_1}{N} & \frac{r_2}{N} & \dots & \frac{r_n}{N} \end{array} \right\} \in \Xi_N \subset \Xi, \tag{3.2}$$

where r_i is the integer number of trials at x_i and $\sum_{i=1}^{n} r_i = N$. Obtaining the optimal design for design (3.1) and (3.2) we need to have design criterion. The most important design criterion in applications is that of D-optimality, in which the generalized variance, or its logarithm $-\log(\det(\mathbf{M}(\xi)))$, is minimized ($\mathbf{M}(\xi)$ is information matrix which is proportional with the inverse of variance of parameters estimator). In this situation, according to $E(\mathbf{Y}|x) = \mathbf{f}^T(x)\boldsymbol{\beta}$, we will have:

$$\mathbf{M}(\xi) = \int_{x\in\mathcal{X}} \mathbf{f}(x)\mathbf{f}^T(x)dx \quad or \quad \mathbf{M}(\xi) = \sum_{i=1}^{n} w_i \mathbf{f}(x_i)\mathbf{f}^T(x_i).$$

Designs which maximize $\det(\mathbf{F}^T\mathbf{F}) = \det(\mathbf{M}(.))$ are called D-optimum (for determinant), where (p is the number of parameters):

$E(\mathbf{Y}|\mathbf{x}) = \mathbf{F}^T\boldsymbol{\beta},\ Var(\mathbf{Y}) = \sigma^2\mathbf{I}_n;$

- $\mathbf{F} = (\mathbf{f}(x_1), \dots, \mathbf{f}(x_n))^T, \quad \boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T,$
- $\mathbf{f}(x_i) = (f_1(x_i), \dots, f_p(x_i))^T.$

(Atkinson, et al. 2007) showed that how different designs can be in the values the yield of $\det(\mathbf{F}^T\mathbf{F})$, in the curve of $d(x,\xi) = \mathbf{f}^T(x)\mathbf{M}^{-1}(\xi)\mathbf{f}(x)$ over the design region $\mathcal{X}$, and in the maximum value of the variance over $\mathcal{X}$. An ideal design for these models would simultaneously minimize the generalized variance of the parameter estimates and minimize $d(x,\xi)$ over $\mathcal{X}$. Usually a choice has to be made between these desiderate. Three possible design criteria which relate to these properties are as follows;

- ***D-optimality:*** A design is D-optimum if, it maximizes the value of $\det(\mathbf{M}(\xi))$. The generalized variance of the parameter estimates is minimized.

- ***G-optimality:*** A G-optimum design minimizes the maximum over the design region $\mathcal{X}$ of the standardized variance $d(x,\xi)$. This maximum value equals p (the number of parameters). For continuous designs this optimum design measure ξ^* will also be D-optimum and $\min_{\xi\in\Xi}\max_{x\in\mathcal{X}} d(x;\xi) = p$. Of course, can be showed that (by example) this equivalence may not hold for exact designs. For an exact design, we may have $\min_{\xi\in\Xi}\max_{x\in\mathcal{X}} d(x;\xi) > p$.

- ***V-optimality***: An alternative to G- optimality is V- optimality in which the average of $d(x,\xi)$ over $\mathcal{X}$ is minimized.

A design in which the distribution of trials over $\mathcal{X}$ is specified by a measure ξ, regardless of N, is called continuous or approximate. An approximate design can be rounded to an exact design without losing too much efficiency *(Pukelsheim 1993)*. Without the relaxation to non-integer designs, the design problem is that of a hard integer programming problem. *(Majumdar 1988), (Majumdar 1992)* has derived Bayesian exact designs for a two way analysis of variance model considering a special subclass of prior distributions. This is a particularly useful approach when dealing with the constraints of incomplete blocks. *(Toman 1994)* has derived Bayes optimal exact designs for two- and three level factorial experiments, with and without blocking. One of the important problems she has solved is that of choosing a fraction of the full factorial design.

Already *(Hoel 1958)* noticed that the D- and G- optimum designs coincide in the model of a one-dimensional polynomial regression, and *(Kiefer and Wolfowitz 1959)* proved that this is true for every linear model and in general, only for continuous designs (design for a specified number of trials are called exact). We must note that the designs depend on the number of trials N, the number of factors, the design region, and the permitted number of factor levels.

In practice all designs are exact. For moderate N good exact designs can frequently be found by integer approximation to the optimum continuous measure ξ^*. Often for simple models with p parameters, there will be p designs points with equal weight $\frac{1}{p}$, so that the exact design with $N = p$ trials is optimum. However, if the design weights are not rational, it will not be possible to find an exact design which is identical with the continuous optimum design. Difficulties in finding exact designs usually arise when N is close to the number of support points of the optimum continuous design, leading to a poor approximation to ξ^*. For example, for an N-trials design the information matrix for $\boldsymbol{\beta}$ in the model $E(\mathbf{Y}|\mathbf{x}) = \mathbf{F}\boldsymbol{\beta}$ was defined as $\mathbf{F}^T\mathbf{F}$, where:

$$\mathbf{F}^T\mathbf{F} = \sum_{i=1}^{N} \mathbf{f}(x_i)\mathbf{f}^T(x_i)$$

and $\mathbf{f}^T(x_i)$ is the i^{th} row of $\mathbf{F}$. For the continuous design ξ, the information matrix is ($\mathfrak{m}(x)$ is the information matrix in one point, x) ;

$$\mathbf{M}(\xi) = \int_{\mathcal{X}} \mathfrak{m}(x)\xi(dx) = \int_{\mathcal{X}} \mathbf{f}(x)\mathbf{f}^T(x)\xi(dx) = \sum_{i=1}^{n} \mathbf{f}^T(x_i)\mathbf{f}(x_i)w_i$$

which is summed over the n design points, because of the presence of the weights w_i, becomes a scaled version for the exact design ξ_N ,i.e

$$\mathbf{M}(\xi_N) = \frac{\mathbf{F}^T\mathbf{F}}{N} \quad \Rightarrow \quad Var(\hat{Y}(x)) = \sigma^2\mathbf{f}^T(x)(\mathbf{F}^T\mathbf{F})^{-1}\mathbf{f}(x)$$

and for continuous designs the standardized variance of the predicted response is as follows:

$$d(x, \xi) = \mathbf{f}^T(x)\mathbf{M}^{-1}(\xi)\mathbf{f}(x).$$

3.1 The General Equivalence Theorem

In the theory for continuous designs we consider minimization of the general measure of imprecision $\Psi(\mathbf{M}(\xi))$ (a function of information matrix). Under very mild assumption, the most important of which are the compactness of $\mathcal{X}$ and the convexity and differentiability of Ψ, designs which minimize Ψ also satisfy a second criterion. One example is D-optimality, in which $\Psi(\mathbf{M}(\xi)) = \ln\det(\mathbf{M}^{-1}(\xi)) = -\ln\det(\mathbf{M}(\xi))$ so that the determinant of the information matrix $\mathbf{M}(\xi)$ is maximized. It has been showed *(Schwabe 1996)* that compared to many other criteria the D-criterion has the advantage that it is not affected by re-parameterizations of the model. Continuous designs which are D-optimum are also G-optimum, they minimize the maximum over $\mathcal{X}$ of the variance *(White 1973)*.

The general equivalence theorem can be viewed as an application of the result that the derivatives are zero at a minimum of a function. However, the function depends on the measure ξ through the information matrix $\mathbf{M}(\xi)$. Let the measure $\bar{\xi}$ put unit mass at the point x and let the measure ξ' be given by, $\xi' = (1-\alpha)\xi + \alpha\bar{\xi}$ then, we will have

$$\mathbf{M}(\xi') = (1-\alpha)\mathbf{M}(\xi) + \alpha\mathbf{M}(\bar{\xi}),$$

where $\bar{\xi} = \xi_x$ (ξ_x denotes a design for only one point). Accordingly, the derivative of Ψ in the direction $\bar{\xi}$ is *(Fedorov 1972), pp.71*:

$$\phi(x,\xi;\bar{\xi}) = \lim_{\alpha \longrightarrow 0^+} \frac{\Psi((1-\alpha)\mathbf{M}(\xi) + \alpha\mathbf{M}(\bar{\xi})) - \Psi(\mathbf{M}(\xi))}{\alpha}. \tag{3.3}$$

The general equivalence theorem then states the equivalence of the following three conditions on ξ^*;

1. The design ξ^* minimizes $\Psi(\mathbf{M}(\xi))$.
2. The minimum of $\phi(x,\xi^*) \geq 0; \forall x \in \mathcal{X}$.
3. The derivative $\min_{x\in\mathcal{X}} \phi(x,\xi^*) = 0$ achieves its minimum at the points of the design.

Theorem 3.1. *(Kiefer and Wolfowitz 1959) The following assertion are equivalence;*

1. *The design ξ^* maximizes $\det(\mathbf{M}(\xi))$, (minimizes $\det(Var(\hat{\beta}))$),*
2. *The design ξ^* minimizes $\max_{x\in\mathcal{X}} \lambda(x)d(x,\xi)$,*
3. *$\max_{x\in\mathcal{X}} \lambda(x)d(x,\xi^*) = p$,*

where p is the number of parameters, $\lambda(x)$ is efficiency function (See *(Fedorov 1972)*, pp.71-73) and

$$\mathbf{M}(\xi) = \sum_{i=1}^{n} w_i \lambda(x_i)\mathbf{f}(x_i)\mathbf{f}^T(x_i).$$

This theorem provides methods for the construction and checking of optimum designs. However, it says nothing about n, the number of support points of the design. The information matrix of any design can be represented as a weighted sum of, at most, $\frac{p(p+1)}{2}$ information matrices $\mathfrak{m}(\bar{\xi}_i)$ where $\bar{\xi}_i$ puts unit weight at the design point x_i.

The bound an the number of design points depends on the linear structure of $\mathbf{M}(\xi)$ and so holds for any criterion which is a function of a single information matrix. The general equivalence theorem holds for continuous designs represented by the measure ξ. In general, it does not hold for exact designs. For D-optimality the implication is that there may be some values of N for which on design will be D-optimum and another G-optimum.

As it was told, for the linear model, a design measure ξ^* is called G-optimum if $d(x, \xi^*)$, which is proportional to the variance (for homoscedastic case) of the least squares estimator of the response at x satisfies:

$$\sup_{x \in \mathcal{X}} d(x, \xi^*) = \min_{\xi \in \Xi} \sup_{x \in \mathcal{X}} d(x, \xi).$$

In order to formulate an analogue to the General Equivalence Theorem for the more general model (non-linear model) an analogue of $d(x, \xi)$ is needed. *(White 1973)* has taken the function:

$$d(x, \xi, \boldsymbol{\theta}) = tr\left(\mathbf{I}(x, \boldsymbol{\theta})\mathbf{M}^{-1}(\xi, \boldsymbol{\theta})\right)$$

as this analogue, for non-linear models, whenever $\mathbf{M}(\xi, \boldsymbol{\theta})$ is nonsingular and $\mathbf{I}(x, \boldsymbol{\theta}) = \mathbf{M}(\xi_x, \boldsymbol{\theta})$ is information matrix for only one point. Motivation for this choice is given below but it should be noted that (i) it reduces to $d(x, \xi)$ for the linear model with normally distributed errors, and (ii) it is invariant under nonlinear transformation of Y_x and of $\boldsymbol{\theta}$. Then, a design measure ξ^* is called $G(\boldsymbol{\theta})$-optimum if

$$\sup_{x \in \mathcal{X}} d(x, \xi^*, \boldsymbol{\theta}) = \min_{\xi \in \Xi} \sup_{x \in \mathcal{X}} d(x, \xi, \boldsymbol{\theta}) \tag{3.4}$$

for $\boldsymbol{\theta}$ taking its true value. Let us consider:

$$D(\boldsymbol{\theta}) = -\ln\det\left(\mathbf{M}(\xi, \boldsymbol{\theta})\right), \qquad G(\boldsymbol{\theta}) = \lambda(x)\mathbf{f}^T(x)\mathbf{M}^{-1}(\xi, \boldsymbol{\theta})\mathbf{f}(x)$$

thus analogue of *Kiefer and Wolfowitz*'s result becomes;

Theorem 3.2. (White 1973)*:* The following conditions on a design measure ξ are equivalent*:*

1. *ξ^* is $D(\boldsymbol{\theta})$-optimum : $D(\xi^*, \boldsymbol{\theta}) = \min_{\xi \in \Xi} D(\xi, \boldsymbol{\theta})$*

2. *ξ^* is $G(\boldsymbol{\theta})$-optimum :$tr\left(\mathbf{I}(x, \boldsymbol{\theta})\mathbf{M}^{-1}(\xi^*, \boldsymbol{\theta})\right) \leq p$; $\forall x \in \mathcal{X}$*

3. *$tr\left(\mathbf{I}(x^*, \boldsymbol{\theta})\mathbf{M}^{-1}(\xi^*, \boldsymbol{\theta})\right) = p$*

for $\boldsymbol{\theta}$ taking its true value.

Based on Theorem 3.2, the third condition is necessary but it is not sufficient. That means that, the first and second condition are equivalent and if those be hold the third condition is hold, also. But, if the third condition be hold we can not say the two conditions first and second are hold.

3.2 Other Optimality Criteria

It was told that the most important design criterion in applications is that of D-optimality in which the generalized variance, or its logarithm $-\ln\det(\mathbf{M}(\xi))$, is minimized. But, there exist other optimality criteria that we can check optimality by using them, like; A-optimality which minimizes the average variance of the parameter estimates ($tr\mathbf{M}^{-1}(\xi)$). It was defined that $\phi(x,\xi)$ is derivative of $\Psi(\mathbf{M}(\xi)) = \ln\det(\mathbf{M}^{-1}(\xi))$. Now in order to state the equivalence theorems for D- and A-Optimality criterion it can be rewritten the derivative of $\Psi(\mathbf{M}(\xi))$ as (See Equation (3.3));

$$\phi(x,\xi) = tr\left(\mathbf{M}(\xi).\frac{\partial\Psi(\mathbf{M}(\xi)}{\partial\mathbf{M}(\xi)}\right) - \psi(x,\xi), \tag{3.5}$$

where:

$$\psi(x,\xi) = \mathbf{f}^T(x)\frac{\partial\Psi(\mathbf{M}(\xi))}{\partial\mathbf{M}(\xi)}\mathbf{f}(x). \tag{3.6}$$

With respect to (3.5) and (3.6) we will have (p is the number of parameters):

- if $\Psi(\mathbf{M}(\xi) = \ln(\det(\mathbf{M}^{-1}(\xi)))$ then:

$$\frac{\partial\Psi(\mathbf{M}(\xi))}{\partial\mathbf{M}(\xi)} = \mathbf{M}^{-1}(\xi) \Rightarrow \phi(x,\xi) = p - \mathbf{f}^T(x)\mathbf{M}^{-1}(\xi)\mathbf{f}(x)$$

- if $\Psi(\mathbf{M}(\xi) = tr(\mathbf{M}^{-1}(\xi)))$ then:

$$\frac{\partial\Psi(\mathbf{M}(\xi))}{\partial\mathbf{M}(\xi)} = \mathbf{M}^{-2}(\xi) \Rightarrow \phi(x,\xi) = tr(\mathbf{M}^{-1}(\xi)) - \left(\mathbf{f}^T(x)\mathbf{M}^{-1}(\xi)\right)\left(\mathbf{f}^T(x)\mathbf{M}^{-1}(\xi)\right)^T,$$

An advantage of D-optimality is that the optimum designs for quantitative factors do not depend upon the scale of the variables. Linear transformations leave the D-optimum design unchanged, which is not in general the case for A-optimum design. But, for designs with all factors qualitative, such as block designs the problem of scale does not arise and A-optimum design are frequently employed. Such that, D-optimum designs are more readily constructed for experiments with quantitative factors. We now consider an useful extension to D-optimality.

D_q-Optimality

D_q-optimum designs are appropriate when interest is in estimating a subset of q of the parameters as precisely as possible. Let the terms of the model be divided into two groups;

$$E(\mathbf{Y}|x) = \mathbf{f}^T(x)\boldsymbol{\theta} = \mathbf{f}_1^T(x)\boldsymbol{\theta}_1 + \mathbf{f}_2^T(x)\boldsymbol{\theta}_2,$$

where the $\boldsymbol{\theta}_1$ are the q parameters of interest, the $p-q$ parameters $\boldsymbol{\theta}_2$ are then treated as nuisance parameters. On example is when $\boldsymbol{\theta}_1$ corresponds to the experimental factors and $\boldsymbol{\theta}_2$ corresponds to the parameters for the blocking factors. In this situation, we will have;

$$\mathbf{M}(\xi) = \begin{pmatrix} \mathbf{M}_{11}(\xi) & \mathbf{M}_{12}(\xi) \\ \mathbf{M}_{21}(\xi) & \mathbf{M}_{22}(\xi) \end{pmatrix} \Rightarrow \mathbf{M}^{-1}(\xi) = \begin{pmatrix} \mathbf{M}^{11}(\xi) & \mathbf{M}^{12}(\xi) \\ \mathbf{M}^{21}(\xi) & \mathbf{M}^{22}(\xi) \end{pmatrix},$$

where the covariance matrix for the least squares of $\boldsymbol{\theta}_1$ is $\mathbf{M}^{11}(\xi)$ which is showed by

$$\mathbf{M}^{11}(\xi) = \left(\mathbf{M}_{11}(\xi) - \mathbf{M}_{12}(\xi)\mathbf{M}_{22}^{-1}(\xi)\mathbf{M}_{12}^{T}(\xi)\right)^{-1}. \tag{3.7}$$

The D_q-optimum design for $\boldsymbol{\theta}_1$ accordingly maximizes the determinant:

$$\det\left(\mathbf{M}_{11}(\xi) - \mathbf{M}_{12}(\xi)\mathbf{M}_{22}^{-1}(\xi)\mathbf{M}_{12}^{T}(\xi)\right) = \frac{\det(\mathbf{M}(\xi))}{\det(\mathbf{M}_{22}(\xi))} \tag{3.8}$$

such that;

$$d_q(x,\xi) = \mathbf{f}^T(x)\mathbf{M}^{-1}(\xi)\mathbf{f}(x) - \mathbf{f}_2^T(x)\mathbf{M}_{22}^{-1}(\xi)\mathbf{f}_2(x)$$

and for D_q-optimum design ξ^*, $d_q(x,\xi^*) \leq q$.

These results follow from those for D_A-optimality by taking $\mathbf{A} = (\mathbf{I}_q, 0)$, where $\mathbf{I}_q$ is the $q \times q$ identity matrix *(Atkinson, et al. 2007), pp.137.*

(White 1973) has expanded D_q-optimality for nonlinear model. She has assumed that if the parameters $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_q; (q < p)$ are the only parameters of interest, it can be partitioned information matrix into:

$$\mathbf{M}(\xi,\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{M}_{11}(\xi,\boldsymbol{\theta}) & \mathbf{M}_{12}(\xi,\boldsymbol{\theta}) \\ \mathbf{M}_{12}^T(\xi,\boldsymbol{\theta}) & \mathbf{M}_{22}(\xi,\boldsymbol{\theta}) \end{pmatrix}, \tag{3.9}$$

where $\mathbf{M}_{11}(\xi,\boldsymbol{\theta})$ is an $q \times q$ sub-matrix and $\mathbf{M}_{22}(\xi,\boldsymbol{\theta})$ is assumed nonsingular. Thus a design measure ξ^* is called $D_q(\boldsymbol{\theta})$-optimum if

$$\frac{\det[\mathbf{M}(\xi^*,\boldsymbol{\theta})]}{\det[\mathbf{M}_{22}(\xi^*,\boldsymbol{\theta})]} = \max_{\xi\in\Xi} \frac{\det[\mathbf{M}(\xi,\boldsymbol{\theta})]}{\det[\mathbf{M}_{22}(\xi,\boldsymbol{\theta})]}$$

for $\boldsymbol{\theta}$ taking its true value.

Note 3.1: The quantity $n^{-1}\left(\frac{\mathbf{M}(\xi,\boldsymbol{\theta})}{\mathbf{M}_{22}(\xi,\boldsymbol{\theta})}\right)^{-1}$ is the asymptotic generalized variance of the estimators of $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_q$.

In this situation, by noting to (3.4) and (3.9) can be told that a design measure ξ^* is called $G_q(\boldsymbol{\theta})$-optimum if $\sup_{x\in\Xi} d_q(x,\xi^*,\boldsymbol{\theta}) = \min_{\xi\in\Xi}\max_{x\in\mathcal{X}} d_q(x,\xi,\boldsymbol{\theta})$, where:

$$d_q(x,\xi,\boldsymbol{\theta}) = tr\left(\mathbf{I}(x,\boldsymbol{\theta})\mathbf{M}^{-1}(\xi,\boldsymbol{\theta})\right) - tr\left(\mathbf{I}_{22}(x,\boldsymbol{\theta})\mathbf{M}_{22}^{-1}(\xi,\boldsymbol{\theta})\right)$$

and $\mathbf{I}(x,\boldsymbol{\theta})$ is partitioned in the same way as $\mathbf{M}(\xi,\boldsymbol{\theta})$. That means;

$$\mathbf{I}(x,\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_{11}(x,\boldsymbol{\theta}) & \mathbf{I}_{12}(x,\boldsymbol{\theta}) \\ \mathbf{I}_{12}^T(x,\boldsymbol{\theta}) & \mathbf{I}_{22}(x,\boldsymbol{\theta}) \end{pmatrix}.$$

(White 1973) has also introduced based on Kiefer's theorem *(Kiefer 1961)* the following theorem for the nonlinear models;

Theorem 3.3. (White 1973)*: The following conditions on a design measure ξ are equivalent (for a subset of parameters):*

1. *ξ^* is $D_q(\boldsymbol{\theta})$-optimum,*
2. *ξ^* is $G_q(\boldsymbol{\theta})$ -optimum,*
3. $\sup_{x\in\mathcal{X}} d_q(x, \xi^*, \boldsymbol{\theta}) = q$

for $\boldsymbol{\theta}$ taking its true value.

3.3 The General Properties of D-Optimal Designs

The general properties of D-optimum designs are as follows:

1. The D-optimum design ξ^* maximizes $\det(\mathbf{M}(\xi))$ or, equivalently, minimizes $\det(\mathbf{M}^{-1}(\xi))$.
2. The D-efficiency of an arbitrary design ξ is defined as;

$$D_{eff.} = \left(\frac{\det(\mathbf{M}(\xi))}{\det(\mathbf{M}(\xi^*))}\right)^{\frac{1}{p}},$$

   where p is the number of parameters.

3. A generalized G-optimum design over the region $\Re$ is one for which,

$$\max_{x\in\Re} w(x)d(x,\xi^*) = \min_{\xi\in\Xi} \max_{x\in\Re} w(x)d(x,\xi).$$

   Here, usually $\Re$ is taken as the design region $\mathcal{X}$ and $w(x) = 1$, when the equivalence of D-and G-optimum designs results. Then, with $\bar{d}(\xi) = \max_{x\in\mathcal{X}} d(x,\xi)$ the G-efficiency of a design ξ is defined by $G_{eff.} = \frac{\bar{d}(\xi^*)}{\bar{d}(\xi)} = \frac{p}{\bar{d}(\xi)}$

4. The D-optimum design need not be unique. If ξ_1^* and ξ_2^* are D-optimum designs, the design; $\xi^* = \alpha\xi_1^* + (1-\alpha)\xi_2^*; 0 \leq \alpha \leq 1$ is also D-optimum. Of course, $\mathbf{M}(.)$ is unique, means that $\mathbf{M}(\xi_1) = \mathbf{M}(\xi_2) = \mathbf{M}(\xi)$ *(Atkinson, et al. 2007), pp.152.*
5. The D-optimality criterion is model dependent. However, the design is invariant to non-degenerate linear transformation of the model. Thus, a design D-optimum for the model $\eta = \boldsymbol{\beta}^T\mathbf{f}(x)$ is also D-optimum for the model $\eta = \gamma^T\mathbf{g}(x)$, if $\mathbf{g}(x) = \mathbf{A}\mathbf{f}(x)$ and $\det(\mathbf{A}) \neq 0$, where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are both $p\times 1$ vectors of unknown parameters. *(Schwabe 1996)* has defined that a design ξ is invariant with respect to transformation group, G,if $\xi^g = \xi; \forall g \in G$, where a group G of transformations of $\mathcal{X}$ induces linear transformations of the regression

function $\mathbf{f} : \mathcal{X} \to \Re^p$ if every $g \in G$ ($g : \mathcal{X} \to \mathcal{X}$, one-one function) induces a linear transformation of $\mathbf{f}$. We know that a transformation g of $\mathcal{X}$ induces a linear transformation of the regression function $\mathbf{f} : \mathcal{X} \rightharpoonup \Re^p$ if there exists a $p \times p$ matrix $\mathbf{Q}_g$ with $\mathbf{f}(g(x)) = \mathbf{Q}_g\mathbf{f}(x); \forall x \in \mathcal{X}$. In this case, it is defined that a design ξ is information invariant with respect to G and the linear regression function $\mathbf{f}$, if $\mathbf{Q}_g\mathbf{M}(\xi)\mathbf{Q}_g^T = \mathbf{M}(\xi); \forall g \in G$ criterion function $\Psi : \Xi \to \Re$ is invariant with respect to G if $\Psi(\xi^g) = \Psi(\xi); \forall \xi \in \Xi$, $\forall g \in G$. Based on previous descriptions *(Schwabe 1996)* has proved the following theorem;

Theorem 3.4. (Schwabe 1996)*: Let G induce linear transformation of* $\mathbf{f}$ *then;*

- *The D-criterion is invariant.*
- *If G is orthogonal for* $\mathbf{f}$*, then every Ψ_k-criterion,* $0 \le k \le \infty$*,(*(Atkinson, et al. 2007), pp.136*) including A- and E-criterion, is invariant, where a group G is orthogonal for* $\mathbf{f}$ *if for every $g \in G$ the transformation matrix* $\mathbf{Q}_g$ *is orthogonal, i.e.* $\mathbf{Q}_g^T = \mathbf{Q}_g^{-1}$.

3.4 The Properties of Information Matrices

The properties of information matrices are as follow :

Theorem 3.5. (Fedorov 1972), pp.66*:*

1. *For any design ξ the information matrix* $\mathbf{M}(\xi)$ *is a symmetric positive-semi-definite matrix;* $\mathbf{a}^T\mathbf{M}\mathbf{a} \ge 0; \forall \mathbf{a} \in \Re^p$.
2. *The matrix* $\mathbf{M}(\xi)$ *is degenerate* $(\det(\mathbf{M}(\xi)) = 0)$*, if the support points of the design ξ contains less than p points.*
3. *The family of matrices* $\mathbf{M}(\xi)$*, corresponding to all possible normalized designs, is convex. If the function* $\mathbf{f}(x)$ *and the efficiency function $\lambda(x)$ are continuous in the region $\mathcal{X}$ of possible measurements, and $\mathcal{X}$ is compacted, then, the set of information matrices is compacted.*
4. *For any design ξ the matrix* $\mathbf{M}(\xi)$ *can be represented in the form ($\lambda(x_i)$ denotes the efficiency function based on x_i):*

$$\mathbf{M}(\xi) = \sum_{i=1}^{n} w_i\lambda(x_i)\mathbf{f}(x_i)\mathbf{f}^T(x_i),$$

where $n \le \left(\frac{p(p+1)}{2}\right) + 1, 0 \le w_i \le 1, \sum_{i=1}^{n} w_i = 1.$

3.5 Bayesian Optimal Design Theory

Design is more difficult when the model is not linear or when a nonlinear function of the coefficients of a linear model is of interest. Such problems are referred to as nonlinear design problems. It will be shown that the design problem can be formulated as maximizing expected utility but approximations must typically be used as the exact expected utility is often a complicated integral. Designs can still be denoted by a probability measure ξ over the design space $\mathcal{X}$ and the set of all such measures be denoted Ξ. The measures may be arbitrary probability measures representing approximate, or continuous, designs, or measures corresponding to exact designs which have mass $\frac{1}{n}$ on n, not necessarily distinct, points.

The frequents's strategy for designing a nonlinear model is to assume a best guess of the parameter values. This approach leads to what are termed *local optimal* designs (obtaining optimal design based on special values of parameters). A natural generalization is to use a prior . *(Chaloner and Larntz 1989)* argued that under mild conditions, the joint posterior distribution of parameters is approximately a multivariate normal distribution with mean equals to the maximum likelihood estimate, and variance covariance matrix equals to the inverse of the observed *Fisher* information matrix evaluated at the MLE' s.

Further, the prior distribution of parameters can be used as the predictive distribution of their MLE's. If Ψ is a convex functional of the *Fisher* information matrix $\mathbf{M}(\xi, \boldsymbol{\theta})$, a *Bayesian* optimality given by *(Chaloner 1993)* is;

$$\Psi(\xi) = E_\theta \left(\Psi(\mathbf{M}(\xi, \boldsymbol{\theta})) \right), \tag{3.10}$$

where E_θ denotes the expectation of $\Psi(\mathbf{M}(\xi, \boldsymbol{\theta}))$ with respect to prior distribution of parameters. *Bayesian* optimal design is a logical outgrowth of classical design for cases where the criterion is a function of unknown parameters.

Following *(Lindley, 1956)* suggestion, several authors *(Stone 1959a),(DeGroot, 1962), (DeGroot, 1986)* and *(Bernardo, 1979)* considered the expected gain in Shannon information given by an experiment as a utility function *(Shannon, 1948)*. These authors purposed choosing a design that maximizes the expected gain in Shannon information or, equivalently, maximizes the expected Kullback-Leibler distance *(Kullback and Leibler 1951)* between the posterior and the prior distributions:

$$\int_{\Theta^p, \mathcal{Y}} \ln \frac{p(\boldsymbol{\theta}|y, \xi)}{\pi(\boldsymbol{\theta})} p(y, \boldsymbol{\theta}|\xi) d\boldsymbol{\theta} dy.$$

(Zacks 1977) has considered problems where the data are to be sampled from an experimental family with known scale parameter and where some function of the mean is linear in an explanatory variable. This class of generalized linear models includes quantal response models and models for exponential lifetimes. The Fisher information matrix has a common form for these models and *(Zacks 1977)* has considered designs that maximize the expected value of the determinant of $\mathbf{M}(\xi, \boldsymbol{\theta})$, that is:

$$D(\xi) = \int_{\Theta^p} \left[\det(n\mathbf{M}(\xi, \boldsymbol{\theta})) \right] \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Also, *(Atkinson, et al. 2007), (Chapter 18)* have used the following design criterion to obtain optimal design, for generalized linear models;

$$D(\xi) = \int_{\Theta^p} \{[\det(n\mathbf{M}(\xi, \boldsymbol{\theta}))]\}^{-\frac{1}{p}} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

where p is the number of parameters.

But, a crude approximation to expected utility would be to approximate the marginal distribution of $\hat{\boldsymbol{\theta}}$ by a one point distribution. The one point would represent a best guess. This approach, known as local optimality, has been used extensively in nonlinear design and is due to *(Chernoff 1953)*. It is also used in the pioneering paper of *(Box and Lucas 1959)* where the important issues in design for nonlinear regression were identified. Although they used local optimality, *(Box and Lucas 1959)* suggested extending this by taking into account a prior distribution on the parameter values. *(White 1973), (White 1975)* showed how results from linear design theory can be adapted to apply to local optimality in nonlinear models and she also derived locally optimal designs for binary regression experiments.

As local optimality is a very crude approximation to expected utility, it can be considered as being approximately Bayesian although it is typically not justified in this way and is usually used in a non-Bayesian framework. The experimenter is required to specify a best guess, $\boldsymbol{\theta}_0$ for the unknown parameters $\boldsymbol{\theta}$. Local D-optimality involves choosing the design ξ maximizing $D_{\theta_0}(\xi) = \det(\mathbf{M}(\xi, \boldsymbol{\theta}_0))$ or minimizing $\tilde{D}_{\theta_0}(\xi) = \ln(\det(\mathbf{M}(\xi, \boldsymbol{\theta}_0)))^{-1}$ for fixed value $\boldsymbol{\theta}_0$.

For local optimality there are several papers deriving closed form expressions for designs: for example *(White 1973), (Kitsos, et al. 1988), (Ford, et al. 1992)* and *(Wu 1988)*. For a particular value of the unknown parameters the problem often reduces to an equivalent linear problem. Finding optimal Bayesian designs algebraically is much hard and thus implementing Bayesian design criteria requires that designs be found by numerical optimization. Exceptions to this are simple special cases: these cases are not very useful in practice, but they give insight into properties of the optimal designs for more realistic and practical situations. Exact, algebraic results are quite difficult to derive as none of the tools from local optimality are very helpful.

In *(Chaloner 1993)* for example, in a one parameter problem, with prior distributions with only two support points, it is possible to examine exactly how the transition from a one point optimal design to a two point optimal design occurs as the prior distribution is changed. *(Mukhopadhyay and Haines 1995), (Haines 1995), (Dette and Neugebauer 1996)* and *(Dette and Neugebauer 1997)* all considered some nonlinear regression problems involving an exponential mean function, and gave conditions under which the optimal design is of a particular form. Loosely speaking these results can be generalized to say that if the prior distribution is not too dispersed and does not have heavy tails then an optimal Bayesian design has the same number of support points as there are unknown parameters. *(Haines 1995)* gave an insightful geometric interpretation of this and demonstrated how, for a prior distribution with finite support, the problem reduces to a particular convex programming problem.

3.6 Support Points

In most non-Bayesian linear problem an upper bound on the number of support points in an optimal design is available, see *(Pukelsheim 1993), pp.188-189)*. The D-optimality criterion in linear models (in univariate regression) leads to an optimal number of support points that is the same as the 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), pp.27)*. In contrast for nonlinear models there is no such bound available on the number of support points.

(Chaloner and Larntz 1989) have given the first examples of how the number of support points in an optimal Bayesian design increases as the prior distribution becomes more dispersed. They found that for prior distributions that have supported over a very small region the Bayesian optimal designs are almost the same as the locally optimal design and they have the same number of support points as the number unknown parameters. For more dispersed prior distributions there are more support points. This is a useful feature for a design as if there are more support points than unknown parameters, the model assumptions can be checked with data from the experiment. In locally D-optimal designs for various non-linear models follow *Caratheodory* theorem *(Fedorov 1972), pp.66)* and *(Silvey 1980),appendix 2)* that for p-parameter nonlinear model, the number of support points is between p and $\frac{p(p+1)}{2}$. When we search for a D-optimal design, we only need to search for the optimal design in the class of design measures with number of support points between p and $\frac{p(p+1)}{2}$ for which the information matrices are nonsingular.

A common tool for the construction of efficient designs in nonlinear regression models are Bayesian or maximin criteria. Both optimality criteria require prior information regarding the parameters which enter in the model nonlinearly. It was observed numerically by many authors that the number of support points of Bayesian and maximin D-optimal designs is increasing with the amount of uncertainty about the location of the nonlinear parameters. *(Braess and Dette 2004)* have established sufficient conditions for the nonlinear regression models under which the number of support points of Bayesian and maximin D-optimal designs can become arbitrarily large if the prior information regarding the unknown nonlinear parameters in the optimality criterion is reduced. These conditions apply to many of the commonly used regression models (in fact they did not find any model, where these conditions were not satisfied).

(Braess and Dette 2004) restricted their investigations to one- and two parametric regression models, where at most one parameter appears nonlinearly in the model. However, their approach is a general one and can also be applied to regression models with more nonlinear parameters, where some of technicalities have to be adapted to the specific model under consideration. For example, consider a nonlinear regression model with two parameters, say $\boldsymbol{\theta} = (\theta_1, \theta_2)$; such that the local D-optimal design depends on both components of $\boldsymbol{\theta}$. *(Braess and Dette 2004)* have showed that the number of the support points of the standardized maximin D-optimal design becomes arbitrarily large provided that the sufficient conditions in their

theorem are satisfied. They have obtained a similar result which is also available for the Bayesian D-optimality criterion. *(Braess and Dette 2004)* have considered several examples which in all them the number of support points of the standardized maximin and Bayesian D-optimal designs exceeds any given bound if the knowledge about the underlying parameter space, which is incorporated in the optimality criteria, is diminished. This gives a rigorous proof of a phenomenon which was conjectured in many nonlinear regression models for a long time in the literature.

3.7 Optimal Design for Logit Models

Suppose that there are K attributes each with $L_k; k = 1, 2, \ldots, K$ levels. In this situation, we will face to a population, which includes $\prod_{k=1}^{K} L_k = \mathcal{J}$ possible alternatives. To analyze data, we can consider $\mathcal{C}$ choice sets each with J_c alternatives (See Section 2.1). Obtaining optimal design for logit models, experiments $\mathcal{J}/J/\mathcal{S}$ can be considered, where there are $\mathcal{S}$ choice sets, $\mathbb{C}_1, \ldots, \mathbb{C}_s, \ldots, \mathbb{C}_{\mathcal{S}}$, each with J_s; $\forall s \in \mathcal{S}$ alternatives. In this situation, $\mathcal{S} \subset \mathcal{C}$. This means that there are $\mathcal{S}$ choice sets each with $J_s; \forall s \in \mathcal{S}$ alternatives and we suppose that $J_s = J; \forall s \in \mathcal{S}$, also. In Section 2.1 was told that $\mathcal{C}$ includes $2^{\mathcal{J}} - (\mathcal{J} + 1)$ choice sets each with $J_c > 1; \forall c \in \mathcal{C}$ alternatives, where:

$$\sum_{J_c=2}^{\mathcal{J}} \binom{\mathcal{J}}{J_c}$$

denotes the number of choice sets each with at least two alternatives and:

$$\mathcal{S} = \binom{\mathcal{J}}{J_s}$$

denotes a subset of $\mathcal{C}$, which includes choice sets each with the same number of alternatives, $J_s = J; \forall s \in \mathcal{S}$. Of course, to experiment the number S ($p \leq S \leq \mathcal{S}$) choice sets each with $J_s; \forall s \in S$ alternatives can be considered, where $p = \sum_{k=1}^{K} L_k - 1$ is the number of parameters.

Now, according to the utility $U_{js} = v_{js} + \varepsilon_{js}$ (Section 2.1), which consists the deterministic component (v_{js}) and the probabilistic component (ε_{js}) we consider:

1. Obtaining the optimal design when $J_s = 2; \forall s \in \mathcal{S}$ (Bradley-Terry type logistic model).
2. Obtaining the optimal design when $J_s > 2; \forall s$ (MNL Models (Subsection 3.7.1) and NMNL Models (Chapters 4, 5 and 6)).

Obtaining the Optimal Design when $J_s = 2; \forall s$ (Bradley-Terry type logistic Model)

In this sub-section, suppose that the error terms of the utility function have extreme value distribution (type II). The preference for alternative j over alternative j' can be expressed by a binary variable

$$Y_{j,j's} = \begin{cases} 1 & U_{js} > U_{j's} \\ 0 & otherwise. \end{cases}$$

A classical approach to describe these preferences is given by the Bradley-Terry model (Bradley and Terry 1952). Under EVD for error terms the probability of preference is as follow: (McFadden 1974)

$$P(Y_{j,j's} = 1) = \frac{\exp(v_{js})}{\exp(v_{js}) + \exp(v_{j's})}.$$

In the logistic model based on direct observations under the one-way layout situation with L varieties a design ξ is defined by its nonnegative weight w_j on the settings from the set $\{1, 2, \ldots, L\}; \sum_{j=1}^{L} w_j = 1$. Then the information matrix results in *(Graßhoff and Schwabe 2008)*:

$$\begin{aligned} \mathbf{M}(\xi, \boldsymbol{\beta}) &= \sum_{j=1}^{L} w'_j \lambda(\mathbf{f}_d^T(j)\boldsymbol{\beta})\mathbf{f}_d(j)\mathbf{f}_d^T(j) \\ &= \sum_{s=1}^{\mathcal{S}} w_s.\mathbf{M}(\mathbb{C}_s, \boldsymbol{\beta}), \end{aligned}$$

where $\boldsymbol{\beta}$ denotes the part-worth parameters, d related to design defined and S the number of choice sets ($\mathbb{C}_s$) each with two alternatives ($J_s = J = 2; \forall s \in \mathcal{S}$).

Example 3.1. *Suppose that there are three attributes each comprised of two levels. In this case, the alternatives are determined as follow:*

alternative	*attribute*1	*attribute*2	*attribute*3
$\tilde{a}_1$	+1	+1	+1
$\tilde{a}_2$	+1	+1	−1
$\tilde{a}_3$	+1	−1	+1
$\tilde{a}_4$	+1	−1	−1
$\tilde{a}_5$	−1	+1	+1
$\tilde{a}_6$	−1	+1	−1
$\tilde{a}_7$	−1	−1	+1
$\tilde{a}_8$	−1	−1	−1

,

where $\tilde{a}_j$ denotes the alternative j in choice set $\mathbb{C} = \{\tilde{a}_1, \ldots, \tilde{a}_j, \ldots, \tilde{a}_{\mathcal{J}}\}$ (Section 2.1). In this case, we will have:

$$\mathcal{S} = \binom{8}{2} = 28.$$

This means that, there are $\mathcal{S} = 28$ choice sets each with two alternatives. In this experiment, there are three parameters (there are three attributes each with two levels) then we can consider $3 \leq S \leq 28$. Now, let us consider $S = 4$. In this situation, four choice sets with their two alternatives can be considered as follow:

$$\mathbb{C}_1 = \begin{bmatrix} +1 & +1 & +1 \\ -1 & -1 & -1 \end{bmatrix}, \qquad \mathbb{C}_2 = \begin{bmatrix} +1 & +1 & -1 \\ -1 & -1 & +1 \end{bmatrix}$$

$$\mathbb{C}_3 = \begin{bmatrix} +1 & -1 & +1 \\ -1 & +1 & -1 \end{bmatrix}, \qquad \mathbb{C}_4 = \begin{bmatrix} +1 & -1 & -1 \\ -1 & +1 & 1 \end{bmatrix},$$

where $\mathbf{f}(a_{js}) = (f_1(a_{js}), f_2(a_{js}), f_3(a_{js}))^T$ denotes the characterizes of three attributes related to alternative j of choice set $\mathbb{C}_s$ (Section 2.1). For example, $f_1(a_{11}) = 1, f_2(a_{11}) = 1, f_3(a_{11}) = 1$ and $f_1(a_{21}) = -1, f_2(a_{21}) = -1, f_3(a_{21}) = -1$ and so on. Now, based on the definition of the information matrix of design ξ, which consists four above choice sets as follow:

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

we will have:

$$\mathbf{M}(\xi, \boldsymbol{\beta}) = \sum_{s=1}^{4} w_s \cdot \mathbf{M}(\mathbb{C}_s, \boldsymbol{\beta}),$$

which is calculated by:

$$\mathbf{M}(\xi, \boldsymbol{\beta}) = \begin{bmatrix} \sum_{s=1}^4 w_s\gamma_s & w_1\gamma_1 + w_2\gamma_2 - (w_3\gamma_3 + w_4\gamma_4) & w_1\gamma_1 + w_3\gamma_3 - (w_2\gamma_2 + w_4\gamma_4) \\ w_1\gamma_1 + w_2\gamma_2 - (w_3\gamma_3 + w_4\gamma_4) & \sum_{s=1}^4 w_s\gamma_s & w_1\gamma_1 + w_4\gamma_4 - (w_2\gamma_2 + w_3\gamma_3) \\ w_1\gamma_1 + w_3\gamma_3 - (w_2\gamma_2 + w_4\gamma_4) & w_1\gamma_1 + w_4\gamma_4 - (w_2\gamma_2 + w_3\gamma_3) & \sum_{s=1}^4 w_s\gamma_s \end{bmatrix}$$

with the corresponding determinant:

$$\det(\mathbf{M}(\xi, \boldsymbol{\beta})) = 16 \underbrace{\sum_{\ell=1}^{4}\sum_{m=1}^{4}\sum_{n=1}^{4}}_{\ell<m<n} w_\ell w_m w_n h_{\ell mn}; \qquad h_{\ell mn} = \gamma_\ell\gamma_m\gamma_n,$$

where:

- $\gamma_1 = \left(\exp\left((\beta_1 + \beta_2 + \beta_3)\right) + \exp\left(-(\beta_1 + \beta_2 + \beta_3)\right)\right)^{-2}$,
- $\gamma_2 = \left(\exp\left((\beta_1 + \beta_2 - \beta_3)\right) + \exp\left((-\beta_1 - \beta_2 + \beta_3)\right)\right)^{-2}$,
- $\gamma_3 = \left(\exp\left((\beta_1 - \beta_2 + \beta_3)\right) + \exp\left((-\beta_1 + \beta_2 - \beta_3)\right)\right)^{-2}$,
- $\gamma_4 = \left(\exp\left((\beta_1 - \beta_2 - \beta_3)\right) + \exp\left((-\beta_1 + \beta_2 + \beta_3)\right)\right)^{-2}$.

In this situation, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)^T$ is the full parameters vector with $\boldsymbol{\beta}_1 = (\beta_{1,1}, -\beta_{1,1})^T$ (related to the first attribute), $\boldsymbol{\beta}_2^T = (\beta_{2,1}, -\beta_{2,1})$ (related to the second attribute) and $\boldsymbol{\beta}_3 = (\beta_{3,1}, -\beta_{3,1})^T$ (related to the third attribute). This corresponds to type-effect coding leading to $\beta_{1,1} = -\beta_{1,2}$, $\beta_{2,1} = -\beta_{2,2}$ and $\beta_{3,1} = -\beta_{3,2}$, then we are faced to the three parameters $\beta_{1,1}$, $\beta_{2,1}$ and $\beta_{3,1}$ instead of six. Now for the sake of simplicity, suppose that $\beta_{1,1} = \beta_1$, $\beta_{2,1} = \beta_2$ and $\beta_{3,1} = \beta_3$ (See Section 2.1).

In this situation, we must solve the optimization problem, $\max_{w_1,\ldots,w_4} \det(\mathbf{M}(\xi, \boldsymbol{\beta}))$ subject to the natural constraints $\sum_{s=1}^4 w_s = 1$ by a multiplier $\delta_0 > 0$ and the conditions $w_s \geq 0, s = 1, \cdots, 4$ are imposed by the multipliers $\delta_1, \delta_2, \delta_3, \delta_4 \geq 0$, where $\Psi(\xi, \boldsymbol{\beta}) = \det(\mathbf{M}(\xi, \boldsymbol{\beta})$ as local D-optimality criterion is considered. To obtain a solution $w_1^*, w_2^*, w_3^*, w_4^*$, the weights must correspond to the existence of multipliers solving the equations based on the first order conditions:

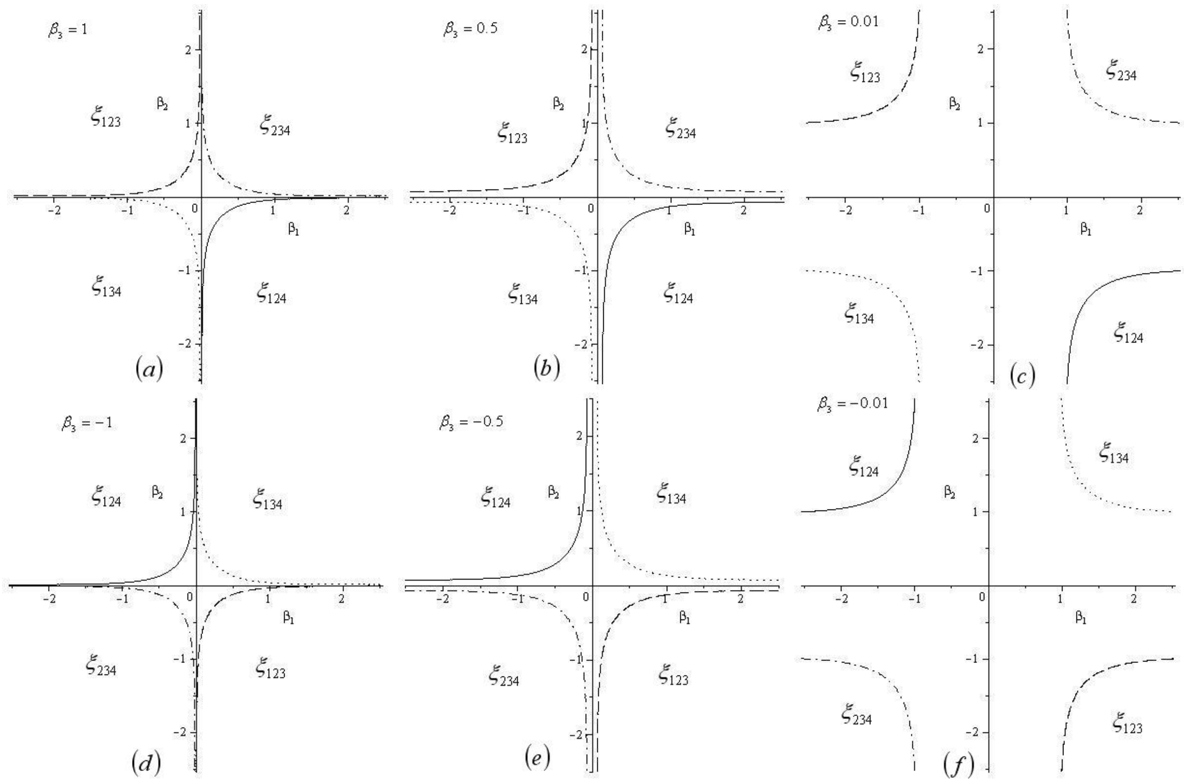

Figure 3.1: Partition of the (β_1, β_2)-plane into regions, where different types of designs are optimal, for different values of β_3, for example, a) $\beta_3 = 1$, b) $\beta_3 = 0.5$, and so on

$$w_2^* w_4^* h_{124} + w_3^* w_4^* h_{134} + w_2^* w_3^* h_{123} + \delta_1 = \delta_0,$$
$$w_1^* w_4^* h_{124} + w_1^* w_3^* h_{123} + w_3^* w_4^* h_{234} + \delta_2 = \delta_0,$$
$$w_1^* w_4^* h_{134} + w_1^* w_2^* h_{123} + w_2^* w_4^* h_{234} + \delta_3 = \delta_0,$$
$$w_1^* w_2^* h_{124} + w_1^* w_3^* h_{134} + w_2^* w_3^* h_{234} + \delta_4 = \delta_0,$$
$$\textstyle\sum_{s=1}^4 \delta_s w_s^* = 0,$$
$$\textstyle\sum_{s=1}^4 w_s = 1.$$

Next consider the case in which three multipliers are equal to zero, $\delta_\ell = \delta_m = \delta_n = 0$, while the other is positive, $\delta_r > 0$. Thus, $w_r^* = 0$ and the reduced system is solved by $\delta_0 = \frac{h_{\ell mn}}{9}$ and $w_\ell^* = w_m^* = w_n^* = \frac{1}{3}$, if the condition:

$$h_{\ell mn} > (h_{\ell mr} + h_{\ell nr} + h_{mnr})$$

hold out. The above is equivalent to:

$$\gamma_\ell \gamma_m \gamma_n > \gamma_r(\gamma_\ell \gamma_m + \gamma_\ell \gamma_n + \gamma_m \gamma_n).$$

In this Example 3.1, this notation $\xi_{\ell mn}$ is used to denote a design including three support points (choice sets) $\mathbb{C}_\ell$, $\mathbb{C}_m$ and $\mathbb{C}_n$ as follow:

$$\xi_{\ell mn} = \left\{ \begin{array}{ccc} \mathbb{C}_\ell & \mathbb{C}_m & \mathbb{C}_n \\ w_\ell & w_m & w_n \end{array} \right\} \in \Xi_{\ell mn}; \forall \ell \neq m \neq n,$$

where $\Xi_{\ell mn} \subset \Xi$.

Figure 3.1 exhibits the contours which separate the parameter regions for the optimal three-point designs, with respect to parameter β_3. When β_3 is positive, it can be used for four different designs $\xi_{123}, \xi_{124}, \xi_{234}$ and ξ_{134} to obtain optimal design nevertheless when considering different regions for β_1 and β_2. To obtain a locally optimal design based on the design ξ_{123}, we must assume $\beta_2 > 0$ and $\beta_1 < 0$, while for the $\beta_1 > 0$ and $\beta_2 > 0$ design, ξ_{234} must be used. Also, we compare design ξ_{134} with the negative values for both β_1, β_2 while based on $\beta_1 > 0, \beta_2 < 0$ the locally optimal design ξ_{234} can be obtained. Note that we must consider condition $\gamma_\ell\gamma_m\gamma_n > \gamma_r(\gamma_\ell\gamma_m + \gamma_\ell\gamma_n + \gamma_m\gamma_n)$ in the all of the described cases. There is a similar result when β_3 is negative (Figure 3.1 (d),(e),(f)). For $\beta_3 < 0$ can be described in a similar way based on the parameters spaces (for all of designs) with the difference that the regions related to ξ_{123}, ξ_{124} and ξ_{134}, ξ_{234} will be replaced respectively (Figure 3.1). In the other case, let β_3 tends to zero from the right. In this situation, to obtain a locally optimal design very big positive values of β_2 must be considered and very small negative values for β_1, based on the design ξ_{123}. In addition, very large positive values for β_1, β_2 must be considered in order to obtain a locally optimal design based on design ξ_{234}. Now, if β_3 tends to zero from left, the large positive values of β_1 and very low negative values for β_2 must be considered to obtain a locally optimal design based on the design ξ_{123}. By Figure 3.1, we can observe all the different conditions required to obtain a locally optimal design when β_3 tends to zero from the right or left. Note that the condition $\gamma_\ell\gamma_m\gamma_n > \gamma_r(\gamma_\ell\gamma_m + \gamma_\ell\gamma_n + \gamma_m\gamma_n)$ does not hold when $\beta_3 = 0$.

For the next case, consider all of the multipliers are equal to zero, $\delta_1 = \delta_2 = \delta_3 = \delta_4 = 0$, but with the assumption that $\beta_3 = 0$. This situation will result in $\gamma_1 = \gamma_2$, $\gamma_3 = \gamma_4$ with the system begging solved by $\delta_0 = \frac{\gamma_1^4\gamma_3^4 D^2}{(\gamma_1\gamma_3^3 - D^2(\gamma_3 - 2\gamma_1))^2}$ and:

$$w_1^* = w_2^* = \frac{\gamma_3(\gamma_1\gamma_3^2 - D^2)}{2(\gamma_1\gamma_3^3 - D^2(\gamma_3 - 2\gamma_1))}\ ,\ w_3^* = w_4^* = \frac{\gamma_1 D^2}{\gamma_1\gamma_3^3 - D^2(\gamma_3 - 2\gamma_1)},$$

where $\gamma_1\gamma_3^2 > D^2$ and $\gamma_1\gamma_3^3 > D^2(\gamma_3 - 2\gamma_1)$ with

$$D = -\sqrt{\gamma_1^2\gamma_3} + \sqrt{\gamma_1\gamma_3(\gamma_1 + \gamma_3)};\quad D > 0.$$

In particular, let $\beta_2 = \beta_1$. See Figure 3.2(c) denoting that $w_3^* = w_4^* > 0$ when $|\beta_1| < 3$, but for the values of β_1 out of the interval $(-3, 3)$, it can be showed that $w_3^* = w_4^* = 0$. In this case, the minimum of w_1^* and the maximum of w_3^* occurs for $\beta_1 = 0$, naturally w_1^* is always greater than w_3^*. But, when $\beta_2 = -\beta_1$, Figure 3.2(d) demonstrates that the maximum of w_1^* and the minimum value of w_3^* occur for $\beta_1 = 0$ (See Figure 3.2(e), where two Figures 3.2(c) and 3.2(d) have been overlapped).

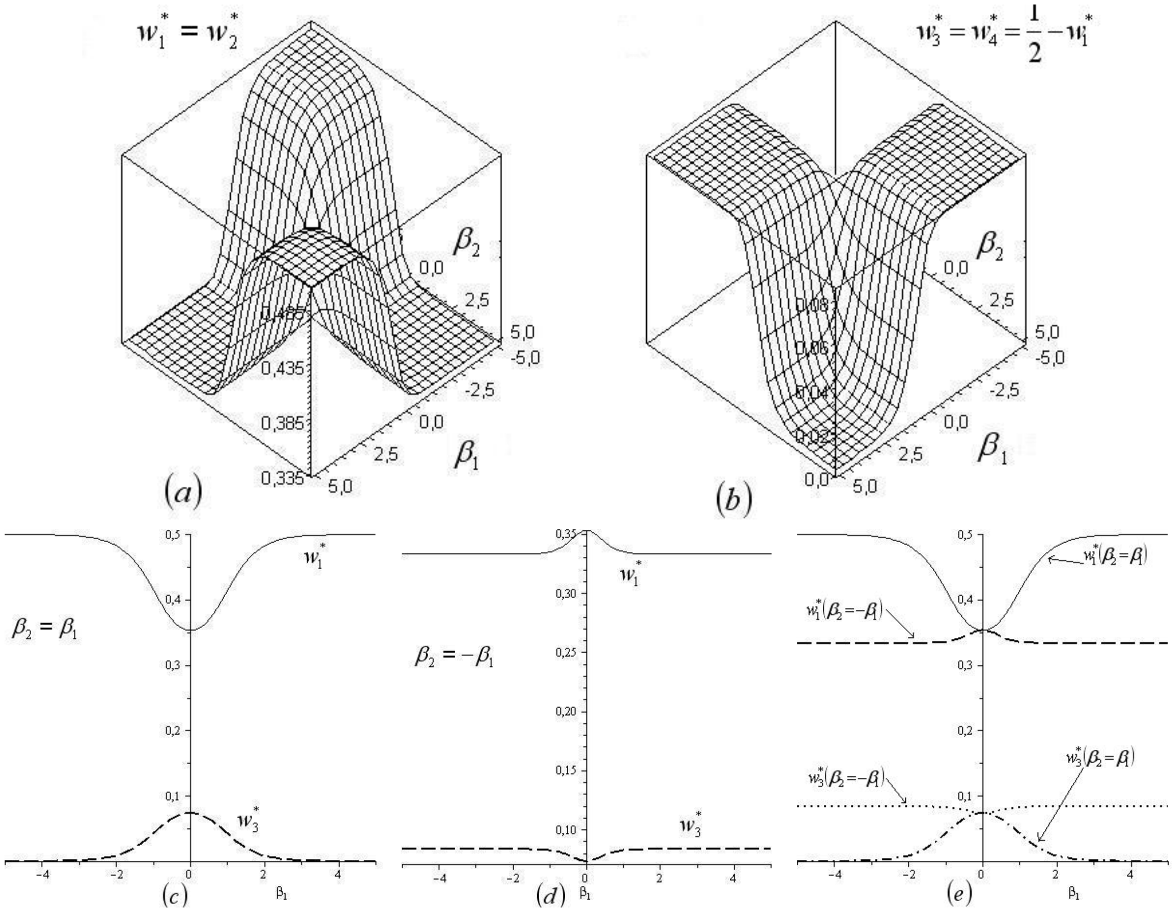

Figure 3.2: $\beta_3 = 0$: optimal weight (a) $w_1^* = w_2^*$, (b) $w_3^* = w_4^*$ based on (β_1, β_2) and two special cases; (c) $\beta_2 = \beta_1$ and (d) $\beta_2 = -\beta_1$, where (e) is overlapped Figure w.r.t Figures (c), (d)

Obtaining the Optimal design when $J_s > 2; \forall s$ (MNL)

The design of a choice experiment comprises a select number of choice sets administered to each respondent. The aim of a choice experiment is to estimate the importance of each attribute and their levels based on the respondent's preferences. The estimates are then used to mimic real marketplace choices by making predictions about consumer future purchases. At present, two design approaches are prevalent; (i) *The Linear design approach* and (ii) *The Baysian design approach*.

Bayesian choice designs have so far been constructed for the Logit models. Since the Logit models are nonlinear in the parameters, the quality of a given design depends on the unknown parameter vector. The *Bayesian* design approach deals with this problem by assuming a prior distribution of likely parameters. To date, most of the *Bayesian* research focus has been on designs for main-effects models. *(Sandor and Wedel 2001)* were the first to introduce the *Bayesian* design procedure in the choice design literature. They generated *Bayesian* designs using the D-optimality criterion for the MNL model. This design criterion seeks to minimize the determinate of the variance-covariance matrix of the parameter estimators. In the *Bayesian* framework, it is referred to as the D_b-optimality criterion.

Optimum design for nonlinear models depends on the values of the vector of unknown parameters $\boldsymbol{\theta}$. Of course, we can solve this problem by replacing $\boldsymbol{\theta}$, by a prior point estimate, $\boldsymbol{\theta}_0$, use which yield locally optimum designs. Now, we want to the cases consider that there is a prior distribution for $\boldsymbol{\theta}$, which may be either discrete or continuous.

An efficient algorithm for constructing *Bayesian* optimal choice designs conjoint choice experiments or more succinctly, choice experiments, are widely used in marketing to measure how the attributes of a product or service jointly affect consumer preferences.

As before was told the aim of a choice experiment is to estimate the importance of each attribute and its levels based on the respondents preferences. The four optimality criteria in the *Bayesian* context are labeled the D_b- ,A_b-,G_b- and V_b-optimality criteria. In this thesis, we use D-optimality criterion, since, *(Yu, et al. Preprint)* have written that D-optimality criterion is invariance to the scale or coding of the attributes. Also, the relative efficiency of the designs does not change when different codings of the attributes are used *(Goos 2002)* and *(Kessels, et al. 2006b)* have denoted that D-and A-optimal designs are nearly as good as the G-and V-optimal designs in terms of prediction quality but much faster to compute compared to G- and V-optimal designs.

The task of the analyst is to find a parameter estimate for $\boldsymbol{\beta}$ in p_{jc} (choice probabilities, Section 2.1) that maximizes the Likelihood given the data. Under very general conditions, the maximum likelihood estimator is consistent and asymptotically normal with covariance matrix $\mathbf{M}^{-1}(\mathbb{C}, \boldsymbol{\beta})$, the inverse of the information matrix.

The fact the information on the parameters depends on the unknown values of those parameters through the probabilities. Therefore, are adopted a *Bayesian* design strategy that integrates the design criteria over a prior parameter distribution $\pi(\boldsymbol{\beta})$. The multivariate normal distribution $N(\boldsymbol{\beta}_0, \Sigma_0)$ (part-worth parameters) was chosen for this purpose.

The D-and A-optimality criteria both are concerned with a precise estimation of the parameters $\boldsymbol{\beta}$ in the Multinomial *logit* models. But the G-and V-optimality criteria were developed to make precise response predictions. In this situation the D-optimality criterion aims at designs that minimize the determinant of the variance-covariance matrix of the parameters estimators, while the A-optimality criterion aims at designs that minimize the trace of the variance-covariance matrix. Thus, the criteria have distinguished related to; (i) *The parameter space* and (ii) *The predicted response*. In this situation, the *Baysian* D-optimality criterion is *(Kessels, et al. 2006b)*:

$$D_b = \int_{\Re^p} \{\det(\mathbf{M}^{-1}(\xi, \boldsymbol{\beta}))\}^{\frac{1}{p}} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta} \tag{3.11}$$

with the D_b-optimal design minimizing integral (3.11). A widely accepted one dimensional measure of information is the determinant of the information matrix. It is motivated from the confidence ellipsoid for $\boldsymbol{\beta}$ (MNL model) that equals:

$$\{\boldsymbol{\beta} : (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{M}(\mathbb{C}, \hat{\boldsymbol{\beta}})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \leq constant\},$$

where $\hat{\boldsymbol{\beta}}$ is the ML estimator of $\boldsymbol{\beta}$. Researchers usually employ the D_p-error:

$$D_p - error = \det\left(\mathbf{M}(\mathbb{C}, \boldsymbol{\beta})\right)^{-\frac{1}{p}}$$

as a one dimensional measure of the efficiency of a design. Here, p is the dimensionality of the parameter vector and the exponent serves to "adjust" the information for the dimensionality of the parameter vector. The power $\frac{1}{p}$ normalizes the determinate of the information matrix, making it proportional to the number of respondents. But, the A_b-optimal design minimizes:

$$A_b = \int_{\Re^p} tr(\mathbf{M}^{-1}(\xi, \boldsymbol{\beta}))\pi(\boldsymbol{\beta}) d\boldsymbol{\beta}. \tag{3.12}$$

The G- and V-optimality criteria are also important, since predicting consumer responses is the goal of choice experiments. In this situation, the G-optimal design minimizes the maximum prediction variance over the design region $\mathcal{X}$, while a V-optimal design minimizes the average prediction variance over this region.

Adopting a *Bayesian* approach to design construction, it is used the prior distribution of the *logit* coefficients $\pi(\boldsymbol{\beta})$ thus obtained to reflect subjective beliefs in the probabilities that particular parameter values occur. For example, the optimal design is the one that minimizes the D_b criterion is, the expectation of the D_p-error over the prior distribution of the parameter values (3.11).

We note that criterion (3.11) is necessarily approximate, as it is based on an asymptotic approximation to the posterior distribution. The expected information is approximated by drawing R times from $\pi(\boldsymbol{\beta})$, and computing:

$$\tilde{D}_b(\xi) \simeq \frac{1}{R} \sum_{r=1}^{R} \det\left(\mathbf{M}(\xi, \boldsymbol{\beta}_r)\right)^{-\frac{1}{p}}. \tag{3.13}$$

The main difficulty in the construction of a proper choice design is that the probabilistic choice models are non-linear in the parameters, implying that the efficiency of the design depends on the unknown parameter vector.

Consequently, researchers need to assume values for the parameters before deriving the experimental design. To circumvent this circular problem, three approaches have been introduced that, *(Kessels, et al. 2006b)* are discussed them for logit choice model, the best known of which is the MNL model. *The first* approach is to use zero prior parameter values so that methods of linear experimental design can be applied. It is implicitly assumed that the respondents prefer all attribute levels and, thus, all alternatives equally *(Großmann, et al. 2002). The second* approach, attributed to the work of *(Huber and Zwerina 1996)* advocates the use of nonzero prior values. The resulting locally D_b-optimal designs prove to be more efficient than the D-optimal designs based on zero prior values to generate the D-optimal designs. *Finally*, the most recent approach has been introduced by *(Sandor and Wedel 2001)* and consists of integrating the associated uncertainty on the assumed parameter values by the use of *Bayesian* design techniques if there is substantial uncertainty about the unknown parameters, the so-called *Bayesian* D_b-optimal designs outperform the locally D_b-optimal designs.

(Kessels, et al. 2006b) discussed the D_b-, A_b-, G_b- and V_b-optimality criteria for the multinomial logit model. After that, they described the approach to generate the optimal designs with the modified Fedorov algorithm, because it is faster than the adjusted RSC algorithm *(Huber and Zwerina 1996)* in generating *Bayesian* optimal designs. They have also constructed D_b-, A_b-, G_b- and V_b- optimal designs of two class with the *Bayesian* Modified *Fedorov* choice algorithm.

According to above explanations, it has been told that conjoint choice experiments are widely used in marketing to measure how the attributes of a product or service jointly affect consumer preferences. In a choice experiment, a product or service is characterized by a combination of attribute levels called a profile or an alternative. Respondents then choose one from a group of profiles called a choice set. They respect this task for several other choice sets presented to them. All submitted choice sets make up the experimental design. Designing an efficient choice experiment involves selecting those choice sets that result in an accurately estimated model providing precise predictions. *(Kessels, et al. 2006b)* compared four different design criteria based on the multinomial logit model to reach this goal. Because, the MNL model is nonlinear in the parameters, the computation of the optimality criteria depends on the unknown parameter vector. To solve this problem, *(Kessels, et al. 2006a)* adopted a Bayesian design procedure as proposed by *(Sandor and Wedel 2001)*. They approximated the design criteria using a Monte Carlo sample (Monte Carlo sampling involves taking a large number of random draws from a probability distribution as a surrogate for that distribution) from a multivariate normal prior parameter distribution.

The most serious criticism of optimal design is typically, that with nonlinear models the researcher must know the parameter values before deriving his or her design.

Optimal design addresses only the statistical aspects of the experimental design problem. Yet, the placement of attributes and combinations of attributes that are presented in choice sets can affect respondent behavior in ways that are not necessarily addressed in standard choice models. For example, when alternatives within a choice set offer similar utility levels but contain large attribute differences, respondents can have a hard time distinguishing among them and identifying their most preferred. This might lead to *heteroskedasticity* among responses. Because of, perhaps of most concern is feature of *logit* models that utility is a function only of attribute level differences, so that the change in utility for a price change from 5 (Euro) to 15 (Euro), for example, is equivalent to price change from 50 (Euro) to 60 (Euro).

More efficient designs enable a reduction in the number of equations asked from a respondent as well as a reduction in the number of respondents. We are interested in generating designs for conjoint choice experiments. Complications in the construction of the these designs arise from the analysis of the data from conjoint choice experiments with the multinomial logit model (MNL) contrary to experimental design methods for linear regression, for the MNL the consternation of an efficient experimental design requires knowledge of the values of the parameters. This is so because the information on the parameters provided by the design is dependent on the value of those parameters.

Unfortunately, the parameter values are unknown at the time the design is constructed, and researchers need to assume values to enable a design to be generated. Often, researchers construct designs by assuming that the parameters are zero. This construction can be motivated by the argument that the design achieves optimality under the null hypothesis of no effect of the attribute level in question.

(Huber and Zwerina 1996) have provided a first and important effort to construct designs with improved efficiency when the parameters are assume to be nonzero. They have argued that in practice, conjoint questionnaires are often pretested on small samples, the results of which may provide reasonable priors for the construction of the design. Researchers must obtain designs that take the uncertainty about the assumed parameter values into account.

3.7.1 Optimal Design in MNL Model

Suppose that there is a population including $\mathcal{J}$ possible alternatives. In Chapter 2 the choice probability related to choosing alternative j by individual i, which has the highest utility have been introduced. Based on the utility function $U_{js} = v_{js} + \varepsilon_{js}$ (for Choice set $\mathbb{C}_s$, which includes $J_s = J; \forall s \in \mathcal{S}$) that its error terms have *i.i.d* extreme value distribution (type II), the model (2.6) was called the standard MNL model.

Theorem 3.6. *According to the model* (2.6) *and the likelihood function* (2.13)*, the information matrix for the choice set* $\mathbb{C}_s$ *is calculated as follows:*

$$\mathbf{I}(\mathbb{C}_s, \boldsymbol{\beta}) = \mathbf{F}_s^T(\mathbf{P}_s - \mathbf{p}_s\mathbf{p}_s^T)\mathbf{F}_s. \tag{3.14}$$

Proof:

$$\begin{aligned}
\mathbf{I}(\mathbb{C}_s, \boldsymbol{\beta}) &= -E\left(\frac{\partial^2 \ell\left(\mathbb{C}_s, \boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right) = -\sum_{j=1}^{J} p_{js} \cdot \frac{\partial^2 \ln(p_{js})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \\
&= \sum_{j=1}^{J}\left[p_{js} \cdot \left(\sum_{j=1}^{J} \mathbf{f}(a_{js}) p_{js} \mathbf{f}^T(a_{js})\right) - p_{js} . \left(\sum_{j=1}^{J} \mathbf{f}(a_{js}) p_{js}\right)\left(\sum_{j=1}^{J} p_{js} \mathbf{f}^T(a_{js})\right)\right] \\
&= \left[\left(\sum_{j=1}^{J} \mathbf{f}(a_{js}) p_{js} \mathbf{f}^T(a_{js})\right) - \left(\sum_{j=1}^{J} \mathbf{f}(a_{js}) p_{js}\right)\left(\sum_{j=1}^{J} p_{js} \mathbf{f}^T(a_{js})\right)\right] \sum_{j=1}^{J} p_{js} \\
&= \mathbf{F}_s^T \mathbf{P}_s \mathbf{F}_s - \mathbf{F}^T \mathbf{p}_s \mathbf{p}_s^T \mathbf{F} = \mathbf{F}_s^T (\mathbf{P}_s - \mathbf{p}_s \mathbf{p}_s^T) \mathbf{F}_s,
\end{aligned}$$

where $E(Y_{js}) = p_{js}$, $\sum_{j=1}^{J} p_{js} = 1; \forall s \in \mathcal{S}$. Let us, $\sum_{l=1}^{L_k} \beta_{k,l} = 0$ thus $\beta_{k,L_k} = -\sum_{l=1}^{L_k-1} \beta_{k,l}$. In this situation, we can introduce the following elements of the information matrix:

- $\mathbf{F}_s = (\mathbf{f}(a_{1s}), \ldots, \mathbf{f}(a_{js}), \ldots, \mathbf{f}(a_{Js}))^T$; $\mathbf{f}(a_{js}) = (\mathbf{f}_1(a_{js}), \ldots, \mathbf{f}_k(a_{js}), \ldots, \mathbf{f}_K(a_{js}))^T$;

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

- $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k, \ldots, \boldsymbol{\beta}_K)^T$; $\boldsymbol{\beta}_k = (\beta_{k,1}, \ldots, \beta_{k,l}, \ldots, \beta_{k,L_k-1})^T$,
- $\mathbf{p}_s = (p_{1s}, \ldots, p_{js}, \ldots, p_{Js})^T$, $\mathbf{P}_s = diag(p_{1s}, \ldots, p_{js}, \ldots, p_{Js})$,
- $p_{js} = \frac{\exp(\mathbf{f}^T(a_{js})\boldsymbol{\beta})}{\sum_{l=1}^{J} \exp(\mathbf{f}^T(a_{ls})\boldsymbol{\beta})}; \forall j = 1, 2, \ldots, J.$

According to the above descriptions, let us consider $\beta_{1,1} = \beta_1, \ldots, \beta_{K,L_K-1} = \beta_p$ and $f_{11}(a_{js}) = f_1(a_{js}), \ldots, f_{KL_K-1}(a_{js}) = f_p(a_{js})$ then we will have:

- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_h, \ldots, \beta_p)^T$,
- $\mathbf{f}(a_{js}) = (f_1(a_{js}), \ldots, f_h(a_{js}), \ldots, f_p(a_{js}))^T$.

Corollary 3.1. *Based on above descriptions, the elements of the information matrix* (3.14) *are calculated as follow:*

$$\mathbf{I}(\mathbb{C}_s, \boldsymbol{\beta}) = \begin{pmatrix}
I_{\beta_1\beta_1} & \cdots & I_{\beta_1\beta_h} & \cdots & I_{\beta_1\beta_{h'}} & \cdots & I_{\beta_1\beta_p} \\
\vdots & \ddots & \vdots & \cdots & \vdots & \cdots & \vdots \\
I_{\beta_h\beta_1} & \cdots & I_{\beta_h\beta_h} & \cdots & I_{\beta_h\beta_{h'}} & \cdots & I_{\beta_h\beta_p} \\
\vdots & \cdots & \vdots & \ddots & \vdots & \cdots & \vdots \\
I_{\beta_{h'}\beta_1} & \cdots & I_{\beta_{h'}\beta_h} & \cdots & I_{\beta_{h'}\beta_{h'}} & \cdots & I_{\beta_{h'}\beta_h} \\
\vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
I_{\beta_p\beta_1} & \cdots & I_{\beta_p\beta_h} & \cdots & I_{\beta_p\beta_{h'}} & \cdots & I_{\beta_p\beta_p}
\end{pmatrix},$$

where $-E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\beta})}{\partial\beta_h\partial\beta_{h'}}\right) = I_{\beta_h\beta_{h'}}$ *and:*

$$I_{\beta_h\beta_{h'}} = \left(\sum_{j=1}^{J} f_h(a_{js})p_{js}f_{h'}(a_{js})\right) - \left(\sum_{j=1}^{J} f_h(a_{js})p_{js}\right)\left(\sum_{j=1}^{J} f_{h'}(a_{js})p_{js}\right).$$

Now, to introduce a design the experiments $\mathcal{J}/J/S$ are considered. This means that J alternatives from a population with $\mathcal{J}$ possible alternatives will be selected, where S ($p \leq S \leq \mathcal{S}$) denotes the number of choice sets, which include $J_s = J; \forall s$ alternatives, each. Based on experiments $\mathcal{J}/J/S$ (selecting J alternatives from population with $\mathcal{J}$ possible alternatives), the following design will be defined:

$$\xi = \left\{ \begin{array}{cccc} \mathbb{C}_1 & \mathbb{C}_2 & \cdots & \mathbb{C}_S \\ w_1 & w_2 & \cdots & w_S \end{array} \right\} \in \Xi, \quad S \geq p, \tag{3.15}$$

where $\Xi = \left\{\xi | 0 \leq w_s \leq 1; \sum_{s=1}^{S} w_s = 1 \text{ and } \mathbb{C}_s \in \mathcal{X}\right\}$, $\mathcal{X}$ denotes the design space. In this case, the information matrix of design (3.15) is obtained by:

$$\mathbf{M}(\xi,\boldsymbol{\beta}) = \sum_{s=1}^{S} w_s \cdot \mathbf{M}(\mathbb{C}_s,\boldsymbol{\beta}), \tag{3.16}$$

where w_s is the weight (frequency) of the choice set $\mathbb{C}_s$, respectively. Moreover, $\mathbf{M}(\mathbb{C}_s,\boldsymbol{\beta})$ denotes the information matrix of the choice set $\mathbb{C}_s$ and the local D-criterion at $\boldsymbol{\beta}$ is denoted by:

$$\Psi(\xi,\boldsymbol{\beta}) = \det(\mathbf{M}(\xi,\boldsymbol{\beta})). \tag{3.17}$$

Based on (3.17), the ξ^* which maximizes the $\Psi(\xi,\boldsymbol{\beta})$ criterion is called locally D-optimal design, where $\xi^* = \arg\max_{\xi\in\Xi} \Psi(\xi,\boldsymbol{\beta}_0)$ (true value of parameter), and

$$\xi^* = \left\{ \begin{array}{cccc} \mathbb{C}_1^* & \mathbb{C}_2^* & \cdots & \mathbb{C}_S^* \\ w_1^* & w_2^* & \cdots & w_S^* \end{array} \right\}. \tag{3.18}$$

The MNL model is a model of the non-linear models family. In this kind of models, there is any boundary for support points (to obtain optimal design), opposite linear models which their support points must be in interval $[p, \frac{p(p+1)}{2}]$ (See Caratheodory's Theorem, *(Silvey 1980)*, where p denotes the number of the part-worth parameters). In design (3.15) choice sets have the role of support points. In this case, to avoid singularity for the information matrix of design (3.15), consider $S \geq p$ (S denotes the number of choice set which has been selected from population to fit model). Based on the possible alternatives in population, $\mathcal{J}$, and the random sample alternatives, J, which will be selected from population, thus:

$$\mathcal{S} = \binom{\mathcal{J}}{J}$$

is the total number of choice sets (each with $J_s = J; \forall s$ alternatives) that can be considered to define a design. In this chapter, we consider $p \leq S \leq \mathcal{S}$. In this case, it must be considered S choice sets to analyze data. Moreover, based on the total number of choice sets (each with J alternatives), we will face to:

$$N_S = \begin{pmatrix} \mathcal{S} \\ S \end{pmatrix},$$

where N_S denotes the number of classes, which make the designs each with S choice sets. In this case, we consider $S = p, p+1, \ldots, \mathcal{S}$, for example, when $S = p$ we will face to $\begin{pmatrix} \mathcal{S} \\ p \end{pmatrix}$ designs each with p choice sets (support points) and so on. Corresponding to the information matrix (it depends on unknown parameters) two cases may be occurred: **Firstly**, obtaining D-optimal design for different designs based on the same parameter space. In this case (the same parameter space) we calculate locally D-optimal design (weight) for each design (the number of support points are equal) as follow:

$$\Psi(\xi_n, \boldsymbol{\beta}) = \det(\mathbf{M}(\xi_n, \boldsymbol{\beta})),$$

where:

$$\xi_n = \left\{ \begin{array}{cccc} \mathbb{C}_{n1} & \mathbb{C}_{n2} & \cdots & \mathbb{C}_{nS} \\ w_{n1} & w_{n2} & \cdots & w_{nS} \end{array} \right\} \in \Xi_n; n = 1, 2, \ldots, N_S, \tag{3.19}$$

where $\Xi = \bigcup_{n=1}^{N_S} \Xi_n$.

Now, if $\exists n' \in N_S;\ \Psi(\xi_{n'}, \boldsymbol{\beta}) \geq \Psi(\xi_n, \boldsymbol{\beta})$ thus $\xi_{n'}$ is the most suitable design with S support points (choice sets) to fit model. But to compare two designs with different support points relative efficiency can be used as a measure *(Tekle, et al. 2008)*. In this case, suppose that ξ_S and ξ_p denote two designs with S ($S \geq p$) and p support points, respectively. Then the relative efficiency (RE) of ξ_S compared to ξ_p is given by:

$$RE(\xi_S, \xi_p, \boldsymbol{\beta}) = \frac{p}{S} \left(\frac{\det(\mathbf{M}(\xi_S, \boldsymbol{\beta}))}{\det(\mathbf{M}(\xi_p, \boldsymbol{\beta}))} \right)^{\frac{1}{p}}.$$

The concept of the RE(ξ_S, ξ_p) is equal to the relative combination of alternatives in extra choice sets that must be taken under ξ_S to obtain the same efficiency as under the ξ_p. Here, ξ_p (a design with the number of support points equal to the number of parameters) is arbitrarily selected as a reference design to compare with the other designs. In the optimal design subject, a design with the number of support point equal to the number of parameters is called a saturated design, however, the saturated design may not always be the optimal design. **Secondly**, in this case we do not able to compare two designs with together, because the optimal design will be obtained based on the partitioned space of parameters. This means that the optimal design for each design (with equality support points or not equality) can be obtained based on a special part of parameters space (Example 3.2), for example, when a design with two support points in a part of parameters space can be optimal, the other designs in the same region of parameters space can not be optimal.

Table 3.1: MNL Model: Discrete choice experiment with four choice sets, $\mathbb{C}_1, \ldots, \mathbb{C}_4$ each with three alternatives ($J_s = 3; \forall s \in \mathcal{S}$), there are two attributes each with two levels; a_{js} denotes the j^{th} alternative of the choice set $\mathbb{C}_s$

Choice set	Alternatives(a_{js})	Attribute(I) ($f_1(a_{js})$)	Attribute(II) ($f_2(a_{js})$)
	$a_{14} = \tilde{a}_1$	$+1$	$+1$
$\mathbb{C}_4$	$a_{24} = \tilde{a}_2$	$+1$	-1
	$a_{34} = \tilde{a}_3$	-1	$+1$
	$a_{12} = \tilde{a}_1$	$+1$	$+1$
$\mathbb{C}_2$	$a_{22} = \tilde{a}_2$	$+1$	-1
	$a_{32} = \tilde{a}_4$	-1	-1
	$a_{13} = \tilde{a}_1$	$+1$	$+1$
$\mathbb{C}_3$	$a_{23} = \tilde{a}_3$	-1	$+1$
	$a_{33} = \tilde{a}_4$	-1	-1
	$a_{11} = \tilde{a}_2$	$+1$	-1
$\mathbb{C}_1$	$a_{21} = \tilde{a}_3$	-1	$+1$
	$a_{31} = \tilde{a}_4$	-1	-1

Example 3.2. *Consider two attributes both of them with two levels, where* $\mathcal{J} = \prod_{k=1}^{2} = L_1 \times L_2 = 2 \times 2$*;* $\mathbb{C} = \{\tilde{a}_1, \tilde{a}_2, \tilde{a}_3, \tilde{a}_4\}$*. In this situation, the experiments* $2 \times 2/3/S$ *has been considered, this means that there are* S *choice sets each with three alternatives (Table 3.1), where* $2 \leq S \leq 4$*. Therefore, to obtain local D-optimal design with different support points the following general design has been considered:*

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

The information matrix of above design is calculated by: $\mathbf{M}(\xi, \boldsymbol{\beta}) = \sum_{s=1}^{4} w_s \cdot \mathbf{M}(\mathbb{C}_s, \boldsymbol{\beta})$, where $\mathbb{C}_s; s = 1, 2, 3, 4$ denote the choice sets with three alternatives. With regards to the definition (3.14) to obtain the information matrix of the choice set $\mathbb{C}_s$, will be (Table 3.1):

$$\mathbf{M}(\mathbb{C}_1, \boldsymbol{\beta}) = \begin{bmatrix} 4(\gamma_{11} + \gamma_{21}) & -4\gamma_{11} \\ -4\gamma_{11} & 4(\gamma_{11} + \gamma_{31}) \end{bmatrix}, \quad \mathbf{M}(\mathbb{C}_2, \boldsymbol{\beta}) = \begin{bmatrix} 4(\gamma_{22} + \gamma_{32}) & 4\gamma_{22} \\ 4\gamma_{22} & 4(\gamma_{12} + \gamma_{22}) \end{bmatrix}$$

$$\mathbf{M}(\mathbb{C}_3, \boldsymbol{\beta}) = \begin{bmatrix} 4(\gamma_{13} + \gamma_{23}) & 4\gamma_{23} \\ 4\gamma_{23} & 4(\gamma_{23} + \gamma_{33}) \end{bmatrix}, \quad \mathbf{M}(\mathbb{C}_4, \boldsymbol{\beta}) = \begin{bmatrix} 4(\gamma_{24} + \gamma_{34}) & -4\gamma_{34} \\ -4\gamma_{34} & 4(\gamma_{14} + \gamma_{34}) \end{bmatrix},$$

where:

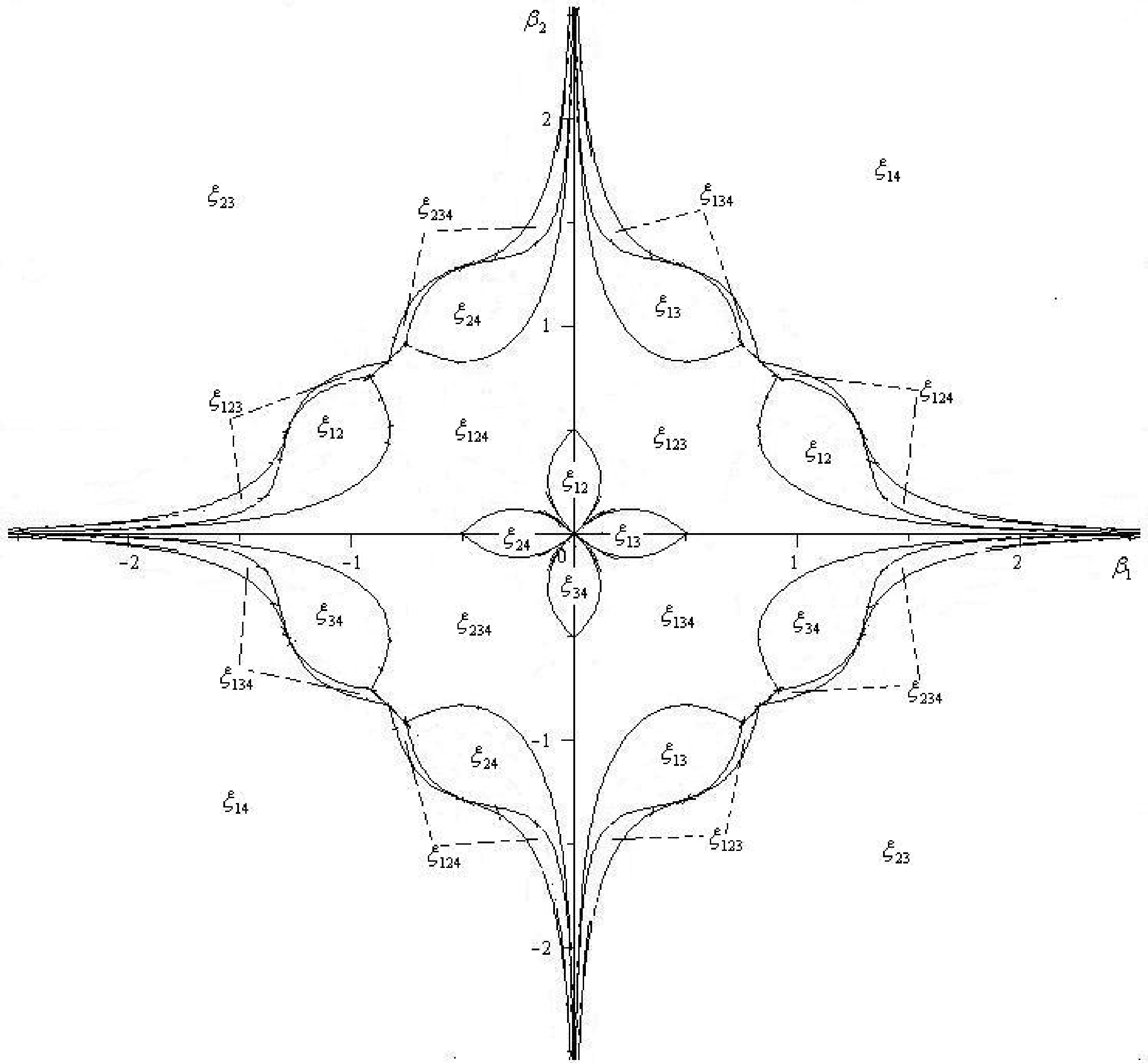

Figure 3.3: MNL Model: Partition of the (β_1, β_2)-plane into regions, where different type designs are optimal

- $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)^T$; $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$, where (w.r.t effect type codding):
  - $\boldsymbol{\beta}_1 = (\beta_{1,1}, \beta_{1,2})^T; \beta_{1,2} = -\beta_{1,1}, \quad \beta_{1,1} = \beta_1,$
  - $\boldsymbol{\beta}_2 = (\beta_{2,1}, \beta_{2,2})^T; \beta_{2,2} = -\beta_{2,1}, \quad \beta_{2,1} = \beta_2,$
- $\gamma_{1s} = p_{1s} \cdot p_{2s}, \quad \gamma_{2s} = p_{1s} \cdot p_{3s}$ and $\gamma_{3s} = p_{2s} \cdot p_{3s}$;
  - $p_{js} = \dfrac{\exp(\mathbf{f}^T(a_{js})\boldsymbol{\beta})}{\sum_{j'=1}^{3} \exp(\mathbf{f}^T(a_{j's})\boldsymbol{\beta})}$

and the choice sets $\mathbb{C}_s; s = 1, 2, 3, 4$ include the design matrix corresponding to each choice set and its alternatives:

$$\mathbf{F}_s^T = (\mathbf{f}(a_{1s}), \mathbf{f}(a_{2s}), \mathbf{f}(a_{3s})); \quad \mathbf{f}(a_{js}) = (f_1(a_{js}), f_2(a_{js}))^T; j = 1, 2, 3.$$

For every design ξ, the determinant of $\mathbf{M}(\xi, \boldsymbol{\beta})$ becomes:

$\det(\mathbf{M}(\xi, \boldsymbol{\beta})) = \sum_{m=1}^{3} \sum_{n=m+1}^{4} b_{mn} w_m w_n + \sum_{m=1}^{4} b_{mm} w_m^2,$

where:

$b_{11} = 16(\gamma_{11}(\gamma_{21} + \gamma_{31}) + \gamma_{21}\gamma_{31}),$

$b_{22} = 16(\gamma_{12}(\gamma_{22} + \gamma_{32}) + \gamma_{22}\gamma_{32}),$

$b_{33} = 16(\gamma_{13}(\gamma_{23} + \gamma_{33}) + \gamma_{23}\gamma_{33}),$

$b_{44} = 16(\gamma_{14}(\gamma_{24} + \gamma_{34}) + \gamma_{24}\gamma_{34}),$

$b_{12} = 16[(\gamma_{11} + \gamma_{21})(\gamma_{12} + \gamma_{22}) + (\gamma_{11} + \gamma_{31})(\gamma_{22} + \gamma_{32}) + 2\gamma_{11}\gamma_{22}],$

$b_{13} = 16[(\gamma_{11} + \gamma_{31})(\gamma_{13} + \gamma_{23}) + (\gamma_{11} + \gamma_{21})(\gamma_{23} + \gamma_{33}) + 2\gamma_{11}\gamma_{23}],$

$b_{14} = 16[\gamma_{11}(\gamma_{14} + \gamma_{24}) + \gamma_{21}(\gamma_{14} + \gamma_{34}) + \gamma_{31}(\gamma_{24} + \gamma_{34})],$

$b_{23} = 16[\gamma_{12}(\gamma_{13} + \gamma_{23}) + \gamma_{22}(\gamma_{13} + \gamma_{33}) + \gamma_{32}(\gamma_{23} + \gamma_{33})],$

$b_{24} = 16[(\gamma_{12} + \gamma_{22})(\gamma_{24} + \gamma_{34}) + (\gamma_{22} + \gamma_{32})(\gamma_{14} + \gamma_{34}) + 2\gamma_{22}\gamma_{34}],$

$b_{34} = 16[(\gamma_{13} + \gamma_{23})(\gamma_{14} + \gamma_{34}) + (\gamma_{23} + \gamma_{33})(\gamma_{24} + \gamma_{34}) + 2\gamma_{23}\gamma_{34}].$

To obtain a local D-optimality design for $\boldsymbol{\beta}$ the maximization problem $\max \det(\mathbf{I}(\xi, \boldsymbol{\beta}))$ must be solved. In this situation, we consider the following function:

$G(\lambda_0, \lambda_s, w) = \det(\mathbf{M}(\xi, \boldsymbol{\beta})) + \sum_{s=1}^{4} \lambda_s w_s - \lambda_0(\sum_{s=1}^{4} w_s - 1),$

where $\sum_{s=1}^{4} \lambda_s w_s = 0$; $\lambda_s \geq 0$, $w_s \geq 0 \forall s = 1,2,3,4)$ and $\sum_{s=1}^{4} w_s = 1$.

Now, corresponding to the existence of multipliers $(\lambda_s,\ \lambda_0)$ and two above restrictions, we consider the following first order conditions to obtain local D-optimality design:

$$b_{12}w_2 + b_{13}w_3 + b_{14}w_4 + 2b_{11}w_1 + \lambda_1 = \lambda_0,$$

$$b_{12}w_1 + b_{23}w_3 + b_{24}w_4 + 2b_{22}w_2 + \lambda_2 = \lambda_0,$$

$$b_{13}w_1 + b_{23}w_2 + b_{34}w_4 + 2b_{33}w_3 + \lambda_3 = \lambda_0,$$

$$b_{14}w_1 + b_{24}w_2 + b_{34}w_3 + 2b_{44}w_4 + \lambda_4 = \lambda_0,$$

$$\textstyle\sum_{s=1}^{4} \lambda_s w_s = 0,$$

$$\textstyle\sum_{s=1}^{4} w_s = 1.$$

For the case that $\lambda_\ell = \lambda_m = 0$, $\lambda_n > 0$, $\lambda_r > 0$, we obtain $w_n^* = w_r^* = 0$ and:

- $w_\ell^* = \frac{b_{\ell m} - 2b_{mm}}{2(b_{\ell m} - b_{\ell\ell} - b_{mm})}$,
- $w_m^* = \frac{b_{\ell m} - 2b_{\ell\ell}}{2(b_{\ell m} - b_{\ell\ell} - b_{mm})}$,
- $\lambda_0 = \frac{b_{\ell m}^2 - 4b_{\ell\ell}b_{mm}}{2(b_{\ell m} - b_{\ell\ell} - b_{mm})}$; $b_{ij} = b_{ji}$; $i, j \in \{1,2,3,4\}$.

According to $\lambda_n > 0$, $\lambda_r > 0$ two conditions:

1. $(b_{\ell m}^2 - 4b_{\ell\ell}b_{mm}) > b_{\ell n}(b_{\ell m} - 2b_{mm}) + b_{mn}(b_{\ell m} - 2b_{\ell\ell})$,
2. $(b_{\ell m}^2 - 4b_{\ell\ell}b_{mm}) > b_{\ell r}(b_{\ell m} - 2b_{mm}) + b_{mr}(b_{\ell m} - 2b_{\ell\ell})$

must be held. By comparing two quantities in the right of two above inequality, we can consider one of them instead of both of them.

For the case that $\lambda_\ell = \lambda_m = \lambda_n = 0$, $\lambda_r > 0$ $(w_r^* = 0)$, we obtain $w_\ell^* = \frac{Q_\ell}{Q}$, $w_m^* = \frac{Q_m}{Q}$, $w_n^* = \frac{Q_n}{Q}$ and $\lambda_0 = \frac{Q_0}{Q}$. In this situation Q_0 is always positive $(Q_0 > 0)$. Thus with respect to assumptions $\lambda_0 > 0$ and $\lambda_r > 0$ the following conditions must be held:

1. $Q > 0, \quad Q_\ell > 0, \quad Q_m > 0, \quad Q_n > 0,$
2. $Q_0 - (b_{nr}Q_n + b_{mr}Q_m + b_{\ell r}Q_\ell) > 0,$

where:

$$Q = 4(b_{\ell\ell}b_{mm} + b_{\ell\ell}b_{nn} + b_{mm}b_{nn}) - 4(b_{\ell\ell}b_{mn} + b_{mm}b_{\ell n} + b_{nn}b_{\ell m}) - (b_{\ell m} - b_{\ell n} - b_{mn})^2 + 4b_{\ell n}b_{mn},$$

$$Q_0 = 8b_{\ell\ell}b_{mm}b_{nn} + 2b_{\ell m}b_{\ell n}b_{mn} - 2(b_{\ell\ell}b_{mn}^2 + b_{mm}b_{\ell n}^2 + b_{nn}b_{\ell m}^2),$$

$$Q_\ell = 4b_{mm}b_{nn} - 2(b_{mm}b_{\ell n} + b_{nn}b_{\ell m}) + b_{mn}(b_{\ell m} + b_{\ell n}) - b_{mn}^2,$$

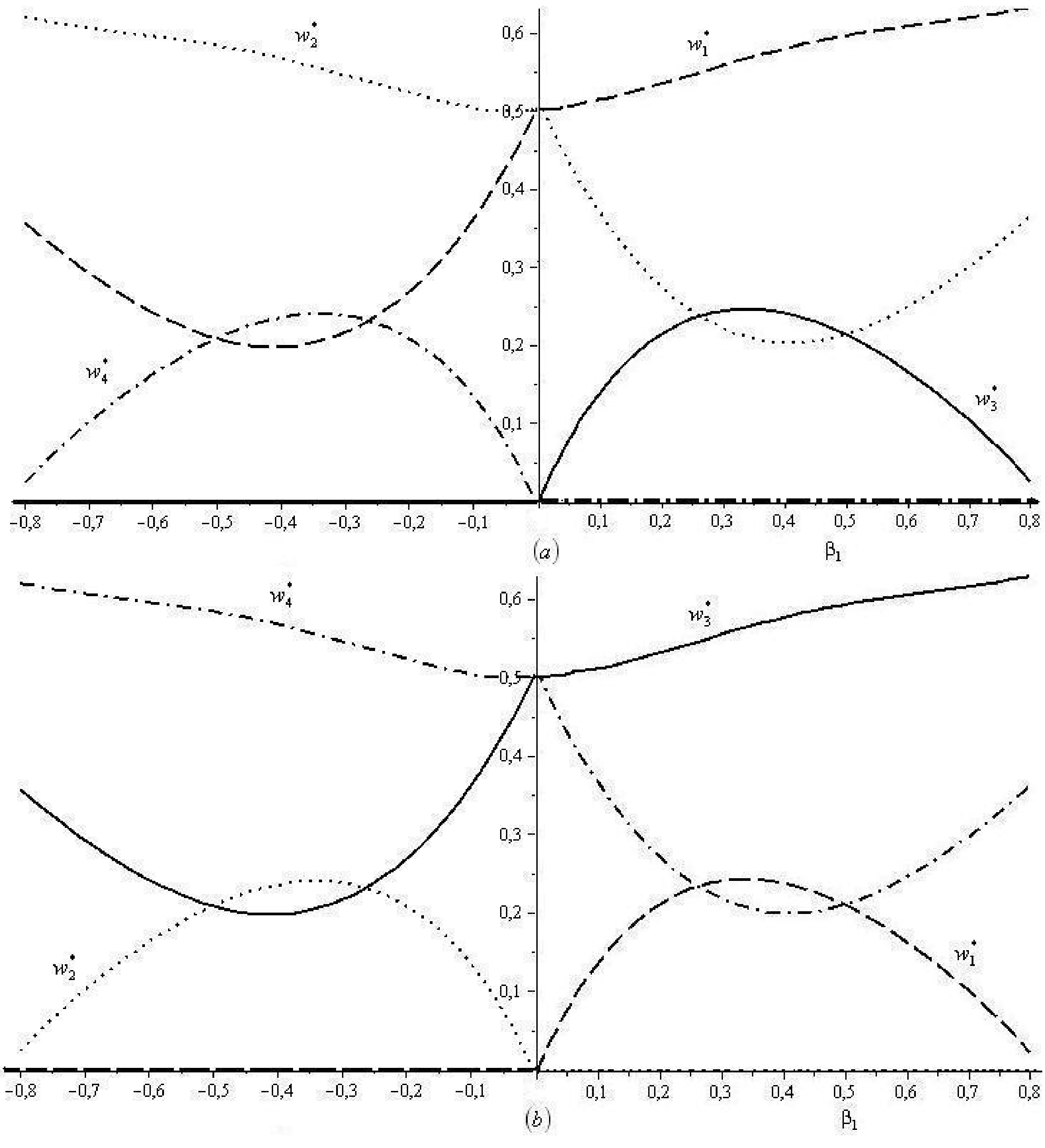

Figure 3.4: MNL Model: Locally D-optimal design; (a); $\beta_2 = 0.5$ and (b); $\beta_2 = -0.5$, based on three-point ($\xi_{\ell mr}$) optimal designs (w.r.t Example 3.2) and Figure 3.3)

$$Q_m = 4b_{\ell\ell}b_{nn} - 2(b_{\ell\ell}b_{mn} + b_{nn}b_{\ell m}) + b_{\ell n}(b_{\ell m} + b_{nm}) - b_{\ell n}^2,$$

$$Q_n = 4b_{\ell\ell}b_{mm} - 2(b_{\ell\ell}b_{nm} + b_{mm}b_{\ell n}) + b_{\ell m}(b_{\ell n} + b_{mn}) - b_{\ell m}^2.$$

Figure 3.3 denotes the contours which separate parameter regions for optimal two-and three-point designs. In this Figure 3.3, regions were marked by $\xi_{\ell m}$ denote that two-point designs are optimal. Also, for parameter setting in regions marked by $\xi_{\ell mn}$ three-point designs are optimal. It can be seen that the regions which two-point designs are locally D-optimal do not have common regions, means that, there are regions for three-points designs between them.

Specially, let $\beta_2 = 0$. In this situation we can find optimal design to estimate parameters based on four two-point designs ξ_{12}, ξ_{13}, ξ_{24} and ξ_{34} as follow:

- For $0.0 < \beta_1 \leq 0.5$:
$$w_1^* = w_3^* = 0.5, w_2^* = w_4^* = 0.0,$$

- For $-0.5 \leq \beta_1 < 0.0$:
$$w_1^* = w_3^* = 0.0, w_2^* = w_4^* = 0.5$$

- For $\beta_1 = 0.0$:
$$w_1^* = w_2^* = w_3^* = w_4^* = 0.25.$$

In particular, let $\beta_2 = 0.5$. We investigate this particulate case in order to denote locally D-optimal solutions which are dependent on parameter β_1. In this situation we consider two three-point designs ξ_{123} and ξ_{124}, where their optimal design conditions in interval $(0, 0.82)$ (w.r.t, ξ_{123}) and $(-0.82, 0)$ (w.r.t, ξ_{124}) for parameter β_1 are hold (Figure 3.3). By Figure 3.4 we have similar situation for two optimal designs. For example, w_1^* and w_2^* tend to $\frac{1}{2}$ when β_1 tends to zero (both from right and left) while w_3^* and w_4^* tend to zero (Figure 3.4 (a)). This means that the optimal design ignores choice set $\mathbb{C}_3$ for low values (close to zero from right) of β_1 (with respect to design ξ_{123}) and don't use of choice set $\mathbb{C}_4$ to fit model (with respect to design ξ_{124}) when β_1 tends to zero from left. By Figure 3.4(a) has been showed that the optimal weight for choice set $\mathbb{C}_1$ increases as β_1 increases and it is always greater than the two others. In interval $(0, 0.25)$, w_2^* decreases and w_3^* increases as β_1 increases, but, the optimal weight for choice set $\mathbb{C}_2$ is greater than the optimal wight of choice set $\mathbb{C}_3$. Also, it is seen that at two points $\beta_1 = 0.24, 0.5$ the two optimal weights w_2^* and w_3^* are equal and for $\beta_1 > 0.5$, w_2^* is greater than the optimal weight for choice set $\mathbb{C}_3$. Based on design ξ_{124}, there exist similar description about the optimal weights for choice sets $\mathbb{C}_1$, $\mathbb{C}_2$ and $\mathbb{C}_4$ (Figure 3.4 (a)).

To discuss about $\beta_2 = -0.5$ and $\beta_1 \in (-0.82, 0.82)$ we consider two three-point designs ξ_{134} and ξ_{234} to obtain the optimal design (Figures 3.4(b)).

In the other hand, let $\beta_2 = 0$ again. In this case, the equations $p_{12} = p_{14} = p_{22} = p_{24}$, $p_{11} = p_{13}$, $p_{21} = p_{23}$, $p_{11} = 1 - 2p_{21}$ will be held, where p_{js} is the choice probability related to choosing alternative j from choice set s, so that $\gamma_{12} = \gamma_{14}$, $\gamma_{22} = \gamma_{32} = \gamma_{24} = \gamma_{34}$ and

$\gamma_{11} = \gamma_{13}, \gamma_{21} = \gamma_{23}, \gamma_{31} = \gamma_{33}$. Due to symmetry considerations we can derive an optimal solution with wights $w_1 = w_3$ and $w_2 = w_4$, where $w_2 = \frac{1}{2} - w_1$, thus:

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

In this situation, the determinate of the the information matrix of the above design is calculated by:

$$\det(\mathbf{M}(\xi, \boldsymbol{\beta})) = 16\,(-w_1 \cdot p_{11} + 2 \cdot p_{12} \cdot w_1 + w_1 \cdot {p_{11}}^2 - 4 \cdot w_1 \cdot {p_{12}}^2 - p_{12} + 2 \cdot {p_{12}}^2) \times$$
$$(-w_1 + 4 \cdot p_{12} \cdot w_1 + w_1 \cdot {p_{11}}^2 - 4 \cdot w_1 \cdot {p_{12}}^2 - 2 \cdot p_{12} + 2 \cdot {p_{12}}^2)\,,$$

where:

$$w_1^* = \frac{1}{2} \cdot \frac{(-4 \cdot p_{12} \cdot p_{11} + 3 \cdot p_{11} - 8 \cdot p_{12} + 1 + 8 \cdot {p_{12}}^2)\, p_{12}}{{p_{11}}^3 - 2 \cdot p_{12} \cdot {p_{11}}^2 + 4 \cdot p_{12} \cdot p_{11} - p_{11} - 4 \cdot {p_{12}}^2 \cdot p_{11} - 8 \cdot {p_{12}}^2 + 2 \cdot p_{12} + 8 \cdot {p_{12}}^3}$$

maximizes $\Psi(\xi, \boldsymbol{\beta}) = \det(\mathbf{M}(\xi, \boldsymbol{\beta}))$. In this result, w_1^* can be written according to $p_{11} = \frac{\exp(\beta_1)}{2 \cdot \exp(-\beta_1) + \exp(\beta_1)}$ and $p_{12} = \frac{\exp(\beta_1)}{\exp(-\beta_1) + 2 \cdot \exp(\beta_1)}$ as follow:

$$w_1^* = \frac{1}{12} \cdot \frac{(2 \cdot \exp(2 \cdot \beta_1) + 2 - \exp(-2 \cdot \beta_1))\,(2 \cdot \exp(-\beta_1) + \exp(\beta_1))^2}{\exp(4 \cdot \beta_1) + \exp(2 \cdot \beta_1) - \exp(-4 \cdot \beta_1) - \exp(-2 \cdot \beta_1)}.$$

Hence, when $\beta_1 \in (-\infty, -0.503] \bigcup [0.503, +\infty)$ the condition $0 \leq w_1^* \leq \frac{1}{2}$ holds. In this situation, w_1^* decreases as β_1 increases in each of these segments. It has been seen that $w_1^* = 0$ for $\beta_1 \in (-0.503, 0)$ and $w_1^* = \frac{1}{2}$ for $\beta_1 \in (0, 0.503)$.

In Example 3.2, we introduced a MNL model with two attributes each with two levels. To estimate parameters four choice sets has been considered, which each include three alternatives. In the other case, we can consider six choice sets each with two alternatives to fit model (See Table 3.2).

G-Optimality Criterion:

Based on Theorem 3.2 the G-optimality criterion for Example 3.2 can be defined as follow:

$$tr(\mathbf{M}(\mathbb{C}_s, \boldsymbol{\beta})\mathbf{M}^{-1}(\xi_{\ell m}^*, \boldsymbol{\beta})), \qquad tr(\mathbf{M}(\mathbb{C}_s, \boldsymbol{\beta})\mathbf{M}^{-1}(\xi_{\ell mn}^*, \boldsymbol{\beta})).$$

The above criteria for two-point (denoted by $\xi_{\ell m}$) and three-point design (denoted by $\xi_{\ell mn}$) has been defined, for example, based on the two-point design ξ_{12} we will have:

$$tr(\mathbf{M}(\mathbb{C}_s, \boldsymbol{\beta})\mathbf{M}^{-1}(\xi_{12}^*, \boldsymbol{\beta})) = 2;\ s = 1, 2.$$

Thus it can be told, $\mathbb{C}_1$ and $\mathbb{C}_2$ in design ξ_{12} are support points. Also, for three-point design ξ_{123} it has been calculated:

$$tr(\mathbf{M}(\mathbb{C}_\ell, \boldsymbol{\beta})\mathbf{M}^{-1}(\xi_{123}^*, \boldsymbol{\beta})) = 2; \ell = 1, 2$$

and with this order we can calculate the G-optimality criterion for the other choice sets and the other designs, also.

Table 3.2: MNL model with two attributes each with two levels: Locally D-optimal design based on experiments $2 \times 2/3/4$ ($\Psi_1(\xi_1, \boldsymbol{\beta})$, Four choice sets each with three alternatives) and $2 \times 2/2/6$ ($\Psi_2(\xi_2, \boldsymbol{\beta})$, Six choice stes each with two alternatives) and locally D-optimal criterion $\Psi_r(\xi_r, \boldsymbol{\beta}) = (\det \mathbf{M}(\xi_r, \boldsymbol{\beta}))^{-1}$; r=1,2

β_1	β_2	w_1^*	w_2^*	w_3^*	w_4^*	$\Psi_1(\xi_1^*, \boldsymbol{\beta}_0)$	$w_1^{'*}$	$w_2^{'*}$	$w_3^{'*}$	$w_4^{'*}$	$w_5^{'*}$	$w_6^{'*}$	$\Psi_2(\xi_2^*, \boldsymbol{\beta}_0)$
	-1	.627	.000	.000	.373	1.939	.140	.140	.000	.440	.140	.140	2.920
	-.8	.601	.000	.000	.399	1.835	.202	.074	.000	.449	.074	.202	2.691
	-.1	.425	.102	.473	.000	1.770	.072	.000	.414	.441	.000	.072	2.358
-1	0.0	.268	.268	.232	.232	1.788	.069	.000	.431	.431	.000	.069	2.350
	.4	.000	.573	.000	.427	1.725	.131	.000	.459	.278	.000	.131	2.458
	1	.000	.627	.373	.000	1.939	.140	.140	.441	.000	.140	.140	2.920
	-1	.665	.335	.000	.000	1.790	.050	.220	.000	.460	.220	.050	2.608
	-.8	.660	.249	.091	.000	1.611	.000	.034	.442	.490	.034	.000	2.361
	-.1	.465	.232	.303	.000	1.400	.000	.000	.500	.500	.000	.000	1.585
-.7	0.0	.368	.368	.132	.132	1.410	.000	.000	.500	.500	.000	.000	1.575
	.4	.099	.579	.000	.322	1.411	.000	.000	.500	.500	.000	.000	1.744
	1	.335	.665	.000	.000	1.790	.050	.220	.460	.000	.220	.050	2.608
	-1	.268	.232	.268	.232	1.788	.000	.069	.431	.431	.069	.000	2.350
	-.8	.326	.174	.326	.174	1.519	.000	.000	.500	.500	.000	.000	1.789
	-.1	.500	.000	.500	.000	1.114	.000	.000	.500	.500	.000	.000	1.010
0.0	**0.0**	**.250**	**.250**	**.250**	**.250**	1.125	.000	.000	.500	.500	.000	.000	1.000
	.4	.000	.500	.000	.500	1.178	.000	.000	.500	.500	.000	.000	1.169
	1	.232	.268	.232	.268	1.788	.000	.069	.431	.431	.069	.000	2.350
	-1	.000	.000	.573	.427	1.725	.000	.131	.459	.278	.131	.000	2.458
	-.8	.032	.000	.587	.381	1.500	.000	.000	.500	.500	.000	.000	1.957
	-.1	.041	.000	.520	.439	1.182	.000	.000	.500	.500	.000	.000	1.179
.4	0.0	.000	.000	.500	.500	1.178	.000	.000	.500	.500	.000	.000	1.169
	.4	.000	.221	.221	.557	1.234	.000	.000	.500	.500	.000	.000	1.337
	1	.000	.000	.427	.573	1.725	.000	.131	.278	.459	.131	.000	2.458
	-1	.000	.373	.627	.000	1.939	.140	.140	.440	.000	.140	.140	2.920
	-.8	.000	.399	.601	.000	1.835	.202	.074	.450	.000	.074	.202	2.691
	-.1	.473	.000	.425	.102	1.770	.073	.000	.441	.414	.000	.073	2.358
1	0.0	.232	.232	.268	.268	1.788	.069	.000	.431	.431	.000	.069	2.350
	.4	.000	.427	.000	.573	1.725	.131	.000	.278	.459	.000	.131	2.458
	1	.373	.000	.000	.627	1.939	.140	.140	.000	.441	.140	.140	2.920

3.7.2 Maximin Efficient Designs

We know that a design, ξ, will be called locally Ψ-optimal if it maximizes $\Psi(\mathbf{M}(\xi,\boldsymbol{\beta}))$ for a given $\boldsymbol{\beta}$. *(Melas 2006)* has written that a design will be called maximin efficient Ψ-optimal if it maximizes:

$$\mathrm{eff}_{\boldsymbol{\beta}}(\xi) = \inf_{\boldsymbol{\beta}} \left(\frac{\Psi(\mathbf{M}(\xi,\boldsymbol{\beta}))}{\Psi(\mathbf{M}(\xi^*,\boldsymbol{\beta}))} \right)^{\frac{1}{p}},$$

where $\boldsymbol{\beta}$ is a given set of possible values of the vector parameter, p is the number of parameters and $\Psi(\mathbf{M}(\xi,\boldsymbol{\beta})) = \det(\mathbf{M}(\xi,\boldsymbol{\beta}))$. $\mathrm{eff}_{\boldsymbol{\beta}}(\xi)$ is the efficiency of the design ξ with respect to a locally Ψ-optimal design for a least favorable value of $\boldsymbol{\beta}$ in parameter space. In this case, maximin efficient means that how many more experiments will be needed under the design ξ with respect to an ideal design to achieve the same accuracy of estimating in the worst case. Therefore, maximin efficient designs can be considered to obtain optimal design equally well for all possible parameter values.

Now, for our Example 3.2 we can rewrite $\mathrm{eff}_{\boldsymbol{\beta}}(\xi)$ as follow,

$$\mathrm{eff}^2(\xi,\boldsymbol{\beta}) \leq \frac{\det(\mathbf{M}(\xi,\boldsymbol{\beta}))}{\det(\mathbf{M}(\xi^*,\boldsymbol{\beta}))},$$

where (Example 3.2):

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

Based on two-point optimal designs ξ^*_{12} we have:

$$\det(\mathbf{M}(\xi^*_{12},\boldsymbol{\beta})) = \frac{b_{12}(b_{12}-2b_{22})(b_{12}-2b_{11}) + b_{11}(b_{12}-2b_{22})^2 + b_{22}(b_{12}-2b_{11})^2}{4(b_{12}-b_{11}-b_{22})^2}.$$

In this case, let $\beta_2 = 0$ and β_1 tends to minus infinity, hence,

$$\inf_{\boldsymbol{\beta}} \mathrm{eff}^2(\xi;\boldsymbol{\beta}) \leq \frac{3}{4}(w_1+w_3)^2 + 3(w_1w_2 + w_1w_4 + w_2w_3 + w_3w_4). \tag{3.20}$$

Similarity we can obtain previous inequality for two-points optimal designs ξ^*_{14}, ξ^*_{23} and ξ^*_{34}, too. Thus the right sid of previous inequality is maximized for $w_1 = \frac{1}{3}$, $w_2 = \frac{1}{6}$, $w_3 = \frac{1}{3}$ and $w_4 = \frac{1}{6}$. Since the optimal weights satisfy $w^*_\ell \leq \frac{1}{dim\boldsymbol{\beta}} = \frac{1}{2}$, equality is achieved *(Kiefer and Wolfowitz 1959)* and $\inf_{\boldsymbol{\beta}} \mathrm{eff}^2(\xi;\boldsymbol{\beta}) = 1$. Then, design ξ on all four points $\mathbb{C}_1$, $\mathbb{C}_2$, $\mathbb{C}_3$ and $\mathbb{C}_4$ with wights $w_1 = \frac{1}{3}$, $w_2 = \frac{1}{6}$, $w_3 = \frac{1}{3}$ and $w_4 = \frac{1}{6}$ is maximin efficient.

Now, let $\beta_2 = 0$ but β_1 tends to plus infinity. In this case, with respect to four two-points optimal design ξ^*_{12}, ξ^*_{14}, ξ^*_{23} and ξ^*_{34} we have:

$$\inf_{\boldsymbol{\beta}} \mathrm{eff}^2(\xi,\boldsymbol{\beta}) \leq \frac{3}{4}(w_2+w_4)^2 + 3(w_1w_2 + w_1w_4 + w_2w_3 + w_3w_4). \tag{3.21}$$

In this case, the right side of previous inequality is maximized for $w_1 = \frac{1}{6}$, $w_2 = \frac{1}{3}$, $w_3 = \frac{1}{6}$ and $w_4 = \frac{1}{3}$.

Consequently, let $\beta_2 = 0$ and β_1 tends to infinity (minus and plus). Thus, according to Equations (3.20) and (3.21) we will have:

$$\inf_{\boldsymbol{\beta}} \text{eff}^2(\xi, \boldsymbol{\beta}) \leq \frac{3}{4} \min\left((w_2 + w_4)^2, (w_1 + w_3)^2\right) + 3(w_1w_2 + w_1w_4 + w_2w_3 + w_3w_4). \quad (3.22)$$

The right side of Equation (3.22) is maximized when $w_1 = w_2 = w_3 = w_4 = \frac{1}{4}$, where $\inf_{\boldsymbol{\beta}} \text{eff}^2(\xi, \boldsymbol{\beta}) = 0.9375$.

In continuation, we can obtain maximin efficient for the vice versa of previous case. Then let $\beta_1 = 0$ and β_2 tends to minus infinity. In this case and based on four two-points optimal designs ξ^*_{13}, ξ^*_{14}, ξ^*_{23} and ξ^*_{24} we will have:

$$\inf_{\boldsymbol{\beta}} \text{eff}^2(\xi, \boldsymbol{\beta}) \leq \frac{3}{4}(w_1 + w_2)^2 + 3(w_1w_3 + w_1w_4 + w_2w_3 + w_2w_4) \quad (3.23)$$

and the right hand side is maximized for $w_1 = w_2 = \frac{1}{3}$ and $w_3 = w_4 = \frac{1}{6}$, and based on these optimal weights equality is achieved, $\inf_{\boldsymbol{\beta}} \text{eff}^2(\xi, \boldsymbol{\beta}) = 1$.

Now, let $\beta_1 = 0$ and β_2 tends to plus infinity. Similarity:

$$\inf_{\boldsymbol{\beta}} \text{eff}^2(\xi, \boldsymbol{\beta}) \leq \frac{3}{4}(w_3 + w_4)^2 + 3(w_1w_3 + w_1w_4 + w_2w_3 + w_2w_4) \quad (3.24)$$

with respect to four two-points designs ξ^*_{13}, ξ^*_{14}, ξ^*_{23} and ξ^*_{24}. According to weights $w_1 = w_2 = \frac{1}{6}$ and $w_3 = w_4 = \frac{1}{3}$, the right side of previous inequality will be maximized, $\inf_{\boldsymbol{\beta}} \text{eff}^2(\xi; \boldsymbol{\beta}) = 1$.

According to Equations (3.23) and (3.24), we consider $\beta_1 = 0$ and β_2 tends to infinity. In this situation, we will have:

$$\inf_{\boldsymbol{\beta}} \text{eff}^2(\xi, \boldsymbol{\beta}) \leq \frac{3}{4} \min\left((w_3 + w_4)^2, (w_1 + w_2)^2\right) + 3(w_1w_3 + w_1w_4 + w_2w_3 + w_2w_4). \quad (3.25)$$

The right side of Equation (3.25) will be maximized when $w_1 = w_2 = w_3 = w_4 = \frac{1}{4}$, where $\inf_{\boldsymbol{\beta}} \text{eff}^2(\xi, \boldsymbol{\beta}) = 0.9375$.

In total, based on Equations (3.22) and (3.25) the following inequality can be considered:

$$\begin{aligned}\inf_{\boldsymbol{\beta}} \text{eff}^2(\xi, \boldsymbol{\beta}) \quad &\leq \quad \frac{3}{4} \min\left((w_1 + w_3)^2, (w_2 + w_4)^2, (w_1 + w_2)^2, (w_3 + w_4)^2\right) \\ &\quad +3 \min((w_1w_2 + w_1w_4 + w_2w_3 + w_3w_4), (w_1w_3 + w_1w_4 + w_2w_3 + w_2w_4)),\end{aligned}$$

The previous inequality will be changed to equality, because of $w^*_s \leq \frac{1}{2}$, means that:

$$\begin{aligned}\inf_{\boldsymbol{\beta}} \text{eff}^2(\xi, \boldsymbol{\beta}) \quad &= \quad \frac{3}{4} \min\left((w_1 + w_3)^2, (w_2 + w_4)^2, (w_1 + w_2)^2, (w_3 + w_4)^2\right) \\ &\quad +3 \min((w_1w_2 + w_1w_4 + w_2w_3 + w_3w_4), (w_1w_3 + w_1w_4 + w_2w_3 + w_2w_4)),\end{aligned}$$

for any design ξ. The uniform four point design $\bar{\xi}$ with $w^*_1 = w^*_2 = w^*_3 = w^*_4 = \frac{1}{4}$ maximizes $\inf_{\boldsymbol{\beta}} \text{eff}^2(\xi, \boldsymbol{\beta})$ with $\text{eff}(\bar{\xi}, \boldsymbol{\beta}) = 0.9682$ and is, hence, maximin efficient.

3.7.3 Invariance

Based on linear model $Y(x) = \mathbf{f}^T(x)\boldsymbol{\beta} + \varepsilon$ ($E(Y|x) = \mathbf{F}^T\boldsymbol{\beta}$), where ε are homoscedastic with $var(\varepsilon) = \sigma^2$, $E(\varepsilon) = 0$ (uncorrelated) and according to the following design:

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \cdots & x_p \\ w_1 & w_2 & \cdots & w_p \end{array} \right\}$$

the information matrix is calculated by $\mathbf{M}(\xi) = \sum_{i=1}^{p} w_i \mathbf{f}(x_i)\mathbf{f}^T(x_i)$. With regards to the definition of the local D-optimality criterion, means that $\Psi(\mathbf{M}(\xi)) = -\ln(\det(\mathbf{M}(\xi)))$, consider the following definitions *(Schwabe 1996)*;

1. i) A one-to-one mapping $g : \mathcal{X} \to \mathcal{X}$ is called a transformation of the design region $\mathcal{X}$.

   ii) A transformation g of $\mathcal{X}$ induces a linear transformation of the regression function $\mathbf{F} : \mathcal{X} \to \Re^p$ if there exists a $p \times p$-matrix $\mathbf{Q}_g$ with $\mathbf{F}(g(x)) = \mathbf{Q}_g\mathbf{F}(x); \forall x \in \mathcal{X}$ where $\mathbf{F}(x) = (\mathbf{f}(x_1), \ldots, \mathbf{f}(x_n))^T$.

2. A group G of transformations of $\mathcal{X}$ induces linear transformation of the regression function $\mathbf{F} : \mathcal{X} \to \Re^p$ if every $g \in G$ induces a linear transformation of $\mathbf{F}$.

3. A deign ξ is invariant with respect to G if $\xi^g = \xi; \forall g \in G$.

4. A design ξ is information invariant with respect to G and the linear regression function $\mathbf{F}$, if $\exists \mathbf{Q}_g \in \Re^{p\times p}; \mathbf{Q}_g\mathbf{M}(\xi)\mathbf{Q}_g^T = \mathbf{M}(\xi); \forall g \in G$

5. For every design ξ the symmetrization $\bar{\xi} = \frac{1}{|G|}\sum_{g\in G}\xi^g$ is a design (the summarized design of ξ with respect to G). Moreover, every symmetrized design $\bar{\xi}$ is invariant. Hence, a design ξ coincides with its symmetrization $\bar{\xi}$ if ξ is invariant. Also, if $\Psi : \Xi \to \Re$ is convex and invariant with respect to G, then $\Psi(\bar{\xi}) \leq \Psi(\xi); \forall \xi \in \Xi$. The concept of invariant is helpful if the transformations do not affect the value of the criterion function under consideration.

6. A criterion function $\Psi : \Xi \to \Re$ is invariant with respect to G if $\Psi(\xi^g) = \Psi(\xi); \forall \xi \in \Xi, \forall g \in G$

7. i) A group G is orthogonal for $\mathbf{F}$ if $\forall g \in G$ the transformation matrix $\mathbf{Q}_g$ is orthogonal i.e $\mathbf{Q}_g^T = \mathbf{Q}_g^{-1}$.

   ii) A group G is uni-modal for $\mathbf{F}$ if $\forall g \in G$ the transformation matrix $\mathbf{Q}_g$ is uni-modal i.e $\det(\mathbf{Q}_g) = 1$.

   iii) If G is orthogonal for $\mathbf{F}$, then G is also uni-modal.

8. If $\Xi' \subset \Xi$ is a class of invariant design, means that $\Xi' = \{\xi' | \xi' = \xi'^g, \forall g \in G\}$ and there exists an invariant design in Ξ' as ξ'^* such that $\Psi(\xi'^*) \leq \Psi(\xi'), \forall \xi' \in \Xi'$ then $\Psi(\xi'^*) \leq \Psi(\xi), \forall \xi \in \Xi$.

Now, suppose that G is a group of transformations $g : \mathcal{X} \to \mathcal{X}$ (one-to-one), then g is linear transformation with respect to $\mathbf{f}(x)$ if:

$$\exists \mathbf{Q'}_g \in \Re^{1\times p}; \forall x \in \mathcal{X}, \mathbf{f}(g(x)) = \mathbf{Q'}_g \mathbf{f}(x).$$

We define the following design (by noting to group transformation):

$$\xi^g = \left\{ \begin{array}{cccc} g(x_1) & g(x_2) & \cdots & g(x_p) \\ w_1 & w_2 & \cdots & w_p \end{array} \right\}$$

with information matrix:

$$\begin{aligned}
\mathbf{M}(\xi^g) &= \sum_{i=1}^{p} w_i \mathbf{f}(g(x_i))\mathbf{f}^T(g(x_i)) \\
&= \sum_{i=1}^{p} w_i \mathbf{Q'}_g \mathbf{f}(x_i)\mathbf{f}^T(x_i)\mathbf{Q'}^T_g \\
&= \mathbf{Q}_g \mathbf{M}(\xi)\mathbf{Q}^T_g.
\end{aligned}$$

In hence, Ψ is linear invariant for g (and $\mathbf{f}$) if *(Schwabe 1996)*

$$\Psi(\mathbf{M}(\xi^g)) = \Psi(\mathbf{M}(\xi)).$$

Invariance D-Optimal Criterion for MNL Model

Obtaining D-optimal design for estimating parameters, a random sample with J size are selected from a papulation with $\mathcal{J}$ alternatives (Subsection 3.7.1). In this situation $\binom{\mathcal{J}}{J} = \mathcal{S}$ choice sets each with J alternatives are considered. Now, based on experiments $\mathcal{J}/J/S$ (based on $p \leq S \leq \mathcal{S}$ choice sets each with J alternatives) the following design is defined (Equation (3.15)):

$$\xi = \left\{ \begin{array}{cccc} \mathbb{C}_1 & \mathbb{C}_2 & \cdots & \mathbb{C}_S \\ w_1 & w_2 & \cdots & w_S \end{array} \right\} \in \Xi,$$

where choice set $\mathbb{C}_s$ includes the characterizes of K attributes as follow (Subsection 3.7.1):

$$\mathbf{F}_s = \begin{pmatrix} \mathbf{f}_1^T(a_{1s}) & \mathbf{f}_2^T(a_{1s}) & \cdots & \mathbf{f}_K^T(a_{1s}) \\ \mathbf{f}_1^T(a_{2s}) & \mathbf{f}_2^T(a_{2s}) & \cdots & \mathbf{f}_K^T(a_{2s}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{f}_1^T(a_{Js}) & \mathbf{f}_2^T(a_{Js}) & \cdots & \mathbf{f}_K^T(a_{Js}) \end{pmatrix}, \ \mathbf{f}_k(a_{js}) = \begin{pmatrix} f_{k1}(a_{js}) \\ f_{k2}(a_{js}) \\ \vdots \\ f_{kL_k-1}(a_{js}) \end{pmatrix}, \ f_{k\ell}(a_{js}) \in \{-1, 0, 1\}.$$

As previously stated, the information matrix corresponding to the design ξ (the MNL model) has been calculated by:

$$\mathbf{M}(\xi) = \sum_{s=1}^{S} w_s \cdot \mathbf{F}_s^T \mathbf{D}_s \mathbf{F}_s,$$

where $\mathbf{D}_s = \mathbf{P}_s - \mathbf{p}_s\mathbf{p}_s^T$ (Equation (3.14)).

Suppose that G is a group of transformations, where $g : \mathcal{X} \to \mathcal{X}$ (one-to-one). According to this transformation group (G), the following design is considered:

$$\xi^g = \left\{ \begin{array}{cccc} g(\mathbf{F}_1) & g(\mathbf{F}_2) & \cdots & g(\mathbf{F}_S) \\ w_1 & w_2 & \cdots & w_S \end{array} \right\} \in \Xi^g.$$

The information matrix of design ξ^g is calculated by (Let $g(\mathbf{X}_s) = \mathbf{X}_s^g$):

$$\mathbf{M}(\xi^g) = \sum_{s=1}^{S} w_s \cdot \mathbf{F}_s^{gT}\mathbf{D}_s^g\mathbf{F}_s^g,$$

where $g_k : \mathbf{f}_k^T(a_{js}) \to g_k(\mathbf{f}_k^T(a_{js})); k = 1, 2, \ldots, K$ (g_k induced by permutation of levels of an attribute), $\mathbf{D}_s^g = \mathbf{P}_s^g - \mathbf{p}_s^g\mathbf{p}_s^{gT}$ and :

$$\mathbf{X}_s^g = g(\mathbf{X}_s) = \begin{pmatrix} g_1(\mathbf{f}_1^T(a_{1s})) & g_2(\mathbf{f}_2^T(a_{1s})) & \cdots & g_K(\mathbf{f}_K^T(a_{1s})) \\ g_1(\mathbf{f}_1^T(a_{2s})) & g_2(\mathbf{f}_2^T(a_{2s})) & \cdots & g_K(\mathbf{f}_K^T(a_{2s})) \\ \vdots & \vdots & \ddots & \vdots \\ g_1(\mathbf{fx}_1^T(a_{Js})) & g_2(\mathbf{f}_2^T(a_{Js})) & \cdots & g_K(\mathbf{f}_K^T(a_{Js})) \end{pmatrix}$$

is the design matrix of choice set $\mathbb{C}_s^g$. Now, if:

$$\exists \mathbf{Q}_k \in \Re^{(L_k-1)\times(L_k-1)}; \mathbf{Q}_{k,g_k} = \mathbf{f}_k^T(a_{js})\mathbf{Q}_k^T, \forall \mathbf{f}_k(a_{js})$$

then it can be written:

$$\mathbf{Q}^g\mathbf{F}_s^T = \begin{pmatrix} \mathbf{Q}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{Q}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Q}_K \end{pmatrix} \begin{pmatrix} \mathbf{f}_1(a_{1s}) & \mathbf{f}_1(a_{2s}) & \cdots & \mathbf{f}_1(a_{Js}) \\ \mathbf{f}_2(a_{1s}) & \mathbf{f}_2(a_{2s}) & \cdots & \mathbf{f}_2(a_{Js}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{f}_K(a_{1s}) & \mathbf{f}_K(a_{2s}) & \cdots & \mathbf{f}_K(a_{Js}) \end{pmatrix},$$

for example, $\mathbf{Q}_k$ which depends on grope G may be a permutation matrix.

Consequently, if $\exists \mathbf{Q}^g \in \Re^{p\times p}; \forall \mathbf{F}_s \in \mathcal{X}; \mathbf{F}_s^{gT} = \mathbf{Q}^g\mathbf{F}_s^T; s = 1, 2, \ldots, S$, then:

$$\begin{aligned} \mathbf{M}(\xi^g) &= \sum_{s=1}^{S} w_s \cdot \mathbf{F}_s^{gT}\mathbf{D}_s^g\mathbf{F}_s^g \\ &= \sum_{s=1}^{S} w_s \cdot \mathbf{Q}^g\mathbf{F}_s^T\mathbf{D}_s^g\mathbf{F}_s\mathbf{Q}^{Tg} \\ &= \mathbf{Q}^g \left(\sum_{s=1}^{S} w_s \cdot \mathbf{F}_s^T\mathbf{D}_s^g\mathbf{F}_s \right) \mathbf{Q}^{Tg}. \end{aligned}$$

Now, if $\boldsymbol{\beta} = \mathbf{0}$ then:

$$\begin{aligned}\mathbf{D}_s^g &= \mathbf{D}_s \\ &= \left(\frac{1}{J} \cdot \mathbf{I}_J - \frac{1}{J^2} \cdot \mathbf{1}_J \mathbf{1}_J^T\right); \quad \forall s = 1, 2, \ldots, S.\end{aligned}$$

In this situation, we will have:

$$\begin{aligned}\mathbf{M}(\xi^g) &= \mathbf{Q}^g \left(\sum_{s=1}^{S} w_s \cdot \mathbf{F}_s^T \mathbf{D}_s \mathbf{F}_s\right) \mathbf{Q}^{Tg} \\ &= \mathbf{Q}^g \mathbf{M}(\xi) \mathbf{Q}^{Tg},\end{aligned}$$

where $\sum_{s=1}^{S} w_s \cdot \mathbf{F}_s^T \mathbf{D}_s \mathbf{F}_s = \mathbf{M}(\xi)$. According to the definition of the D-optimality criterion, $\Psi(\mathbf{M}(\xi)) = -\ln(\det(\mathbf{M}(\xi)))$, it can be told that Ψ is linear invariant for g (and $\mathbf{X}$), if $\Psi(\mathbf{M}(\xi^g)) = \Psi(\mathbf{M}(\xi))$. In this case, we will have:

$$\begin{aligned}\Psi(\mathbf{M}(\xi^g)) &= -\ln(\det(\mathbf{M}(\xi^g))) \\ &= -\ln\left(\det(\mathbf{Q}^g \mathbf{M}(\xi) \mathbf{Q}^{Tg})\right) \\ &= -2\ln(\det(\mathbf{Q}^g)) - \ln(\det(\mathbf{M}(\xi))).\end{aligned}$$

In this result,$|\det(\mathbf{Q}^g)| = 1$ (uni-modal) then $\Psi(\mathbf{M}(\xi^g)) = \Psi(\mathbf{M}(\xi))$.

4 OPTIMAL DESIGN IN TWO-LEVEL NMNL MODELS

Conjoint analysis or more precisely, discrete choice experiments are widely used in marketing to measure how the attributes of a product or service jointly affect consumer preferences.

In a choice experiment, a product or service is characterized by a profile or an alternative which is combination of different attribute levels and respondents are asked to choose one of these profiles from the choice set. This task is repeated several times for different choice sets.

The set of choice sets presented constitute the experimental design. The aim of a choice experiment is to estimate the importance of each attribute and its levels based on the respondent's preferences. The estimates are then used to mimic real marketplace decisions by making predictions about consumers'future purchasing behavior.

Designing an efficient choice experiment involves selecting those choice sets which result most precise predictions in accurately estimated models.

Because conjoint choice experiments have become preferred tools for the collection of information on consumers' preference structures *(Louviere and Woodworth 1983)*, the question of how to improve the design of such experiments is of growing importance. More recently, there is some progress in choice experiments that improve the efficiency of designs (see *(Burgess and Street 2003); (Huber and Zwerina 1996); (Sandor and Wedel 2001), (Sandor and Wedel 2002)*; besides others).

(Sandor and Wedel 2005) showed that the construction of heterogeneous designs is preferable, because they produce more accurate estimates of conjoint choice parameters. Heterogeneous designs consist of several sub-designs that are offered to different consumers and can be constructed with relative ease for a wide range of conjoint analysis models.

The chapter is organized as follows. Section 4.1, discusses the model specifications of nested multinomial logit (NMNL) models which has been discussed in subsection 2.2.1 previously. Section 4.2 presents the D-optimality criterion for two-level NMNL models (with M nests).

4.1 Model Specifications

The simplest model in conjoint analysis is called Multinomial Logit model (MNL, Section 3.7). We will take a sample of $\mathcal{I}$ consumers with the choice of $\mathcal{J}$ discrete alternatives in $\mathcal{C}$ choice sets, each of them with J_c ($J_c > 1; \forall c \in \mathcal{C}$) alternatives (Section 2.1). The MNL model has the property that p_{jc} (2.9)) is Independence from Irrelevant Alternative(IIA) (Subsection 2.1.3).

In this situation, to define a design and obtain optimal design we consider some experiments which consist $\mathcal{S} \subset \mathcal{C}$ choice sets each with $J_s = J; \forall s \in \mathcal{S}$ alternatives (Section 3.7).

As stated previously told, the MNL model has a restriction which is called IIA. The most widely used relaxation of the independence assumption is the nested multinomial Logit (NMNL) model *(McFadden 1978a)*, *(Williams 1977)* and *(Daly and Zachary 1978)*. A NMNL model is appropriate when the set of alternatives faced by a decision maker can be partitioned into subsets, called nests, in such a way that the following properties hold:

- For any two alternatives that are in the same nest, the ratio of probabilities $(\frac{P(Y_{j|mc}=1)}{P(Y_{j'|mc}=1)}; j, j' \in \mathbb{C}_{mc})$ is independent of the attributes or existence of all other alternatives where C_m is the set of alternatives in nest m. That is, IIA holds within each nest.

- For any two alternatives in different nests, the ratio of probabilities $(\frac{P(Y_{j|mc}=1)}{P(Y_{j'|m'c}=1)}; j \in \mathbb{C}_{mc}, j' \in \mathbb{C}_{m'c})$ can depend on the attributes of other alternatives in the two nests. IIA does not hold in general for alternatives in different nests.

- For any two nests, the ratio of probabilities $(\frac{P(Y_{mc}=1)}{P(Y_{m'c}=1)}; m \neq m')$ is independent of the attributes of all other nests. Then a form of IIA holds, therefore, even for alternatives in different nests. This form of IIA can be loosely described as " independent from irrelevant nest" or IIN. Thus, with a nested logit model, IIA holds over alternatives in each nest and IIN holds over alternatives in different nests.

As stated (Subsection 2.2.1) in the NMNL model, all possible alternatives ($\mathcal{J} = \prod_{k=1}^{K} L_k$) are divided into groups which are called nests ($m = 1, 2, \ldots, M$), dividing these alternatives into nests depends upon dissimilarly parameters, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m, \ldots, \lambda_M)^T$ (M-dimensional vector) and part-worth parameter, $\boldsymbol{\beta}$ (p-dimensional vector, where $p=\sum_{k=1}^{K}(L_k - 1)$, Section 2.1). In this case, we are considering the utility $U_{jmc} = U_{mc} + U_{j|mc}$ ($U_{j|mc} = v_{j|mc} + \varepsilon_{j|mc}$ and $U_{mc} = v_{mc} + \varepsilon_{mc}$, Subsection 2.2.1) related to choosing the alternative j in nest m *(Train 2003)*, where $\varepsilon_{j|mc}$ and ε_{mc} are independent and they have EVD.

4.2 Information Matrix for The NMNL Model

For the NMNL model, the local information matrix at $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\lambda})^T$ of the choice set $\mathbb{C}_s$, (for one individual) is calculated by (Considering $\mathcal{S}$ choice sets each with $J_s; \forall s \in \mathcal{S} \subset \mathcal{C}$ alternatives):

$$\mathbf{I}(\mathbb{C}_s, \boldsymbol{\theta}) = -E\left(\frac{\partial^2 \ell(\mathbb{C}_s, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right) = \begin{pmatrix} \mathbf{I}_\beta & \mathbf{I}_{\beta\lambda} \\ \mathbf{I}_{\beta\lambda}^T & \mathbf{I}_\lambda \end{pmatrix}, \tag{4.1}$$

where

$$\ell(\mathbb{C}_s, \boldsymbol{\theta}) = \sum_{m=}^{M} \sum_{j=1}^{J_{ms}} y_{jms} \ln(p_{jms})$$

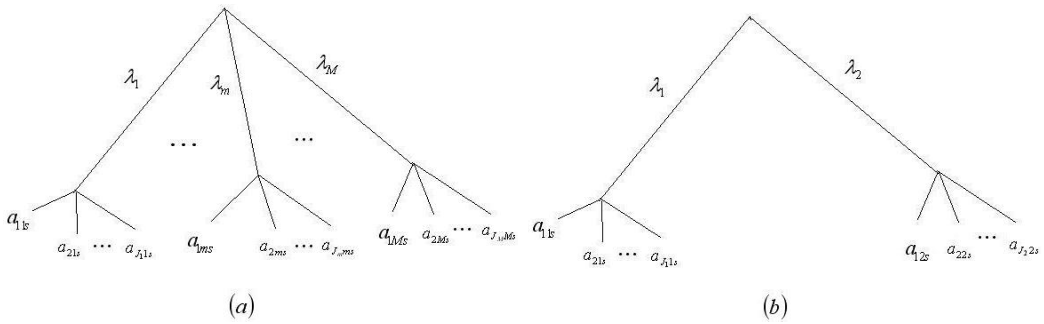

Figure 4.1: (a):The NMNL Model with M nests each with J_{ms} alternatives (w.r.t Choice Set $\mathbb{C}_s$) and (b): when $M = 2$ (Lemma 4.1)

is the log-likelihood function for a nested MNL model with M nests and J_{ms} is the number of alternatives in nest m of choice set s (Figure 4.1(a)) such that $\mathbb{C}_s = \bigcup_{s=1}^{\mathcal{S}} \mathbb{C}_{ms}$,

$$\mathbb{C}_{ms} = \{a_{1ms}, \dots, a_{jms}, \dots, a_{J_{ms}ms}\},$$

where a_{jms} denotes the j^{th} alternative of the m^{th} nest. Based on the choice probability (2.18) (for choice set $\mathbb{C}_s$ and one individual) we will have:

$$\ell(\mathbb{C}_s, \boldsymbol{\theta}) = \sum_{m=1}^{M}\sum_{j=1}^{J_{ms}} y_{jms}\left[\frac{\mathbf{f}^T(a_{jms})\boldsymbol{\beta}}{\lambda_m} + (\lambda_m - 1)\ln\left(\sum_{l=1}^{J_{ms}}\exp\left(\frac{\mathbf{f}^T(a_{jms})\boldsymbol{\beta}}{\lambda_m}\right)\right) - \ln\sum_{n=1}^{M}\left(\sum_{l=1}^{J_{ns}}\exp(\frac{\mathbf{f}^T(a_{jns})\boldsymbol{\beta}}{\lambda_n})\right)^{\lambda_n}\right].$$

As we know that to obtain the elements of the information matrix we must first calculate the partial derivatives of degree two of the likelihood function, which are calculated as follow;

$$\frac{\partial^2\ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\beta_h\partial\beta_{h'}} = \sum_{m=1}^{M}\frac{y_{ms}(\lambda_m-1)-\lambda_m p_{ms}}{\lambda_m^2}\left[\left(\sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms}f_{h'}(a_{jms})\right) - \left(\sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms}\right)\left(\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms})p_{j|ms}\right)\right]$$
$$+\left(\sum_{m=1}^{M} p_{ms}(\sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms})\right)\left(\sum_{m=1}^{M} p_{ms}(\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms})p_{j|ms})\right) - \sum_{m=1}^{M} p_{ms}\left(\sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms}\right)\left(\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms})p_{j|ms}\right),$$

$$\frac{\partial^2\ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\beta_h\partial\lambda_m} = \sum_{j=1}^{J_{ms}}\frac{f_h(a_{jms})(y_{ms}p_{j|ms}-y_{jms})}{\lambda_m^2} - \frac{\beta_h(y_{ms}(\lambda_m-1)+\lambda_m p_{ms})}{\lambda_m^3}\left(\sum_{j=1}^{J_{ms}} f_h^2(a_{jms})p_{j|ms} - (\sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms})^2\right)$$
$$-p_{ms}\left(\sum_{n=1}^{M} p_{ns}(\sum_{j=1}^{J_{ns}} f_h(a_{jns})p_{j|ns}) - \sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms}\right)\times\left(\frac{\beta_h}{\lambda_m}\sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms} - \ln\left(\sum_{j=1}^{J_{ms}}\exp\left(\frac{\mathbf{f}^T(a_{jms})\boldsymbol{\beta}}{\lambda_m}\right)\right)\right),$$

$$\frac{\partial^2\ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\lambda_m^2} = \frac{y_{ms}(\lambda_m-1)-\lambda_m p_{ms}}{\lambda_m^4}\boldsymbol{\beta}^T\left(\sum_{j=1}^{J_{ms}}\mathbf{f}(a_{jms})p_{j|ms}\mathbf{f}^T(a_{jms}) - (\sum_{j=1}^{J_{ms}}\mathbf{f}(a_{jms})p_{j|ms})(\sum_{j=1}^{J_{ms}} p_{j|ms}\mathbf{f}^T(a_{jms}))\right)\boldsymbol{\beta}$$
$$-p_{ms}(1-p_{ms})\left(\ln\left(\sum_{j=1}^{J_{ms}}\exp\left(\frac{\mathbf{f}^T(a_{jms})\boldsymbol{\beta}}{\lambda_m}\right)\right) - \frac{1}{\lambda_m}\boldsymbol{\beta}^T\sum_{j=1}^{J_{ms}}\mathbf{f}(a_{jms})p_{j|ms}\right)^2 + \frac{2\boldsymbol{\beta}^T}{\lambda_m^3}\sum_{j=1}^{J_{ms}}\mathbf{f}^T(a_{jms})(y_{jms}-y_{ms}p_{j|ms})$$

$$\frac{\partial^2 \ell(\mathbb{C}_s, \boldsymbol{\theta})}{\partial \lambda_m \partial \lambda_{m'}} = \frac{p_{ms}p_{m's}}{\lambda_m \lambda_{m'}} \left(\lambda_m \ln \left(\sum_{j=1}^{J_{ms}} \exp \left(\frac{\mathbf{f}^T(a_{jms})\boldsymbol{\beta}}{\lambda_m} \right) \right) - \boldsymbol{\beta}^T \sum_{j=1}^{J_{ms}} \mathbf{f}(a_{jms}) p_{j|ms} \right)$$
$$\times \left(\lambda_{m'} \ln \left(\sum_{j=1}^{J_{m's}} \exp \left(\frac{\mathbf{f}^T(a_{jm's})\boldsymbol{\beta}}{\lambda_m} \right) \right) - \left(\sum_{j=1}^{J_{m's}} p_{j|ms} \mathbf{f}^T(a_{jms}) \right) \boldsymbol{\beta} \right); \quad m \neq m',$$

where $Y_{jms} = Y_{j|ms} \times Y_{ms}$ (Subsection 2.2.1) with $\sum_{j=1}^{J_{ms}} y_{j|ms} = 1$ and $\sum_{m=1}^{M} y_{ms} = 1$. The elements of information matrix, $\mathbf{I}(\mathbb{C}_s, \boldsymbol{\theta})$ are obtained by $-E(\frac{\partial^2 \ell(\mathbb{C}_s, \boldsymbol{\theta})}{\partial \theta \partial \theta^T})$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\lambda})^T$;

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

and for simplicity we suppose that $\beta_{1,1} = \beta_1, \ldots, \beta_{K,L_K-1} = \beta_p$ ($p = \sum_{k=1}^{K} L_k - 1$) then we can rewrite:

$$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_h, \ldots, \beta_p)^T.$$

Corresponding to the part-worth parameters vector, $\boldsymbol{\beta}$, we denote the design matrix for alternative j by:

- $\mathbf{f}(a_{jms}) = (\mathbf{f}_1(a_{jms}), \ldots, \mathbf{f}_k(a_{jms}), \ldots, \mathbf{f}_K(a_{jms}))^T$;

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

Now, let us consider $f_{11}(a_{jms}) = f_1(a_{jms}), \ldots, f_{KL_K-1}(a_{jms}) = f_p(a_{jms})$, thus we can write:

$$\mathbf{f}(a_{jms}) = (f_1(a_{jms}), \ldots, f_h(a_{jms}), \ldots, f_p(a_{jms}))^T$$

. In this situation based on dissimilarity parameters vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m, \ldots \lambda_M)^T$, the full parameters vector is

$$\boldsymbol{\theta} = (\beta_1, \ldots, \beta_h, \ldots, \beta_p, \lambda_1, \ldots, \lambda_m, \ldots, \lambda_M)^T.$$

Theorem 4.1. *According to Subsection 2.2.1 ($E(Y_{j|ms}) = p_{j|ms}$, $E(Y_{ms}) = p_{ms}$) and with respect to the partial derivatives of degree two of the likelihood function (previous page), the elements of information matrix* (4.1) *are calculated as follow:*

$$\mathbf{I}_{\beta} = \begin{pmatrix} I_{\beta_1\beta_1} & \cdots & I_{\beta_1\beta_h} & \cdots & I_{\beta_1\beta_{h'}} & \cdots & I_{\beta_1\beta_p} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \cdots & \vdots \\ I_{\beta_h\beta_1} & \cdots & I_{\beta_h\beta_h} & \cdots & I_{\beta_h\beta_{h'}} & \cdots & I_{\beta_h\beta_p} \\ \vdots & \cdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ I_{\beta_{h'}\beta_1} & \cdots & I_{\beta_{h'}\beta_h} & \cdots & I_{\beta_{h'}\beta_{h'}} & \cdots & I_{\beta_{h'}\beta_h} \\ \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ I_{\beta_p\beta_1} & \cdots & I_{\beta_p\beta_h} & \cdots & I_{\beta_p\beta_{h'}} & \cdots & I_{\beta_p\beta_p} \end{pmatrix}; \; -E\left(\frac{\partial^2 \ell(\mathbb{C}_s, \boldsymbol{\theta})}{\partial \beta_h \partial \beta_{h'}}\right) = I_{\beta_h\beta_{h'}},$$

$$\mathbf{I}_{\lambda} = \begin{pmatrix} I_{\lambda_1\lambda_1} & \cdots & I_{\lambda_1\lambda_m} & \cdots & I_{\lambda_1\lambda_{m'}} & \cdots & I_{\lambda_1\lambda_M} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \cdots & \vdots \\ I_{\lambda_m\lambda_1} & \cdots & I_{\lambda_m\lambda_m} & \cdots & I_{\lambda_m\lambda_{m'}} & \cdots & I_{\lambda_m\lambda_M} \\ \vdots & \cdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ I_{\lambda_{m'}\lambda_1} & \cdots & I_{\lambda_{m'}\lambda_m} & \cdots & I_{\lambda_{m'}\lambda_{m'}} & \cdots & I_{\lambda_{m'}\lambda_M} \\ \vdots & \cdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ I_{\lambda_M\lambda_1} & \cdots & I_{\lambda_M\lambda_m} & \cdots & I_{\lambda_M\lambda_{m'}} & \cdots & I_{\lambda_M\lambda_M} \end{pmatrix}; -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\lambda_m\partial\lambda_{m'}}\right) = I_{\lambda_m\lambda_{m'}},$$

$$\mathbf{I}_{\beta\lambda} = \begin{pmatrix} I_{\beta_1\lambda_1} & \cdots & I_{\beta_1\lambda_m} & \cdots & I_{\beta_1\lambda_M} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ I_{\beta_h\lambda_1} & \cdots & I_{\beta_h\lambda_m} & \cdots & I_{\beta_h\lambda_M} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ I_{\beta_p\lambda_1} & \cdots & I_{\beta_p\lambda_m} & \cdots & I_{\beta_p\lambda_M} \end{pmatrix}; -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\beta_h\partial\lambda_m}\right) = I_{\beta_h\lambda_m}$$

Proof:

$$\begin{aligned} -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\beta_h\partial\beta_{h'}}\right) &= \sum_{m=1}^{M}\frac{p_{ms}}{\lambda_m^2}\left[\left(\sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms}f_{h'}(a_{jms})\right) - \left(\sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms}\right)\left(\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms})p_{j|ms}\right)\right] \\ &- \left(\sum_{m=1}^{M} p_{ms}\left(\sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms}\right)\right)\left(\sum_{m=1}^{M} p_{ms}\left(\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms})p_{j|ms}\right)\right) \\ &+ \sum_{m=1}^{M} p_{ms}\left(\sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms}\right)\left(\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms})p_{j|ms}\right). \end{aligned}$$

$$-E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\beta_h\partial\lambda_m}\right) = -\frac{\beta_h p_{ms}}{\lambda_m^3}\left(\sum_{j=1}^{J_{ms}} f_h^2(a_{jms})p_{j|ms} - \left(\sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms}\right)^2\right)$$
$$+p_{ms}\left(\sum_{n=1}^{M} p_n\left(\sum_{j=1}^{J_n} f_h(a_{jns})p_{j|n}\right) - \sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms}\right)\left(\frac{\beta_h}{\lambda_m}\sum_{j=1}^{J_{ms}} f_h(a_{jms})p_{j|ms} - \ln\left(\sum_{j=1}^{J_{ms}}\exp\left(\frac{\mathbf{f}^T(a_{jms})\boldsymbol{\beta}}{\lambda_m}\right)\right)\right),$$

$$\begin{aligned} -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\lambda_m^2}\right) = \frac{p_{ms}}{\lambda_m^4}\boldsymbol{\beta}^T\left(\sum_{j=1}^{J_{ms}}\mathbf{f}(a_{jms})p_{j|ms}\mathbf{f}^T(a_{jms}) - \left(\sum_{j=1}^{J_{ms}}\mathbf{f}(a_{jms})p_{j|ms}\right)\left(\sum_{j=1}^{J_{ms}} p_{j|ms}\mathbf{f}^T(a_{jms})\right)\right)\boldsymbol{\beta} \\ +p_{ms}(1-p_{ms})\left(\ln\left(\sum_{j=1}^{J_{ms}}\exp\left(\frac{\mathbf{f}^T(a_{jms})\boldsymbol{\beta}}{\lambda_m}\right)\right) - \frac{1}{\lambda_m}\boldsymbol{\beta}^T\sum_{j=1}^{J_{ms}}\mathbf{f}(a_{jms})p_{j|ms}\right)^2, \end{aligned}$$

$$\begin{aligned} -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\lambda_m\partial\lambda_{m'}}\right) &= -\frac{p_{ms}p_{m's}}{\lambda_m\lambda_{m'}}\left(\lambda_m\ln\left(\sum_{j=1}^{J_{ms}}\exp\left(\frac{\mathbf{f}^T(a_{jms})\boldsymbol{\beta}}{\lambda_m}\right)\right) - \boldsymbol{\beta}^T\sum_{j=1}^{J_{ms}}\mathbf{f}(a_{jms})p_{j|ms}\right) \\ &\times\left(\lambda_{m'}\ln\left(\sum_{j=1}^{J_{m's}}\exp\left(\frac{\mathbf{f}^T(a_{jm's})\boldsymbol{\beta}}{\lambda_m}\right)\right) - \left(\sum_{j=1}^{J_{m's}} p_{j|m's}\mathbf{f}^T(a_{jm's})\right)\boldsymbol{\beta}\right), \quad m \neq m'. \end{aligned}$$

Lemma 4.1. *As stated previously was told, the NMNL model will be reduced to the MNL model when $\lambda_m = 1; \forall m = 1, 2, \ldots, M$ (Subsection 2.2.1). In this situation, according to the elements of the information matrix* (4.1) *and Theorem 4.1, the sub-information matrix $\mathbf{I}_\beta = -E\left(\frac{\partial^2 \ell(\mathbb{C}_s, \boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right)$ will be equal to the information matrix depend on the MNL model (Equation* (3.14)*) with the following elements:*

$$I_{\beta_h \beta_{h'}} = \left(\sum_{j=1}^{J} f_h(a_{js}) p_{js} f_{h'}(a_{js})\right) - \left(\sum_{j=1}^{J} f_h(a_{js}) p_{js}\right)\left(\sum_{j=1}^{J} f_{h'}(a_{js}) p_{js}\right).$$

Proof: According to Theorem 4.1 and $\lambda_m = 1; \forall m$ we will have:

$$\begin{aligned}
-E\left(\frac{\partial^2 \ell(\mathbb{C}_s, \boldsymbol{\theta})}{\partial \beta_h \partial \beta_{h'}}\right) &= \sum_{m=1}^{M} \frac{p_{ms}}{1} \left(\sum_{j=1}^{J_{ms}} f_h(a_{jms}) p_{j|ms} f_{h'}(a_{jms})\right) - \sum_{m=1}^{M} \frac{p_{ms}}{1} \left(\sum_{j=1}^{J_{ms}} f_h(a_{jms}) p_{j|ms}\right)\left(\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms}) p_{j|ms}\right) \\
&- \left(\sum_{m=1}^{M}\left(\sum_{j=1}^{J_{ms}} f_h(a_{jms}) p_{j|ms} p_{ms}\right)\right)\left(\sum_{m=1}^{M}\left(\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms}) p_{j|ms} p_{ms}\right)\right) \\
&+ \sum_{m=1}^{M} p_{ms} \left(\sum_{j=1}^{J_{ms}} f_h(a_{jms}) p_{j|ms}\right)\left(\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms}) p_{j|ms}\right).
\end{aligned}$$

In this situation, we can write:

$$\begin{aligned}
-E\left(\frac{\partial^2 \ell(\mathbb{C}_s, \boldsymbol{\theta})}{\partial \beta_h \partial \beta_{h'}}\right) &= \sum_{m=1}^{M}\sum_{j=1}^{J_{ms}} f_h(a_{jms}) p_{j|ms} p_{ms} f_{h'}(a_{jms}) - \sum_{m=1}^{M} p_{ms} \left(\sum_{j=1}^{J_{ms}} f_h(a_{jms}) p_{j|ms}\right)\left(\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms}) p_{j|ms}\right) \\
&- \left(\sum_{m=1}^{M}\sum_{j=1}^{J_{ms}} f_h(a_{jms}) p_{j|ms} p_{ms}\right)\left(\sum_{m=1}^{M}\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms}) p_{j|ms} p_{ms}\right) \\
&+ \sum_{m=1}^{M} p_{ms} \left(\sum_{j=1}^{J_{ms}} f_h(a_{jms}) p_{j|ms}\right)\left(\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms}) p_{j|ms}\right).
\end{aligned}$$

Then:

$$\begin{aligned}
-E\left(\frac{\partial^2 \ell(\mathbb{C}_s, \boldsymbol{\theta})}{\partial \beta_h \partial \beta_{h'}}\right) &= \sum_{m=1}^{M}\sum_{j=1}^{J_{ms}} f_h(a_{jms}) p_{j|ms} p_{ms} f_{h'}(a_{jms}) \\
&- \left(\sum_{m=1}^{M}\sum_{j=1}^{J_{ms}} f_h(a_{jms}) p_{j|ms} p_{ms}\right)\left(\sum_{m=1}^{M}\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms}) p_{j|ms} p_{ms}\right).
\end{aligned}$$

Now, we can write:

- $\sum_{m=1}^{M}\sum_{j=1}^{J_{ms}} f_h(a_{jms}) p_{j|ms} p_{ms} f_{h'}(a_{jms}) = \sum_{j=1}^{J} f_h(a_{js}) p_{js} f_{h'}(a_{js})$,
- $\left(\sum_{m=1}^{M}\sum_{j=1}^{J_{ms}} f_h(a_{jms}) p_{j|ms} p_{ms}\right) = \sum_{j=1}^{J} f_h(a_{js}) p_{js}$,

- $\left(\sum_{m=1}^{M}\sum_{j=1}^{J_{ms}} f_{h'}(a_{jms})p_{j|ms}p_{ms}\right) = \sum_{j=1}^{J} f_{h'}(a_{js})p_{js}$

thus:

$$I_{\beta_h\beta_{h'}} = -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\beta_h\partial\beta_{h'}}\right) = \left(\sum_{j=1}^{J} f_h(a_{js})p_{js}f_{h'}(a_{js})\right) - \left(\sum_{j=1}^{J} f_h(a_{js})p_{js}\right)\left(\sum_{j=1}^{J} f_{h'}(a_{js})p_{js}\right),$$

where $J_s = \sum_{m=1}^{M} J_{ms}$ denotes the number of alternatives in choice set $\mathbb{C}_s$ and (The MNL model, Section 2.1):

$$p_{js} = \frac{\exp\left(\mathbf{f}^T(a_{js})\boldsymbol{\beta}\right)}{\sum_{j'=1}^{J}\exp\left(\mathbf{f}^T(a_{j's})\boldsymbol{\beta}\right)}.$$

4.3 D-Optimal Design for NMNL Model

The information matrix of a design (ξ) with S choice sets, $(\mathbb{C}_1,\ldots,\mathbb{C}_s,\ldots,\mathbb{C}_S)$ is calculated by (Equation (4.1)):

$$\mathbf{I}(\xi,\boldsymbol{\theta}) = \sum_{s=1}^{S} w_s.\mathbf{I}(\mathbb{C}_s,\boldsymbol{\theta}), \tag{4.2}$$

which depends on unknown parameters, $\boldsymbol{\theta}$, where w_s is the weight (frequency) of the choice set $\mathbb{C}_s$. Moreover, the local D-optimality criterion at $\boldsymbol{\theta}$ is $\Psi(\xi,\boldsymbol{\theta}) = \ln\det\left(\mathbf{I}(\xi,\boldsymbol{\theta})\right)$. In this situation, the ξ^* which maximizes the local D-optimality criterion:

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

is called locally D-optimal design where:

$$\xi^* = \left\{\begin{array}{cccc} \mathbb{C}_1^* & \mathbb{C}_2^* & \cdots & \mathbb{C}_S^* \\ w_1^* & w_2^* & \cdots & w_S^* \end{array}\right\}.$$

Thus, in this chapter also the local D-optimality criterion $\Psi(\xi,\boldsymbol{\theta})$ for true values of parameters will be considered.

To do experiment based on the NMNL model, we select $J_s = J; \forall s \in \mathcal{S}$ alternatives from the population with $\mathcal{J}$ possible alternatives ($\mathcal{S}\subset\mathcal{C}$, See Section 3.7) which have been divided into M nests, where each nest consists of $\mathcal{J}_m$ alternatives, i.e $\sum_{m=1}^{M}\mathcal{J}_m = \mathcal{J}$ (Table4.1). In an arbitrary choice set, the number of alternatives which are selected from each nest may vary. The set of all choice sets of a given size $J_s = J; \forall s \in \mathcal{S}$, may be split up into N different classes, which are characterized by the corresponding number J_{nms} of alternatives coming from each nest m, i.e. $\sum_{m=1}^{M} J_{nms} = J_{ns}; n \in N,\ s \in \mathcal{S}$. Note that some of the numbers J_{nms} may be equal to zero, where J_{nms} denotes the number of alternatives, which are selected from nest m of size $\mathcal{J}_m$ in the n^{th} class (w.r.t choice set $\mathbb{C}_s$).

Table 4.1: NMNL Model: The total number of alternatives has been divided into M nests each with $\mathcal{J}_m$ alternatives; $\tilde{a}_{jm}$ denotes alternative j in nest m

1^{st} Nest	...	m^{nd} Nest	...	M^{th} Nest
$\underbrace{\tilde{a}_{11}, \tilde{a}_{21}, \ldots, \tilde{a}_{\mathcal{J}_1 1}}_{size \mathcal{J}_1}$	...	$\underbrace{\tilde{a}_{1m}, \tilde{a}_{2m}, \ldots, \tilde{a}_{\mathcal{J}_m m}}_{size \mathcal{J}_m}$	...	$\underbrace{\tilde{a}_{1M}, \tilde{a}_{2M}, \ldots, \tilde{a}_{\mathcal{J}_M M}}_{size \mathcal{J}_M}$

In order to obtain the optimal design, we first select the $J_{n1s}, \ldots, J_{nms}, \ldots, J_{nMs}$ alternatives of the $1^{st}, \ldots, m^{th}, \ldots, M^{th}$ nest in n^{th} class, in order where N is the number of classes, which form the experiment $(n = 1, 2, \ldots, N)$. The number $\mathcal{S}_n$ of all possible choice sets $(\mathbb{C}_{n1}, \ldots, \mathbb{C}_{ns}, \ldots, \mathbb{C}_{n\mathcal{S}_n})$ in each class n is calculated by:

$$\mathcal{S}_n = \left(\begin{array}{c} \mathcal{J}_1 \\ J_{n1s} \end{array} \right) \cdots \left(\begin{array}{c} \mathcal{J}_m \\ J_{nms} \end{array} \right) \cdots \left(\begin{array}{c} \mathcal{J}_M \\ J_{nMs} \end{array} \right) = \prod_{m=1}^{M} \left(\begin{array}{c} \mathcal{J}_m \\ J_{nms} \end{array} \right), \tag{4.3}$$

where $\sum_{m=1}^{M} J_{nms} = J_s = J; \forall n \in N, s \in \mathcal{S}_n$. This means that the number of alternatives which are selected from the population in each of class (n) and for each choice set (s) are equal. In this situation, $J_{nms} = J_{nms'}; \forall s \neq s' \in \mathcal{S}_n$, but J_{nms} and $J_{n'ms'}$ may or may not be equal.

According to Equation (4.3) we can define the following designs (on class n):

$$\xi_n = \left\{ \begin{array}{cccc} \mathbb{C}_{n1} & \mathbb{C}_{n2} & \cdots & \mathbb{C}_{n\mathcal{S}_n} \\ w_{n1} & w_{n2} & \cdots & w_{n\mathcal{S}_n} \end{array} \right\} \in \Xi_n, \tag{4.4}$$

where $\Xi = \bigcup_{n=1}^{N} \Xi_n$. Based on these designs (above), the information matrixes of designs ξ_n can be calculated as follows:

$$\mathbf{I}(\xi_n, \boldsymbol{\theta}) = \sum_{s=1}^{\mathcal{S}_n} w_{ns}.\mathbf{I}(\mathbb{C}_{ns}, \boldsymbol{\theta}); \quad \forall n = 1, 2, \ldots, N, \tag{4.5}$$

where $\mathbb{C}_{ns}$ denotes choice set s, which has been created in the n^{th} class with $J_{ns} = \sum_{m=1}^{M} J_{nms}$ alternatives (J_{nms} is the number of alternatives in class n and the m^{th} nest of choice set s). Naturally in this chapter the same number of alternatives will be considered in each of the choice sets, resulting in $J_{ns} = J; \forall n \in N, \forall s \in \mathcal{S}_n$.

In Section 3.7 was discussed about number S choice sets ($p \leq S \leq \mathcal{S}$) for doing experiment and obtaining optimal design. Here, we can consider a similar situation for NMNL models, for example, considering S_n choice sets ($p + M \leq S_n \leq \mathcal{S}_n$) based on class n to obtain optimal design within class n, where $p + M$ is the number of parameters. The design ξ_n can not be considered when attempting to obtain an optimal design if $S_n < p + M$ or more generality its information matric is singular (To obtain optimal design within class n).

In this situation, to obtain optimal design based on all of classes, $\xi_n; \forall n \in N$ must be combined to produce a design. For example, consider the two following designs ($N = 2$):

$$\xi_1 = \left\{ \begin{array}{cccc} \mathbb{C}_{11} & \mathbb{C}_{12} & \cdots & \mathbb{C}_{1S_1} \\ w_{11} & w_{12} & \cdots & w_{1S_1} \end{array} \right\}, \quad \xi_2 = \left\{ \begin{array}{cccc} \mathbb{C}_{21} & \mathbb{C}_{22} & \cdots & \mathbb{C}_{2S_2} \\ w_{21} & w_{22} & \cdots & w_{2S_2} \end{array} \right\}.$$

Now, a new design can be defined by $\xi = \alpha\xi_1 + (1-\alpha)\xi_2; 0 \le \alpha \le 1$ as follows:

$$\xi = \left\{ \begin{array}{cccccccc} \mathbb{C}_{11} & \mathbb{C}_{12} & \cdots & \mathbb{C}_{1S_1} & \mathbb{C}_{21} & \mathbb{C}_{22} & \cdots & \mathbb{C}_{2S_2} \\ \alpha w_{11} & \alpha w_{12} & \cdots & \alpha w_{1S_1} & (1-\alpha)w_{12} & (1-\alpha)w_{22} & \cdots & (1-\alpha)w_{2S_2} \end{array} \right\},$$

where $\mathbf{I}(\xi,\boldsymbol{\theta}) = \alpha\mathbf{I}(\xi_1,\boldsymbol{\theta}) + (1-\alpha)\mathbf{I}(\xi_2,\boldsymbol{\theta})$. More general $\mathbf{I}(\xi,\boldsymbol{\theta}) = \sum_{n=1}^{N} \alpha_n \mathbf{I}(\xi_n,\boldsymbol{\theta})$, where:

$$\Xi = \left\{ \sum_{n=1}^{N} \alpha_n \xi_n |\ \xi_n \in \Xi_n, \sum_{n=1}^{N} \alpha_n = 1, \alpha_n \ge 0 \right\}$$

and with respect to local D-optimality criterion, where $\Psi'(\xi,\boldsymbol{\theta}) = -\ln(\det(\mathbf{I}(\xi,\boldsymbol{\theta})))$ there will be convexity:

$$\Psi'(\xi,\boldsymbol{\theta}) \le \alpha\Psi'(\xi_1,\boldsymbol{\theta}) + (1-\alpha)\Psi'(\xi_2,\boldsymbol{\theta}).$$

In this case, we can say that:

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

is locally D-optimal design for $\boldsymbol{\theta}$ takes true values.

Lemma 4.2. *Consider a NMNL with two nests ($M = 2$). The information matrix of a choice set $\mathbb{C}_s$ and each with J_{1s} and J_{2s} alternatives in the first and second nest, respectively, is given by (Figure 4.1(b)):*

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

According to Section 4.2, and Theorem 4.1, we assume that:

- $\mathbf{A}_{ms} = \mathbf{F}_{ms}^T \mathbf{p}_{.|ms};\ m = 1, 2,$
- $\mathbf{B}_{ms} = \mathbf{F}_{ms}^T \mathbf{P}_{.|ms} \mathbf{F}_{ms};,$
- $\mathbf{p}_{.|ms} = (p_{1|ms}, \ldots, p_{j|ms}, \ldots, p_{J_{ms}|ms})^T,$
- $\mathbf{P}_{.|ms} = diag[p_{1|ms}, \ldots, p_{j|ms}, \ldots, p_{J_{ms}|ms}],$
- $\mathbf{X}_{ms} = (\mathbf{f}(a_{1ms}), \ldots, \mathbf{f}(a_{jms}), \ldots, \mathbf{f}(a_{J_{ms}ms}))^T,$
- $\mathbf{f}(a_{jms}) = (f_1(a_{jms}), \ldots, f_h(a_{jms}), \ldots, f_p(a_{jms}))^T,$
- $v_{ms} = \ln\left(\sum_{j=1}^{J_{ms}} \exp\left(\frac{\mathbf{f}^T(a_{jms})\boldsymbol{\beta}}{\lambda_m}\right)\right),$
- $p_{j|ms} = \dfrac{\exp\left(\frac{\mathbf{f}^T(a_{jms})\boldsymbol{\beta}}{\lambda_m}\right)}{\sum_{l=1}^{J_{ms}} \exp\left(\frac{\mathbf{f}^T(a_{lms})\boldsymbol{\beta}}{\lambda_1}\right)},$

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

Thus (with respect to the elements of information matric by Theorem 4.1)

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

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

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

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

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

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

Here $\mathbf{F}_{1s}^T$, $\mathbf{F}_{2s}^T$ denote the $p \times J_{1s}$ and $p \times J_{2s}$-design matrices with respect to the first and second nest. $\mathbf{p}_{.|1s}^T$, $\mathbf{p}_{.|2s}^T$ are $1 \times J_{1s}$ and $1 \times J_{2s}$ vectors, which consist of the probabilities related to choosing alternatives for the first and second nest (See Subsection 2.2.1). J_{1s} and J_{2s} are the number of alternatives that have been selected from the first and second nest.
Proof: To proof Lemma 4.2 we use Section 4.2.

The information matrix (Lemma 4.2) was created for one choice set ($\mathbb{C}_s$) with J_{1s} and J_{2s} alternatives selected from the two nests, each with $\mathcal{J}_1$ and $\mathcal{J}_2$ alternatives. There are

$$\mathcal{S} = \binom{\mathcal{J}_1}{J_{1s}} \times \binom{\mathcal{J}_2}{J_{2s}}$$

choice sets each with $J_{1s} + J_{2s}$ alternatives, where $J_{1s} + J_{2s} = J_{1s'} + J_{2s'}; \forall s \neq s' \in \mathcal{S}$.

Corollary 4.1. *For $\boldsymbol{\beta} = \mathbf{0}$, the information matrix of choice set $\mathbb{C}_s$ (Lemma 4.2) was calculated as follows:*

$$[\mathbf{I}_{11s}] = \left(\frac{{J_{1s}}^{\lambda_1}}{J_{1s}\lambda_1^2({J_{1s}}^{\lambda_1} + {J_{2s}}^{\lambda_2})}\right)[\mathbf{F}_{1s}^T(\mathbf{I}_{J_{1s}} - \tfrac{1}{J_{1s}}\mathbf{1}_{J_{1s}}\mathbf{1}_{J_{1s}}^T)\mathbf{F}_{1s}] + \left(\frac{{J_{2s}}^{\lambda_2}}{J_{2s}\lambda_2^2({J_{1s}}^{\lambda_1} + {J_{2s}}^{\lambda_2})}\right)[\mathbf{F}_{2s}^T(\mathbf{I}_{J_{2s}} - \tfrac{1}{J_{2s}}\mathbf{1}_{J_{2s}}\mathbf{1}_{J_{2s}}^T)\mathbf{F}_{2s}]$$
$$-\left(\frac{{J_{2s}}^{\lambda_2}{J_{1s}}^{\lambda_1}}{({J_{1s}}^{\lambda_1} + {J_{2s}}^{\lambda_2})^2}\right)\left(\frac{1}{J_{2s}^2}\mathbf{F}_{1s}^T\mathbf{1}_{J_{1s}}\mathbf{1}_{J_{1s}}^T\mathbf{F}_{1s} + \frac{1}{J_{2s}^2}\mathbf{F}_{2s}^T\mathbf{1}_{J_{2s}}\mathbf{1}_{J_{2s}}^T\mathbf{F}_{2s} - \frac{1}{J_{1s}J_{2s}}\mathbf{F}_{1s}^T\mathbf{1}_{J_{1s}}\mathbf{1}_{J_{2s}}^T\mathbf{F}_{2s} - \frac{1}{J_{1s}J_{2s}}\mathbf{F}_{2s}^T\mathbf{1}_{J_{2s}}\mathbf{1}_{J_{1s}}^T\mathbf{F}_{1s}\right)$$

$$[\mathbf{I}_{12s}] = \frac{{J_{2s}}^{\lambda_2}{J_{1s}}^{\lambda_1}\ln(J_{1s})}{\left({J_{1s}}^{\lambda_1} + {J_{2s}}^{\lambda_2}\right)^2}\left[\frac{1}{J_{1s}}\mathbf{F}_{1s}^T\mathbf{1}_{J_{1s}} - \frac{1}{J_{2s}}\mathbf{F}_{2s}^T\mathbf{1}_{J_{2s}}\right]$$

$$[\mathbf{I}_{13s}] = \frac{{J_{2s}}^{\lambda_2}{J_{1s}}^{\lambda_1}\ln(J_{2s})}{\left({J_{1s}}^{\lambda_1} + {J_{2s}}^{\lambda_2}\right)^2}\left[\frac{1}{J_{2s}}\mathbf{F}_{2s}^T\mathbf{1}_{J_{2s}} - \frac{1}{J_{1s}}\mathbf{F}_{1s}^T\mathbf{1}_{J_{1s}}\right]$$

$$[I_{22s}] = \frac{{J_{1s}}^{\lambda_1}{J_{2s}}^{\lambda_2}}{\left({J_{1s}}^{\lambda_1} + {J_{2s}}^{\lambda_2}\right)^2}(\ln(J_{1s}))^2,$$

$$[I_{23s}] = -\frac{{J_{1s}}^{\lambda_1}{J_{2s}}^{\lambda_2}}{\left({J_{1s}}^{\lambda_1} + {J_{2s}}^{\lambda_2}\right)^2}\ln(J_{1s})\ln(J_{2s}),$$

$$[I_{33s}] = \frac{{J_{1s}}^{\lambda_1}{J_{2s}}^{\lambda_2}}{\left({J_{1s}}^{\lambda_1} + {J_{2s}}^{\lambda_2}\right)^2}(\ln(J_{2s}))^2,$$

Corollary 4.2. *For $\boldsymbol{\beta} = \mathbf{0}$ and $\lambda_1 = \lambda_2 = \lambda$, the information matrix of choice set $\mathbb{C}_s$ (Lemma 4.2) was calculated as follows:*

$$[\mathbf{I}_{11s}] = (\frac{J_{1s}{}^{\lambda}}{J_{1s}\lambda^2(J_{1s}{}^{\lambda}+J_{2s}{}^{\lambda})})[\mathbf{F}_{1s}^T(\mathbf{I}_{J_{1s}} - \frac{1}{J_{1s}}\mathbf{1}_{J_{1s}}\mathbf{1}_{J_{1s}}^T)\mathbf{F}_{1s}] + (\frac{J_{2s}{}^{\lambda}}{J_{2s}\lambda^2(J_{1s}{}^{\lambda}+J_{2s}{}^{\lambda})})[\mathbf{F}_{2s}^T(\mathbf{I}_{J_{2s}} - \frac{1}{J_{2s}}\mathbf{1}_{J_{2s}}\mathbf{1}_{J_{2s}}^T)\mathbf{F}_{2s}]$$

$$-(\frac{J_{2s}{}^{\lambda}J_{1s}{}^{\lambda}}{(J_{1s}{}^{\lambda}+J_{2s}{}^{\lambda})^2})(\frac{1}{J_{2s}^2}\mathbf{F}_{1s}^T\mathbf{1}_{J_{1s}}\mathbf{1}_{J_{1s}}^T\mathbf{F}_{1s} + \frac{1}{J_{2s}^2}\mathbf{F}_{2s}^T\mathbf{1}_{J_{2s}}\mathbf{1}_{J_{2s}}^T\mathbf{F}_{2s} - \frac{1}{J_{1s}J_{2s}}\mathbf{F}_{1s}^T\mathbf{1}_{J_{1s}}\mathbf{1}_{J_{2s}}^T\mathbf{F}_{2s} - \frac{1}{J_{1s}J_{2s}}\mathbf{F}_{2s}^T\mathbf{1}_{J_{2s}}\mathbf{1}_{J_{1s}}^T\mathbf{F}_{1s})$$

$$[\mathbf{I}_{12s}] = \frac{J_{2s}{}^{\lambda}J_{1s}{}^{\lambda}\ln(J_{1s})}{\left(J_{1s}{}^{\lambda}+J_{2s}{}^{\lambda}\right)^2}\left[\frac{1}{J_{1s}}\mathbf{F}_{1s}^T\mathbf{1}_{J_{1s}} - \frac{1}{J_{2s}}\mathbf{F}_{2s}^T\mathbf{1}_{J_{2s}}\right]$$

$$[\mathbf{I}_{13s}] = \frac{J_{2s}{}^{\lambda}J_{1s}{}^{\lambda}\ln(J_{2s})}{\left(J_{1s}{}^{\lambda}+J_{2s}{}^{\lambda}\right)^2}[\frac{1}{J_{2s}}\mathbf{F}_{2s}^T\mathbf{1}_{J_{2s}} - \frac{1}{J_{1s}}\mathbf{F}_{1s}^T\mathbf{1}_{J_{1s}}]$$

$$[I_{22s}] = \frac{J_{1s}{}^{\lambda}J_{2s}{}^{\lambda}}{\left(J_{1s}{}^{\lambda}+J_{2s}{}^{\lambda}\right)^2}(\ln(J_{1s}))^2,$$

$$[I_{23s}] = -\frac{J_{1s}{}^{\lambda}J_{2s}{}^{\lambda}}{\left(J_{1s}{}^{\lambda}+J_{2s}{}^{\lambda}\right)^2}\ln(J_{1s})\ln(J_{2s}),$$

$$[I_{33s}] = \frac{J_{1s}{}^{\lambda}J_{2s}{}^{\lambda}}{\left(J_{1s}{}^{\lambda}+J_{2s}{}^{\lambda}\right)^2}(\ln(J_{2s}))^2,$$

Corollary 4.3. *For $\boldsymbol{\beta} = \mathbf{0}$, $\lambda_1 = \lambda_2 = \lambda$ and $J_{1s} = J_{2s} = a$ the information matrix of choice set $\mathbb{C}_s$ (Lemma 4.2) was calculated as follows:*

$$[\mathbf{I}_{11s}] = \frac{1}{2a\lambda^2}\left[\mathbf{F}_{1s}^T\left(\mathbf{I}_a - \frac{1}{a}\mathbf{1}_a\mathbf{1}_a^T\right)\mathbf{F}_{1s} + \mathbf{F}_{2s}^T\left(\mathbf{I}_a - \frac{1}{a}\mathbf{1}_a\mathbf{1}_a^T\right)\mathbf{F}_{2s}\right]$$

$$-\frac{1}{4a^2}\left(\mathbf{F}_{1s}^T\mathbf{1}_a\mathbf{1}_a^T\mathbf{F}_{1s} + \mathbf{F}_{2s}^T\mathbf{1}_a\mathbf{1}_a^T\mathbf{F}_{2s} - \mathbf{F}_{1s}^T\mathbf{1}_a\mathbf{1}_a^T\mathbf{F}_{2s} - \mathbf{F}_{2s}^T\mathbf{1}_a\mathbf{1}_a^T\mathbf{F}_{1s}\right)$$

$$[\mathbf{I}_{12s}] = \frac{\ln(a)}{4a}\left[\mathbf{F}_{1s}^T\mathbf{1}_a - \mathbf{F}_{2s}^T\mathbf{1}_a\right]$$

$$[\mathbf{I}_{13s}] = \frac{\ln(a)}{4a}\left[\mathbf{F}_{2s}^T\mathbf{1}_a - \mathbf{F}_{1s}^T\mathbf{1}_a\right]$$

$$[I_{22s}] = \frac{1}{4}(\ln(a))^2$$

$$[I_{23s}] = -\frac{1}{4}(\ln(a))^2$$

$$[I_{33s}] = \frac{1}{4}(\ln(a))^2$$

where $\mathbf{I}_r$ and $\mathbf{1}_r$ denote the identity matrix with the dimensional r and a vector $r \times 1$ when all of its elements are one, respectively.

4.4 Example

Now consider subclasses, which are invariant with respect to a certain group of translations, i.e. orbits of the group on the design region. Moving on, let us consider two nested logit models. Consider one of them with the assumption $\boldsymbol{\beta} = \mathbf{0}$ (Example 4.1) and the other with $\boldsymbol{\beta} \neq \mathbf{0}$ (Example 4.2).

Example 4.1. *There is a two-level NMNL model with two attributes (one attribute has three levels and the other attribute contains two). The alternatives are divided into two nests, each of them with three alternatives as follows:*

	First Nest				*Second Nest*		
Alt.	*At.*(11)($f_1(\tilde{a}_{j1})$)	*At.*(12)($f_2(\tilde{a}_{j1})$)	*At*(21)($f_3(\tilde{a}_{j1})$)	*Alt.*	*At.*(11)($f_1(\tilde{a}_{j2})$)	*At.*(12)($f_2(\tilde{a}_{j2})$)	*At.*(21)($f_3(\tilde{a}_{j2})$)
$\tilde{a}_{11}$	$+1$	0	$+1$	$\tilde{a}_{12}$	$+1$	0	-1
$\tilde{a}_{21}$	0	$+1$	$+1$	$\tilde{a}_{22}$	0	$+1$	-1
$\tilde{a}_{31}$	-1	-1	$+1$	$\tilde{a}_{32}$	-1	-1	-1

,

where $\tilde{a}_{jm}$ denotes the j^{th} alternative of the m^{th} nest. In above Table At.$(k\ell)$ shows the level ℓ of the attribute k, where $\ell = 1, 2, \ldots, L_k - 1; \forall k = 1, 2$.

In this model, there are 15 choice sets of size $J = 4$ in three classes, where (Table 4.2)). In this situation, there are three cases ($N = 3$), where $\mathcal{S}_1 = 9, \mathcal{S}_2 = 3, \mathcal{S}_3 = 3$ and $J_{11s} = 2, J_{12s} = 2; \forall s \in S_1$ and $J_{21s} = 3, J_{22s} = 3; \forall s \in \mathcal{S}_2$ and $J_{31s} = 3, J_{32s} = 3; \forall s \in \mathcal{S}_3$, while $J_{n1s} + J_{n2s} = J = 4; \forall n \in N, \forall s\ \mathcal{S}_n$ (Table 4.2, where a_{jnms} denotes alternative j w.r.t class n in the m^{th} nest of choice set $\mathbb{C}_s$). Therefore, according to Table 4.2 the three different designs and their information matrices are as follows:

- $\xi_1 = \left\{ \begin{array}{cccc} \mathbb{C}_{11} & \mathbb{C}_{12} & \cdots & \mathbb{C}_{19} \\ w_{11} & w_{12} & \cdots & w_{19} \end{array} \right\} \in \Xi_1;\quad \mathbf{I}(\xi_1, \boldsymbol{\theta}) = \sum_{s=1}^{9} w_{1s} \cdot \mathbf{I}(\mathbb{C}_{1s}, \boldsymbol{\theta}),$

- $\xi_2 = \left\{ \begin{array}{ccc} \mathbb{C}_{21} & \mathbb{C}_{22} & \mathbb{C}_{23} \\ w_{21} & w_{22} & w_{23} \end{array} \right\} \in \Xi_2;\quad \mathbf{I}(\xi_2, \boldsymbol{\theta}) = \sum_{s=1}^{3} w_{2s} \cdot \mathbf{I}(\mathbb{C}_{2s}, \boldsymbol{\theta}),$

- $\xi_3 = \left\{ \begin{array}{ccc} \mathbb{C}_{31} & \mathbb{C}_{32} & \mathbb{C}_{33} \\ w_{31} & w_{32} & w_{33} \end{array} \right\} \in \Xi_3;\quad \mathbf{I}(\xi_3, \boldsymbol{\theta}) = \sum_{s=1}^{3} w_{3s} \cdot \mathbf{I}(\mathbb{C}_{3s}, \boldsymbol{\theta}),$

where $\theta = (\boldsymbol{\beta}, \lambda_1, \lambda_2)^T$ and $\boldsymbol{\beta} = (\beta_{1,1}, \beta_{1,2}, -\beta_{1,1} - \beta_{1,2}, \beta_{2,1}, -\beta_{2,1})^T$ and for the sake of clarity, let it be assumed that: $\beta_1 = \beta_{1,1}, \beta_2 = \beta_{1,2}, \beta_3 = \beta_{2,1}$, based on the effects-type coding since $\beta_{1,3} = -\beta_{1,1} - \beta_{1,2}$ and $\beta_{2,2} = -\beta_{2,1}$ thus it can be written as $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$. Due to λ_1 dose not occur in design ξ_2 then that is not identifiable, λ_2 is not also identifiable because that dose not occur in design ξ_1 then the determinant of the information matrices $\mathbf{I}(\xi_2, \boldsymbol{\theta}), \mathbf{I}(\xi_3, \boldsymbol{\theta})$ will be equal to zero. Now, we must combine the three designs ξ_1, ξ_2 and ξ_3 to obtain optimal design in a new design as below:

$$\xi = \left\{ \begin{array}{cccccccccc} \mathbb{C}_{11} & \mathbb{C}_{12} & \cdots & \mathbb{C}_{19} & \mathbb{C}_{21} & \mathbb{C}_{22} & \mathbb{C}_{23} & \mathbb{C}_{31} & \mathbb{C}_{32} & \mathbb{C}_{33} \\ w'_{11} & w'_{12} & \cdots & w'_{19} & w'_{21} & w'_{22} & w'_{23} & w'_{31} & w'_{32} & w'_{33} \end{array} \right\} \in \Xi,$$

where $w'_{1s} = \alpha_1 \cdot w_{1s}; s \in \mathcal{S}_1$ and $w'_{ns} = \alpha_n \cdot w_{ns}; n = 2, 3,\ s \in \mathcal{S}_n$ ($\alpha_1 + \alpha_2 + \alpha_3 = 1,\quad \alpha_n \geq 0; \forall n \in \mathcal{S}_n$). With respect to assumption $\boldsymbol{\beta} = \mathbf{0}$, and the permutation consideration between the levels of the first attribute we can consider equality between the weights of above design, ξ, in the following two cases:

Firstly: Let us suppose that $\lambda_1 = \lambda_2 = \lambda$. In this situation, we consider the following equality between the weights of design ξ (based on permutation considerations):

1. $w'_{11} = w'_{15} = w'_{19} = w'_1,$

Table 4.2: NMNL Model with two nests and the Choice sets related to Example 4.1: There are three ($N = 3$) classes , $J_{11s} = J_{12s} = 2; \forall s \in \mathcal{S}_1$, $J_{21s} = 1, J_{22s} = 3; \forall s \in \mathcal{S}_2$ and $J_{31s} = 3, J_{32s} = 1; \forall s \in \mathcal{S}_3$ (a_{jnms} denotes j^{th} alternative by class n from nest m w.r.t choice set s.)

	$\mathbb{C}_{1s}$	First Nest	Second Nest
First Class	$\mathbb{C}_{11}$	$a_{1111} = \tilde{a}_{11}, a_{2111} = \tilde{a}_{21}$	$a_{1121} = \tilde{a}_{12}, a_{2121} = \tilde{a}_{22}$
	$\mathbb{C}_{12}$	$a_{1112} = \tilde{a}_{11}, a_{2112} = \tilde{a}_{21}$	$a_{1122} = \tilde{a}_{12}, a_{2122} = \tilde{a}_{32}$
	$\mathbb{C}_{13}$	$a_{1113} = \tilde{a}_{11}, a_{2113} = \tilde{a}_{21}$	$a_{1123} = \tilde{a}_{22}, a_{2123} = \tilde{a}_{32}$
	$\mathbb{C}_{14}$	$a_{1114} = \tilde{a}_{11}, a_{2114} = \tilde{a}_{31}$	$a_{1124} = \tilde{a}_{12}, a_{2124} = \tilde{a}_{22}$
	$\mathbb{C}_{15}$	$a_{1115} = \tilde{a}_{11}, a_{2115} = \tilde{a}_{31}$	$a_{1125} = \tilde{a}_{12}, a_{2125} = \tilde{a}_{32}$
	$\mathbb{C}_{16}$	$a_{1116} = \tilde{a}_{11}, a_{2116} = \tilde{a}_{31}$	$a_{1126} = \tilde{a}_{22}, a_{2126} = \tilde{a}_{32}$
	$\mathbb{C}_{17}$	$a_{1117} = \tilde{a}_{21}, a_{2117} = \tilde{a}_{31}$	$a_{1127} = \tilde{a}_{12}, a_{2127} = \tilde{a}_{22}$
	$\mathbb{C}_{18}$	$a_{1118} = \tilde{a}_{21}, a_{2118} = \tilde{a}_{31}$	$a_{1128} = \tilde{a}_{12}, a_{2128} = \tilde{a}_{32}$
	$\mathbb{C}_{19}$	$a_{1119} = \tilde{a}_{21}, a_{2119} = \tilde{a}_{31}$	$a_{1129} = \tilde{a}_{22}, a_{2129} = \tilde{a}_{32}$
	$\mathbb{C}_{2s}$	First Nest	Second Nest
Second Class	$\mathbb{C}_{21}$	$a_{1211} = \tilde{a}_{11}$	$a_{1221} = \tilde{a}_{12}, a_{2221} = \tilde{a}_{22}, a_{3221} = \tilde{a}_{32}$
	$\mathbb{C}_{22}$	$a_{1212} = \tilde{a}_{21}$	$a_{1222} = \tilde{a}_{12}, a_{2222} = \tilde{a}_{22}, a_{3222} = \tilde{a}_{32}$
	$\mathbb{C}_{23}$	$a_{1213} = \tilde{a}_{31}$	$a_{1223} = \tilde{a}_{12}, a_{2223} = \tilde{a}_{22}, a_{3223} = \tilde{a}_{32}$
	$\mathbb{C}_{3s}$	First Nest	Second Nest
Third Class	$\mathbb{C}_{31}$	$a_{1311} = \tilde{a}_{11}, a_{2311} = \tilde{a}_{21}, a_{3311} = \tilde{a}_{31}$	$a_{1321} = \tilde{a}_{12}$
	$\mathbb{C}_{32}$	$a_{1312} = \tilde{a}_{11}, a_{2312} = \tilde{a}_{21}, a_{3312} = \tilde{a}_{31}$	$a_{1322} = \tilde{a}_{22}$
	$\mathbb{C}_{33}$	$a_{1313} = \tilde{a}_{11}, a_{2313} = \tilde{a}_{21}, a_{3313} = \tilde{a}_{31}$	$a_{1323} = \tilde{a}_{32}$

2. $w'_{12} = w'_{13} = w'_{14} = w'_{16} = w'_{17} = w'_{18} = w'_2$,

3. $w'_{21} = w'_{22} = w'_{23} = w'_{31} = w'_{32} = w'_{33} = w'_3$,

where $3 \cdot w'_1 + 6 \cdot w'_2 + 6 \cdot w'_3 = 1$, $0 \leq w'_1 \leq \frac{1}{3}$ and $0 \leq w'_i \leq \frac{1}{6}; \forall i = 2, 3$. Table 4.3 shows that $w_1'^{*} = 0.000$ for all of values of λ and $w_2'^{*}$ decreases as λ increases. Also, we can see that $w_3'^{*}$ has an increasing trend when λ increases.

Now, according to the obtained results of Table 4.3 we define a new design so that:

1. $w'_{11} = w'_{15} = w'_{19} = w'_1 = 0.00$,

2. $w'_{12} = w'_{13} = w'_{14} = w'_{16} = w'_{17} = w'_{18} = w'_2$,

3. $w'_{21} = w'_{22} = w'_{23} = w'_{31} = w'_{32} = w'_{33} = w'_3$.

Table 4.3: NMNL model, $\boldsymbol{\beta} = \mathbf{0}$, $\lambda_1 = \lambda_2 = \lambda$ (two nests) with two attributes, one of them with three levels and the other with two: Locally D-optimal design for $\mathbb{C}_{ns}; n = 1, s = 1, 2, \ldots, 9$ and $\mathbb{C}_{ns}; n = 2, 3,\ s = 1, 2, 3$, where $3w_1' + 6w_2' + 6w_3' = 1$, with Local D-optimality criterion: $\Psi(\xi, \boldsymbol{\theta}_0) = \ln(\det(\mathbf{I}(\xi, \boldsymbol{\theta}_0)))$ for Example 4.1 (with initial value $w_1' = 0.1, w_2' = 0.1, w_3' = \frac{1}{60}$ so that all of results converge), Here $w_1'^*, w_2'^*, w_3'^*$ have been rounded to four digits.

λ	$w_1'^*$	$w_2'^*$	$w_3'^*$
0.100	0.0000	0.0650	0.1020
0.200	0.0000	0.0630	0.1030
0.300	0.0000	0.0620	0.1050
0.400	0.0000	0.0610	0.1060
0.500	0.0000	0.0590	0.1070
0.600	0.0000	0.0580	0.1090
0.700	0.0000	0.0572	0.1093
0.800	0.0000	0.0565	0.1100
0.900	0.0000	0.0556	0.1111
1.000	0.0000	0.0556	0.1111

where $6 \cdot w_2' + 6 \cdot w_3' = 1$ and $0 \leq w_i' \leq \frac{1}{6}; \forall i = 2, 3$. In this situation, the determinant of the information matrix:

$$\mathbf{I}(\xi, \boldsymbol{\theta}) = w_2' \sum_{s \in \{2,3,4,6,7,8\}} \mathbf{I}(\mathbb{C}_{1s}, \boldsymbol{\theta}) + w_3' \left(\sum_{s=1}^{3} \mathbf{I}(\mathbb{C}_{2s}, \boldsymbol{\theta}) + \sum_{s=1}^{3} \mathbf{I}(\mathbb{C}_{3s}, \boldsymbol{\theta}) \right),$$

will be calculated by:

$$\det(\mathbf{I}(\xi, \boldsymbol{\theta})) = c(\lambda) \cdot w_2' \cdot w_3'^2 \left(B_2 \cdot w_2' + B_3 \cdot w_3'\right)^2,$$

where:

- $c(\lambda) = \frac{81}{32} \cdot \frac{\ln^2(3)(\ln(3) - 2\ln(2))^2 (p_{22}(1 - p_{22}))^2}{\lambda^4}$,
- $B_2 = 3(\lambda^2 + 4)$,
- $B_3 = 16 p_{22}(\lambda^2(1 - p_{22}) + 1)$

and $p_{22} = p_{22s} = \frac{3^\lambda}{1+3^\lambda}; \forall s \in \mathcal{S}_2$, where p_{22s} denotes the marginal choice probability of the second nest ($m = 2$) with respect to the second class ($n = 2$) in choice set $\mathbb{C}_{2s}$.

Now, the optimization problem can be solved $\max_{w_2', w_3'} \ln\left(\det(\mathbf{I}(\xi, \boldsymbol{\theta}))\right)$, subject to the natural restrictions of $6 \cdot w_2' + 6 \cdot w_3' = 1,\ w_2', w_3' \geq 0$ by a multiplier $\delta > 0$ (Lagrange coefficient). Thus, $w_2'^*$, $w_3'^*$ can be a solution for the weights, based on first order conditions

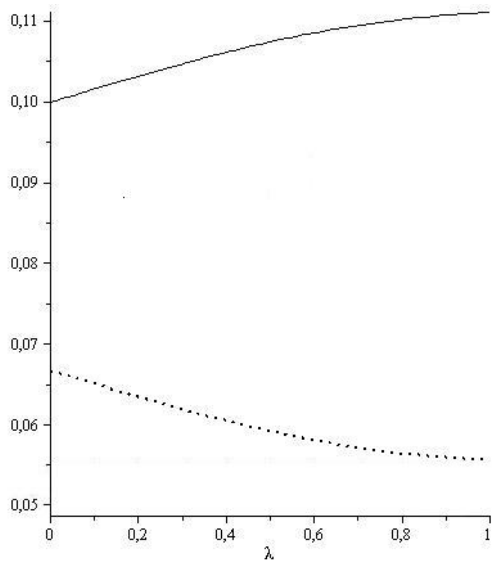

Figure 4.2: NMNL Model (two nests), $\boldsymbol{\beta} = \mathbf{0}$, $\lambda_1 = \lambda_2 = \lambda$ $(0 < \lambda < 1)$: Optimal weights $w_2^{'*}$ (dotted line) and $w_3^{'*}$ (solid).

- $\frac{1}{w_2^{'*}} + \frac{2 \cdot B_2}{B_2 \cdot w_2^{'*} + B_3 \cdot w_3^{'*}} = \delta$,
- $\frac{2}{w_3^{'*}} + \frac{2 \cdot B_3}{B_2 \cdot w_2^{'*} + B_3 \cdot w_3^{'*}} = \delta$,
- $w_2^{'*} + w_3^{'*} = \frac{1}{6}$.

Thus this system is solved by $(0 < \lambda < 1)$:

- $\delta = 5$,
- $w_2^{'*} = \frac{1}{60} \cdot \frac{B_3\left(-7\,B_2+4\,B_3+\sqrt{9\,{B_2}^2-16\,B_2B_3+16\,{B_3}^2}\right)\left(-2-\frac{1}{4}\,\frac{-7\,B_2+4\,B_3+\sqrt{9\,{B_2}^2-16\,B_2B_3+16\,{B_3}^2}}{B_2-B_3}\right)}{(B_2-B_3)B_2\left(-1-\frac{1}{4}\,\frac{-7\,B_2+4\,B_3+\sqrt{9\,{B_2}^2-16\,B_2B_3+16\,{B_3}^2}}{B_2-B_3}\right)}$,
- $w_3^{'*} = \frac{1}{60} \cdot \frac{7\,B_2-4\,B_3-\sqrt{9\,{B_2}^2-16\,B_2B_3+16\,{B_3}^2}}{B_2-B_3}$.

The model, which has been defined by Example 4.1, will be consistent with the Random Utility Maximization dependent on the following conditions (Theorem 2.3 and Corollary 2.3):

1. $(\lambda - 2)2^{\lambda} \leq 0$,
2. $\lambda \leq (1 + 3^{\lambda})$,
3. $\lambda \leq \frac{4}{\frac{3}{1+3^{\lambda}} + \sqrt{(1+7\lambda)(1-\lambda)}}$.

In this case, the above conditions will hold up if $\lambda \in (0,1)$. Figure 4.2 denotes the optimal weights based on design ξ, where $w_2^{'*}$ tends (decreases) to 0.055 as λ tends (increases) to one. Also, $w_3^{'*}$ is always greater than $w_2^{'*}$:

$$w_3^{'*} > w_2^{'*}; \quad \forall \lambda \in (0,1).$$

Secondly: Let us suppose that $\lambda_1 \neq \lambda_2$. In this situation, we consider the following equality between the weights of design ξ (permutation considerations):

1. $w'_{11} = w'_{15} = w'_{19} = w_1$,
2. $w'_{12} = w'_{13} = w'_{14} = w'_{16} = w'_{17} = w'_{18} = w_2$,
3. $w'_{21} = w'_{22} = w'_{23} = w_3$,
4. $w'_{31} = w'_{32} = w'_{33} = w_4$,

where $3w_1 + 6w_2 + 3w_3 + 3w_4 = 1$, $0 \leq w_i \leq \frac{1}{3}; \forall i = 1,3,4$ and $0 \leq w_2 \leq \frac{1}{6}$. Table 4.4 shows that $w_1^* = 0.000$ for all values of λ_1, λ_2 and we can observe different trends for w_2^*, w_3^* and w_4^* when λ_2 is fixed and λ_1 increases. In this situation we can say that (Table 4.4):

$$w_3^* > w_4^* \Leftrightarrow \lambda_1 > \lambda_2.$$

Now, according to the obtained results of Table 4.4 we define a new design so that:

1. $w'_{11} = w'_{15} = w'_{19} = w_1 = 0.00$,
2. $w'_{12} = w'_{13} = w'_{14} = w'_{16} = w'_{17} = w'_{18} = w_2$,
3. $w'_{21} = w'_{22} = w'_{23} = w_3$,
4. $w'_{31} = w'_{32} = w'_{33} = w_4$

where $6w_2 + 3w_3 + 3w_4 = 1$, $0 \leq w_i \leq \frac{1}{3}; \forall i = 3,4$ and $0 \leq w_2 \leq \frac{1}{6}$.

Similarly, the determinant of the information matrix (based on three weights w_2, w_3, w_4):

$$\mathbf{I}(\xi, \boldsymbol{\theta}) = w_2 \sum_{s \in \{2,3,4,6,7,8\}} \mathbf{I}(\mathbb{C}_{1s}, \boldsymbol{\theta}) + w_3 \sum_{s=1}^{3} \mathbf{I}(\mathbb{C}_{2s}, \boldsymbol{\theta}) + w_4 \sum_{s=1}^{3} \mathbf{I}(\mathbb{C}_{3s}, \boldsymbol{\theta}),$$

is obtained by:

$$\det(\mathbf{I}(\xi, \boldsymbol{\theta})) = c(\lambda_1, \lambda_2) \cdot w_2 \cdot w_3 \cdot w_4 \left(A_2 \cdot w_2 - A_3 \cdot w_3 - A_4 \cdot w_4\right)^2,$$

where:

- $c(\lambda_1, \lambda_2) = \frac{81}{2} \cdot \frac{\ln^2(3)(\ln(3) - 2\ln(2))^2 p_{11} \cdot p_{22} \cdot p_{31} \cdot (1-p_{11})(1-p_{22})(1-p_{31})}{\lambda_1^4 \cdot \lambda_2^4}$,

Table 4.4: NMNL model, $\boldsymbol{\beta} = \mathbf{0}$, $\lambda_1 \neq \lambda_2$ (two nests) with two attributes, one of them with three levels and the other with two: Locally D-optimal design for $\mathbb{C}_{ns}; n = 1, s = 1, 2, \ldots, 9$ and $\mathbb{C}_{ns}; n = 2, 3,\ \ s = 1, 2, 3$, where $3w_1 + 6w_2 + 3w_3 + 3w_4 = 1$, with Locally D-optimal criterion: $\Psi(\xi, \boldsymbol{\theta}_0) = \ln(\det(\mathbf{I}(\xi, \boldsymbol{\theta}_0)))$ for Example 4.1 (with initial value $w_1 = 0.1, w_2 = \frac{1}{60}, w_3 = 0.1, w_4 = 0.1$ so that all of results are converge), Here $w_1^*, w_2^*, w_3^*, w_4^*$ have been rounded to our digits.

	λ_2= 0.050				λ_2= 0.250			
λ_1	w_1^*	w_2^*	w_3^*	w_4^*	w_1^*	w_2^*	w_3^*	w_4^*
0.100	0.0000	0.0612	0.1344	0.0766	0.0000	0.0580	0.0736	0.1437
0.200	0.0000	0.0567	0.1507	0.0693	0.0000	0.0624	0.0927	0.1158
0.300	0.0000	0.0550	0.1555	0.0679	0.0000	0.0619	0.1144	0.0952
0.400	0.0000	0.0537	0.1585	0.0674	0.0000	0.0592	0.1304	0.0845
0.500	0.0000	0.0526	0.1610	0.0672	0.0000	0.0567	0.1410	0.0790
0.600	0.0000	0.0516	0.1631	0.0671	0.0000	0.0547	0.1481	0.0758
0.700	0.0000	0.0506	0.1651	0.0670	0.0000	0.0530	0.1533	0.0739
0.800	0.0000	0.0497	0.1670	0.0669	0.0000	0.0517	0.1574	0.0725
0.900	0.0000	0.0488	0.1688	0.0669	0.0000	0.0505	0.1608	0.0716
1.000	0.0000	0.0480	0.1706	0.0668	0.0000	0.0494	0.1637	0.0709
	λ_2= 0.550				λ_2= 0.850			
λ_1	w_1^*	w_2^*	w_3^*	w_4^*	w_1^*	w_2^*	w_3^*	w_4^*
0.100	0.0000	0.0526	0.0684	0.1597	0.0000	0.0495	0.0676	0.1668
0.200	0.0000	0.0545	0.0735	0.1508	0.0000	0.0504	0.0702	0.1624
0.300	0.0000	0.0567	0.0816	0.1383	0.0000	0.0518	0.0742	0.1556
0.400	0.0000	0.0583	0.0917	0.1249	0.0000	0.0533	0.0796	0.1472
0.500	0.0000	0.0588	0.1024	0.1130	0.0000	0.0546	0.0859	0.1382
0.600	0.0000	0.0583	0.1131	0.1037	0.0000	0.0556	0.0928	0.1292
0.700	0.0000	0.0571	0.1223	0.0966	0.0000	0.0562	0.1000	0.1210
0.800	0.0000	0.0559	0.1302	0.0914	0.0000	0.0563	0.1071	0.1137
0.900	0.0000	0.0545	0.1368	0.0875	0.0000	0.0560	0.1138	0.1075
1.000	0.0000	0.0532	0.1424	0.0845	0.0000	0.0555	0.1200	0.1023

- $A_2 = 6\left[(1-p_{11})\lambda_1^2 - p_{11}\cdot\lambda_2^2\left(1+(1-p_{11})\lambda_1^2\right)\right]$,
- $A_3 = 4\lambda_1^2 \cdot p_{22}\left(1+(1-p_{22})\lambda_2^2\right)$,
- $A_4 = 4\lambda_2^2 \cdot p_{31}\left(1+(1-p_{31})\lambda_1^2\right)$,
- $p_{11} = \frac{2^{\lambda_1}}{2^{\lambda_1}+2^{\lambda_2}}, \quad p_{22} = \frac{3^{\lambda_1}}{1+3^{\lambda_1}}, \quad p_{31} = \frac{3^{\lambda_2}}{1+3^{\lambda_2}}$

and p_{nms} denotes the marginal choice probability of the m^{th} nest, according to the n^{th} class in choice set $\mathbb{C}_{ns}$, where $p_{11s} = p_{11}; \forall s \in \mathcal{S}_1$, $p_{22s} = p_{22}; \forall s \in \mathcal{S}_2$ and $p_{31s} = p_{31}; \forall s \in \mathcal{S}_3$.

We have to solve the optimization problem, $\max_{w_2,w_3,w_4} \ln\left(\det(\mathbf{I}(\xi,\boldsymbol{\theta}))\right)$ subject to the natural restrictions $6w_2+3w_3+3w_4 = 1,\ w_2, w_3, w_4 \geq 0$ by a multiplier $\delta > 0$. Thus, it can be supposed that w_2^*, w_3^*, w_4^* is a solution for the weights based on first order conditions

- $\frac{1}{w_2^*} + \frac{2A_2}{\sum_{s=2}^4 A_s \cdot w_s^*} = 6\delta$,
- $\frac{1}{w_3^*} + \frac{2A_3}{\sum_{s=2}^4 A_s \cdot w_s^*} = 3\delta$,
- $\frac{1}{w_4^*} + \frac{2A_4}{\sum_{s=2}^4 A_s \cdot w_s^*} = 3\delta$,
- $6w_2^* + 3w_3^* + 3w_4^* = 1$.

According to the inequality between the two dissimilarity parameters, λ_1 and λ_2, the following conditions must hold for consistency (Model) with RUM (Theorem 2.3 and Corollary 2.3):

1. $\lambda_1 \cdot 2^{\lambda_2} \leq (2^{\lambda_1}+2^{\lambda_2})\ , \quad \lambda_2 \cdot 2^{\lambda_1} \leq (2^{\lambda_1}+2^{\lambda_2})$,
2. $\lambda_1 \leq (1+3^{\lambda_1})\ , \quad \lambda_2 \leq (1+3^{\lambda_2})$,
3. $\lambda_1 \leq \frac{4(1+3^{\lambda_1})}{3+(1+3^{\lambda_1})\sqrt{(1+7\lambda_1)(1-\lambda_1)}}\ , \quad \lambda_2 \leq \frac{4(1+3^{\lambda_2})}{3+(1+3^{\lambda_2})\sqrt{(1+7\lambda_2)(1-\lambda_2)}}$.

In this situation, if $0 < \lambda_1 \leq 1$ and $0 < \lambda_2 \leq 1$, the above six RUM conditions are satisfied. The Locally D-optimal weights w_2^*, w_3^*, w_4^* can be similar the first part obtained .

Example 4.2. *For a NMNL model (two nests) with two attributes each with two levels as follow:*

First Nest			*Second Nest*		
Alternative	$At.(1)(f_1(\tilde{a}_{j1}))$	$At.(2)(f_2(\tilde{a}_{j1}))$	*Alternative*	$At.(1)(f_1(\tilde{a}_{j2}))$	$At.(2)(f_2(\tilde{a}_{j2}))$
$\tilde{a}_{11}$	$+1$	$+1$	$\tilde{a}_{12}$	$+1$	-1
$\tilde{a}_{21}$	-1	$+1$	$\tilde{a}_{22}$	-1	-1

,

where $\tilde{a}_{jm}$ denotes the j^{th} alternative of the m^{th} nest. In above Table, $At.(k)$ shows the attribute $k = 1, 2$. In this situation, it has been considered experiments $2 \times 2/3/4$. Then there are two classes, $N = 2$ each with two choice sets (Table 4.5):

Table 4.5: NMNL Model (two nests): There are two classes ($N = 2$), each class with two choice sets ($\mathbb{C}_{ns}; n = 1, 2, s = 1, 2$) which include three alternatives in two nests, where a_{jnms} denote the j^{th} alternative of the m^{th} nest in choice set s by class n

Choice set	First Nest(I)	Second Nest(II)
$\mathbb{C}_{11}$	$a_{1111} = \tilde{a}_{11}$, $a_{2111} = \tilde{a}_{21}$	$a_{1121} = \tilde{a}_{12}$
$\mathbb{C}_{12}$	$a_{1112} = \tilde{a}_{11}$, $a_{2112} = \tilde{a}_{21}$	$a_{1122} = \tilde{a}_{22}$
$\mathbb{C}_{21}$	$a_{1213} = \tilde{a}_{11}$	$a_{1223} = \tilde{a}_{12}$, $a_{2223} = \tilde{a}_{22}$
$\mathbb{C}_{22}$	$a_{1214} = \tilde{a}_{21}$	$a_{1223} = \tilde{a}_{12}$, $a_{2223} = \tilde{a}_{22}$

Table 4.6: NMNL model (two nests), $\beta_2 = 0$, $\lambda_1 = \lambda_2 = \lambda$: Locally D-optimal weights w_1^*, w_2^*, w_3^* and w_4^*, according to local D-optimality criterion $\Psi(\xi, \boldsymbol{\theta}) = \ln(\det(\mathbf{I}(\xi, \boldsymbol{\theta})))$ and w.r.t $\beta_1 \in (-1, 1)$, $\lambda \in (0, 1]$ (based on RUM conditions), for design ξ Example 4.2 with initial values $w_1 = w_2 = w_3 = 0.2, w_4 = 0.4$ (all of results are converge)

	λ=0.100				λ=0.500				λ=1.000			
β_1	w_1^*	w_2^*	w_3^*	w_4^*	w_1^*	w_2^*	w_3^*	w_4^*	w_1^*	w_2^*	w_3^*	w_4^*
0.900	0.000	0.459	0.376	0.165	0.000	0.461	0.380	0.159	0.000	0.472	0.421	0.107
0.700	0.000	0.464	0.294	0.242	0.000	0.465	0.303	0.232	0.000	0.473	0.364	0.163
0.500	0.000	0.477	0.176	0.347	0.000	0.477	0.200	0.323	0.000	0.478	0.302	0.220
0.300	0.000	0.500	0.000	0.500	0.000	0.495	0.048	0.457	0.037	0.454	0.247	0.262
0.100	0.000	0.500	0.000	0.500	0.000	0.500	0.000	0.500	0.184	0.330	0.232	0.254
0.001	0.000	0.500	0.000	0.500	0.246	0.257	0.244	0.253	0.257	0.259	0.242	0.242
-0.001	0.500	0.000	0.500	0.000	0.257	0.245	0.253	0.245	0.258	0.258	0.242	0.242
-0.010	0.500	0.000	0.500	0.000	0.500	0.000	0.500	0.000	0.330	0.184	0.254	0.232
-0.300	0.500	0.000	0.500	0.000	0.495	0.000	0.457	0.048	0.454	0.037	0.262	0.247
-0.500	0.477	0.000	0.348	0.175	0.477	0.000	0.323	0.200	0.478	0.000	0.220	0.302
-0.700	0.464	0.000	0.243	0.293	0.465	0.000	0.232	0.303	0.473	0.000	0.163	0.364
-0.900	0.460	0.000	0.164	0.376	0.461	0.000	0.159	0.380	0.472	0.000	0.108	0.420

Table 4.7: NMNL Model with two nests and the Choice sets related to Example 4.2: There are two ($N = 2$) classes , $J_{11s} = 2, J_{12s} = 1; \forall s \in \mathcal{S}_1$, $J_{21s} = 1, J_{22s} = 2; \forall s \in \mathcal{S}_2$ (a_{jnms} denotes j^{th} alternative by class n from nest m w.r.t choice set s).

	$\mathbb{C}_{1s}$	First Nest ($J_{11s} = 2; \forall s \in \mathcal{S}_1$)	Second Nest ($J_{12s} = 1; \forall s \in \mathcal{S}_1$)
First Class	$\mathbb{C}_{11}$	$a_{1111} = \tilde{a}_{11}, a_{2111} = \tilde{a}_{21}$	$a_{1121} = \tilde{a}_{12}$
	$\mathbb{C}_{12}$	$a_{1112} = \tilde{a}_{11}, a_{2112} = \tilde{a}_{21}$	$a_{2122} = \tilde{a}_{22}$
	$\mathbb{C}_{2s}$	First Nest ($J_{21s} = 1; \forall s \in \mathcal{S}_2$)	Second Nest ($J_{22s} = 2; \forall s \in \mathcal{S}_2$)
Second Class	$\mathbb{C}_{21}$	$a_{1211} = \tilde{a}_{11}$	$a_{1221} = \tilde{a}_{12}, a_{2221} = \tilde{a}_{22}$
	$\mathbb{C}_{22}$	$a_{2212} = \tilde{a}_{21}$	$a_{1222} = \tilde{a}_{12}, a_{2222} = \tilde{a}_{22}$

- $\mathcal{S}_1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$,
- $\mathcal{S}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix}$,

Now, according to above table and Table 4.5 four choice sets $\mathbb{C}_{11}, \mathbb{C}_{12}, \mathbb{C}_{21}, \mathbb{C}_{22}$ and their alternatives has been showed in Table 4.7:

In this Example 4.2 the parameters are as follow:

- $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_1, \lambda_2)^T$; $\boldsymbol{\beta}_1 = (\beta_{1,1}, -\beta_{1,1})^T$, $\boldsymbol{\beta}_2 = (\beta_{2,1}, -\beta_{2,1})^T$.

Let us, suppose that $\beta_{1,1} = \beta_1$ and $\beta_{2,1} = \beta_2$ thus parameter vector can be rewritten as:

- $\boldsymbol{\theta} = (\beta_1, \beta_2, \lambda_1, \lambda_2)^T$.

According to Table 4.5 we can define:

- $\xi_1 = \left\{ \begin{matrix} \mathbb{C}_{11} & \mathbb{C}_{12} \\ w_{11} & w_{12} \end{matrix} \right\} \in \Xi_1;\quad \mathbf{I}(\xi_1, \boldsymbol{\theta}) = w_{11} \cdot \mathbf{I}(\mathbb{C}_{11}, \boldsymbol{\theta}) + w_{12} \cdot \mathbf{I}(\mathbb{C}_{12}, \boldsymbol{\theta})$
- $\xi_2 = \left\{ \begin{matrix} \mathbb{C}_{21} & \mathbb{C}_{22} \\ w_{21} & w_{22} \end{matrix} \right\} \in \Xi_2;\quad \mathbf{I}(\xi_2, \boldsymbol{\theta}) = w_{21} \cdot \mathbf{I}(\mathbb{C}_{21}, \boldsymbol{\theta}) + w_{22} \cdot \mathbf{I}(\mathbb{C}_{22}, \boldsymbol{\theta})$.

Thus, we will have:

$$\xi = \left\{ \begin{matrix} \mathbb{C}_{11} & \mathbb{C}_{12} & \mathbb{C}_{21} & \mathbb{C}_{22} \\ w'_{11} & w'_{12} & w'_{21} & w'_{22} \end{matrix} \right\} \in \Xi,$$

where $w'_{ns} = \alpha_n \cdot w_{ns}; \forall n \in N, s \in \mathcal{S}_n$, $\alpha_1 + \alpha_2 = 1; 0 \leq \alpha_n \leq 1$ and $\Xi_n \subset \Xi$, $\forall n \in N$. In this situation we will have:

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

Now, for simplicity, we consider $\mathbb{C}_{11} = \mathbb{C}_1$, $\mathbb{C}_{12} = \mathbb{C}_2$, $\mathbb{C}_{21} = \mathbb{C}_3$, $\mathbb{C}_{22} = \mathbb{C}_4$ and w_1, w_2, w_3, w_4 instead of w'_{11}, w'_{12}, w'_{21}, w'_{22}, respectively. According to the new notations the following design is defined:

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

where $\mathcal{S}_1 + \mathcal{S}_2 = p = 4$ (Some locally D-optimal designs has been calculated in Table 4.6, when $\beta_2 = 0$, $\lambda_1 = \lambda_2 = \lambda$ where $\beta_1 \in (-1, 1)$, $\lambda \in (0, 1]$ (based on RUM conditions)).

In particular, let us assume that $\beta_2 = 0$ and $\lambda_1 = \lambda_2 = \lambda$. In this situation based on Table 4.5, it will follow that $p_{I1} = p_{II3}$ and $p_{I2} = p_{II4}$, where p_{ms} denotes the marginal choice probability related to choose nest m with respected to choice set $\mathbb{C}_s$ (Subsection 2.2.1). Now, according to the permutation of the levels of the second attribute, consider the following design:

$$\xi' = \left\{ \begin{array}{cccc} \mathbb{C}_1 & \mathbb{C}_2 & \mathbb{C}_3 & \mathbb{C}_4 \\ w & \frac{1}{2} - w & w & \frac{1}{2} - w \end{array} \right\} \in \Xi', \tag{4.6}$$

where $\Xi' \subset \Xi$ and $w_1 = w_3 = w$, $w_2 = w_4 = \frac{1}{2} - w$; $0 \leq w \leq \frac{1}{2}$. In this case, the RUM conditions are as follows (Theorem 2.3, Corollary 2.3):

1. $\lambda \cdot \exp(\beta_1) \leq \Big(\exp(\beta_1) + \big(\exp(\frac{\beta_1}{\lambda}) + \exp(\frac{-\beta_1}{\lambda})\big)^{\lambda}\Big)$,

2. $\lambda \cdot \exp(-\beta_1) \leq \Big(\exp(-\beta_1) + \big(\exp(\frac{\beta_1}{\lambda}) + \exp(\frac{-\beta_1}{\lambda})\big)^{\lambda}\Big)$.

These inequations support, uphold $\beta_1 \in (-1, 1)$, $\lambda \in (0, 1]$. In this case, w^* decreases when β_1 ($\beta_1 < 0$) decreases, agreeing with the fixed values of λ. The optimal weight, w^* increases in fixed amounts of λ when β_1 ($\beta_1 > 0$) increases. Furthermore, Table 4.8 also denotes that the optimal weight, w^* is equal to 0.5 for low values of β_1 ($-0.5 \leq \beta_1 < -0.05$) but that is equal to 0.0 when β_1 belongs to the interval $(0.05, 0.5]$. Notwithstanding, these cases will occur with low values of λ ($0.0 < \lambda \leq 0.5$). Then, as such a situation, two kind of optimal designs can be considered as follows:

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

where $0.0 < \lambda \leq 0.5$ (Table 4.8). On the other hand, the optimal weight (w^*) decreases (based on the fixed negative values of β_1 ($\beta_1 < 0$)) as λ increases. But, for the positive fixed amounts of β_1, we can say that optimal weight, w^*, increases when λ increases.

The other important note is about locally D-optimal design when β_1 tends to zero and $\lambda = 1$. In this case, w^* tends to value $\frac{1}{4}$ as β_1 tends to zero (from both left and right) when $\lambda = 1$. As previously stated, the NMNL model collapse to the MNL model when $\lambda_m = 1; \forall m$. In this situation, it has been achieved the same result as MNL model when $\lambda = 1$ and β_1 tends to zero, means that (Table 3.2):

$$\xi^*_{\lambda=1} = \left\{ \begin{array}{cccc} \mathbb{C}_1 & \mathbb{C}_2 & \mathbb{C}_3 & \mathbb{C}_4 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right\}$$

as β_1 tends to zero from both left and right.

Table 4.8: NMNL model (two nests), $\beta_2 = 0$, $\lambda_1 = \lambda_2 = \lambda$: Locally optimal weight w^*, w.r.t local D-optimality criterion $\Psi(\xi, \boldsymbol{\theta}) = \ln(\det(\mathbf{I}(\xi, \boldsymbol{\theta})))$ for design (4.6), where $\beta_1 \in (-1, 1)$, $\lambda \in (0, 1]$ (based on RUM conditions)

		λ										
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	1.0
	0.9	0.183	0.183	0.183	0.184	0.186	0.189	0.193	0.200	0.208	0.213	0.218
	0.8	0.150	0.150	0.150	0.151	0.154	0.159	0.167	0.177	0.188	0.195	0.202
	0.7	0.104	0.104	0.105	0.107	0.113	0.121	0.133	0.149	0.166	0.175	0.185
	0.6	0.042	0.042	0.043	0.048	0.058	0.073	0.093	0.116	0.142	0.155	0.168
	0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.082	0.119	0.136	0.153
	0.4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.052	0.103	0.126	0.146
	0.3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.042	0.105	0.130	0.152
	0.2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.072	0.133	0.155	0.173
	0.1	0.000	0.000	0.000	0.000	0.000	0.000	0.087	0.148	0.185	0.198	0.208
	0.05	0.000	0.000	0.000	0.000	0.000	0.108	0.166	0.198	0.217	0.223	0.229
	0.005	0.000	0.000	0.121	0.197	0.224	0.236	0.241	0.245	0.247	0.247	0.248
β_1	0.001	0.000	0.161	0.224	0.239	0.245	0.247	0.248	0.249	0.249	0.249	0.249
	-0.001	0.500	0.339	0.274	0.261	0.255	0.253	0.252	0.251	0.251	0.251	0.250
	-0.005	0.500	0.500	0.379	0.303	0.276	0.264	0.259	0.253	0.253	0.253	0.252
	-0.05	0.500	0.500	0.500	0.500	0.500	0.392	0.334	0.302	0.283	0.277	0.271
	-0.1	0.500	0.500	0.500	0.500	0.500	0.500	0.413	0.352	0.315	0.302	0.292
	-0.2	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.428	0.367	0.345	0.327
	-0.3	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.458	0.395	0.367	0.348
	-0.4	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.448	0.397	0.374	0.354
	-0.5	0.500	0.500	0.500	0.500	0.500	0.488	0.455	0.418	0.381	0.364	0.347
	-0.6	0.459	0.458	0.457	0.452	0.442	0.427	0.407	0.384	0.358	0.345	0.332
	-0.7	0.396	0.396	0.395	0.393	0.387	0.379	0.367	0.351	0.334	0.325	0.315
	-0.8	0.350	0.350	0.350	0.349	0.346	0.341	0.333	0.323	0.311	0.305	0.298
	-0.9	0.316	0.316	0.316	0.316	0.314	0.311	0.307	0.300	0.292	0.287	0.282

5 OPTIMAL DESIGN IN A THREE-LEVEL NMNL MODEL

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 (Independence 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. *(Börsch-Supan 1990)* argues that the *DZM* (Daly, Zachary and McFadden) 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 Nested logit 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 5.1.

The rest of this chapter is structured as follows. Section 5.1 discusses the model specifications of three-level NMNL models. Section 5.2 presents the information matrix for a three-level NMNL models (with two nests). We will introduce the D-optimal criterion by section 5.4.

5.1 Model Specifications

Following *(Gil-Molton and Hole 2004)*, let us consider a sample of $\mathcal{I}$ individuals with $\mathcal{J}$ discrete possible alternatives (as in choice set $\mathbb{C}$), produced by K attributes, each with L_k levels. In this chapter, for a three-level NMNL model, the total number of alternatives is showed by $\sum_{m=1}^{M}\sum_{h=1}^{H_m}\mathcal{J}_{hm}$, where $\mathcal{J}_{hm}$ is the number of alternatives in the sub-set h of nest m (Figure 5.1). In this case, there are $\mathcal{C}$ choice sets each containing J_c alternatives to fit model, where if $\mathbb{C}_c$ is a choice set with J_c alternatives then $\mathbb{C} = \bigcup_{c=1}^{\mathcal{C}}\mathbb{C}_c$, where $J_c = \sum_{m=1}^{M}\sum_{h=1}^{H_m} J_{mhc}$, J_{mhc} denotes the number of alternatives in sub-nest of the nest m with respect to choice set $\mathbb{C}_c$ (Subsection 2.2.2). Certainly, in such a model, the total numbers of alternatives in choice set $\mathbb{C}$ 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 three-level nested logit model (w.r.t choice set $\mathbb{C}_c$), where the individual i is derived when choosing alternative j as denoted by U_{jhms}. This utility is partitioned into systematic component, v_{jhmc}, and a random component, ε_{jhmc} (c denotes the choice set $\mathbb{C}_c$), to produce (ignoring index i):

$$U_{jhmc} = U_{j|hmc} + U_{h|mc} + U_{mc}, \tag{5.1}$$

where:

$$U_{j|hmc} = v_{j|hmc} + \varepsilon_{j|hmc}, \quad U_{h|mc} = v_{h|mc} + \varepsilon_{h|mc}, \quad U_{mc} = v_{mc} + \varepsilon_{mc}, \tag{5.2}$$

where $\varepsilon_{j|hms}$ has EVD (extreme value distribution (type II)) with variance σ^2_{hm} (They are correlated in the same sub-nest, $\rho_{hm} = corr(\varepsilon_{j|hmc}, \varepsilon_{j'|hmc})$) and the distributions of $\varepsilon_{h|mc}$ is such that variable $\max_{j\in\mathbb{C}_{hmc}} U_{j|hmc}$ with variance σ^2_m (They are correlated in the same nest, $\rho_m = corr(\varepsilon_{h|mc}, \varepsilon_{h'|mc})$) and ε_{mc} is such that variable $\max_{h\in H_m} U_{h|mc}$ will have EVD (type II) with variance σ^2 *(McFadden 1978b)*, where $corr(\varepsilon_{mc}, \varepsilon_{m'c}) = 0;\ m \neq m'$. Naturally, these three error terms are independent. Now, with consideration to utility (5.2) observation variables as follows (subsection 2.2.2) can be introduced:

$$Y_{j|h_mc} = \begin{cases} 1, & U_{j|hms} = \max_{j'\in\mathbb{C}_{hmc}} U_{j'|hmc}; \\ 0, & otherwise. \end{cases} \quad , \quad Y_{h|mc} = \begin{cases} 1, & U_{h|mc} = \max_{h'\in H_m} U_{h'|mc}; \\ 0, & otherwise. \end{cases}$$

$$Y_{mc} = \begin{cases} 1, & U_{mc} = \max_{m'} U_{m'c}; \\ 0, & otherwise. \end{cases}$$

Thus, when the variables $Y_{j|hmc}$, $Y_{h|mc}$ and Y_{mc} are independent:

$$p_{jhmc} = p_{j|hmc} \times p_{h|mc} \times p_{mc}, \tag{5.3}$$

where $P(A)$ denotes the probability of event A and $p_{j|hmc} = P(Y_{j|hmc} = 1)$ is the conditional probability of choosing alternative j, given that sub-nest h and nest m have been chosen, $p_{h|mc} = P(Y_{h|mc} = 1)$ is the conditional probability of choosing sub-nest h when nest m is chosen and $p_{mc} = P(Y_{mc} = 1)$ is the marginal probability of choosing nest m (with respect to

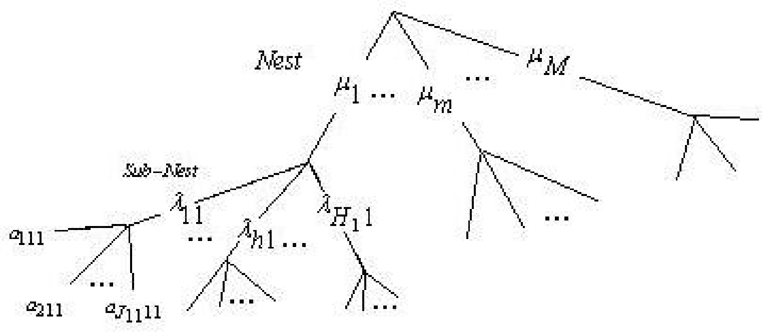

Figure 5.1: NMNL Model: There are M nests each with $H_m; m = 1, 2, \ldots, M$ sub-nests and each sub-nest consists $\mathcal{J}_{hm}$ alternatives

choice set $\mathbb{C}_c$). Based on the distribution of the error terms of utility, these probabilities can be calculated by (McFadden 1981):

$$p_{j|hmc} = \frac{\exp\left(\frac{v_{j|hmc}}{\lambda_{hm}}\right)}{\sum_{j'=1}^{J_{hmc}} \exp\left(\frac{v_{j'|hmc}}{\lambda_{hm}}\right)}, \; p_{h|mc} = \frac{\exp\left(\frac{\lambda_{hm}}{\mu_m} IV_{hmc}\right)}{\sum_{h'=1}^{H_m} \exp\left(\frac{\lambda_{h'm}}{\mu_m} IV_{h'mc}\right)}, \; p_{mc} = \frac{\exp\left(\mu_m IV_{mc}\right)}{\sum_{m'=1}^{M} \exp\left(\mu_{m'} IV_{m'c}\right)}, \tag{5.4}$$

where IV_{mc} and IV_{hmc} (Subsection 2.2.2) are the inclusive values of nest m and sub-nest hm, respectively and

- $IV_{mc} = E\left(\max_{h\in H_m} U_{h|mc}\right) = \ln\left(\sum_{h=1}^{H_m} \exp\left(\frac{\lambda_{hm}}{\mu_m} IV_{hmc}\right)\right)$,
- $IV_{hmc} = E\left(\max_{j\in\mathbb{C}_{hmc}} U_{j|hmc}\right) = \ln\left(\sum_{j=1}^{J_{hmc}} \exp\left(\frac{v_{j|hmc}}{\lambda_{hm}}\right)\right)$,
- $v_{j|hmc} = \mathbf{f}^T(a_{jhmc})\boldsymbol{\beta} = \sum_{k=1}^{K} \mathbf{f}_k^T(a_{jhmc})\boldsymbol{\beta}_k$;
- $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k, \ldots, \boldsymbol{\beta}_K)^T$;
- $\boldsymbol{\beta}_k = (\beta_{k,1}, \ldots, \beta_{k,\ell}, \ldots, \beta_{k,L_k})^T$,
- $\mathbf{f}(a_{jhmc}) = (\mathbf{f}_1(a_{jhmc}), \ldots, \mathbf{f}_k(a_{jhmc}), \ldots, \mathbf{f}_K(a_{jhmc}))^T$;
- $\mathbf{f}_k(a_{jhmc}) = (f_{k1}(a_{jhmc}), \ldots, f_{k\ell}(a_{jhmc}), \ldots, f_{kL_k}(a_{jhmc}))^T$,

where $\mathbf{f}_k(a_{jhmc})$ denotes the characteristics of the attribute k related to choosing alternative j by individual i (ignored) in the sub-nest h of the nest m according to choice set $\mathbb{C}_c$. In Subsection 2.2.2, it was demonstrated that the three-level nested logit model is consistent with RUM (Theorem 2.4), where it has been used the three RUM conditions (Subsection 2.1.2) and based on model (5.3).

In this chapter, to define design and obtain optimal design, $\mathcal{S}$ ($\mathcal{S} \subset \mathcal{C}$) choice sets each with J_s ($J_s = J; \forall s \in \mathcal{S}$) alternatives will be considered. As stated previously told, the number of choice sets $\mathcal{C}$ are as follow:

$$\mathcal{C} = \sum_{J_c \in \mathcal{J}} \left(\begin{array}{c} \mathcal{J} \\ J_c \end{array} \right),$$

where the number of alternatives J_c depend on the number of nests and sub-nests. In this situation, the number of choice sets $\mathcal{S}$, which is considered to define a design is as follows:

$$\mathcal{S} = \left(\begin{array}{c} \mathcal{J} \\ J_s \end{array} \right),$$

where $J_s = \sum_{m=1}^{M} \sum_{h=1}^{H_m} J_{mhs}$.

5.2 Information Matrix

As in the previous chapters, 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 a three-level nested logit model. In this situation, a log-likelihood function is required, defined for the choice set $\mathbb{C}_s$ and one individual as follow:

$$\ell(\mathbb{C}_s, \boldsymbol{\theta}) = \sum_{m=1}^{M} \sum_{h=1}^{H_m} \sum_{j=1}^{J_{hms}} Y_{jhms} \ln\left(p_{jhms}\right),$$

where J_{hms} denotes the number of alternatives in sub-nest hm corresponding to choice set $\mathbb{C}_s$ and Y_{jhms} as defined in Subsection (2.2.2). In this situation, based on Equation (5.4) and the number choice sets $\mathcal{S}$ each with J_s alternatives we can write:

$$\begin{aligned} \ell(\mathbb{C}_s, \boldsymbol{\theta}) \quad = \quad & \sum_{m=1}^{M} \sum_{h=1}^{H_m} \sum_{j=1}^{J_{hms}} Y_{jhms} \left(\left(\frac{v_{j|hmc}}{\lambda_{hm}} \right) + \left(\frac{\lambda_{hm}}{\mu_m} IV_{hms} \right) + \left(\mu_m IV_{ms} \right) \right) \\ & - \sum_{m=1}^{M} \sum_{h=1}^{H_m} \sum_{j=1}^{J_{hms}} Y_{jhms} \ln \left(\sum_{j'=1}^{J_{hmc}} \exp\left(\frac{v_{j'|hmc}}{\lambda_{hm}} \right) \right) \left(\sum_{h'=1}^{H_m} \exp\left(\frac{\lambda_{h'm}}{\mu_m} IV_{h'ms} \right) \right) \left(\sum_{m'=1}^{M} \exp\left(\mu_{m'} IV_{m's} \right) \right). \end{aligned}$$

Now, based on the definition of the information matrix (based on choice set $\mathbb{C}_s$) and $E(Y_{jhms}) = p_{jhms}$ we will have:

$$-E\left(\frac{\partial^2 \ell(\mathbb{C}_s, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right) = \sum_{m=1}^{M} \sum_{h=1}^{H_m} \sum_{j=1}^{J_{hms}} p_{j|hms} p_{h|ms} p_{ms} \left(\frac{-\partial^2 \ln(p_{j|hms})}{\partial \theta \partial \theta^T} + \frac{-\partial^2 \ln(p_{h|ms})}{\partial \theta \partial \theta^T} + \frac{-\partial^2 \ln(p_{ms})}{\partial \theta \partial \theta^T} \right),$$

where

$$\mathbf{I}(\mathbb{C}_s, \boldsymbol{\theta}) = \left(\begin{array}{ccc} \mathbf{I}_{\beta} & \mathbf{I}_{\beta\mu} & \mathbf{I}_{\beta\lambda} \\ \mathbf{I}_{\beta\mu}^T & \mathbf{I}_{\mu} & \mathbf{I}_{\mu\lambda} \\ \mathbf{I}_{\beta\lambda}^T & \mathbf{I}_{\mu\lambda}^T & \mathbf{I}_{\lambda} \end{array} \right) \tag{5.5}$$

is the information matrix of choice set $\mathbb{C}_s$ and $\boldsymbol{\theta}$ is the full parameters vector, so that:

- $\mathbf{I}_{\beta} = -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right)$, $\mathbf{I}_{\beta\mu} = -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\mu}^T}\right)$, $\mathbf{I}_{\beta\lambda} = -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\lambda}^T}\right)$,
- $\mathbf{I}_{\mu} = -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^T}\right)$, $\mathbf{I}_{\mu\lambda} = -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\lambda}^T}\right)$, $\mathbf{I}_{\lambda} = -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^T}\right)$,
- $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\lambda})^T$;
- $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k, \ldots, \boldsymbol{\beta}_K)^T$
- $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m, \ldots, \mu_M)^T$,
- $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \ldots, \boldsymbol{\lambda}_M)^T$;
- $\boldsymbol{\lambda}_m = (\lambda_{1m}, \ldots, \lambda_{hm}, \ldots, \lambda_{H_m m})^T$,

where $k = 1, 2, \ldots, K$, $h = 1, 2, \ldots, H_m$, $m = 1, 2, \ldots, M$.

According to effect type coding ($\sum_{\ell=1}^{L_k} \beta_{k,\ell} = 0$; $\beta_{k,\ell} = -\sum_{\ell=1}^{L_k-1} \beta_{k,\ell}$) we can write:

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

In this case, we suppose that $\beta_{1,1} = \beta_1, \ldots, \beta_{K,L_K-1} = q_1$ then we will have:

- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r, \ldots, \beta_{q_1})^T$.

This means that parameter $\beta_{k,\ell}$ is related to the ℓ^{th} level of attribute k, μ_m, the dissimilarity parameter of nest m and $\boldsymbol{\lambda}_m$ is the dissimilarity parameters vector of nest m, where λ_{hm} denotes the dissimilarity parameter of sub-nest hm in nest m, thus we will have:

- $\mathbf{I}_{\beta} = \begin{pmatrix} I_{\beta_1} & \cdots & I_{\beta_1\beta_r} & \cdots & I_{\beta_1\beta_{q_1}} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ I_{\beta_r\beta_1} & \cdots & I_{\beta_r} & \cdots & I_{\beta_r\beta_{q_1}} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ I_{\beta_{q_1}\beta_1} & \cdots & I_{\beta_{q_1}\beta_r} & \cdots & I_{\beta_{q_1}} \end{pmatrix}$; $-E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\beta_r\partial\beta_{r'}}\right) = \begin{cases} I_{\beta_r\beta_{r'}}, & r \neq r'; \\ I_{\beta_r}, & r = r'. \end{cases}$
- $\mathbf{I}_{\beta\mu} = \begin{pmatrix} I_{\beta_1\mu_1} & \cdots & I_{\beta_1\mu_m} & \cdots & I_{\beta_1\mu_M} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ I_{\beta_r\mu_1} & \cdots & I_{\beta_r\mu_m} & \cdots & I_{\beta_r\mu_M} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ I_{\beta_{q_1}\mu_1} & \cdots & I_{\beta_{q_1}\mu_m} & \cdots & I_{\beta_{q_1}\mu_M} \end{pmatrix}$; $-E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\beta_r\partial\mu_m}\right) = I_{\beta_r\mu_m}$,
- $\mathbf{I}_{\beta\lambda} = \begin{pmatrix} \mathbf{I}_{\beta\lambda_1} & \cdots & \mathbf{I}_{\beta\lambda_m} & \cdots & \mathbf{I}_{\beta\lambda_M} \end{pmatrix}$;
- $\mathbf{I}_{\beta\lambda_m} = \begin{pmatrix} I_{\beta_1\lambda_{1m}} & \cdots & I_{\beta_1\lambda_{hm}} & \cdots & I_{\beta_1\lambda_{H_m m}} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ I_{\beta_r\lambda_{1m}} & \cdots & I_{\beta_r\lambda_{hm}} & \cdots & I_{\beta_r\lambda_{H_m m}} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ I_{\beta_{q_1}\lambda_{1m}} & \cdots & I_{\beta_{q_1}\lambda_{hm}} & \cdots & I_{\beta_{q_1}\lambda_{H_m m}} \end{pmatrix}$; $-E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\beta_r\partial\lambda_{hm}}\right) = I_{\beta_r\lambda_{hm}}$,

- $\mathbf{I}_{\mu} = \begin{pmatrix} I_{\mu_1} & \cdots & I_{\mu_1\mu_m} & \cdots & I_{\mu_1\mu_M} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ I_{\mu_m\mu_1} & \cdots & I_{\mu_m} & \cdots & I_{\mu_m\mu_M} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ I_{\mu_M\mu_1} & \cdots & I_{\mu_M\mu_m} & \cdots & I_{\mu_M} \end{pmatrix}; \quad -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\mu_m\partial\mu_{m'}}\right) = \begin{cases} I_{\mu_m\mu_{m'}}, & m \neq m'; \\ I_{\mu_m}, & m = m'. \end{cases},$
- $\mathbf{I}_{\mu\lambda} = \begin{pmatrix} \mathbf{I}_{\mu\lambda_1} & \cdots & \mathbf{I}_{\mu\lambda_m} & \cdots & \mathbf{I}_{\mu\lambda_M} \end{pmatrix};$
- $\mathbf{I}_{\mu\lambda_m} = \begin{pmatrix} I_{\mu_1\lambda_{1m}} & \cdots & I_{\mu_1\lambda_{hm}} & \cdots & I_{\mu_1\lambda_{H_m m}} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ I_{\mu_m\lambda_{1m}} & \cdots & I_{\mu_m\lambda_{hm}} & \cdots & I_{\mu_m\lambda_{H_m m}} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ I_{\mu_M\lambda_{1m}} & \cdots & I_{\mu_M\lambda_{hm}} & \cdots & I_{\mu_M\lambda_{H_m m}} \end{pmatrix}; \quad -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\mu_m\partial\lambda_{hm}}\right) = I_{\mu_m\lambda_{hm}},$
- $\mathbf{I}_{\lambda} = \begin{pmatrix} \mathbf{I}_{\lambda_1} & \cdots & \mathbf{I}_{\lambda_1\lambda_m} & \cdots & \mathbf{I}_{\lambda_1\lambda_M} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ \mathbf{I}_{\lambda_m\lambda_1} & \cdots & \mathbf{I}_{\lambda_m} & \cdots & \mathbf{I}_{\lambda_m\lambda_M} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ \mathbf{I}_{\lambda_M\lambda_1} & \cdots & \mathbf{I}_{\lambda_M\lambda_m} & \cdots & \mathbf{I}_{\lambda_M} \end{pmatrix}; \quad -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}_m\partial\boldsymbol{\lambda}_{m'}^T}\right) = \begin{cases} \mathbf{I}_{\lambda_m\lambda_{m'}}, & m \neq m'; \\ \mathbf{I}_{\lambda_m}, & m = m'. \end{cases};$
- $\mathbf{I}_{\lambda_m} = \begin{pmatrix} I_{\lambda_{1m}} & \cdots & I_{\lambda_{1m}\lambda_{hm}} & \cdots & I_{\lambda_{1m}\lambda_{H_m m}} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ I_{\lambda_{hm}\lambda_{1m}} & \cdots & I_{\lambda_{hm}} & \cdots & I_{\lambda_{hm}\lambda_{H_m m}} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ I_{\lambda_{H_m m}\lambda_{1m}} & \cdots & I_{\lambda_{H_m m}\lambda_{hm}} & \cdots & I_{\lambda_{H_m m}} \end{pmatrix}; \quad -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\lambda_{hm}\partial\lambda_{h'm}}\right) = \begin{cases} I_{\lambda_{hm}\lambda_{h'm}} =, & h \neq h'; \\ I_{\lambda_{hm}}, & h = h'. \end{cases}$
- $\mathbf{I}_{\lambda_m\lambda_{m'}} = \begin{pmatrix} I_{\lambda_{1m}\lambda_{1m'}} & \cdots & I_{\lambda_{1m}\lambda_{h'm'}} & \cdots & I_{\lambda_{1m}\lambda_{H_{m'}m'}} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ I_{\lambda_{hm}\lambda_{1m'}} & \cdots & I_{\lambda_{hm}\lambda_{h'm'}} & \cdots & I_{\lambda_{hm}\lambda_{H_{m'}m'}} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ I_{\lambda_{H_m m}\lambda_{1m'}} & \cdots & I_{\lambda_{H_m m}\lambda_{h'm'}} & \cdots & I_{\lambda_{H_m m}\lambda_{H_{m'}m'}} \end{pmatrix}; \quad -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\lambda_{hm}\partial\lambda_{h'm'}}\right) = I_{\lambda_{hm}\lambda_{h'm'}}.$

According to above descriptions, the number of parameters in three-level nested logit model are as follows:

$$q = \sum_{k=1}^{K}(L_k - 1) + \sum_{m=1}^{M} H_m + M = q_1 + q_2 + M,$$

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 the dissimilarity parameters of the nests, hence, the information matrix (5.5) is a symmetric positive semi definite $q \times q$-matrix.

In order to define a design based on the three-level NMNL model, let us consider the following experiments:

$$\sum_{m=1}^{M}\sum_{h=1}^{H_m}\mathcal{J}_{hm}/J_s/S, \tag{5.6}$$

where $\mathcal{J}=\sum_{m=1}^{M}\sum_{h=1}^{H_m}\mathcal{J}_{hm}$ denotes the total number of alternatives in population and $J_s=\sum_{m=1}^{M}\sum_{h=1}^{H_m}J_{hms}$ denotes the number of alternatives in choice set s, selected from population, $\mathcal{J}$, randomly. In particular, suppose that $J_s=J;\ \forall s$ then in this case, there will be $\mathcal{S}$ choice sets each with J alternatives. However, based on (5.6), the $q\leq S\leq\mathcal{S}$ choice sets can be considered instead of $\mathcal{S}$. Also, the number of alternatives, which will be selected from subnests, may vary. Thus, there are different classes can be used in order to obtain a sample with size J from the population by (as similar as Chapter 4):

$$\mathcal{S}_n=\begin{pmatrix}\mathcal{J}_{11}\\ J_{n11s}\end{pmatrix}\cdots\begin{pmatrix}\mathcal{J}_{H_11}\\ J_{nH_11s}\end{pmatrix}\cdots\begin{pmatrix}\mathcal{J}_{hm}\\ J_{nhms}\end{pmatrix}\cdots\begin{pmatrix}\mathcal{J}_{1M}\\ J_{n1Ms}\end{pmatrix}\cdots\begin{pmatrix}\mathcal{J}_{H_MM}\\ J_{nH_MMs}\end{pmatrix}=\prod_{m=1}^{M}\prod_{h=1}^{H_m}\begin{pmatrix}\mathcal{J}_{hm}\\ J_{nhms}\end{pmatrix}, \tag{5.7}$$

where $\sum_{m=1}^{M}\sum_{h=1}^{H_m}J_{nhms}=J;\forall n\in N,\forall s\in\mathcal{S}_n$ and $\mathcal{S}_n$ is the number of choice sets, each including $J_s=J$ alternatives. Based on class n to create an experiment, J_s can be rewrite as $J_{ns}=\sum_{m=1}^{M}\sum_{h=1}^{H_m}J_{nhms}$, where $J_{ns}=J;\forall s\in\mathcal{S}_n,\forall n\in N$ and $J_{nhms}=J_{nhms'};\forall s\neq s'\in\mathcal{S}_n$ but J_{nhms} and $J_{n'hms'}$ (for different class and different choice set) may be equal or not equal. According to reduce the total number of choice sets ($\mathcal{S}$) to a reasonable number (S), we reduce $\mathcal{S}_n$ to S_n in each class, where $q\leq S_n\leq\mathcal{S}_n;\forall n\in N$ (avoiding singular information matrix) then consider $\sum_{m=1}^{M}\sum_{h=1}^{H_m}\mathcal{J}_{hm}/J/S_n\ ,\ \ q\leq S_n;\ \forall n\in N$ instead of (5.6). This involves choosing S_n choice sets each of them with J alternatives in each class.

According to the type of experiment, it is possible that $\mathcal{S}_n<q;\forall n\in N$. In this situation, the information matrix of design ξ_n ($\xi_n\in\Xi_n;\forall n\in N$, where $\Xi_n\subset\Xi$), which consists of choice sets $\mathbb{C}_{n1},\ldots,\mathbb{C}_{ns},\ldots,\mathbb{C}_{n\mathcal{S}_n}$ may be singular. In such a case and in order to avoid a singularity information matrix, we must combine them (ξ_n) together to create a new design. Although, to obtain optimal design (totally) we must combine all of $\xi_n;\forall n\in N$ in a design as follow:

$$\xi=\sum_{n=1}^{N}\alpha_n\xi_n,$$

where the information matrix of design ξ is calculated by:

$$\mathbf{I}(\xi,\boldsymbol{\theta})=\sum_{n=1}^{N}\alpha_n\cdot\mathbf{I}(\xi_n,\boldsymbol{\theta});\quad\sum_{n=1}^{N}\alpha_n=1,\ \alpha_n\geq 0;\forall n\in N.$$

Calculating the information matrix $\mathbf{I}(\xi,\boldsymbol{\theta})$, we must first calculate the information matrix of each choice set, means that $\mathbf{I}(\mathbb{C}_{ns},\boldsymbol{\theta})$, because of $\mathbf{I}(\xi_n,\boldsymbol{\theta})=\sum_{s=1}^{\mathcal{S}_n}w_{ns}\cdot\mathbf{I}(\mathbb{C}_{ns},\boldsymbol{\theta})$. Therefore, to obtain the information matrix related of choice set $\mathbb{C}_s$ for a special class (ignoring index n), Equation (5.5) will be considered.

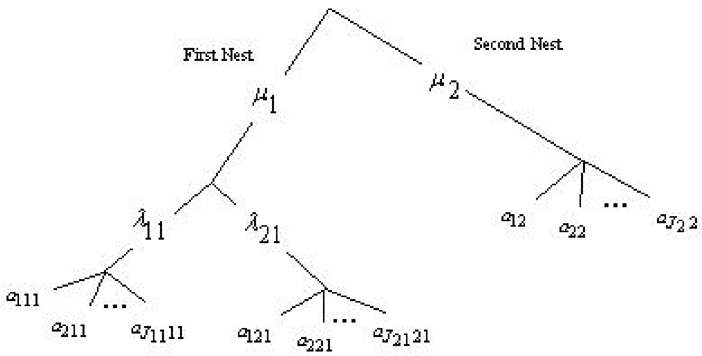

Figure 5.2: NMNL model: There are two nests, the first nest includes two sub-nests (with $\mathcal{J}_{11}$ and $\mathcal{J}_{21}$ alternatives) and the second does not have any sub-nest with $\mathcal{J}_2$ alternatives

Lemma 5.1. *The information matrix is related to a three-level nested logit model (choice set $\mathbb{C}_s$) with two nests, the first nest has two sub-nests with J_{11s} and J_{21s} alternatives and the second, J_{12s} alternatives (Table 5.1 and Figure 5.2 denote a population with $\mathcal{J}_{11}+\mathcal{J}_{21}+\mathcal{J}_{12}$ alternatives, where $\tilde{a}_{jhm}$ denotes the j^{th} alternative in sub-nest h of nest m) is calculated as follows:*

$$\mathbf{I}(\mathbb{C}_s,\boldsymbol{\theta}) = -E(\frac{\partial^2\ell(\mathbb{C}_s,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}) = \begin{pmatrix} \mathbf{I}_{11s} & \mathbf{I}_{12s} & \mathbf{I}_{13s} & \mathbf{I}_{14s} & \mathbf{I}_{15s} \\ \mathbf{I}_{12s}^T & I_{22s} & I_{23s} & I_{24s} & I_{25s} \\ \mathbf{I}_{13s}^T & I_{23s} & I_{33s} & I_{34s} & I_{35s} \\ \mathbf{I}_{14s}^T & I_{24s} & I_{34s} & I_{44s} & I_{45s} \\ \mathbf{I}_{15s}^T & I_{25s} & I_{35s} & I_{45s} & I_{55s} \end{pmatrix},$$

where (with respect to Information matrix (5.5)):

- $\mathbf{I}_\beta = \mathbf{I}_{11s},\ \mathbf{I}_{\beta\mu_1} = \mathbf{I}_{12s},\ \mathbf{I}_{\beta\mu_2} = \mathbf{I}_{13s},\ \mathbf{I}_{\beta\lambda_1} = \mathbf{I}_{14s},\ \mathbf{I}_{\beta\lambda_2} = \mathbf{I}_{15s},$
- $\mathbf{I}_\mu = \begin{pmatrix} I_{\mu_1} = I_{22s} & I_{\mu_1\mu_2} = I_{23s} \\ I_{\mu_1\mu_2} = I_{23s} & I_{\mu_2} = I_{33s} \end{pmatrix},\ \ I_{\mu_1\lambda_1} = I_{24s},\ I_{\mu_1\lambda_2} = I_{24s},$
- $\mathbf{I}_\lambda = \begin{pmatrix} I_{\lambda_1} = I_{44s} & I_{\lambda_1\lambda_2} = I_{45s} \\ I_{\lambda_1\lambda_2} = I_{45s} & I_{\lambda_2} = I_{55s} \end{pmatrix}.$

In this situation, $\boldsymbol{\theta} = (\boldsymbol{\beta},\mu_1,\mu_2,\lambda_{11},\lambda_{21})^T$. For simplicity, suppose that $\lambda_1 = \lambda_{11}, \lambda_2 = \lambda_{21}$ then $\boldsymbol{\theta} = (\beta_1,\ldots,\beta_{q_1},\mu_1,\mu_2,\lambda_1,\lambda_2)^T$, where $q_2 = 2$ and $M = 2$.

Now, with respect to the following considerations:

- $\mathbf{B}_{h|1s} = \mathbf{F}_{h1s}^T\mathbf{P}_{.|h1s}\mathbf{F}_{h1s};\ h = 1,2,$

- $\mathbf{B}_{1|2s} = \mathbf{F}_{12s}^T \mathbf{P}_{.|12s} \mathbf{F}_{12s} = \mathbf{B}_{2s}$,
- $\mathbf{A}_{h|1s} = \mathbf{F}_{h1s}^T \mathbf{p}_{.|h1s};\ h = 1, 2$,
- $\mathbf{A}_{1|2s} = \mathbf{F}_{12s}^T \mathbf{p}_{.|12s} = \mathbf{A}_{2s}$,
- $\mathbf{F}_{h1s} = \left(\mathbf{f}(a_{1h1s}), \ldots, \mathbf{f}(a_{jh1s}), \ldots, \mathbf{f}(a_{J_{h1s}h1s})\right)^T;\ h = 1, 2$,
  $\mathbf{f}(a_{jh1s}) = (\mathbf{f}_1(a_{jh1s}), \ldots, \mathbf{f}_k(a_{jh1s}), \ldots, \mathbf{f}_K(a_{jh1s}))^T$,
  $\mathbf{f}_k(a_{jh1s}) = (f_{k1}(a_{jh1s}), \ldots, f_{k\ell}(a_{jh1s}), \ldots, f_{kL_k-1}(a_{jh1s}))^T;\ q_1 = \sum_{k=1}^K L_k - 1$,
  Let $f_{11}(a_{jh1s}) = f_1(a_{jh1s}), \ldots, f_{KL_K-1}(a_{jh1s}) = f_{q_1}(a_{jh1s})$ then:
- $\mathbf{f}(a_{jh1s}) = (f_1(a_{jh1s}), \ldots, f_r(a_{jh1s}), \ldots, f_{q_1}(a_{jh1s}))^T;\quad h = 1, 2$
- $\mathbf{F}_{12s} = \left(\mathbf{f}(a_{112s}), \ldots, \mathbf{f}(a_{j12s}), \ldots, \mathbf{f}(a_{J_{12s}12s})\right)^T;$
- $\mathbf{f}(a_{j12s}) = \left(f_1(a_{j12s}), \ldots, f_r(a_{j12s}), \ldots, f_{q_1}(a_{j12s})\right)^T$
- $\mathbf{P}_{.|h1s} = diag\left(p_{1|h1s}, \ldots, p_{j|h1s}, \ldots, p_{J_{h1s}|h1s}\right);\ h = 1, 2$,
- $\mathbf{P}_{.|12s} = diag\left(p_{1|12s}, \ldots, p_{j|12s}, \ldots, p_{J_{12s}|12s}\right)$,
- $\mathbf{p}_{.|h1s} = \left(p_{1|h1s}, \ldots, p_{j|h1s}, \ldots, p_{J_{h1s}|h1s}\right)^T;\ h = 1, 2$,
- $\mathbf{p}_{.|12s} = \left(p_{1|12s}, \ldots, p_{j|12s}, \ldots, p_{J_{12s}|12s}\right)^T$,
- $b_{h|1s} = \ln\left(\sum_{j=1}^{J_{h1s}} \exp\left(\frac{\mathbf{f}^T(a_{jh1s})\boldsymbol{\beta}}{\lambda_1}\right)\right);\ h = 1, 2$,
- $b_{2s} = \ln\left(\sum_{j=1}^{J_{12s}} \exp\left(\frac{\mathbf{f}^T(a_{j12s})\boldsymbol{\beta}}{\mu_2}\right)\right)$
- $p_{j|11s} = \dfrac{\exp\left(\frac{\mathbf{f}^T(a_{j11s})\boldsymbol{\beta}}{\lambda_1}\right)}{\sum_{l=1}^{J_{11s}} \exp\left(\frac{\mathbf{f}^T(a_{l11s})\boldsymbol{\beta}}{\lambda_1}\right)};\ j = 1, 2, \ldots, J_{11s}$,
- $p_{j|21s} = \dfrac{\exp\left(\frac{\mathbf{f}^T(a_{j21s})\boldsymbol{\beta}}{\lambda_2}\right)}{\sum_{l=1}^{J_{21s}} \exp\left(\frac{\mathbf{f}^T(a_{l21s})\boldsymbol{\beta}}{\lambda_2}\right)};\ j = 1, 2, \ldots, J_{21s}$,
- $p_{j|12s} = \dfrac{\exp\left(\frac{\mathbf{f}^T(a_{j12s})\boldsymbol{\beta}}{\mu_2}\right)}{\sum_{l=1}^{J_{12s}} \exp\left(\frac{\mathbf{f}^T(a_{l12s})\boldsymbol{\beta}}{\mu_2}\right)};\ j = 1, 2, \ldots, J_{12s}$,
- $b_{21|1s} = \ln\left(\left(\sum_{j=1}^{J_{11s}} \exp\left(\frac{\mathbf{f}^T(a_{j11s})\boldsymbol{\beta}}{\lambda_1}\right)\right)^{\frac{\lambda_1}{\mu_1}} + \left(\sum_{j=1}^{J_{21s}} \exp\left(\frac{\mathbf{f}^T(a_{j21s})\boldsymbol{\beta}}{\lambda_2}\right)\right)^{\frac{\lambda_2}{\mu_1}}\right)$

Table 5.1: NMNL Model: There are two nests, the first nest includes two sub-nests (with $\mathcal{J}_{11}$ and $\mathcal{J}_{21}$ alternatives) and the second does not have any sub-nest with $\mathcal{J}_2$ alternatives, where $\tilde{a}_{jhm}$ denotes the j^{th} alternative in sub-nest h of nest m.

First Nest(1)		Second Nest(2)
Sub-Nest(1)	Sub-Nest(2)	$\tilde{a}_{112},\ldots,\tilde{a}_{j12},\ldots,\tilde{a}_{\mathcal{J}_{12}12}$
$\tilde{a}_{111},\ldots,\tilde{a}_{j11},\ldots,\tilde{a}_{\mathcal{J}_{11}11}$	$\tilde{a}_{121},\ldots,\tilde{a}_{j21},\ldots,\tilde{a}_{\mathcal{J}_{21}21}$	

- $p_{1s} = \dfrac{\left(\sum_{h=1}^{2}\left(\sum_{l=1}^{J_{h1s}}\exp\left(\frac{\mathbf{f}^T(a_{lh1s})\boldsymbol{\beta}}{\lambda_h}\right)\right)^{\frac{\lambda_h}{\mu_1}}\right)^{\mu_1}}{\left(\sum_{j=1}^{J_{12s}}\exp\left(\frac{\mathbf{f}^T(a_{j12s})\boldsymbol{\beta}}{\mu_2}\right)\right)^{\mu_2}+\left(\sum_{h=1}^{2}\left(\sum_{l=1}^{J_{h1s}}\exp\left(\frac{\mathbf{f}^T(a_{lh1s})\boldsymbol{\beta}}{\lambda_h}\right)\right)^{\frac{\lambda_h}{\mu_1}}\right)^{\mu_1}}; \quad p_{2s} = 1 - p_{1s},$

- $p_{1|1s} = \dfrac{\left(\sum_{l=1}^{J_{11s}}\exp\left(\frac{\mathbf{f}^T(a_{l11s})\boldsymbol{\beta}}{\lambda_1}\right)\right)^{\frac{\lambda_1}{\mu_1}}}{\sum_{h=1}^{2}\left(\sum_{l=1}^{J_{h1s}}\exp\left(\frac{\mathbf{f}^T(a_{lh1s})\boldsymbol{\beta}}{\lambda_h}\right)\right)^{\frac{\lambda_h}{\mu_1}}}; \quad p_{2|1s} = 1 - p_{1|1s}$

the elements of the above information matrix, defined by Lemma 5.1 are as follow:

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

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

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

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

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

$\boxed{I_{22s}} = \frac{1}{\mu_1^4} \cdot p_{1s} \cdot p_{1|1s} \cdot p_{2|1s} \cdot (\lambda_1 \cdot b_{1|1s} - \lambda_2 \cdot b_{2|1s})^2 + \frac{1}{\mu_1^2} \cdot p_{1s} \cdot p_{2s}(\lambda_1 \cdot p_{1|1s} \cdot b_{1|1s} + \lambda_2 \cdot p_{2|1s} \cdot b_{2|1s})^2 - \frac{2}{\mu_1} \cdot p_{1s} \cdot p_{2s} \cdot b_{21|1s}(\lambda_1 \cdot p_{1|1s} \cdot b_{1|1s} + \lambda_2 \cdot p_{2|1s} \cdot b_{2|1s}) + p_{1s} \cdot p_{2s} \cdot b_{21|1s}^2,$

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

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

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

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

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

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

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

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

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

Here, $\mathbf{I}_{11s}$ is a $q_1 \times q_1$-matrix and $\mathbf{I}_{12s}$, $\mathbf{I}_{13s}$, $\mathbf{I}_{14s}$ and $\mathbf{I}_{15s}$ are $q_1 \times 1$-matrix and the other elements, which have been denoted by $I_{..s}$ are 1×1 matrix (scalar).

In this situation and with respect to class n we will have:

$$\mathcal{S}_n = \begin{pmatrix} \mathcal{J}_{11} \\ J_{n11s} \end{pmatrix}\begin{pmatrix} \mathcal{J}_{21} \\ J_{n21s} \end{pmatrix}\begin{pmatrix} \mathcal{J}_{21} \\ J_{n21s} \end{pmatrix}; \forall n \in N.$$

Corollary 5.1. *When $\boldsymbol{\beta} = \mathbf{0}$ then the above information matrix (Lemma 5.1) should be rewritten by (Considering a special class, we ignore index n):*

- $\mathbf{B}_{h|1s} = \frac{1}{J_{h1s}} \left(\mathbf{F}^T_{h1s} \mathbf{I}_{J_{h1s}} \mathbf{F}_{h1s} \right); \quad h = 1, 2,$
- $\mathbf{A}_{h|1s} = \frac{1}{J_{h1s}} \mathbf{F}^T_{h1s} \mathbf{1}_{J_{h1s}}; \quad h = 1, 2,$
- $\mathbf{B}_{1|2s} = \frac{1}{J_{12s}} \left(\mathbf{F}^T_{12s} \mathbf{I}_{J_{12s}} \mathbf{F}_{12s} \right) = \mathbf{B}_{2s},$
- $\mathbf{A}_{1|2s} = \frac{1}{J_{12s}} \mathbf{F}^T_{12s} \mathbf{1}_{J_{12s}} = \mathbf{A}_{2s},$
- $\mathbf{F}_{h1s} = (\mathbf{f}(a_{1h1s}), \ldots, \mathbf{f}(a_{jh1s}), \ldots, \mathbf{f}(a_{J_{h1s}h1s}))^T; \quad h = 1, 2,$
- $\mathbf{f}(a_{jh1s}) = (f_1(a_{jh1s}), \ldots, f_r(a_{jh1s}), \ldots, f_{q_1}(a_{jh1s}))^T; \quad h = 1, 2, \; j = 1, 2, \ldots, J_{h1s},$
- $\mathbf{F}_{12s} = \left(\mathbf{f}(a_{112s}), \ldots, \mathbf{f}(a_{j12s}), \ldots, \mathbf{f}_(a_{J_{12s}12s}) \right)^T,$
- $\mathbf{f}^T(a_{j12s}) = (f_1(a_{j12s}), \ldots, f_r(a_{j12s}), \ldots, f_{q_1}(a_{j12s}))\,; j = 1, 2, \ldots, J_{12s},$
- $b_{1|1s} = \ln(J_{11s}), \;\; b_{2|1s} = \ln(J_{21s}), \;\; b_{2s} = \ln(J_{12s}), \; b_{21|1s} = \ln\left((J_{11s})^{\frac{\lambda_1}{\mu_1}} + (J_{21s})^{\frac{\lambda_2}{\mu_1}} \right),$
- $p_{1s} = \dfrac{\left(\sum_{h=1}^2 (J_{h1s})^{\frac{\lambda_h}{\mu_1}} \right)^{\mu_1}}{(J_{2s})^{\mu_2} + \left(\sum_{h=1}^2 (J_{h1s})^{\frac{\lambda_h}{\mu_1}} \right)^{\mu_1}}; \qquad p_{2s} = 1 - p_{1s},$
- $p_{1|1s} = \dfrac{(J_{11s})^{\frac{\lambda_1}{\mu_1}}}{\sum_{h=1}^2 (J_{h1s})^{\frac{\lambda_h}{\mu_1}}}; \quad p_{2|1s} = 1 - p_{1|1s}.$

Thus we will have that:

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

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

$$\boxed{\mathbf{I}_{13s}} = -b_{2s} \cdot p_{1s} \cdot p_{2s} \left(p_{1|1s} \cdot \mathbf{A}_{1|1s} + p_{2|1s} \cdot \mathbf{A}_{\mathbf{2|1s}} - \mathbf{A}_{\mathbf{2s}} \right),$$

$\boxed{\mathbf{I}_{14s}}$
$= \frac{1}{\mu_1^2} \cdot b_{1|1s} \cdot p_{1s} \cdot p_{1|1s} \cdot p_{2|1s} \left(\mathbf{A}_{1|1s} - \mathbf{A}_{2|1s}\right) + b_{1|1s} \cdot p_{1s} \cdot p_{2s} \cdot p_{1|1s} \left(p_{1|1s} \cdot \mathbf{A}_{1|1s} + p_{2|1s} \cdot \mathbf{A}_{2|1s} - \mathbf{A}_{2s}\right),$

$\boxed{\mathbf{I}_{15s}}$
$= \frac{1}{\mu_1^2} \cdot b_{2|1s} \cdot p_{1s} \cdot p_{1|1s} \cdot p_{2|1s} \cdot \left(\mathbf{A}_{2|1s} - \mathbf{A}_{1|1s}\right) + b_{2|1s} \cdot p_{1s} \cdot p_{2s} \cdot p_{2|1s} \cdot \left(p_{2|1s} \cdot \mathbf{A}_{2|1s} + p_{1|1s} \cdot \mathbf{A}_{1|1s} - \mathbf{A}_{2s}\right),$

$\boxed{I_{22s}} = \frac{1}{\mu_1^4} \cdot p_{1s} \cdot p_{1|1s} \cdot p_{2|1s} \cdot (\lambda_1 \cdot b_{1|1s} - \lambda_2 \cdot b_{2|1s})^2 + \frac{1}{\mu_1^2} \cdot p_{1s} \cdot p_{2s}(\lambda_1 \cdot p_{1|1s} \cdot b_{1|1s} + \lambda_2 \cdot p_{2|1s} \cdot b_{2|1s})^2 - \frac{2}{\mu_1} \cdot p_{1s} \cdot p_{2s} \cdot b_{21|1s}(\lambda_1 \cdot p_{1|1s} \cdot b_{1|1s} + \lambda_2 \cdot p_{2|1s} \cdot b_{2|1s}) + p_{1s} \cdot p_{2s} \cdot b_{21|1s}^2,$

$\boxed{I_{23s}} = -\frac{1}{\mu_1} \cdot b_{2s} \cdot p_{1s} \cdot p_{2s} \cdot (\mu_1 \cdot b_{21|1s} - \lambda_1 \cdot p_{1|1s} \cdot b_{1|1s} - \lambda_2 \cdot p_{2|1s} \cdot b_{2|1s}),$

$\boxed{I_{24s}} = -\frac{1}{\mu_1^3} \cdot b_{1|1s} \cdot p_{1s} \cdot p_{1|1s} \cdot p_{2|1s}(\lambda_1 \cdot b_{1|1s} - \lambda_2 \cdot b_{2|1s}) - \frac{1}{\mu_1} \cdot b_{1|1s} \cdot p_{1s} \cdot p_{2s} \cdot p_{1|1s} \left(\lambda_1 \cdot p_{1|1s} \cdot b_{1|1s} + \lambda_2 \cdot p_{2|1s} \cdot b_{2|1s} - \mu_1 \cdot b_{21|1s}\right),$

$\boxed{I_{25s}} = -\frac{1}{\mu_1^3} \cdot b_{2|1s} \cdot p_{1s} \cdot p_{1|1s} \cdot p_{2|1s}(\lambda_2 \cdot b_{2|1s} - \lambda_1 \cdot b_{1|1s}) - \frac{1}{\mu_1} \cdot b_{2|1s} \cdot p_{1s} \cdot p_{2s} \cdot p_{2|1s}(\lambda_1 \cdot p_{1|1s} \cdot b_{1|1s} + \lambda_2 \cdot p_{2|1s} \cdot b_{2|1s} - \mu_1 \cdot b_{21|1s}),$

$\boxed{I_{33s}} = b_{2s}^2 \cdot p_{1s} \cdot p_{2s},$

$\boxed{I_{34s}} = -b_{2s} \cdot b_{1|1s} \cdot p_{1s} \cdot p_{2s} \cdot p_{1|1s},$

$\boxed{I_{35s}} = -b_{2s} \cdot b_{2|1s} \cdot p_{1s} \cdot p_{2s} \cdot p_{2|1s},$

$\boxed{I_{44s}} = \frac{1}{\mu_1^2} \cdot b_{1|1s}^2 \cdot p_{1s} \cdot p_{1|1s} \cdot (p_{2|1s} + \mu_1^2 \cdot p_{2s} \cdot p_{1|1s}),$

$\boxed{I_{45s}} = -\frac{1}{\mu_1^2} \cdot b_{1|1s} \cdot b_{2|1s} \cdot p_{1s} \cdot p_{1|1s} \cdot p_{2|1s} \cdot (1 - p_{2s} \cdot \mu_1^2),$

$\boxed{I_{55s}} = \frac{1}{\mu_1^2} \cdot b_{2|1s}^2 \cdot p_{1s} \cdot p_{2|1s} \cdot (p_{1|1s} + \mu_1^2 \cdot p_{2s} \cdot p_{2|1s}),$

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.

5.3 D-Optimal Criterion

Taking into account (5.6) and (5.7), consider the following designs to fit the model, which was introduced in Table 5.1:

$$\xi_n = \left\{ \begin{array}{cccc} \mathbb{C}_{n1} & \mathbb{C}_{n2} & \cdots & \mathbb{C}_{n\mathcal{S}_n} \\ w_{n1} & w_{n2} & \cdots & w_{n\mathcal{S}_n} \end{array} \right\} \in \Xi_n; n = 1, 2, \ldots, N. \tag{5.8}$$

Table 5.2: NMNL Model: There are two nests, the first nest includes two sub-nest each with two alternatives and the second nest does not sub-nest and contains two alternatives.

First nest(1)		Second nest(2)
Sub-nest(1)	Sub-nest(2)	
$\tilde{a}_{111}$, $\tilde{a}_{211}$	$\tilde{a}_{121}$, $\tilde{a}_{221}$	$\tilde{a}_{112}$, $\tilde{a}_{212}$

The information matrix of the design (5.8) is calculated as follows:

$$\mathbf{I}(\xi_n, \boldsymbol{\theta}) = \sum_{s=1}^{\mathcal{S}_n} w_{ns} \cdot \mathbf{I}(\mathbb{C}_{ns}, \boldsymbol{\theta}), \tag{5.9}$$

where w_{ns} is the weight (frequency) of the choice set $\mathbb{C}_{ns}$. Moreover, the local D-optimality criterion at $\boldsymbol{\theta}$ is considered by:

$$\Psi(\xi, \boldsymbol{\theta}) = (\det(\mathbf{I}(\xi, \boldsymbol{\theta})))^{-1},$$

where $\mathbf{I}(\xi, \boldsymbol{\theta})) = \sum_{n=1}^{N} \alpha_n \cdot \mathbf{I}(\xi_n, \boldsymbol{\theta}))$; $\sum_{n=1}^{N} \alpha_n = 1, \quad 0 \leq \alpha_n \leq 1; \forall n \in N$. This case, in which ξ^* minimizes the local D-optimal criterion, is called the locally D-optimal design and will be obtained by the solution:

$$\xi^* = \arg\min_{\xi \in \Xi} \Psi(\xi, \boldsymbol{\theta}), \tag{5.10}$$

where $\Xi_n \subset \Xi$ and

$$\Xi_n = \left\{ \xi_n | \sum_{s=1}^{\mathcal{S}_n} w_{ns} = 1, 0 \leq w_{ns} \leq 1; s \in \mathcal{S}_n \right\}.$$

5.4 Example

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 5.2; $\sum_{m=1}^{2} \sum_{h=1}^{H_m} \mathcal{J}_{hm} = 2 + 2 + 2 = \mathcal{J}$, where $(1, 1, 1)$, $(1, 1, -1)$, $(1, -1, 1)$, $(1, -1, -1)$, $(-1, 1, 1)$, $(-1, 1, -1)$ characterize alternatives $\tilde{a}_{111}, \tilde{a}_{211}, \tilde{a}_{121}, \tilde{a}_{221}, \tilde{a}_{112}$ and $\tilde{a}_{212}$, respectively). In this situation, we consider experiments, which include six choice sets each with five alternatives (Table 5.3, Where a_{jnhms} denotes j^{th} alternatives in class n of sub-nest h of the nest m in choice set s), where there are three classes each with two choice sets as follow:

- $\mathcal{S}_1 = \binom{2}{2}\binom{2}{2}\binom{2}{1}$,
- $\mathcal{S}_2 = \binom{2}{2}\binom{2}{1}\binom{2}{2}$,

Table 5.3: NMNL Model (two nests, the first nest with two sub-nests and the second does not sub-nest): There are three classes, $N = 3$, each with two choice sets ($\mathbb{C}_{ns} = \mathbb{C}_{s'}; \forall n \in N, s \in \mathcal{S}_n, s' = 1, \ldots, 6$), based on a_{jnhms}, which denotes the j^{th} alternative in class n of the sub-nest h of the nest m in choice set s.

Choice Set ($\mathbb{C}_{ns}$)	First Nest(1)		Second Nest(2)
	Sub-Nest(1)	Sub-Nest(2)	
$\mathbb{C}_{11} = \mathbb{C}_1$	$a_{11111}(\tilde{a}_{111}), a_{21111}(\tilde{a}_{211})$	$a_{11211}(\tilde{a}_{121}), a_{21211}(\tilde{a}_{221})$	$a_{11121}(\tilde{a}_{112})$
$\mathbb{C}_{12} = \mathbb{C}_2$	$a_{11112}(\tilde{a}_{111}), a_{21112}(\tilde{a}_{211})$	$a_{11212}(\tilde{a}_{121}), a_{21212}(\tilde{a}_{221})$	$a_{11122}(\tilde{a}_{212})$
$\mathbb{C}_{21} = \mathbb{C}_3$	$a_{12111}(\tilde{a}_{111}), a_{22111}(\tilde{a}_{211})$	$a_{12211}(\tilde{a}_{121})$	$a_{12121}(\tilde{a}_{112}), a_{22121}(\tilde{a}_{212})$
$\mathbb{C}_{22} = \mathbb{C}_4$	$a_{12112}(\tilde{a}_{111}), a_{22112}(\tilde{a}_{211})$	$a_{12212}(\tilde{a}_{221})$	$a_{12122}(\tilde{a}_{112}), a_{22122}(\tilde{a}_{212})$
$\mathbb{C}_{31} = \mathbb{C}_5$	$a_{13111}(\tilde{a}_{111})$	$a_{13211}(\tilde{a}_{121}), a_{23211}(\tilde{a}_{221})$	$a_{13121}(\tilde{a}_{112}), a_{23121}(\tilde{a}_{212})$
$\mathbb{C}_{32} = \mathbb{C}_6$	$a_{13112}(\tilde{a}_{211})$	$a_{13212}(\tilde{a}_{121}), a_{23212}(\tilde{a}_{221})$	$a_{13122}(\tilde{a}_{112}), a_{23122}(\tilde{a}_{212})$

- $\mathcal{S}_3 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix}$.

In this case, three classes ($N = 3$) we found to define the design and because of $\mathcal{S}_n < 7; \forall n = 1, 2, 3$, we have to combine them in order to define a suitable design. Thus there are six choice sets (Table 5.3) with their design matrixes, as shown by Table 5.4. In this situation, $\boldsymbol{\theta} = (\beta_1, \beta_2, \beta_3, \mu_1, \mu_2, \lambda_1, \lambda_2)^T$ is full parameters vector. In this case and keeping to RUM conditions (Subsection 2.2.2), we will encounter the two conditions as follows:

1. $\mu_m \leq \frac{1}{1-p_{ms}}; m = 1, 2$
2. $\lambda_{h1} \leq \frac{\mu_1}{(1+\mu_1 p_{h|1s})(1-p_{1s}))}; h = 1, 2$ and $\forall s \in \mathcal{S}$,

where $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}$.
For estimating the parameters of the model, which have been described on Table 5.2 and based on experiments $2 + 2 + 2/5/6$ (considering six choice sets each with five alternatives), consider the following design:

$$\xi = \left\{ \begin{array}{cccccc} \mathbb{C}_{11} & \mathbb{C}_{12} & \mathbb{C}_{21} & \mathbb{C}_{22} & \mathbb{C}_{31} & \mathbb{C}_{32} \\ w_{11} & w_{12} & w_{21} & w_{22} & w_{31} & w_{32} \end{array} \right\} \in \Xi. \tag{5.11}$$

The information matrix of design (5.11) is calculated by $\mathbf{I}(\xi, \boldsymbol{\theta}) = \sum_{n=1}^{3} \sum_{s=1}^{2} w_{ns} \cdot \mathbf{I}(\mathbb{C}_{ns}, \boldsymbol{\theta})$.

Specifically, let $\boldsymbol{\beta} = \mathbf{0}$. Now, according to Lemma 5.1 and Corollary 5.1, the following assumptions are used to calculate the elements of the information matrix $\mathbf{I}(\mathbb{C}_{ns}, \boldsymbol{\theta}); \forall s \in \mathcal{S}_n$, $n = 1, 2, 3$ (To adapt to Lemma 5.1 we consider $\mathbb{C}_{ns} = \mathbb{C}_{s'}$; $s' = 1, 2, \ldots, \sum_{n=1}^{N} \sum_{s=1}^{\mathcal{S}_n}$):

For $\mathbb{C}_1 = \mathbb{C}_{11}$:

Table 5.4: MNL Model (two nests, the first nest with two sub-nests and the second does not sub-nest): The characterizes of three attributes each with two levels; considering six choice sets each with five alternatives, $\mathbb{C}_{ns} = \mathbb{C}_{s'}; \forall n \in N, s \in \mathcal{S}_n, s' = 1\ldots, 6$.

Choice Set($\mathbb{C}_{ns} = \mathbb{C}_{s'}$)	$(\mathbf{F}_{hms'}) = \mathbf{F}_{11s'}$	$(\mathbf{F}_{hms'}) = \mathbf{F}_{21s'}$	$(\mathbf{F}_{ms}) = \mathbf{F}_{2s}$
$\mathbb{C}_{11} = \mathbb{C}_1$	$\begin{bmatrix} +1 & +1 & +1 \\ +1 & +1 & -1 \end{bmatrix}$	$\begin{bmatrix} +1 & -1 & +1 \\ +1 & -1 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & +1 & +1 \end{bmatrix}$
$\mathbb{C}_{12} = \mathbb{C}_2$	$\begin{bmatrix} +1 & +1 & +1 \\ +1 & +1 & -1 \end{bmatrix}$	$\begin{bmatrix} +1 & -1 & +1 \\ +1 & -1 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & +1 & -1 \end{bmatrix}$
$\mathbb{C}_{21} = \mathbb{C}_3$	$\begin{bmatrix} +1 & +1 & +1 \\ +1 & +1 & -1 \end{bmatrix}$	$\begin{bmatrix} +1 & -1 & +1 \end{bmatrix}$	$\begin{bmatrix} -1 & +1 & +1 \\ -1 & +1 & -1 \end{bmatrix}$
$\mathbb{C}_{22} = \mathbb{C}_4$	$\begin{bmatrix} +1 & +1 & +1 \\ +1 & +1 & -1 \end{bmatrix}$	$\begin{bmatrix} +1 & -1 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & +1 & +1 \\ -1 & +1 & -1 \end{bmatrix}$
$\mathbb{C}_{31} = \mathbb{C}_5$	$\begin{bmatrix} +1 & +1 & +1 \end{bmatrix}$	$\begin{bmatrix} +1 & -1 & +1 \\ +1 & -1 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & +1 & +1 \\ -1 & +1 & -1 \end{bmatrix}$
$\mathbb{C}_{32} = \mathbb{C}_6$	$\begin{bmatrix} +1 & +1 & -1 \end{bmatrix}$	$\begin{bmatrix} +1 & -1 & +1 \\ +1 & -1 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & +1 & +1 \\ -1 & +1 & -1 \end{bmatrix}$

$$\mathbf{A}_{1|11} = \begin{bmatrix} +1 \\ +1 \\ 0 \end{bmatrix}, \quad \mathbf{A}_{2|11} = \begin{bmatrix} +1 \\ -1 \\ 0 \end{bmatrix}, \quad \mathbf{A}_{21} = \begin{bmatrix} -1 \\ +1 \\ +1 \end{bmatrix}, \quad \mathbf{B}_{21} = \begin{bmatrix} +1 & -1 & -1 \\ -1 & +1 & +1 \\ -1 & +1 & +1 \end{bmatrix},$$

$$\mathbf{B}_{1|11} = \begin{bmatrix} +1 & +1 & 0 \\ +1 & +1 & 0 \\ 0 & 0 & +1 \end{bmatrix}, \quad \mathbf{B}_{2|11} = \begin{bmatrix} +1 & -1 & 0 \\ -1 & +1 & 0 \\ 0 & 0 & +1 \end{bmatrix}, \quad b_{1|11} = \ln(2), \quad b_{2|11} = \ln(2),$$

$$b_{21} = 0, \quad b_{21|11} = \ln\left(2^{\frac{\lambda_1}{\mu_1}} + 2^{\frac{\lambda_2}{\mu_1}}\right), \quad p_2^{(1)} = \frac{1}{1+\left(2^{\frac{\lambda_1}{\mu_1}}+2^{\frac{\lambda_2}{\mu_1}}\right)^{\mu_1}}, \quad p_{1|11} = \frac{2^{\frac{\lambda_1}{\mu_1}}}{2^{\frac{\lambda_1}{\mu_1}}+2^{\frac{\lambda_2}{\mu_1}}},$$

$$p_{1|111} = \tfrac{1}{2}, \quad p_{1|211} = \tfrac{1}{2}.$$

For $\mathbb{C}_2 = \mathbb{C}_{12}$:

$$\mathbf{A}_{1|12} = \mathbf{A}_{1|11}, \quad \mathbf{A}_{2|12} = \mathbf{A}_{2|11}, \quad \mathbf{A}_{22} = \begin{bmatrix} -1 \\ +1 \\ -1 \end{bmatrix}, \quad \mathbf{B}_{22} = \begin{bmatrix} +1 & -1 & +1 \\ -1 & +1 & -1 \\ +1 & -1 & +1 \end{bmatrix},$$

$$\mathbf{B}_{1|12} = \mathbf{B}_{1|11}, \quad \mathbf{B}_{2|12} = \mathbf{B}_{2|11}, \quad b_{1|12} = b_{1|11}, \quad b_{2|12} = b_{2|11}, b_{22} = 0,$$

$b_{21|12} = b_{21|11}, \quad p_{1|12} = p_{1|11}, \quad p_{22} = p_{21}, \quad p_{1|111} = p_{1|211} = \frac{1}{2}.$

For $\mathbb{C}_3 = \mathbb{C}_{21}$:

$$\mathbf{A}_{1|13} = \mathbf{A}_{1|11}, \quad \mathbf{A}_{2|13} = \begin{bmatrix} +1 \\ -1 \\ +1 \end{bmatrix}, \quad \mathbf{A}_{23} = \begin{bmatrix} -1 \\ +1 \\ 0 \end{bmatrix}, \quad \mathbf{B}_{2|13} = \begin{bmatrix} +1 & -1 & +1 \\ -1 & +1 & -1 \\ +1 & -1 & +1 \end{bmatrix},$$

$$\mathbf{B}_{23} = \begin{bmatrix} +1 & -1 & 0 \\ -1 & +1 & 0 \\ 0 & 0 & +1 \end{bmatrix}, \quad \mathbf{B}_{1|13} = \mathbf{B}_{1|11}, \quad b_{1|13} = b_{1|11}, \quad b_{2|13} = 0,$$

$$b_{23} = \ln(2), \quad b_{21|13} = \ln\left(2^{\frac{\lambda_1}{\mu_1}} + 1\right), \quad p_{1|113} = p_{1|111}, \quad p_{1|23} = \tfrac{1}{2},$$

$$p_{23} = \frac{2^{\mu_2}}{2^{\mu_2} + \left(2^{\frac{\lambda_1}{\mu_1}} + 1\right)^{\mu_1}}, \quad p_{1|13} = \frac{2^{\frac{\lambda_1}{\mu_1}}}{2^{\frac{\lambda_1}{\mu_1}} + 1}.$$

For $\mathbb{C}_4 = \mathbb{C}_{22}$:

$$\mathbf{A}_{1|14} = \mathbf{A}_{1|11}, \quad \mathbf{A}_{2|14} = \begin{bmatrix} +1 \\ -1 \\ -1 \end{bmatrix}, \quad \mathbf{A}_{24} = \mathbf{A}_{23}, \quad \mathbf{B}_{24} = \mathbf{B}_{23},$$

$$\mathbf{B}_{1|14} = \mathbf{B}_{1|11}, \quad \mathbf{B}_{2|13} = \begin{bmatrix} +1 & -1 & -1 \\ -1 & +1 & +1 \\ -1 & +1 & +1 \end{bmatrix}, \quad b_{1|14} = b_{1|11}, \quad b_{2|14} = 0,$$

$$b_{24} = b_{23}, \quad b_{21|14} = \ln\left(2^{\frac{\lambda_1}{\mu_1}} + 1\right), \quad p_{1|114} = p_{1|111}, \quad p_{1|24} = p_{1|23},$$

$$p_{1|14} = \frac{2^{\frac{\lambda_1}{\mu_1}}}{2^{\frac{\lambda_1}{\mu_1}} + 1}, \quad p_{24} = \frac{2^{\mu_2}}{2^{\mu_2} + \left(2^{\frac{\lambda_1}{\mu_1}} + 1\right)^{\mu_1}}.$$

For $\mathbb{C}_5 = \mathbb{C}_{31}$:

$$\mathbf{A}_{1|15} = \begin{bmatrix} +1 \\ +1 \\ +1 \end{bmatrix}, \quad \mathbf{A}_{2|15} = \mathbf{A}_{2|11}, \quad \mathbf{A}_{25} = \mathbf{A}_{23}, \quad \mathbf{B}_{25} = \mathbf{B}_{23},$$

$$\mathbf{B}_{1|15} = \begin{bmatrix} +1 & +1 & +1 \\ +1 & +1 & +1 \\ +1 & +1 & +1 \end{bmatrix}, \quad \mathbf{B}_{2|15} = \mathbf{B}_{2|11}, \quad b_{1|15} = 0, \quad b_{25} = b_{23},$$

Table 5.5: NMNL Model, $\mu_1 = 2\lambda, \mu_2 = 4\lambda$ (two nests, first nest with two sub-nests and the second does not sub-nest): Locally D-optimal design when $0 < \lambda \leq 0.25$ (based on RUM conditions) with initial values $w_1 = 0.1, w_2 = w_3 = 0.2$ and w.r.t local D-optimality criterion $\Psi(\xi'', \boldsymbol{\theta}_0) = (\det(\mathbf{I}(\xi'', \boldsymbol{\theta}_0)))^{-1}$.

λ	0.01000	0.05000	0.10000	0.15000	0.17000	0.20000	0.25000
w_1^*	0.30920	0.31200	0.31500	0.31800	0.31950	0.32100	0.32500
w_2^*	0.09540	0.09400	0.09250	0.09100	0.09050	0.08990	0.08900
w_3^*	0.09540	0.09400	0.09250	0.09100	0.09000	0.08910	0.08600
$\Psi(\xi'', \boldsymbol{\theta}_0)$	0.00368	0.08985	0.35081	0.77476	0.98940	1.36021	2.11334

$$b_{2|15} = \ln(2), \quad b_{21|15} = \ln\left(2^{\frac{\lambda_2}{\mu_1}} + 1\right), \quad p_{1|215} = \tfrac{1}{2}, \quad p_{1|25} = p_{1|23},$$

$$p_{25} = \frac{2^{\mu_2}}{2^{\mu_2} + \left(2^{\frac{\lambda_2}{\mu_1}} + 1\right)^{\mu_1}}, \quad p_{1|15} = \frac{1}{2^{\frac{\lambda_2}{\mu_1}} + 1}.$$

For $\mathbb{C}_6 = \mathbb{C}_{32}$:

$$\mathbf{A}_{1|16} = \begin{bmatrix} +1 \\ +1 \\ -1 \end{bmatrix}, \quad \mathbf{A}_{2|16} = \mathbf{A}_{2|11}, \quad \mathbf{A}_{26} = \mathbf{A}_{23}, \quad \mathbf{B}_{26} = \mathbf{B}_{23},$$

$$\mathbf{B}_{1|16} = \begin{bmatrix} +1 & +1 & -1 \\ +1 & +1 & -1 \\ -1 & -1 & +1 \end{bmatrix}, \quad \mathbf{B}_{2|16} = \mathbf{B}_{2|11}, \quad b_{1|16} = 0, \quad b_{26} = a_{23},$$

$$b_{2|16} = b_{2|11}, \quad b_{21|16} = \ln\left(2^{\frac{\lambda_2}{\mu_1}} + 1\right), \quad p_{1|216} = p_{1|215}, \quad p_{1|26} = p_{1|23},$$

$$p_{26} = \frac{2^{\mu_2}}{2^{\mu_2} + \left(2^{\frac{\lambda_2}{\mu_1}} + 1\right)^{\mu_1}}, \quad p_{1|16} = \frac{1}{2^{\frac{\lambda_2}{\mu_1}} + 1}.$$

According to the rule of permutation, the levels of third attribute in choice sets $\mathbb{C}_1$ and $\mathbb{C}_2$ will acquire permutation between these two choice sets. Also, a permutation between the two choice sets $\mathbb{C}_3$ and $\mathbb{C}_4$ exists with respect to the permutation of the levels of the third attribute. By permutation, the levels of the third attribute, we will encounter permutation between the two choice sets $\mathbb{C}_5$ and $\mathbb{C}_6$. Thus, we can define a new design to fit the model, already introduced by Table 5.2 and according to Table 5.4, as follows:

$$\xi' = \left\{ \begin{array}{cccccc} \mathbb{C}_2 & \mathbb{C}_1 & \mathbb{C}_4 & \mathbb{C}_3 & \mathbb{C}_6 & \mathbb{C}_5 \\ w_1 & w_2 & w_3 & w_4 & w_5 & w_6 \end{array} \right\} \in \Xi', \tag{5.12}$$

Table 5.6: NMNL Model, $\mu_1 = 0.15, \mu_2 = 0.25$ and $\lambda_1 = 0.1$ (two nests, first nest with two sub-nests and the second does not sub-nest): Locally D-optimal design when $0 < \lambda_2 \leq 0.150$ (based on RUM conditions) with initial values $w_1 = 0.1, w_2 = w_3 = 0.2$ and w.r.t local D-optimality criterion $\Psi(\xi'', \boldsymbol{\theta}_0) = (\det(\mathbf{I}(\xi'', \boldsymbol{\theta}_0)))^{-1}$.

λ_2	0.01000	0.05000	0.06000	0.08000	0.10000	0.12000	0.15000
w_1^*	0.33100	0.33380	0.32440	0.31330	0.30950	0.31030	0.31650
w_2^*	0.00000	0.00010	0.02820	0.06570	0.09530	0.11830	0.14580
w_3^*	0.16900	0.16610	0.14740	0.12100	0.09520	0.07140	0.03770
$\Psi(\xi'', \boldsymbol{\theta}_0)$	0.05879	0.15900	0.17522	0.19926	0.21502	0.22534	0.23478

Table 5.7: NMNL Model, $\mu_1 = \mu_2 = \lambda_1 = \lambda_2 = \lambda$ (two nests, first nest with two sub-nests and the second does not sub-nest): Locally D-optimal design when $0 < \lambda \leq 1$ (based on RUM conditions) with initial values $w_1 = 0.1, w_2 = w_3 = 0.2$ and w.r.t local D-optimality criterion $\Psi(\xi'', \boldsymbol{\theta}_0) = (\det(\mathbf{I}(\xi'', \boldsymbol{\theta}_0)))^{-1}$.

λ	0.05000	0.10000	0.15000	0.20000	0.30000	0.40000	0.50000	0.60000
w_1^*	0.28890	0.29080	0.29260	0.29430	0.29790	0.30110	0.30400	0.30680
w_2^*	0.10560	0.10460	0.10370	0.10290	0.10120	0.09980	0.09880	0.09820
w_3^*	0.10550	0.10460	0.10370	0.10280	0.10090	0.09910	0.09710	0.09500
$\Psi(\xi'', \boldsymbol{\theta}_0)$	0.02689	0.10466	0.22959	0.39868	0.86051	1.47810	2.24747	3.17183

where $\Xi' \subset \Xi$; $\Xi' = \left\{\xi' | \sum_{s'=1}^{6} w_{s'} = 1, \quad w_{s'} \geq 0; \forall s' = 1, 2, \ldots, 6\right\}$.

In this situation, in order to have an equation between the two designs ξ (5.11) and ξ' (5.12), the following design can be considered:

$$\xi'' = \left\{ \begin{array}{cccccc} \mathbb{C}_1 & \mathbb{C}_2 & \mathbb{C}_3 & \mathbb{C}_4 & \mathbb{C}_5 & \mathbb{C}_6 \\ w_1 & w_1 & w_2 & w_2 & w_3 & w_3 \end{array} \right\} \in \Xi. \tag{5.13}$$

Here $w_1 + w_2 + w_3 = \frac{1}{2}$.

For simplicity, suppose that $\lambda_1 = \lambda_2 = \lambda$. Then, it is seen that $p_{21} = p_{22}$, $p_{23} = p_{2s'}; \forall s' = 4, 5, 6$. According to the RUM conditions and the six choice sets, $\mathbb{C}_1$ to $\mathbb{C}_6$, we will encounter the following conditions:

1) $\mu_1 \leq (1 + 2^{\lambda+\mu_1})$,

2) $\mu_1 \cdot 2^{\mu_2} \leq (2^{\mu_2} + (1 + 2^{\frac{\lambda}{\mu_1}})^{\mu_1})$,

3) $\mu_2 \cdot 2^{\lambda+\mu_1} \leq (1 + 2^{\lambda+\mu_1})$,

4) $\mu_2 \cdot (1 + 2^{\frac{\lambda}{\mu_1}})^{\mu_1} \leq (2^{\mu_2} + (1 + 2^{\frac{\lambda}{\mu_1}})^{\mu_1})$,

5) $\lambda(1 + \frac{\mu_1}{2}) \leq \mu_1 \cdot (1 + 2^{\lambda+\mu_1})$,

6) $\lambda \cdot 2^{\mu_2} \cdot (1 + \frac{\mu_1}{2}) \leq \mu_1 \cdot (2^{\mu_2} + (1 + 2^{\frac{\lambda}{\mu_1}})^{\mu_1})$.

Moreover, we know that $\lambda \leq \mu_1$ and according to Table 5.2, it is to be expected that $\mu_2 \geq \mu_1$ (there is any sub-nest, it has been supposed that there are not the alternatives more similar (there are not enough similarity) than the others to make sub-nests). According to above conditions, let us consider $\mu_1 = 2\lambda$ and $\mu_2 = 4\lambda$, thus $\det(\mathbf{I}(\xi, \boldsymbol{\theta})$ will be changed to a more function of λ, w_1 and w_2 where $w_3 = \frac{1}{2} - (w_1 + w_2)$. In this situation, the six above conditions (RUM) will be upheld when $0 < \lambda \leq 0.25$. According to this condition for λ, some locally optimal design has been calculated in Table 5.5. Table 5.5 shows that w_1^* increases as λ increases but w_2^* and w_3^* decrease when λ increases because of the combination of alternatives (and attributes) in the two choice sets $\mathbb{C}_1$ and $\mathbb{C}_2$ are less similar than in the other choice sets. According to Table 5.4, we can observe that two sub-nests of the first nest in the choice sets $\mathbb{C}_1$ and $\mathbb{C}_2$ 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 w_1^* increases as λ increases. In choice sets $\mathbb{C}_3$ and $\mathbb{C}_4$, there are two different alternatives in second sub-nest of the first nest. We can see a similar situation for choice sets $\mathbb{C}_5$ and $\mathbb{C}_6$, naturally, there are two different alternatives in the first sub-nest of the first nest (there is no change in the second nest for choice sets $\mathbb{C}_3$ to $\mathbb{C}_6$). With respect to the combination of the alternatives in the four choice sets $\mathbb{C}_3$ to $\mathbb{C}_6$, then a similar result for w_2^* and w_3^* will be obtained, so that these two weights are almost equal and decrease as λ increases ($0 < \lambda \leq 0.15$). But, the decreasing trend of w_3^* is faster than w_2^* as $\lambda \geq 0.17$, then the combination of these attributes and their levels in the two $\mathbb{C}_5$ and $\mathbb{C}_6$ are more similar than the choice sets $\mathbb{C}_3$ and $\mathbb{C}_4$.

Table 5.8: NMNL Model, $\mu_1 = 0.1, \lambda_1 = \lambda_2 = 0.08$ (two nests, first nest with two sub-nests and the second does not sub-nest): Locally D-optimal design when $0.1 \le \mu_2 \le 1$ (based on RUM conditions) with initial values $w_1 = 0.1, w_2 = w_3 = 0.2$ and w.r.t local D-optimality criterion $\Psi(\xi'', \boldsymbol{\theta}_0) = (\det(\mathbf{I}(\xi'', \boldsymbol{\theta}_0)))^{-1}$.

μ_2	0.10000	0.15000	0.20000	0.25000	0.30000	0.40000	0.50000	0.70000
w_1^*	0.29640	0.30380	0.30740	0.30950	0.31100	0.31340	0.31520	0.31860
w_2^*	0.10180	0.09810	0.09630	0.09530	0.09450	0.09330	0.09240	0.09070
w_3^*	0.10180	0.09810	0.09630	0.09520	0.09450	0.09330	0.09240	0.09070
$\Psi(\xi'', \boldsymbol{\theta}_0)$	0.09941	0.10191	0.10319	0.10409	0.10484	0.10619	0.10750	11025

Table 5.9: NMNL Model, $\mu_2 = 0.5, \lambda_1 = 0.1, \lambda_2 = 0.2$ (two nests, first nest with two sub-nests and the second does not sub-nest): Locally D-optimal design when $0.2 \le \mu_1 \le 0.5$ (based on RUM conditions) with initial values $w_1 = 0.1, w_2 = w_3 = 0.2$ and w.r.t local D-optimality criterion $\Psi(\xi'', \boldsymbol{\theta}_0) = (\det(\mathbf{I}(\xi'', \boldsymbol{\theta}_0)))^{-1}$.

μ_1	0.20000	0.25000	0.30000	0.35000	0.40000	0.45000	0.50000
w_1^*	0.33330	0.33470	0.33550	0.33600	0.33630	0.33640	0.33640
w_2^*	0.16670	0.16530	0.16450	0.16400	0.16370	0.16360	0.16360
w_3^*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$\Psi(\xi'', \boldsymbol{\theta}_0)$	0.39644	0.56598	0.75789	0.97082	1.20394	1.45688	1.72959

Now, suppose that $\mu_1 = 0.15$, $\mu_2 = 0.25$ and $\lambda_1 = 0.1$, then the RUM conditions hold if $0 < \lambda_2 \le 0.15$. In Table 5.6 several locally optimal designs based on Table 5.4 were obtained. In this situation, w_2^* increases as λ_2 increases (Table 5.6) but w_3^* decreases. That means the alternatives in the second sub-nest (first nest) of choice sets $\mathbb{C}_5$ and $\mathbb{C}_6$ are much similar, but the alternatives in choice sets $\mathbb{C}_3$ and $\mathbb{C}_4$ (w.r.t, second sub-nest) are much more dissimilar.

Another Table 5.7 which includes was calculated some locally optimal designs based on $\mu_1 = \mu_2 = \lambda$ and $0 < \lambda \le 1$. In this case, RUM conditions hold. Table 5.7 denotes: w_1^* increases as λ increases, but w_2^* and w_3^* decrease. Noting the decreasing trend of w_2^* and w_3^*, we can observe that the decreasing trend of w_3^* is faster than w_2^*, because of more similarity (alternatives) in the choice sets $\mathbb{C}_5$ and $\mathbb{C}_6$ in contrast of that between $\mathbb{C}_3$ and $\mathbb{C}_4$.

With respect to fixed amounts for $\mu_1 = 0.1$ and $\lambda_1 = \lambda_2 = 0.08$ (Table 5.8), w_2^* and w_3^* are equal and they decrease as μ_2 increases, but w_1^* increases. Then, the alternatives in the second nest (choice sets $\mathbb{C}_3$ to $\mathbb{C}_6$) are more similar than the alternatives in the second nest of the choice sets $\mathbb{C}_1$ and $\mathbb{C}_2$.

Suppose that $\mu_2 = 0.5, \lambda_1 = 0.1, \lambda_2 = 0.2$. In this situation, RUM conditions hold if $0.2 \le \mu_1 \le 0.5$. Table 5.9 showed that w_1^* increases (almost as always, with a decreasing trend) as μ_1

increases. The third row of Table 5.9 denotes, w_2^* decreases (with a very weak decreasing trend) and w_3^* is equal zero as μ_1 increases. That means that the alternatives in the choice sets $\mathbb{C}_5$ and $\mathbb{C}_6$ are much more similar than are the others. And we can say, if $\mu_2 = 0.5, \lambda_1 = 0.1, \lambda_2 = 0.2$ and $0.45 \leq \mu_1$, then:

$$\xi''^* = \left\{ \begin{array}{cccccc} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 \\ 0.3364 & 0.3364 & 0.1636 & 0.1636 & 0 & 0 \end{array} \right\}$$

is a locally D-optimal design in Ξ.

According to the results which were obtained in the different classes in Table 5.5 to Table 5.9, we can say that the alternatives in the two choice sets $\mathbb{C}_1$ and $\mathbb{C}_2$ are less dissimilar than the others (because, the optimal weights of these two choice sets increase when dissimilarity parameters increase, although, we had sometimes faced to decreasing trend) and the alternatives in the choice sets $\mathbb{C}_5$ and $\mathbb{C}_6$ are more similar than the others (the optimal weights of these two choice sets decrease when dissimilarity parameters increase).

Note: To obtain a locally D-optimal $(\Psi(\xi'', \boldsymbol{\theta}_0) = (\det(\mathbf{I}(\xi'', \boldsymbol{\theta}_0)))^{-1})$ design (Tables 5.5 to 5.9), Maple has been used with the initial values $w_1 = 0.1$ and $w_2 = w_3 = 0.2$ (all of solutions are converge). The sequential Quadratic Programming (SQP) method was also used and naturally the number 1000 was implemented for the iteration limit.

6 OPTIMAL DESIGN IN THE RANK-ORDER TWO-LEVEL NMNL MODEL

A discrete choice experiment measures the importance of the features of a goods or service in making a purchase decision. This is achieved by asking each respondent to choose his/her preferred alternative from a number of choice sets. In stated-preference experiments, respondents may be asked to rank the alternatives instead of just identifying the one alternative that they would choose. This ranking can be requested in a variety of ways. The respondents can be asked to state which alternative they would choose and then, after they have made this choice, can be asked which of the remaining alternatives they would choose, continuing through all the alternatives. Instead, respondents can simply be asked to rank the alternatives from best to worst. In any case, the data that the researcher obtains constitute a ranking of the alternatives that presumably reflects the utility that the respondent obtains from each alternative.

A rank-order conjoint experiment measures the importance of the features of a goods or service 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 (MNL, NMNL, . . .) models *(Beggs, et al. 1981), (Hausman and Ruud 1987)*.

The design of an experiment has a significant impact on the accuracy of the estimated parameters of the fitted model. Choosing the appropriate alternatives and grouping them in choice sets in the best possible way according to an optimality criterion, yields an optimal design which guarantees precise parameter estimates and therefore an accurate view on the preference of the customer.

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 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. In some cases there may be too many alternatives to rank. Secondly, the respondent may not be able to distinguish between his less-preferred alternatives. In any case, straightforwardly using reported rankings may lead to a substantial bias in the parameter estimates in the rank-order logit model, see *(Chapman and Staelin 1982)*. To solve this issue, *(Chapman and Staelin 1982)* suggest to only use the first few ranks in the estimation. They consider 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 the parameter estimates based on different rank information. *(Hausman and Ruud 1987)* proposed an alternative method to test the number of ranks to use in the estimation. However, in both approaches this number is assumed to be the same for all respondents. If ranking capabilities differ across individuals, this may lead to an efficiency loss.

(Vermeulen, et al. (2007) have proposed to use the D-optimality criterion which focuses on the accuracy of the estimates of the rank-ordered MNL model (its parameters).

In this chapter we study the use of the D-optimality criterion to estimate a rank-ordered NMNL model (full parameters). The central question is then whether the corresponding Bayesian D_b-optimal ranking design results in significantly more precise estimates and predictions than commonly used design strategies in marketing.

In the next section 6.1, we review the rank-ordered multinomial logit model *(Vermeulen, et al. (2007)*, then in section 6.2 we obtain the information matrix related to the rank-ordered NMNL model and we define a special class of design and the method of obtaining locally D-optimal design for that.

6.1 Rank-Order MNL (RO.MNL) Model

The Rank-Order Logit Model was introduced into literature by *(Beggs, et al. 1981)*. The model can be used to analyze the preferences of individuals over a set of alternatives,where the preferences are partially observed through surveys or conjoint studies. Any rank order can be regarded as a sequence of choices made by the respondent. This was used as the starting point for the extension of the multinomial logit model to the rank-order multinomial logit model by *(Beggs, et al. 1981)* where the alternatives with lower ranking are considered. 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 classical MNL model there are $\mathcal{J}$ alternatives. Then we consider $\mathcal{C}$ choice sets each with $J_c > 1; \forall c$ alternatives (Section 2.1).

The rank of an alternative is determined by its utility. The utility of the alternative j in choice set $\mathbb{C}_c$ experienced by respondent i is modeled as (by effects-type coding, Section 2.1),

$$U_{jc} = \mathbf{f}^T(a_{jc})\boldsymbol{\beta} + \varepsilon_{jc} = \sum_{k=1}^{K}\sum_{\ell=1}^{L_k-1} f_{k\ell}(a_{jc})\beta_{k,\ell} + \varepsilon_{jc}; j = 1, 2, \ldots, J_c, \tag{6.1}$$

where:

- $\mathbf{f}(a_{jc}) = (\mathbf{f}_1(a_{jc}), \ldots, \mathbf{f}_k(a_{jc}), \ldots, \mathbf{f}_K(a_{js}))^T$;

$$\mathbf{f}_k(a_{jc}) = (f_{k1}(a_{jc}), \ldots, f_{k\ell}(a_{jc}), \ldots, f_{kL_k-1}(a_{jc}))^T \quad \text{(See Section 2.1)}$$

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

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

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) (Section 2.1).

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 $r_{(1)}$; ($r_{(1)} \in \{1, 2, \ldots, J_c\}$) 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 $r_{(2)}$ has the second rank, its utility is less than the utility of $r_{(1)}$ and greater than 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 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 situation $\mathbb{C}_{c(j)}$ denotes a choice set (choice set c), which excludes alternative j. In this situation and to obtain choice probabilities, we can also define the observation variables as follows:

$$Y_{r_{(1)}c} = \begin{cases} 1, & U_{r_{(1)}c} = \max_{j \in \mathbb{C}_c} U_{jc}; \\ 0, & \text{otherwise.} \end{cases} \quad , \quad Y_{r_{(2)}s} = \begin{cases} 1, & U_{r_{(2)}c} = \max_{j \in C_{c(r_{(1)})}} U_{jc}; \\ 0, & \text{otherwise.} \end{cases}$$

and so on. Now, we can define the probabilities of rank-order alternatives as follows:

$$\begin{aligned}
P(Y_{(1)} = r_{(1)}) = P_{r_{(1)}c} &= \frac{\exp\left(\mathbf{f}^T(a_{r_{(1)}c})\boldsymbol{\beta}\right)}{\sum_{a_\ell \in \mathbb{C}_s} \exp\left(\mathbf{f}^T(a_{\ell c})\boldsymbol{\beta}\right)};\ a_{r_{(1)}} \in \mathbb{C}_c, \\
P(Y_{(2)} = r_{(2)}) = P_{r_{(2)}c} &= \frac{\exp\left(\mathbf{f}^T(a_{r_{(2)}c})\boldsymbol{\beta}\right)}{\sum_{a_\ell \in \mathbb{C}_{c(r_{(1)})}} \exp\left(\mathbf{f}^T(a_{\ell c})\boldsymbol{\beta}\right)};\ a_{r_{(2)}} \in \mathbb{C}_{c(r_{(1)})}, \\
P(Y_{(3)} = r_{(3)}) = P_{r_{(3)c}} &= \frac{\exp\left(\mathbf{f}^T(a_{r_{(3)}c})\boldsymbol{\beta}\right)}{\sum_{a_\ell \in \mathbb{C}_{c(r_{(1)}, r_{(2)})}} \exp\left(\mathbf{f}^T(a_{\ell c})\boldsymbol{\beta}\right)};\ a_{r_{(3)}} \in \mathbb{C}_{c(r_{(1)}, r_{(2)})}, \\
&\vdots \\
P(Y_{(J_c)} = r_{(J_c)}) = P_{r_{(J_c)}c} &= 1;\ a_{r_{(J_c)}} \in \mathbb{C}_{c(r_{(1)}, r_{(2)}, \ldots, r_{(J_c - 1)})},
\end{aligned}$$

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) then 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, consider a combination of alternatives like $(r_{(1)}, r_{(2)}, \ldots, r_{(J_c)})$; $r_{(1)}, r_{(2)}, \ldots, r_{(J_c)} \in \{1, 2, \ldots, J_c\}$ to analyze data, means that the alternative $r_{(1)}$ has the first rank, alternative $r_{(2)}$ has the second rank and so on $(r_{(j)} \neq r_{(j')}, j \neq j' = 1, 2, \ldots, J_c)$. Because 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 variable observations in rank-order MNL model:

$$\mathcal{Y}_{(r_{(1)}r_{(2)}\ldots r_{(J_c)})} = \begin{cases} 1, & if\ Y_{(1)} = r_{(1)}, Y_{(2)} = r_{(2)}, \ldots, Y_{(J_c)} = r_{(J_c)}; \\ 0, & otherwise, \end{cases} \tag{6.2}$$

where:

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

As stated previously was told, we consider $\mathcal{S}$ $(\mathcal{S} \subset \mathcal{C})$ choice sets each with $J_s = J; \forall s$ alternatives to obtain optimal design. In this chapter, we act as similar previous Chapters (Chapters 3, 4 and 5). Thus, according to the choice rank probabilities and Equation (6.2), the Likelihood function can be defined as follows:

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

and the log-likelihood function as:

$$\begin{aligned} \ell(\mathbb{C}_s, \boldsymbol{\beta}) &= \underbrace{\sum_{r_{(1)}=1}^{J} \sum_{r_{(2)}=1}^{J} \cdots \sum_{r_{(J)}=1}^{J}}_{r_{(1)} \neq r_{(2)} \cdots \neq r_{(J)}} \mathcal{Y}_{(r_{(1)}r_{(2)}\ldots r_{(J)})} \cdot \ln \left(P_{r_{(1)}s} \cdot P_{r_{(2)}s} \ldots P_{r_{(J)}s} \right) \\ &= \underbrace{\sum_{r_{(1)}=1}^{J} \sum_{r_{(2)}=1}^{J} \cdots \sum_{r_{(J)}=1}^{J}}_{r_{(1)} \neq r_{(2)} \cdots \neq r_{(J)}} \mathcal{Y}_{(r_{(1)}r_{(2)}\ldots r_{(J)})} \cdot \left(\ln P_{r_{(1)}s} + \ln P_{r_{(2)}s} + \ldots + \ln P_{r_{(J)}s} \right). \end{aligned} \tag{6.3}$$

Lemma 6.1. *The information matrix based on Log-Likelihood* (6.3) *is obtained by:*

$$
\begin{aligned}
\mathbf{I}_{R(MNL)}(\mathbb{C}_s,\boldsymbol{\beta}) &= \mathbf{I}_{MNL}(\mathbb{C}_s,\boldsymbol{\beta}) + \sum_{r_{(1)}=1}^{J} P_{r_{(1)}s}\cdot \mathbf{I}_{MNL}(\mathbb{C}_{s(r_{(1)})},\boldsymbol{\beta}) \\
&+ \sum_{r_{(1)}\neq r_{(2)}}^{J} P_{r_{(1)}s}P_{r_{(2)}}\cdot \mathbf{I}_{MNL}(\mathbb{C}_{s(r_{(1)},r_{(2)})},\boldsymbol{\beta}) + \\
&+ \ldots + \sum_{r_{(1)}\neq r_{(2)}\neq\ldots\neq r_{(J-2)}}^{J} P_{r_{(1)}}P_{r_{(2)}}\ldots P_{r_{(J-2)}}\cdot \mathbf{I}_{MNL}(\mathbb{C}_{s(r_{(1)},r_{(2)},\ldots,r_{(J-2)})},\boldsymbol{\beta}),
\end{aligned}
$$

Proof:

$$
\begin{aligned}
\mathbf{I}_{R(MNL)}(\mathbb{C}_s,\boldsymbol{\beta}) &= -E\left(\frac{\partial^2 \ell(\mathbb{C}_s,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right) \\
&= \underbrace{\sum_{r_{(1)}=1}^{J}\sum_{r_{(2)}=1}^{J}\ldots\sum_{r_{(J)}=1}^{J}}_{r_{(1)}\neq r_{(2)}\ldots\neq r_{(J)}} P_{r_{(1)}s}\cdot P_{r_{(2)}s}\ldots P_{r_{(J)}s}\cdot\left(-\frac{\partial^2 \ln P_{r_{(1)}s}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} - \frac{\partial^2 \ln P_{r_{(2)}s}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} - \frac{\partial^2 \ln P_{r_{(J)}s}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right)
\end{aligned}
$$

$$
\begin{aligned}
= &\sum_{r_{(1)}=1}^{J} P_{r_{(1)}s}\left(\frac{-\partial^2 \ln P_{r_{(1)}s}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right) + \sum_{r_{(1)}=1}^{J} P_{r_{(1)}s}\sum_{r_{(2)}=1}^{J} P_{r_{(2)}s}\cdot\left(\frac{-\partial^2 \ln P_{r_{(2)}s}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right) \\
&+ \underbrace{\sum_{r_{(1)}=1}^{J}\sum_{r_{(2)}=1}^{J}}_{r_{(1)}\neq r_{(2)}} P_{r_{(1)}s}\cdot P_{r_{(2)}s}\sum_{r_{(3)}=1}^{J} P_{r_{(3)}s}\left(\frac{-\partial^2 \ln P_{r_{(3)}s}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right) + \ldots + \\
&+ \underbrace{\sum_{r_{(1)}=1}^{J}\sum_{r_{(2)}=1}^{J}\ldots\sum_{r_{(J-2)}=1}^{J}}_{r_{(1)}\neq r_{(2)}\ldots\neq r_{(J-2)}} P_{r_{(1)}s}\cdot P_{r_{(2)}s}\ldots P_{r_{(J-2)}s}\sum_{r_{(J-1)}=1}^{J} P_{r_{(J-1)}s}\left(-\frac{\partial^2 \ln P_{r_{(J-1)}s}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right),
\end{aligned}
$$

where:

- $\sum_{r_{(1)}=1}^{J} P_{r_{(1)}s}\left(\frac{-\partial^2 \ln P_{r_{(1)}s}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right) = \mathbf{I}_{MNL}(\mathbb{C}_s,\boldsymbol{\beta})$,
- $\sum_{r_{(2)}=1}^{J} P_{r_{(2)}s}\cdot\left(\frac{-\partial^2 \ln P_{r_{(2)}s}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right) = \mathbf{I}_{MNL}(\mathbb{C}_{s(r_{(1)})},\boldsymbol{\beta})$,
- $\sum_{r_{(3)}=1}^{J} P_{r_{(3)}s}\left(\frac{-\partial^2 \ln P_{r_{(3)}s}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right) = \mathbf{I}_{MNL}(\mathbb{C}_{s(r_{(1)},r_{(2)})},\boldsymbol{\beta})$,
- $\sum_{r_{(J-1)}=1}^{J} P_{r_{(J-1)}s}\left(-\frac{\partial^2 \ln P_{r_{(J-1)}s}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right) = \mathbf{I}_{MNL}(\mathbb{C}_{s(r_{(1)},r_{(2)},\ldots,r_{(J-2)})},\boldsymbol{\beta})$.

where:

- $\mathbf{I}_{MNL}(\mathbb{C}_s, \boldsymbol{\beta}) = \mathbf{F_s}^T(\mathbf{P_s} - \mathbf{p_s p_s}^T)\mathbf{F_s}$: The information matrix of a discrete choice experiment (classic) (3.14) by choice set with size J (Sandor and Wedel 2001),
- $\mathbf{I}_{MNL}(\mathbb{C}_{s(j)}, \boldsymbol{\beta}) = \mathbf{F}^T_{s(j)}(\mathbf{P}_{s(j)} - \mathbf{p}_{s(j)}\mathbf{p}^T_{s(j)})\mathbf{F}_{s(j)}$;
- $C_{s(j)}$ denotes a choice set without alternative 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 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. Because this difference $\mathbf{I}_{R_{(MNL)}}(\mathbb{C}_s, \boldsymbol{\beta}) - \mathbf{I}_{MNL}(\mathbb{C}_s, \boldsymbol{\beta})$ is a nonnegative definite matrix, which 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.

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

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

to obtain the local D-optimal criterion, in this case (rank-order MNL) is also used a similar situation, means that:

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

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

6.2 Rank-Order Two-Level 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 the others and which have been divided into M nests. In this case, the utility of choosing an alternative j and nest m by individual i is calculated as follows (subsection 2.2.1; ignored index i):

$$U_{jms} = U_{j|ms} + U_{ms},$$

where $U_{j|ms} = v_{j|ms} + \varepsilon_{j|ms}$ and $U_{ms} = v_{ms} + \varepsilon_{ms}$ (see Chapters 2 and 4).

Now, consider that $(Y_{(1)}, Z_{(1)})$,..., $(Y_{(j)}, Z_{(j)})$, ..., $(Y_{(J_s)}, Z_{(J_s)})$ denote the joint rank alternative variables of a choice set with J_s alternatives, where variable $Y_{(r)}$ and $Z_{(r)}$ denote an alternative and a nest with rank r. For example, $(Y_{(1)} = r_{(1)}, Z_{(1)} = m_{(1)})$ means that the alternative $r_{(1)} \in \mathbb{C}_{m_{(1)}s}$ 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 $r_{(2)}$ from nest $m_{(2)}$ have the second rank. Certainly in the original choice set, but in the new choice set (it is denoted by $\mathbb{C}_{s(r_{(1)})}$ after removing the alternative $r_{(1)}$) alternative $r_{(2)}$ and nest $m_{(2)}$ will have the first rank, where $m_{(2)}$ may be the same as nest $m_{(1)}$ or not. In this situation, if there is $m_{(1)} = m_{(2)}$ then $r_{(1)}$ and $r_{(2)}$ 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 j.

We know that if $corr(\varepsilon_{j|m_{(1)}s}, \varepsilon_{j'|m_{(1)}s}) = \rho_{m_{(1)}}; j \neq j'$ and $corr(\varepsilon_{j|m_{(2)}s}, \varepsilon_{j'|m_{(2)}s}) = \rho_{m_{(2)}}; j, j' \neq j_{(1)}$ then $\rho_{m_{(1)}} = \rho_{m_{(2)}}$ if $m_{(1)} = m_{(2)}$. In this case, keeping to the relation between $\rho_.$ and $\lambda_.$ (Subsection 2.2.1), 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.

In this situation, we denote the choice probability related to choosing an alternative with the rank r as follows (w.r.t the choice set $\mathbb{C}_s$):

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

Now, for alternative with the first rank we will have:

$$P_{r_{(1)}m_{(1)}s} = \frac{\exp\left(\frac{\mathbf{f}^T(a_{r_{(1)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)}{\sum_{j=1}^{J_{m_{(1)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)} \cdot \frac{\left(\sum_{a_j\in C_{m_{(1)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)\right)^{\lambda_{m_{(1)}}}}{\left(\sum_{j\in C_{m_{(1)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)\right)^{\lambda_{m_{(1)}}} + \sum_{m\neq m_{(1)}}^{M} \left(\sum_{a_j\in \mathbb{C}_{ms}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_m}\right)\right)^{\lambda_m}}$$

and $\mathbb{C}_{ms}$ denotes a choice set, which includes all of alternatives in nest m (subsection 2.2.1) 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 $\mathbf{x}_{r_{(j)}s}$ denotes the characterizes of the attributes related to alternative $r_{(j)}$, which has the rank r (w.r.t choice set $\mathbb{C}_s$)

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

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

where:

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

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

and

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

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

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 $r_{(1)}$) so that if $m_{(1)}=m_{(2)}$ then $J_{m_{(2)}s}=J_{m_{(1)}s}-1$.
- $J^{r_{(1)}}_{m_{(1)}s}$ is the number of alternatives in nest $m_{(1)}$ after choosing alternative $r_{(1)}$ and removing it.

Also, the conditional choice probabilities related to the third rank alternative have been calculated by:

$$P_{r_{(3)}|m_{(3)}s}=\frac{\exp\left(\frac{\mathbf{f}^T(a_{r_{(3)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)}{\sum_{j=1}^{J^{r_{(1)},r_{(2)}}_{m_{(1)}}s}\exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)}; \quad m_{(3)}=m_{(2)}=m_{(1)},$$

$$=\frac{\exp\left(\frac{\mathbf{f}^T(a_{r_{(3)}s})\boldsymbol{\beta}}{\lambda_{m_{(2)}}}\right)}{\sum_{j=1}^{J^{r_{(2)}}_{m_{(2)}s}}\exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(2)}}}\right)}; \quad m_{(3)}=m_{(2)}\neq m_{(1)},$$

$$= \frac{\exp\left(\frac{\mathbf{f}^T(a_{r_{(3)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)}{\sum_{j=1}^{J^{r_{(1)}}_{m_{(1)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)}; \quad m_{(3)} = m_{(1)} \neq m_{(2)},$$

$$= \frac{\exp\left(\frac{\mathbf{f}^T(a_{r_{(3)}s})\boldsymbol{\beta}}{\lambda_{m_{(3)}}}\right)}{\sum_{j=1}^{J_{m_{(3)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(3)}}}\right)}; \quad m_{(3)} \neq m_{(2)} = m_{(1)}$$

$$= \frac{\exp\left(\frac{\mathbf{f}^T(a_{r_{(3)}s})\boldsymbol{\beta}}{\lambda_{m_{(3)}}}\right)}{\sum_{j=1}^{J_{m_{(3)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(3)}}}\right)}; \quad m_{(3)} \neq m_{(2)} \neq m_{(1)}.$$

The marginal choice probabilities related to the third rank nest for difference cases are calculated as follows:

$$P_{m_{(3)}s} = \frac{\left(\sum_{j=1}^{J^{r_{(1)},r_{(2)}}_{m_{(1)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)\right)^{\lambda_{m_{(1)}}}}{\left(\sum_{j=1}^{J^{r_{(1)},r_{(2)}}_{m_{(1)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)\right)^{\lambda_{m_{(1)}}} + \sum_{m=1}^{M}\left(\sum_{j=1}^{J_{ms}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_m}\right)\right)^{\lambda_m}}; \; m_{(3)} = m_{(2)} = m_{(1)},$$

$$= \frac{\left(\sum_{j=1}^{J^{r_{(2)}}_{m_{(2)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(2)}}}\right)\right)^{\lambda_{m_{(2)}}}}{\left(\sum_{j=1}^{J^{r_{(2)}}_{m_{(2)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(2)}}}\right)\right)^{\lambda_{m_{(2)}}} + \left(\sum_{j=1}^{J^{r_{(1)}}_{m_{(1)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)\right)^{\lambda_{m_{(1)}}} + \sum_{m=1}^{M}\left(\sum_{j=1}^{J_{ms}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_m}\right)\right)^{\lambda_m}}; \; m_{(3)} = m_{(2)} \neq m_{(1)},$$

$$= \frac{\left(\sum_{j=1}^{J^{r_{(1)}}_{m_{(1)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)\right)^{\lambda_{m_{(1)}}}}{\left(\sum_{j=1}^{J^{r_{(1)}}_{m_{(1)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)\right)^{\lambda_{m_{(1)}}} + \left(\sum_{j=1}^{J^{r_{(2)}}_{m_{(2)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(2)}}}\right)\right)^{\lambda_{m_{(2)}}} + \sum_{m=1}^{M}\left(\sum_{j=1}^{J_{ms}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_m}\right)\right)^{\lambda_m}}; \; m_{(3)} = m_{(1)} \neq m_{(2)},$$

$$= \frac{\left(\sum_{j=1}^{J_{m_{(3)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(3)}}}\right)\right)^{\lambda_{m_{(3)}}}}{\left(\sum_{j=1}^{J_{m_{(3)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(3)}}}\right)\right)^{\lambda_{m_{(3)}}} + \left(\sum_{j=1}^{J^{r_{(1)},r_{(2)}}_{m_{(1)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)\right)^{\lambda_{m_{(1)}}} + \sum_{m=1}^{M}\left(\sum_{j=1}^{J_{ms}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_m}\right)\right)^{\lambda_m}}; \; m_{(3)} \neq m_{(1)} = m_{(2)},$$

$$= \frac{\left(\sum_{j=1}^{J_{m_{(3)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(3)}}}\right)\right)^{\lambda_{m_{(3)}}}}{\left(\sum_{j=1}^{J_{m_{(3)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(3)}}}\right)\right)^{\lambda_{m_{(3)}}} + \left(\sum_{j=1}^{J^{r_{(1)}}_{m_{(1)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)\right)^{\lambda_{m_{(1)}}} + \left(\sum_{j=1}^{J^{r_{(2)}}_{m_{(2)}s}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(2)}}}\right)\right)^{\lambda_{m_{(2)}}} + \sum_{m=1}^{M}\left(\sum_{j=1}^{J_{ms}} \exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_m}\right)\right)^{\lambda_m}};$$

$$m_{(3)} \neq m_{(1)} \neq m_{(2)},$$

where $m \neq m_{(1)}$ (in the first row), $m \neq m_{(1)}, m_{(2)}$ (in the second and third row), $m \neq m_{(1)}, m_{(3)}$ (in the forth row) and $m \neq m_{(1)}, m_{(2)}, m_{(3)}$ (in the last row). Also, to obtain the other choice probabilities related to the alternatives with lower ranking like $P_{r_{(4)}m_{(4)}}, \ldots, P_{r_{(J)}m_{(J)}}$, we act as similar way, where $J_s = \sum_{m=1}^{M} J_{ms}$ for each choice set, $\mathbb{C}_s$, and $J_{ms}^{j,j'}$ denotes the number of alternatives in nest m after removing the alternatives j and j'.

In this situation, we define the following observation variables to analyze the RO.NMNL model:

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

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

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

6.2.1 Information Matrix for RO.NMNL Model

In this subsection, we obtain the information matrix for the two-level RO.NMNL model. Afterwards, 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 $\mathbb{C}_s$):

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

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 was based on the above likelihood function is calculated by:

$$\ell(\mathbb{C}_s, \boldsymbol{\theta}) = \underbrace{\sum_{m_{(1)}=1}^{M} \sum_{m_{(2)}=1}^{M} \cdots \sum_{m_{(J)}=1}^{M}}_{\textit{some nests maybe equal}} \underbrace{\sum_{r_{(1)}=1}^{J_{m_{(1)}s}} \sum_{r_{(2)}=1}^{J_{m_{(2)}s}} \cdots \sum_{r_{(J)}=1}^{J_{m_{(J)}s}}}_{r_{(1)} \neq r_{(2)} \cdots \neq r_{(J)}} \mathcal{Y}' \cdot G_s\left((r_{(1)}, m_{(1)}), (r_{(2)}, m_{(2)}), \ldots, (r_{(J)}, m_{(J)})\right), \tag{6.4}$$

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

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

According to the number of attributes, K, each with $L_k; k = 1, 2, \dots, K$ levels, we define parameters vector and the characterizes of attributes as follow (Chapter 4):

- $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\lambda})^T$;
- $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_k, \dots, \boldsymbol{\beta}_K)^T$;
- $\boldsymbol{\beta}_k = (\beta_{k,1}, \dots, \beta_{k,\ell}, \dots, \beta_{k,L_k-1})^T$ (w.r.t effect type codding, $\sum_{\ell=1}^{L_k} \beta_{k\ell} = 0$),
- $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m, \dots, \lambda_M)^T$,

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 $p = \sum_{k=1}^{K}(L_k - 1)$ (see Chapter 4). Also:

- $v_{r_{(j)}|m_{(r)}s} = \mathbf{f}^T(a_{r_{(j)}m_{(r)}s})\boldsymbol{\beta}$,
- $\mathbf{f}(a_{r_{(j)}m_{(r)}s}) = (\mathbf{f}_1(a_{r_{(j)}m_{(r)}s}), \dots, \mathbf{f}_k(a_{r_{(j)}m_{(r)}s}), \dots, \mathbf{f}_K(a_{r_{(j)}m_{(r)}s}))^T$;
- $\mathbf{f}_k(a_{r_{(j)}m_{(r)}s}) = (f_{k1}(a_{r_{(j)}m_{(r)}s}), \dots, f_{k\ell}(a_{r_{(j)}m_{(r)}s}), \dots, f_{kL_k-1}(a_{r_{(j)}m_{(r)}s}))^T$,

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

Corresponding to Equation (6.4) 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:

$$-E\left(\frac{\partial^2 \ell(\mathbb{C}_s, \boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right) = \sum_{m_{(1)}=1}^{M} \sum_{m_{(2)}=1}^{M} \cdots \sum_{m_{(J)}=1}^{M} \underbrace{\sum_{r_{(1)}=1}^{J_{m_{(1)}s}} \sum_{r_{(2)}=1}^{J_{m_{(2)}s}} \cdots \sum_{r_{(J)}=1}^{J_{m_{(J)}s}}}_{r_{(1)} \neq r_{(2)} \cdots \neq r_{(J)}} E(\mathcal{Y}') \cdot \frac{-\partial^2 G_s(r_{(1)}, r_{(2)} \dots r_{(J)})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T},$$

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

- $E(\mathcal{Y}') = P_{r_{(1)}m_{(1)}s} \cdot P_{r_{(2)}m_{(2)}s} \dots P_{r_{(J)}m_{(J)}s}$
- $\frac{-\partial^2 G_s(r_{(1)}, r_{(2)} \dots r_{(J)})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} = -\frac{\partial^2 \ln P_{r_{(1)}m_{(1)}s}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} - \frac{\partial^2 \ln P_{r_{(2)}m_{(2)}s}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} - \dots - \frac{\partial^2 \ln P_{r_{(J-1)}m_{(J-1)}s}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}$.

Lemma 6.2. *According to above descriptions, the information matrix of RO.NMNL model is calculated as follow:*

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

Proof:

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

$$\begin{aligned}&+\sum_{m_{(1)}=1}^{M}\sum_{m_{(2)}=1}^{M}\cdots\sum_{m_{(J-2)}=1}^{M}\underbrace{\sum_{r_{(1)}=1}^{J_{m_{(1)}s}}\sum_{r_{(2)}=1}^{J_{m_{(2)}s}}\cdots\sum_{r_{(J-2)}=1}^{J_{m_{(J-2)}s}}}_{r_{(1)}\neq r_{(2)}\cdots\neq r_{(J-2)}} P_{r_{(1)}m_{(1)}s}\cdot P_{r_{(2)}m_{(2)}s}\ldots P_{r_{(J-2)}m_{(J-2)}s}\\ &\times\left(\sum_{m_{(J-1)}=1}^{M}\sum_{r_{(J-1)}=1}^{J_{m_{(J-1)}s}} P_{r_{(J-1)}m_{(J-1)}s}\cdot\left(\frac{-\partial^2\ln P_{r_{(J-1)}m_{(J-1)}s}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right)\right),\end{aligned}$$

where:

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

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

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

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

are used in Chapter 4, Section 4.2. For example, according to a NMNL model with two nest each with J_{1s} and J_{2s} alternatives, Lemma 4.2 can be used to obtain above information matrices. In this situation, the information matrix $\mathbf{I}_{NMNL}(\mathbb{C}_s, \boldsymbol{\theta})$ is calculated by Lemma 4.2, directly. But, to obtain the information matrix $\mathbf{I}_{NMNL}(\mathbb{C}_{s(j)}, \boldsymbol{\theta})$, Lemma 4.2 will be considered as follows:

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

where $\mathbf{I}_{..s(j)}$ and $I_{..s(j)}$ denote the elements of above information matrix without considering alternative j and according to the following assumptions (Section 4.2):

- $\mathbf{A}_{ms(j)} = \mathbf{F}_{ms(j)}^T \mathbf{p}_{.|ms(j)}$,
- $\mathbf{B}_{ms(j)} = \mathbf{F}_{ms(j)}^T \mathbf{P}_{.|ms(j)} \mathbf{F}_{ms(j)}$,
- $\mathbf{p}_{.|ms(j)} = (p_{1|ms(j)}, \ldots, p_{j-1|ms(j)}, p_{j+1|ms(j)}, \ldots, p_{J_{ms(j)}|ms(j)})^T$,
- $\mathbf{P}_{.|ms(j)} = diag[p_{1|ms(j)}, \ldots, p_{j-1|ms(j)}, p_{j+1|ms(j)}, \ldots, p_{J_{ms(j)}|ms(j)}]$,
- $\mathbf{F}_{ms(j)} = (\mathbf{f}(a_{1ms}), \ldots, \mathbf{f}(a_{j-1,ms}), \mathbf{f}(a_{j+1,ms}), \ldots, \mathbf{f}(a_{J_{ms}^j ms}))^T$,
- $\mathbf{f}(a_{j'ms}) = (f_1(a_{j'ms}), \ldots, f_h(a_{j'ms}), \ldots, f_p(a_{j'ms}))^T; j' \neq j$,
- $v_{ms(j)} = \ln\left(\sum_{j'=1}^{J_{ms}^j} \exp\left(\frac{\mathbf{f}^T(a_{j'ms})\boldsymbol{\beta}}{\lambda_m}\right)\right)$,
- $p_{j'|ms(j)} = \dfrac{\exp\left(\frac{\mathbf{f}^T(a_{j'ms})\boldsymbol{\beta}}{\lambda_m}\right)}{\sum_{l=1}^{J_{ms}^j} exp\left(\frac{\mathbf{f}^T(a_{lms})\boldsymbol{\beta}}{\lambda_1}\right)}$,
- $p_{ms(j)} = \dfrac{\left(\sum_{l=1}^{J_{ms}^j} \exp\left(\frac{\mathbf{f}^T(a_{lms})\boldsymbol{\beta}}{\lambda_m}\right)\right)^{\lambda_m}}{\left(\sum_{l=1}^{J_{1s}^j} \exp\left(\frac{\mathbf{f}^T(a_{l1s})\boldsymbol{\beta}}{\lambda_1}\right)\right)^{\lambda_1} + \left(\sum_{l=1}^{J_{2s}^j} \exp\left(\frac{\mathbf{f}^T(a_{l2s})\boldsymbol{\beta}}{\lambda_2}\right)\right)^{\lambda_2}}$.

We will have:

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

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

Table 6.1: NMNL Model: There are two Nests each with $\mathcal{J}_1$ and $\mathcal{J}_2$ alternatives (Chapter 4)

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

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

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

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

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

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

6.2.2 D-Optimal Design

Similar to the classical choice experiments, which have been used (Chapter 4):

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

to obtain the local D-optimal criterion, in the RO.NMNL is also used:

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

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

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

for true value of parameters.

6.2.3 Example

Imagine that there is a two-level NMNL model with two nests such that one of them has $\mathcal{J}_1$ alternatives and the other includes $\mathcal{J}_2$ alternatives (Table 6.1), 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 6.1 (It is assumed that just three alternatives have suitable ranking). In this case, we will encounter two classes ($N = 2$), where:

- $\mathcal{S}_1 = \begin{pmatrix} \mathcal{J}_1 \\ 1 \end{pmatrix} \times \begin{pmatrix} \mathcal{J}_2 \\ 2 \end{pmatrix}$,

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

According to the dimension 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, see *(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 (see, *(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 (Chapter 4):

$$\xi_n = \left\{ \begin{array}{cccc} \mathbb{C}_{n1} & \mathbb{C}_{n2} & \cdots & \mathbb{C}_{n\mathcal{S}_n} \\ w_{n1} & w_{n2} & \cdots & w_{n\mathcal{S}_n} \end{array} \right\} \in \Xi_n; n = 1, 2. \tag{6.5}$$

Similarity (Chapter 4), the information matrix corresponding to ξ_n (6.5) is calculated by:

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

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

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

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

$$\xi^* = \arg\min_{\xi \in \Xi} \Psi_R(\xi, \boldsymbol{\theta}) \tag{6.6}$$

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

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

Lemma 6.3. *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 6.1), is calculated by:*

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

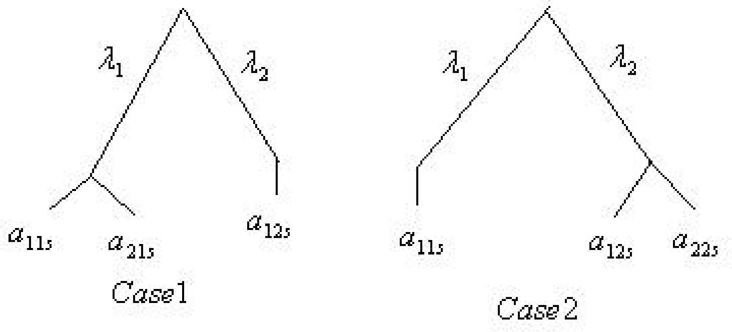

Figure 6.1: NMNL Model: There are two nests(for choice set $\mathbb{C}_s$), one of them with two alternatives and one for another (a_{jms} denotes the j^{th} alternative of the m^{th} nest from choice set s)

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

$$P_{r_{(1)}|m_{(1)}s} = \begin{cases} \dfrac{\exp\left(\frac{\mathbf{f}^T(a_{r_{(1)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)}{\exp\left(\frac{\mathbf{f}^T(a_{r_{(1)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)+\exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)} & if\ \mathbb{C}_{m_{(1)}s} = \{a_{r_{(1)}m_{(1)}s}, a_{jm_{(1)}s}\} \\ 1 & if\ \mathbb{C}_{m_{(1)}s} = \{a_{r_{(1)}m_{(1)}s}\} \end{cases}, \tag{6.7}$$

$$P_{m_{(1)}s} = \begin{cases} \dfrac{\left(\exp\left(\frac{\mathbf{f}^T(a_{r_{(1)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)+\exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)\right)^{\lambda_{m_{(1)}}}}{\left(\exp\left(\frac{\mathbf{f}^T(a_{r_{(1)}s})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)+\exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_{m_{(1)}}}\right)\right)^{\lambda_{m_{(1)}}}+\exp\left(\mathbf{f}^T(a_{j's})\boldsymbol{\beta}\right)} & if\ \mathbb{C}_{m_{(1)}s} = \{a_{r_{(1)}m_{(1)}s}, a_{jm_{(1)}s}\} \\ \dfrac{\exp\left(\mathbf{f}^T(a_{r_{(1)}s})\boldsymbol{\beta}\right)}{\left(\exp\left(\frac{\mathbf{f}^T(a_{js})\boldsymbol{\beta}}{\lambda_m}\right)+\exp\left(\frac{\mathbf{f}^T(a_{j's})\boldsymbol{\beta}}{\lambda_m}\right)\right)^{\lambda_m}+\exp\left(\mathbf{f}^T(a_{r_{(1)}s})\boldsymbol{\beta}\right)} & if\ \mathbb{C}_{m_{(1)}s} = \{a_{r_{(1)}m_{(1)}s}\} \end{cases}, \tag{6.8}$$

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 alternative with the highest utility (first rank). In Equations (6.7) and (6.8), notation $\mathbb{C}_{m_{(1)}s}$ denotes the choice set with the first rank nest.

Now, consider a special case (Figure 6.1, Case1). In this case, $\mathbf{I}_{R(NMNL)}(\mathbb{C}_s, \boldsymbol{\theta})$ is calculated as follows:

$$\mathbf{I}_{R_{(NMNL)}}(\mathbb{C}_s,\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I}_{11s(0)} & \mathbf{I}_{12s(0)} & \underbrace{\mathbf{I}_{13s(0)}}_{\mathbf{0}} \\ \mathbf{I}^T_{12s(0)} & I_{22s(0)} & \underbrace{I_{23s(0)}}_{0} \\ \underbrace{\mathbf{I}^T_{13s(0)}}_{\mathbf{0}^T} & \underbrace{I_{23s(0)}}_{0} & \underbrace{I_{33s(0)}}_{0} \end{bmatrix} + P_{a_{11s}1s} \begin{bmatrix} \mathbf{I}_{11s(a_{11s})} & \underbrace{\mathbf{I}_{12s(a_{11s})}}_{\mathbf{0}} & \underbrace{\mathbf{I}_{13s(a_{11s})}}_{\mathbf{0}} \\ \underbrace{\mathbf{I}^T_{12s(a_{11s})}}_{\mathbf{0}^T} & \underbrace{I_{22s(a_{11s})}}_{0} & \underbrace{I_{23s(a_{11s})}}_{0} \\ \underbrace{\mathbf{I}^T_{13s(a_{11s})}}_{\mathbf{0}^T} & \underbrace{I_{23s(a_{11s})}}_{0} & \underbrace{I_{33s(a_{11s})}}_{0} \end{bmatrix}$$

$$+P_{a_{21s}1s} \begin{bmatrix} \mathbf{I}_{11s(a_{21s})} & \underbrace{\mathbf{I}_{12s(a_{21s})}}_{\mathbf{0}} & \underbrace{\mathbf{I}_{13s(a_{21s})}}_{\mathbf{0}} \\ \underbrace{\mathbf{I}^T_{12s(a_{21s})}}_{\mathbf{0}^T} & \underbrace{I_{22s(a_{21s})}}_{0} & \underbrace{I_{23s(a_{21s})}}_{0} \\ \underbrace{\mathbf{I}^T_{13s(a_{21s})}}_{\mathbf{0}^T} & \underbrace{I_{23s(a_{21s})}}_{0} & \underbrace{I_{33s(a_{21s})}}_{0} \end{bmatrix} + P_{a_{12s}2s} \begin{bmatrix} \mathbf{I}_{11s(a_{12s})} & \mathbf{I}_{12s(a_{12s})} & \underbrace{\mathbf{I}_{13s(a_{12s})}}_{\mathbf{0}} \\ \mathbf{I}^T_{12s(a_{12s})} & I_{22s(a_{12s})} & \underbrace{I_{23s(a_{12s})}}_{0} \\ \underbrace{\mathbf{I}^T_{13s(a_{12s})}}_{\mathbf{0}^T} & \underbrace{I_{23s(a_{12s})}}_{0} & \underbrace{I_{33s(a_{12s})}}_{0} \end{bmatrix}$$

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 (6.7) and (6.8), where $P_{r_{(j)}m_{(r)}s}$ has already been defined (See Subsection 6.2), $\mathbf{I}_{\ell\ell's(0)}$ $(I_{\ell\ell'(0)})$ denote the ℓ^{th} row and ℓ'^{th} column of the information matrix $\mathbf{I}$ (based on choice set $\mathbb{C}_s$) with respect to all of alternatives in choice set $\mathbb{C}_s$. But, $\mathbf{I}_{\ell\ell'(j)}$ $(I_{\ell\ell'(j)})$ denotes the ℓ^{th} row and ℓ'^{th} column of the information matrix $\mathbf{I}$ (based on choice set $\mathbb{C}_s$) without alternative j. Moreover,

- $\mathbf{I}_{11s(j)} = p_{1s(j)}p_{2s(j)}(\mathbf{F}^T_{1s(j)}\mathbf{F}_{1s(j)} + \mathbf{F}^T_{2s(j)}\mathbf{F}_{2s(j)} - \mathbf{F}^T_{1s(j)}\mathbf{F}_{2s(j)} - \mathbf{F}^T_{2s(j)}\mathbf{F}_{1s(j)}); \forall j = a_{11s}, a_{21s}$,
- $\mathbf{I}_{11s(a_{12s})} = \frac{1}{\lambda_1^2}(\mathbf{B}_{1s(a_{12s})} - \mathbf{A}_{1s(a_{12s})}\mathbf{A}^T_{1s(a_{12s})})$,
- $\mathbf{I}_{12s(a_{12s})} = -\frac{1}{\lambda_1^3}(\mathbf{B}_{1s(a_{12s})} - \mathbf{A}_{1s(a_{12s})}\mathbf{A}^T_{1s(a_{12s})})\boldsymbol{\beta}$,
- $I_{22s(a_{12s})} = \frac{1}{\lambda_1^4}\boldsymbol{\beta}^T(\mathbf{B}_{1s(a_{12s})} - \mathbf{A}_{1s(a_{12s})}\mathbf{A}^T_{1s(a_{12s})})\boldsymbol{\beta}$,
- $p_{1s(j)} = \frac{\exp\left(\mathbf{f}^T(a_{j's})\boldsymbol{\beta}\right)}{\exp\left(\mathbf{f}^T(a_{j's})\boldsymbol{\beta}\right) + \exp\left(\mathbf{f}^T(a_{12s}s)\beta\right)}; j, j' \in \{a_{11s}, a_{21s}\}; j \neq j'$ (see Section 6.2.1),

where $\mathbf{F}_{ms(j)}$, $p_{ms(j)}$ denote design matrix and the probability of choosing nest (in choice set $\mathbb{C}_s$) when alternative j has been removed, also, $\mathbf{B}$ and $\mathbf{A}$ have same definition with the element of the information matrix calculated in Lemma 4.2.

Table 6.2: Two-level Nested MNL model: Four choice sets, $\mathbb{C}_s; s = 1,2,3,4$, each with three alternatives (a_{jms} denote the j^{th} alternative of the nest m in choice set s)

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

Example 6.1. *For a two-level RO.NMNL model, we have two attributes each with two levels, where* $\mathbb{C} = \{\tilde{a}_{11}, \tilde{a}_{21}, \tilde{a}_{12}, \tilde{a}_{22}\}$. *Let us consider experiments, which include four choice sets each with three alternatives (Table 6.2, Figure 6.1), where* $\boldsymbol{\beta}_1 = (\beta_{1,1}, -\beta_{1,1})^T$, $\boldsymbol{\beta}_2 = (\beta_{2,1}, -\beta_{2,1})^T$ *and for simplicity we consider* $\beta_{1,1} = \beta_1, \beta_{2,1} = \beta_2$ *(Section 4.2) thus we can write* $\boldsymbol{\theta} = (\beta_1, \beta_2, \lambda_1, \lambda_2)^T$. *Also, according to Table 6.2 the characterizes of the alternatives of choice sets are as follow:*

- $\mathbf{f}(\tilde{a}_{11}) = [\ +1 \quad +1\ ]^T$ and $\mathbf{f}(\tilde{a}_{21}) = [\ -1 \quad +1\ ]^T$
- $\mathbf{f}(\tilde{a}_{12}) = [\ +1 \quad -1\ ]^T$ and $\mathbf{f}(\tilde{a}_{22}) = [\ -1 \quad -1\ ]^T$

Thus we define the following general design (See Example 4.2):

$$\xi = \left\{ \begin{array}{cccc} \mathbb{C}_1 & \mathbb{C}_2 & \mathbb{C}_3 & \mathbb{C}_4 \\ w_1 & w_2 & w_3 & w_4 \end{array} \right\} \in \Xi. \tag{6.9}$$

The information matrix of the design (6.9) is calculated by:

$$\mathbf{I}_{R(NMNL)}(\xi, \boldsymbol{\theta}) = \sum_{s=1}^{4} w_s \mathbf{I}_{R_{(NMNL)}}(\mathbb{C}_s, \boldsymbol{\theta}), \tag{6.10}$$

where $\mathbf{I}_{R_{(NMNL)}}(\mathbb{C}_s, \boldsymbol{\theta})$ for each of choice sets are calculated by (Lemma 6.3 and Subsection 6.2.1):

<u>**For** $\mathbb{C}_1$</u>:

$$\mathbf{I}_{R_{(NMNL)}}(\mathbb{C}_1, \boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I}_{111(0)} & \mathbf{I}_{121(0)} & \mathbf{0} \\ \mathbf{I}_{121(0)}^T & I_{221(0)} & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix} + P_{a_{111}11} \begin{bmatrix} \mathbf{I}_{111(a_{111})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix}$$

$$+ P_{a_{211}11} \begin{bmatrix} \mathbf{I}_{111(a_{211})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix} + P_{a_{121}21} \begin{bmatrix} \mathbf{I}_{111(a_{121})} & \mathbf{I}_{121(a_{121})} & \mathbf{0} \\ \mathbf{I}_{121(a_{121})} & I_{221(a_{121})} & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix},$$

where

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

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

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

$$\mathbf{I}_{111(a_{111})} = 4p_{11(a_{111})}(1-p_{11(a_{111})}) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{I}_{111(a_{211})} = 4p_{11(a_{211})}(1-p_{11(a_{211})}) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{I}_{111(a_{121})} = \frac{4p_{1|11}(1-p_{1|11})}{\lambda_1^2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{I}_{121(a_{121})} = -\frac{4p_{1|11}(1-P_{1|11})}{\lambda_1^3} \begin{bmatrix} \beta_1 \\ 0 \end{bmatrix}$$

$$I_{221(a_{121})} = \frac{4p_{1|11}(1-p_{1|11})}{\lambda_1^4}\beta_1^2, \quad p_{11(a_{111})} = \frac{\exp(-\beta_1+\beta_2)}{\exp(-\beta_1+\beta_2)+\exp(\beta_1-\beta_2)}, \quad p_{11(a_{211})} = \frac{\exp(\beta_1+\beta_2)}{\exp(\beta_1+\beta_2)+\exp(\beta_1-\beta_2)}$$

$$p_{11(0)} = \frac{\left(\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)+\exp\left(\frac{-\beta_1+\beta_2}{\lambda_1}\right)\right)^{\lambda_1}}{\left(\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)+\exp\left(\frac{-\beta_1+\beta_2}{\lambda_1}\right)\right)^{\lambda_1}+\exp(\beta_1-\beta_2)}, \quad p_{1|11} = \frac{\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)}{\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)+\exp\left(\frac{-\beta_1+\beta_2}{\lambda_1}\right)}.$$

Now, let $\boldsymbol{\beta} = \mathbf{0}$, then the previous information matrix is calculated by:

$$\mathbf{I}_{R(NMNL)}(\mathbb{C}_1, \boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{2}.\frac{2.2^{\lambda_1+1}+2.4^{\lambda_1}+6.2^{\lambda_1-1}{\lambda_1}^2+{\lambda_1}^2 4^{\lambda_1}+2}{{\lambda_1}^2\left(1+2^{\lambda_1}\right)^2} & -\frac{1}{2}.\frac{10.2^{\lambda_1-1}+4^{\lambda_1}}{\left(1+2^{\lambda_1}\right)^2} & -\frac{2^{\lambda_1}\ln(2)}{\left(1+2^{\lambda_1}\right)^2} & 0 \\ -\frac{1}{2}.\frac{10.2^{\lambda_1-1}+4^{\lambda_1}}{\left(1+2^{\lambda_1}\right)^2} & \frac{5.2^{\lambda_1}+4^{\lambda_1}}{\left(1+2^{\lambda_1}\right)^2} & \frac{2^{\lambda_1+1}\ln(2)}{\left(1+2^{\lambda_1}\right)^2} & 0 \\ -\frac{2^{\lambda_1}\ln(2)}{\left(1+2^{\lambda_1}\right)^2} & \frac{2^{\lambda_1+1}\ln(2)}{\left(1+2^{\lambda_1}\right)^2} & \frac{2^{\lambda_1}(\ln(2))^2}{\left(1+2^{\lambda_1}\right)^2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

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

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

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

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

$$\mathbf{I}_{112(a_{112})} = 4p_{12(a_{112})}(1-p_{12(a_{112})}) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{I}_{112(a_{212})} = 4p_{12(a_{212})}(1-p_{12(a_{212})}) \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

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

$$I_{222(a_{122})} = \frac{4p_{1|11}(1-p_{1|11})}{\lambda_1^4}\beta_1^2, \quad p_{12(a_{112})} = \frac{\exp(-\beta_1+\beta_2)}{\exp(-\beta_1+\beta_2)+\exp(-\beta_1-\beta_2)}, \quad p_{12(a_{212})} = \frac{\exp(\beta_1+\beta_2)}{\exp(\beta_1+\beta_2)+\exp(-\beta_1-\beta_2)}$$

$$p_{1|11} = \frac{\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)}{\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)+\exp\left(\frac{-\beta_1+\beta_2}{\lambda_1}\right)}, \quad p_{12(0)} = \frac{\left(\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)+\exp\left(\frac{-\beta_1+\beta_2}{\lambda_1}\right)\right)^{\lambda_1}}{\left(\exp\left(\frac{\beta_1+\beta_2}{\lambda_1}\right)+\exp\left(\frac{-\beta_1+\beta_2}{\lambda_1}\right)\right)^{\lambda_1}+\exp(-\beta_1-\beta_2)}.$$

When $\boldsymbol{\beta} = \mathbf{0}$, then the information matrix (w.r.t $\mathbb{C}_2$) is calculated as follows:

$$\mathbf{I}_{R_{(NMNL)}}(\mathbb{C}_2, \boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{2}.\frac{2.2^{1+\lambda_1}+2.4^{\lambda_1}+6.2^{\lambda_1-1}{\lambda_1}^2+{\lambda_1}^2 4^{\lambda_1}+2}{{\lambda_1}^2\left(1+2^{\lambda_1}\right)^2} & \frac{1}{2}\frac{10.2^{\lambda_1-1}+4^{\lambda_1}}{\left(1+2^{\lambda_1}\right)^2} & \frac{2^{\lambda_1}\ln(2)}{\left(1+2^{\lambda_1}\right)^2} & 0 \\ \frac{1}{2}\frac{10.2^{\lambda_1-1}+4^{\lambda_1}}{\left(1+2^{\lambda_1}\right)^2} & \frac{5.2^{\lambda_1}+4^{\lambda_1}}{\left(1+2^{\lambda_1}\right)^2} & \frac{2^{1+\lambda_1}\ln(2)}{\left(1+2^{\lambda_1}\right)^2} & 0 \\ \frac{2^{\lambda_1}\ln(2)}{\left(1+2^{\lambda_1}\right)^2} & \frac{2^{1+\lambda_1}\ln(2)}{\left(1+2^{\lambda_1}\right)^2} & \frac{2^{\lambda_1}(\ln(2))^2}{\left(1+2^{\lambda_1}\right)^2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

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

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

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

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

$$\mathbf{I}_{113(a_{223})} = 4p_{13(a_{223})}(1-p_{13(a_{223})})\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{I}_{113(a_{123})} = 4p_{13(a_{123})}(1-p_{13(a_{123})})\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

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

$$I_{333(a_{213})} = \frac{4p_{3|23}(1-p_{3|23})}{\lambda_2^4}\beta_1^2, \quad p_{13(a_{223})} = \frac{\exp(\beta_1+\beta_2)}{\exp(\beta_1+\beta_2)+\exp(\beta_1-\beta_2)}, \quad p_{13(a_{123})} = \frac{\exp(\beta_1+\beta_2)}{\exp(-\beta_1-\beta_2)+\exp(\beta_1+\beta_2)}$$

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

And with the assumption $\boldsymbol{\beta}=\mathbf{0}$ thus:

$$\mathbf{I}_{R_{(NMNL)}}(\mathbb{C}_3,\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{2}.\frac{2.2^{1+\lambda_2}+2.4^{\lambda_2}+6.2^{\lambda_2-1}\lambda_2{}^2+\lambda_2{}^2 4^{\lambda_2}+2}{\lambda_2{}^2\left(1+2^{\lambda_2}\right)^2} & \frac{1}{2}.\frac{10.2^{\lambda_2-1}+4^{\lambda_2}}{\left(1+2^{\lambda_2}\right)^2} & 0 & -\frac{2^{\lambda_2}\ln(2)}{\left(1+2^{\lambda_2}\right)^2} \\ \frac{1}{2}.\frac{10.2^{\lambda_2-1}+4^{\lambda_2}}{\left(1+2^{\lambda_2}\right)^2} & \frac{5.2^{\lambda_2}+4^{\lambda_2}}{\left(1+2^{\lambda_2}\right)^2} & 0 & -\frac{2^{1+\lambda_2}\ln(2)}{\left(1+2^{\lambda_2}\right)^2} \\ 0 & 0 & 0 & 0 \\ -\frac{2^{\lambda_2}\ln(2)}{\left(1+2^{\lambda_2}\right)^2} & -\frac{2^{1+\lambda_2}\ln(2)}{\left(1+2^{\lambda_2}\right)^2} & 0 & \frac{2^{\lambda_2}(\ln(2))^2}{\left(1+2^{\lambda_2}\right)^2} \end{bmatrix}.$$

For $\mathbb{C}_4$:

$$\mathbf{I}_{R_{(NMNL)}}(\mathbb{C}_4,\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I}_{114(0)} & \mathbf{0} & \mathbf{I}_{134(0)} \\ \mathbf{0} & 0 & 0 \\ \mathbf{I}_{134(0)}^T & 0 & I_{334(0)} \end{bmatrix} + P_{a_{114}14}\begin{bmatrix} \mathbf{I}_{114(a_{114})} & \mathbf{0} & \mathbf{I}_{134(a_{114})} \\ \mathbf{0} & 0 & 0 \\ \mathbf{I}_{134(a_{114})}^T & 0 & I_{334(a_{114})} \end{bmatrix}$$
$$+P_{a_{224}24}\begin{bmatrix} \mathbf{I}_{114(a_{224})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix} + P_{a_{124}24}\begin{bmatrix} \mathbf{I}_{114(a_{124})} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & 0 \\ \mathbf{0} & 0 & 0 \end{bmatrix}$$

where

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

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

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

$$\mathbf{I}_{114(a_{224})} = 4p_{14(a_{224})}(1-p_{14(a_{224})}) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{I}_{114(a_{124})} = 4p_{14(a_{124})}(1-p_{14(a_{124})}) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{I}_{114(a_{114})} = \frac{4p_{3|23}(1-p_{3|23})}{\lambda_2^2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{I}_{134(a_{114})} = -\frac{4p_{3|23}(1-p_{3|23})}{\lambda_2^3} \begin{bmatrix} \beta_1 \\ 0 \end{bmatrix} \quad I_{334(a_{114})} = \frac{4p_{3|23}(1-p_{3|23})}{\lambda_2^4}\beta_1^2,$$

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

$$p_{14(0)} = \frac{\exp(-\beta_1+\beta_2)}{\exp(-\beta_1+\beta_2)+\left(\exp\left(\frac{\beta_1-\beta_2}{\lambda_2}\right)+\exp\left(\frac{-\beta_1-\beta_2}{\lambda_2}\right)\right)^{\lambda_2}}, \quad p_{3|23} = \frac{\exp\left(\frac{\beta_1-\beta_2}{\lambda_2}\right)}{\exp\left(\frac{\beta_1-\beta_2}{\lambda_2}\right)+\exp\left(\frac{-\beta_1-\beta_2}{\lambda_2}\right)}.$$

When $\boldsymbol{\beta} = \mathbf{0}$ we will have ;

$$\mathbf{I}_{R(NMNL)}(\mathbb{C}_4, \boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{2}.\frac{2\,2^{1+\lambda_2}+2.4^{\lambda_2}+6.2^{\lambda_2-1}{\lambda_2}^2+{\lambda_2}^2 4^{\lambda_2}+2}{{\lambda_2}^2\left(1+2^{\lambda_2}\right)^2} & -\frac{1}{2}.\frac{10.2^{\lambda_2-1}+4^{\lambda_2}}{\left(1+2^{\lambda_2}\right)^2} & 0 & \frac{2^{\lambda_2}\ln(2)}{\left(1+2^{\lambda_2}\right)^2} \\ -\frac{1}{2}\frac{10.2^{\lambda_2-1}+4^{\lambda_2}}{\left(1+2^{\lambda_2}\right)^2} & \frac{5.2^{\lambda_2}+4^{\lambda_2}}{\left(1+2^{\lambda_2}\right)^2} & 0 & -\frac{2^{1+\lambda_2}\ln(2)}{\left(1+2^{\lambda_2}\right)^2} \\ 0 & 0 & 0 & 0 \\ \frac{2^{\lambda_2}\ln(2)}{\left(1+2^{\lambda_2}\right)^2} & -\frac{2^{1+\lambda_2}\ln(2)}{\left(1+2^{\lambda_2}\right)^2} & 0 & \frac{2^{\lambda_2}(\ln(2))^2}{\left(1+2^{\lambda_2}\right)^2} \end{bmatrix},$$

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 6.3, which includes some locally D-optimal design based on the RO.NMNL model.

Table 6.4 denotes some locally D-optimal designs for C.NMNL (Classical NMNL) model. Based on Table 6.4 and Table 6.5, it is seen that:

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

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 alternatives in choice sets $\mathbb{C}_s; s = 1, 3$, Chapter 4) 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 $w_2 = w_4$ for the design (6.9), where $2w_1 + 2w_2 = 1$ or $w_2 = \frac{1}{2} - w_1$ as follows:

$$\xi = \left\{ \begin{array}{cccc} \mathbb{C}_1 & \mathbb{C}_2 & \mathbb{C}_3 & \mathbb{C}_4 \\ w_1 & \frac{1}{2} - w_1 & w_1 & \frac{1}{2} - w_1 \end{array} \right\} \in \Xi \qquad (6.11)$$

Table 6.3: RO.NMNL model $\lambda_1 = .6, \lambda_2 = .4$ (Two nests): Locally D-optimal design for Design (6.9), 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^*	$\Psi_R(\xi^*, \boldsymbol{\theta}_0)$
-.8	-.7	.489	.000	.297	.214	1.370
	-.4	.534	.000	.466	.000	1.233
	0.0	.500	.000	.500	.000	1.164
	.3	.473	.000	.527	.000	1.183
	.6	.459	.000	.541	.000	1.269
-.3	-.7	.487	.019	.494	.000	1.383
	-.4	.501	.000	.499	.000	1.170
	0.0	.504	.000	.496	.000	1.061
	.3	.505	.000	.495	.000	1.096
	.6	.507	.000	.493	.000	1.230
0.0	-.7	.235	.242	.260	.263	3.680
	-.4	.249	.248	.252	.251	2.974
	0.0	.273	.260	.235	.232	2.702
	.3	.293	.268	.223	.217	2.943
	.6	.309	.273	.212	.206	3.619
.6	-.7	.000	.528	.232	.240	1.183
	-.4	.000	.533	.060	.407	1.102
	0.0	.000	.509	.000	.491	1.069
	.3	.256	.262	.000	.482	1.088
	.6	.467	.079	.000	.454	1.146
.9	-.7	.000	.465	.360	.175	1.551
	-.4	.000	.496	.119	.385	1.493
	0.0	.124	.369	.000	.507	1.447
	.3	.460	.024	.000	.516	1.382
	.6	.521	.000	.000	.479	1.359

Table 6.4: C.NMNL model, $\lambda_1 = .6, \lambda_2 = .4$ (two nests): Locally D-optimal design for Design (6.9), where there are four choice sets each with three alternatives; w.r.t local D-optimal criterion, $\Psi_R(\xi, \boldsymbol{\theta}_0) = (\det(\mathbf{I}_{R(NMNL)}(\xi, \boldsymbol{\theta}_0)))^{-1}$ (Comparing to Table 6.3)

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

Table 6.5: RO.NMNL model, $\beta_1 = \beta_2 = 0.0$ (two nests): Locally D-optimal design for Design (6.9), where there are four choice sets each with three alternatives; w.r.t local D-optimal criterion, $\Psi_R(\xi, \boldsymbol{\theta}_0) = (\det(\mathbf{I}_{R(NMNL)}(\xi, \boldsymbol{\theta}_0)))^{-1}$

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
	0.4	0.378	0.332	0.149	0.140	1.366
0.1	0.5	0.379	0.333	0.148	0.140	1.371
	0.6	0.380	0.333	0.148	0.139	1.356
	0.7	0.380	0.332	0.149	0.139	1.381
	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

In this case, $w_1^* = 0.5$ when $\beta_1 < -0.05$ (Table 6.6) 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 $0 \leq \beta_1 < 0.01$, for too small values of β_1) then increases (when $0.05 \leq \beta_1$) as β_1 increases (Table 6.6). For example, 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 6.6 (the Sequential Quadratic Programming method by MAPLE).

Let us consider $\lambda = 1.0$. In this situation, Table 6.6 denotes that w_1^* decreases as β_1 increases. So that, we will face to (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 tend 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 $\mathbb{C}_4$. In this situation, we consider the following invariant design instead:

$$\xi = \left\{ \begin{array}{cccc} \mathbb{C}_1 & \mathbb{C}_2 & \mathbb{C}_3 & \mathbb{C}_4 \\ w & w & \frac{1}{2} - w & \frac{1}{2} - w \end{array} \right\} \in \Xi. \tag{6.12}$$

Based on design (6.12) and the assumptions $\boldsymbol{\beta} = \mathbf{0}, \lambda_2 = 0.1$, the determinate of the information matrix of design (6.12) is calculated as follows:

$$\det(\mathbf{I}_{R_{(NMNL)}}(\xi, \boldsymbol{\theta})) = \{w[\tfrac{2^{\lambda_1+1}}{(2^{\lambda_1}+1)}(\lambda_1^{-2} + 1 - \tfrac{2^{\lambda_1}}{2^{\lambda_1}+1}) + \tfrac{2^{\lambda_1}(\lambda_1^2+3)+4}{\lambda_1^2(2^{\lambda_1}+1)(2^{\lambda_1}+2)}] + 100.5083608$$
$$-201.0167216\, w\}(-2.961797575\, w 2^{\lambda_1} - 1.997599657\, w 2^{2\lambda_1-1} - 1.997599657\, w 2^{\lambda_1-1}$$
$$-0.2583504347\,(2^{\lambda_1})^2 - 0.5167008694\, 2^{\lambda_1} - 0.2583504347 + 0.5167008694\, w (2^{\lambda_1})^2$$
$$+0.5167008694\, w)(\tfrac{w 2^{\lambda_1}(\ln(2))^4(-1+2\,w)}{(2^{\lambda_1}+1)^4}).$$

Table 6.7 denote 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 showed by Table 6.7 (It has been 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 6.7). Also, with respect to the fixed values for the dissimilarity parameter λ_2, Table 6.7 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

Table 6.6: RO.NMNL model, $\beta_2 = 0$, $\lambda_1 = \lambda_2 = \lambda$ (two nests): Locally D-optimal design, w_1^*, with respect to Design (6.11)

		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 6.7: RO.NMNL model, $\boldsymbol{\beta} = \mathbf{0}$: Locally D-optimal design, w^*, with respect to Design (6.12)

		λ_1									
		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**

increases. That means that 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 6.7 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.

7 DISCUSSION and EXTENSIONS

In this thesis, we have applied optimal design theory to the area of conjoint analysis, in particular to find optimal combination of alternatives in the choice sets. We have used some models from discrete choice models (MNL, NMNL) to analyze data. When the variance of the MNL model for all of decision makers are as the same, then this model is also called Homoscedastic MNL model (Section 2.1). But, when the error terms in the utility function, $U_{ij} = v_{ijc} + \varepsilon_{ijc}$, are in fact heteroscedastic across decision makers, the general form of the extreme value distribution suggests the use of the following distribution (w.r.t Choice $\mathbb{C}_c$):

$$F_{\varepsilon_{ijc}}(\epsilon_{ijc}) = \exp((-\exp((-\epsilon_{ijc}\mu_i)))).$$

Since in this distribution the variance of ε_{ijc} is $\frac{\pi^2}{6\mu_i^2}$ which can vary with i. In practice, μ_i can take the form of a function of an independent variable that takes different values across decision makers. Thus if $\mu_i > 0$ (for example, $\mu_i = e^{\mathbf{z}_i^T \boldsymbol{\delta}}$), the choice probabilities from this distribution can be derived in a simple manner analogous to weighted least squares in the linear model. In the new heteroscedastic utility model, weight both sides of the equation with μ_i:

$$U_{ijc}\mu_i = v_{ijc}\mu_i + \varepsilon_{ijc}\mu_i,$$

where $v_{ijc}\mu_i$ is new deterministic component and $\varepsilon_{ijc}\mu_i$ is new probabilistic component and $\mathbf{z}_i$ (which may or may not be part of $\mathbf{x}_{ijs}$) believed to influence the error variances. Then, the new error terms will have a constant variance $\frac{\pi^2}{6}$, like the *logit* model and:

$$p_{ijc} = Pr\left(U_{ijc} > U_{ij'c}, \forall j \neq j'\right) = Pr\left(U_{ijc}\mu_i > U_{ij'c}\mu_i, \forall j \neq j'\right) = \frac{\exp(v_{ijc}\mu_i)}{\sum_{j'=1}^{J_c} \exp(v_{ij'c}\mu_i)}. \quad (7.1)$$

According to model (7.1) can be considered a specification test for heteroscedasticity model as follow:

$$\begin{cases} H_0 : \mu_i = 1;\ \forall i \\ H_1 : \mu_i \neq 1;\ \exists i \end{cases} \quad or \quad \begin{cases} H_0 : \boldsymbol{\delta} = \mathbf{0} \\ H_1 : \boldsymbol{\delta} \neq \mathbf{0} \end{cases}.$$

Since, the estimator of the heteroscedastic model is not difficult, the likelihood ratio test can be easily performed. In this case, $LRT = -2\ln\frac{L\left(\tilde{\boldsymbol{\beta}}, \boldsymbol{\delta}=\mathbf{0}\right)}{L_h\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}\right)}$ has chi-square distribution (asymptotically) with degree of freedom, which is equal to the number of parameters in $\boldsymbol{\delta}$, where L_h is the likelihood at convergence from the heteroscedastic *logit* model. Thus, if $LRT > \chi^2_{\alpha,p'}$ then H_0 will be rejected, else it is accepted, where p' is dimension of parameter $\boldsymbol{\delta}$.

7.1 Discuss about IIA

In the Chapter 2 it has been described the property of the MNL model, mean that IIA. As has stated previously told, we can use the MNL model to analyze data if IIA be held. In this situation, the statistical tests can be performed to test, which the IIA property hold a particular application. When these tests show that the IIA holds the MNL model can be used. For example, one way can be to test the IIA assumption by estimating (Two-way) interaction effects.

The second way for testing IIA, the Corollary 2.2 can be used, so that the MNL model specification will test by comparing parameter estimates obtained from choice data from the full choice set with estimates obtained from conditional choice data from a restricted choice set.

Also, we can consider a generalization of the MNL model which is called the Nested Logit model, which has been discussed about it in the Section 2.2.

A Test for IIA

As stated previously (Subsection 2.1.2), the MNL model has a property that is called Independence from Irrelevant Alternative, this mean that, if A be a subset of $\mathbb{C}_c$ thus (Corollary 2.1):

$$p_{ijA} = \frac{\exp(\mathbf{f'}^T(a_{ijc})\boldsymbol{\beta})}{\sum_{j' \in A} \exp(\mathbf{f'}^T(a_{j'c})\boldsymbol{\beta})}, \tag{7.2}$$

where $\mathbf{f'}^T(a_{jc})$ denotes the characterizes of attributes in choice set A ($[f'_k(a_{ijc})]_{k=1,2,\ldots,K'}^{j=1,2,\ldots,J'_c} \subset [f_k(a_{ijc})]_{k=1,2,\ldots,K}^{j=1,2,\ldots,J_c}$; $i = 1, 2, \ldots, \mathcal{I}$; $K' \leq K$ and $J'_c \leq J_c$). In this situation, the *Hausman*' test is on the based on eliminating one or more alternatives from the choice set to see if underlying choice behavior from the restricted choice set obeys the IIA property .

We estimate the unknown parameters from both the restricted and unrestricted choice sets. If the parameter estimates are here approximately the same, then we do not reject the MNL model specification, otherwise it will be rejected. In this case, we act as follow:

Suppose that $i = 1, 2, \ldots, \mathcal{I}$ be a random sample of individuals and $\mathbf{f}(a_{ijc})$ be the attractive of the $\mathbb{C}_c$ (choice set) for alternative j, and define $y_{ijc} = 1$ if individual i chooses alternative j (in choice set $\mathbb{C}_c$) and $y_{ijc} = 0$ otherwise. In this case, the normalized log-likelihood of the sample is showed by:

$$\ell(\mathbb{C}_c, \boldsymbol{\beta}) = \sum_{i=1}^{\mathcal{I}} \sum_{j \in \mathbb{C}_c} y_{ijc} \ln(p_{ij\mathbb{C}_c}).$$

If we have:

$$\frac{\partial \ell(\mathbb{C}_c, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0 \Rightarrow \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\mathbb{C}}, \tag{7.3}$$

where $\hat{\boldsymbol{\beta}}$ is the MLE (Maximum Likelihood Estimator) for $\boldsymbol{\beta}$, which is showed by $\hat{\boldsymbol{\beta}}_{\mathbb{C}}$ (w.r.t unrestricted model, means that there are all of alternatives in choice set, $\mathbb{C}_c$).

Note 7.1: We know that:

1. $\hat{\boldsymbol{\beta}}_{\mathbb{C}}$ is consistent asymptotically, means;

$$P\left(\lim_{\mathcal{I}\to+\infty}\hat{\boldsymbol{\beta}}_{\mathbb{C}}\right)=\boldsymbol{\beta}$$

2. $\hat{\boldsymbol{\beta}}_{\mathbb{C}}$ has normal distribution asymptotically;

$$\hat{\boldsymbol{\beta}}_{\mathbb{C}}\sim^{a} N_p(\boldsymbol{\beta},\mathbf{I}^{-1}(\boldsymbol{\beta}))\ \ ;\ \ (\hat{\boldsymbol{\beta}}_{\mathbb{C}}-\boldsymbol{\beta})\sim^{a} N_p(0,\mathbf{I}^{-1}(\beta))$$

where notation $\sim^{a}\equiv$ denotes asymptotically distribution, $\mathbf{I}^{-1}(\boldsymbol{\beta})$ is the inverse of the information matrix; $\mathbf{I}(\boldsymbol{\beta})=-E\left(\frac{\partial^2\ell(\boldsymbol{\beta};\mathbb{C}_c)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right)=-E\begin{pmatrix}\frac{\partial^2\ell(\boldsymbol{\beta};\mathbb{C}_c)}{\partial\beta_1^2} & \cdots & \frac{\partial^2\ell(\boldsymbol{\beta};\mathbb{C}_c)}{\partial\beta_1\partial\beta_p}\\ \vdots & \ddots & \vdots\\ \frac{\partial^2\ell(\boldsymbol{\beta};\mathbb{C}_c)}{\partial\beta_p\partial\beta_1} & \cdots & \frac{\partial^2\ell(\boldsymbol{\beta};\mathbb{C}_c)}{\partial\beta_p^2}\end{pmatrix}$ and p is the number of parameters.

3. $\hat{\boldsymbol{\beta}}_{\mathbb{C}}$ is efficiency asymptotically. This means that, if there exist other consistent estimators with normal distribution asymptotically, for example $\tilde{\boldsymbol{\beta}}_{\mathbb{C}}$, in this case $V(\tilde{\boldsymbol{\beta}}_{\mathbb{C}})\geq V(\hat{\boldsymbol{\beta}}_{C})$ $(V(\tilde{\boldsymbol{\beta}}_{\mathbb{C}})-V(\hat{\boldsymbol{\beta}}_{C})$ is p.s.d matrix).

Now, suppose that $A\subset\mathbb{C}_c$ and the IIA property be hold. The likelihood function is defined by:

$$\ell(A,\boldsymbol{\beta})=\sum_{i=1}^{\mathcal{I}}\sum_{j\in A} y_{ijc}\ \ln(p_{ijA}),\tag{7.4}$$

where (Subsection 2.1.3):

$$p_{ijA}=\frac{p_{ij\mathbb{C}_c}}{p_{iA\mathbb{C}_c}};\forall j\in A.$$

Here, some component of $\boldsymbol{\beta}$, such as the coefficients of alternative specific variables for excluded alternatives are not identified by choice from A. Thus, we suppose that $\mathbf{f}^T(a_{ijc})=(\mathbf{f}''^T(a_{ijc}),\mathbf{f}'^T(a_{ijc}))$ be a partition of the explanatory variable into a vector $\mathbf{f}''^T(a_{ijc})$, which only varies outside A and a vector $\mathbf{f}'^T(a_{ijc})$, which only varies within A, and let $\boldsymbol{\beta}^T=(\boldsymbol{\gamma}^T,\boldsymbol{\delta}^T)$ be a commensurate partition of the parameter vector, thus:

$$p_{ijA}=\frac{\exp(\mathbf{f}'^T(a_{ijc})\boldsymbol{\delta})}{\sum_{j'\in A}\exp(\mathbf{f}'^T(a_{ij'c})\boldsymbol{\delta})};\ \mathbf{f}''(a_{ijc})=\mathbf{f}''(a_{ij'c}),\forall j,j'\in A.$$

In this situation, $\ell(A,\boldsymbol{\beta})$ can be rewritten as follow:

$$\ell(A,\boldsymbol{\delta})=\sum_{i=1}^{\mathcal{I}}\sum_{j\in A} y_{ijc}\ln(p_{ijA}).\tag{7.5}$$

By setting to zero $\frac{\partial \ell(A,\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}$ we can find (if there exist) $\hat{\boldsymbol{\delta}}_A$, which is the MLE of $\boldsymbol{\delta}$ in subset A. The $\hat{\boldsymbol{\delta}}_A$ includes all of properties in Note 7.1.

The specification test statistic is based on the parameter difference $\boldsymbol{\theta} = \boldsymbol{\beta}_A - \boldsymbol{\delta}_{\mathbb{C}}$, where $\boldsymbol{\beta}_{\mathbb{C}} = (\boldsymbol{\gamma}_{\mathbb{C}}, \boldsymbol{\delta}_{\mathbb{C}})^T$ and $\boldsymbol{\beta}_A = \boldsymbol{\delta}_A$ when the regularity assumption hold and the MNL is true, $P(\lim_{\mathcal{I}\longrightarrow\infty} \boldsymbol{\theta}) = 0$ conversely, when the MNL specification is false, the IIA property fails, and:

$$p_{ij\mathbb{C}} \neq p_{ijA}.p_{iA\mathbb{C}_c},$$

that means, (7.5) is not maximized in $\hat{\boldsymbol{\delta}}_A$.

Now, for testing the hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$ or $H_0 : \boldsymbol{\delta}_A - \boldsymbol{\delta}_{\mathbb{C}} = \mathbf{0}$ we need a suitable test statistic. *(McFadden and Hausman 1984)* have demonstrated that the asymptotic covariance matrix of $\sqrt{\mathcal{I}}(\hat{\boldsymbol{\delta}}_A - \hat{\boldsymbol{\delta}}_{\mathbb{C}})$ is $\varphi = \Sigma_A - \Sigma_{\mathbb{C}_{\delta\delta}}$, the difference of the asymptotic covariance matrices of $\hat{\boldsymbol{\delta}}_A$ and $\hat{\boldsymbol{\delta}}_{\mathbb{C}}$, where:

$$\Sigma_{\mathbb{C}} = \begin{pmatrix} \Sigma_{\mathbb{C}_{\gamma\gamma}} & \Sigma_{\mathbb{C}_{\gamma\delta}} \\ \Sigma^T_{\mathbb{C}_{\gamma\delta}} & \Sigma_{\mathbb{C}_{\delta\delta}} \end{pmatrix}.$$

Thus the test statistic:

$$\chi = \mathcal{I}.(\hat{\boldsymbol{\delta}}_A - \hat{\boldsymbol{\delta}}_{\mathbb{C}})^T \varphi^- (\hat{\boldsymbol{\delta}}_A - \hat{\boldsymbol{\delta}}_{\mathbb{C}})_{|H_0} \sim^a \chi^2_p, \tag{7.6}$$

has asymptotically distributed Chi-Square with degrees of freedom equal to the rank of φ, under the null hypothesis, where φ^- is a generalized inverse of φ.

Result:

$$if \begin{cases} \chi > \chi^2_{\alpha,p} & \textit{we reject } H_0 \textit{ with significance level } \alpha, \\ \chi < \chi^2_{\alpha,p} & \textit{we accept } H_0 : \boldsymbol{\delta}_A = \boldsymbol{\delta}_{\mathbb{C}} \textit{ and the MNL is true.} \end{cases}$$

Example 7.1. *Let $\mathbb{C}_c = \{a_{1c}, a_{2c}, a_{3c}\}$ (a choice set) and $f(a_{1c}) = 1$, $f(a_{2c}) = f(a_{3c}) = 0$ (there is an attribute with two levels one and zero), in this situation we assume that $A_1 = \{a_{1c}, a_{2c}\}$, $A_2 = \{a_{1c}, a_{3c}\}$ and $A_3 = \{a_{2c}, a_{3c}\}$ are three subsets of choice set $\mathbb{C}_c$. In this case, we are going to test:*

$$H_0 : \boldsymbol{\delta}_{\mathbb{C}} = \boldsymbol{\delta}_{A_r}; r = 1, 2, 3.$$

Based on all of alternatives in choice set $\mathbb{C}_c$, the log-likelihood function is defined as follow (Section 2.1):

$$\ell(\mathbb{C}_c, \boldsymbol{\beta}) = \sum_{i=1}^{\mathcal{I}} \sum_{j\in\mathbb{C}_c} y_{ijc} \ \ln(p_{ij\mathbb{C}_c}). \tag{7.7}$$

If, we assume $\mathcal{I} = \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3$ are the number of individuals, where $\mathcal{I}_j; j = 1, 2, 3$ denote the number of individuals, which selected the alternative j, thus we can rewrite $\ell(\mathbb{C}_c, \boldsymbol{\beta})$ as follows:

$$\ell(\mathbb{C}_c, \boldsymbol{\beta}) = \mathcal{I}_1 \ln(p_{a_1\mathbb{C}_c}) + \mathcal{I}_2 \ln(p_{a_2\mathbb{C}_c}) + \mathcal{I}_3 \ln(p_{a_3\mathbb{C}_c}).$$

According to $U_{ijc} = \mathbf{f}^T(a_{ijc})\boldsymbol{\beta}+\varepsilon_{ijc}$ (here, $\mathbf{f}^T(a_{ijc})$ has just one element to determine alternative j) and with assumption the IIA we will have:

$$p_{a_1\mathbb{C}_c} = \frac{\exp(\beta)}{(2+\exp(\beta))} \quad , \quad p_{a_2\mathbb{C}_c} = \frac{1}{(2+\exp(\beta))} = p_{a_3\mathbb{C}_c}.$$

Hence, Equation (7.7) will be rewritten by:

$$\ell(\mathbb{C}_c, \boldsymbol{\beta}) = \mathcal{I}_1\beta - \mathcal{I}_1 \ln(2+\exp(\beta)) - \mathcal{I}_2 \ln(2+\exp(\beta)) - \mathcal{I}_3 \ln(2+\exp(\beta)).$$

By setting derivative $\ell(\mathbb{C}_c, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ into zero will be:

$$\exp(\hat{\boldsymbol{\beta}}) = \frac{2\mathcal{I}_1}{\mathcal{I}_2+\mathcal{I}_3} \quad \Rightarrow \quad \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\mathbb{C}} = \ln(\frac{2\mathcal{I}_1}{\mathcal{I}_2+\mathcal{I}_3}).$$

Since $\boldsymbol{\beta}$ has one dimension then, let us $\boldsymbol{\beta} = \beta_{\mathbb{C}}$:

$$I^{-1}(\beta_{\mathbb{C}}) = \frac{\mathcal{I}}{\mathcal{I}_1(\mathcal{I}_2+\mathcal{I}_3)} = \Sigma_{\beta_{\mathbb{C}}} \quad \Rightarrow \quad (\hat{\beta}_{\mathbb{C}} - \beta_{\mathbb{C}}) \sim^a N(0, \frac{\mathcal{I}}{\mathcal{I}_1(\mathcal{I}_2+\mathcal{I}_3)}).$$

Now, according to subset A_1 we can write:

$$\ell(A_1, \delta_1) = \sum_{i=1}^{\mathcal{I}} \sum_{j\in A_1} y_{ijc} \ln(p_{jA_1}) = \mathcal{I}_1\delta_1 - \mathcal{I}_1 \ln(1+e^{\delta_1}) - \mathcal{I}_2 \ln(1+e^{\delta_1}).$$

Withe regards to $\frac{\partial \ell(A_1,\delta_1)}{\partial \delta_1} = 0 \quad \Rightarrow \quad \hat{\delta}_1 = \hat{\delta}_{A_1} = \ln(\frac{\mathcal{I}_1}{\mathcal{I}_2})$, since δ_1 has one dimension thus:

$$I^{-1}(\delta_{A_1}) = \frac{\mathcal{I}_1+\mathcal{I}_2}{\mathcal{I}_1\mathcal{I}_2} = \Sigma_{A_1} \quad \Rightarrow \quad \varphi_1 = \Sigma_{A_1} - \Sigma_{\mathbb{C}_{\delta\delta}} = \frac{\mathcal{I}_3}{\mathcal{I}_2(\mathcal{I}_2+\mathcal{I}_3)}$$

and

$$\chi^{(1)} = (\hat{\delta}_{A_1} - \hat{\beta}_{\mathbb{C}})^T \varphi^{-1} (\hat{\delta}_{A_1} - \hat{\beta}_{\mathbb{C}}) = \left(\ln \left(\frac{\mathcal{I}_2+\mathcal{I}_3}{2\mathcal{I}_2} \right)^2 . \frac{\mathcal{I}_2(\mathcal{I}_2+\mathcal{I}_3)}{\mathcal{I}_3} \right) \sim^a \chi^2_1. \tag{7.8}$$

Hence if $\chi^{(1)} > \chi^2_{\alpha,1}$ then we reject $H_0 : \delta_1 - \beta = 0$, else it will be accepted.

Similarity and based on subset A_2 we will have, $\hat{\delta}_2 = \hat{\delta}_{A_2} = \ln(\frac{\mathcal{I}_1}{\mathcal{I}_3})$, where:

$$I^{-1}(\theta_{A_2}) = \frac{\mathcal{I}_1+\mathcal{I}_3}{\mathcal{I}_1\mathcal{I}_3} = \Sigma_{A_2} \tag{7.9}$$

then:

$$\chi^{(2)} = (\hat{\delta}_{A_2} - \hat{\beta}_{\mathbb{C}})^T \varphi_2^- (\hat{\delta}_{A_2} - \hat{\beta}_{\mathbb{C}}) = (\ln \frac{\mathcal{I}_2+\mathcal{I}_3}{2\mathcal{I}_3})^2 \cdot \frac{\mathcal{I}_3(\mathcal{I}_2+\mathcal{I}_3)}{\mathcal{I}_2} \sim^a \chi^2_1. \tag{7.10}$$

Consequently, if $\chi^{(2)} > \chi^2_{\alpha,1}$, $H_0 : \delta_2 - \beta = 0$ will be rejected, else it will be accepted.

In this case, because of $f''(a_{2c}) = f''(a_{3c}) = 0$, then $\chi^{(3)}$ can not be calculated (dose not exist any parameter). Hence, if either $\chi^{(1)}$ or $\chi^{(2)}$ are not significant, then the MNL model is true, else the MNL is false and we shall not use this model to analyze data. In this situation, the Nested logit model can be used to analyze data (Section 2.2).

7.2 More About NMNL Models

In Section 2.2 have been described about the standard NMNL models. *(Zeng 2000)* has proposed another particularity useful case of the heteroscedastic GEV model, which is the heteroscedastic nested *logit* model. For example if there is a choice set with three alternatives ($\mathbb{C}_c = \{a_{1c}, a_{2c}, a_{3c}\}$, a_{jc} denotes alternative j in choice set c), which have been grouped in two nests, so that first nest includes two alternative $\mathbb{C}_{1c} = \{a_{11c} = a_{2c}, a_{21c} = a_{3c}\}; J_{1c} = 2$ (a_{jmc} the alternative j of the nest m in choice set c) and the second nest includes remain alternative $\mathbb{C}_{2c} = \{a_{12c} = a_{1c}\}; J_{2c} = 1$. Thus, we will have:

- $p_{i2c} = \frac{\exp(v_{ia_1|2c}\mu_{i1})}{\exp(v_{ia_1|2c}\mu_{i1})+(\exp(\frac{v_{ia_2|1c}\mu_{i2}}{\lambda})+\exp(\frac{v_{ia_3|1c}\mu_{i3}}{\lambda}))^{\lambda}}$,
- $p_{i1s} = \frac{(\exp(\frac{v_{ia_2|1c}\mu_{i2}}{\lambda})+\exp(\frac{v_{ia_3|1c}\mu_{i3}}{\lambda}))^{\lambda}}{\exp(v_{ia_1|2c}\mu_{i1})+(\exp(\frac{v_{ia_2|1c}\mu_{i2}}{\lambda})+\exp(\frac{v_{ia_3|1c}\mu_{i3}}{\lambda}))^{\lambda}}$,

where p_{imc} denotes the marginal choice probability related to nest m in choice set $\mathbb{C}_c$ by individual i and the conditional probabilities are as follow:

$$p_{ij|2c} = \frac{\exp(\frac{v_{ij|2c}\mu_{ij}}{\lambda})}{\exp(\frac{v_{ia_2|2c}\mu_{i2}}{\lambda}) + \exp(\frac{v_{ia_3|2c}\mu_{i3}}{\lambda})}; j = a_{2c}, a_{3c}.$$

The standard nested *logit* model avoids the IIA restriction of the standard *logit* model. The heteroscedastic nested logit model further relaxes the assumption of homoscedastic errors (in nests) in the nested logit model, the achieving greater functional flexibility.

In general, estimation of the heteroscedastic GEV model can be carried out using standard maximum likelihood techniques, similar to the heteroscedastic *logit* model.

Now, let μ_{ij} be a function of observable variables that takes positive values only, for example:

$$\mu_{ij} = \exp((1 + \mathbf{z}_j\boldsymbol{\delta}_1 + \mathbf{z}_i\boldsymbol{\delta}_2)), \tag{7.11}$$

or:

$$\mu_{ij} = \exp((\mathbf{z}_j\boldsymbol{\delta}_1 + \mathbf{z}_i\boldsymbol{\delta}_2)), \tag{7.12}$$

where $\mathbf{z}_j$ and $\mathbf{z}_i$ can be vectors varying with choice alternatives and decision makers (individuals), respectively. The Z variables can be part of the independent variables, X, that enter the utility functions of the choice model or they can be other observable variables believed to influence the error variances.

Here, we suppose that have sample data (Y_{ijmc}, X_{ijmc}), where $Y_{ijmc} = 1$ if individual i chooses alternative j and nest m, $Y_{ijmc} = 0$ otherwise, and X_{ijmc} are the independent variables in the utility functions. Then, the log-likelihood function is (based on one choice set, $\mathbb{C}_c$):

$$\ell(y_{ijmc}, x_{ijmc}, \boldsymbol{\theta}) = \sum_{i=1}^{\mathcal{I}} \sum_{m=1}^{M} \sum_{j=1}^{J_{mc}} y_{ijmc} \ln(p_{ijmc}), \tag{7.13}$$

where $\boldsymbol{\theta}$ is the vector of all unknown parameters of p_{ijmc} ($\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$), including the parameters $\boldsymbol{\beta}$ in $v_{ij|mc}$ (when, as usual, $v_{ij|mc}$ take the linear form $\mathbf{x}_{ij|mc}^T\boldsymbol{\beta}$), $\boldsymbol{\delta}$ in μ_{ij} and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_M)^T$. In this situation, the maximum likelihood estimates for $\boldsymbol{\theta}$ can then be obtained using standard numerical maximization methods.

Also, specification test for heteroscedasticity in the GEV model amounts to the test of:

$$\begin{cases} H_0 : \{\boldsymbol{\delta}_1 = \mathbf{0}\} \bigcap \{\boldsymbol{\delta}_2 = \mathbf{0}\} \\ H_0 : \{\boldsymbol{\delta}_1 \neq \mathbf{0}\} \bigcup \{\boldsymbol{\delta}_2 \neq \mathbf{0}\}. \end{cases}$$

Because, the restricted model is a parametric special case of the unrestricted model, classical test such as WALD, Likelihood Ratio and Lagrange Multiplier (LM) can be applied in a straightforward manner, for example:

$$LRT = \ln\left(\frac{L(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\lambda}}, \boldsymbol{\delta}_1 = \boldsymbol{\delta}_2 = \mathbf{0})}{L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\delta}}_1, \hat{\boldsymbol{\delta}}_2)}\right) \sim^a \chi^2_{d_1+d_2},$$

where LRT denotes likelihood ratio statistic, which has asymptotically chi-square with degree of freedom $d_1 + d_2$ (the dimension of parameters δ_1 and δ_2, respectively) and:

$$WALD = \left(\begin{pmatrix} \hat{\boldsymbol{\delta}}_1 \\ \hat{\boldsymbol{\delta}}_2 \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}\right)^T \mathbf{I}^{-1}_{\delta\delta^T} \left(\begin{pmatrix} \hat{\boldsymbol{\delta}}_1 \\ \hat{\boldsymbol{\delta}}_2 \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}\right) \sim^a \chi^2_{d_1+d_2},$$

where according to the information matrix $\mathbf{I}(\mathbb{C}_c, \boldsymbol{\theta})$ we will have:

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{pmatrix} L_{\beta\beta^T} & L_{\beta\lambda^T} & L_{\beta\delta^T} \\ L_{\lambda\beta^T} & L_{\lambda\lambda^T} & L_{\lambda\delta^T} \\ L_{\delta\beta^T} & L_{\delta\lambda^T} & L_{\delta\delta^T} \end{pmatrix} \quad and \quad \mathbf{I}^{-1}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_{\beta\beta^T} & \mathbf{I}_{\beta\lambda^T} & \mathbf{I}_{\beta\delta^T} \\ \mathbf{I}_{\lambda\beta^T} & \mathbf{I}_{\lambda\lambda^T} & \mathbf{I}_{\lambda\delta^T} \\ \mathbf{I}_{\delta\beta^T} & \mathbf{I}_{\delta\lambda^T} & \mathbf{I}_{\delta\delta^T} \end{pmatrix}$$

with

$$\mathbf{I}_{\delta\delta^T} = \begin{pmatrix} \mathbf{I}_{\delta_1\delta_1^T} & \mathbf{I}_{\delta_1\delta_2^T} \\ \mathbf{I}_{\delta_2\delta_1^T} & \mathbf{I}_{\delta_2\delta_2^T} \end{pmatrix}.$$

In particular, the Lagrange Multiplier(LM) test does not require the estimation of the new model, which can be convenient in the same case.

7.3 About Optimal Design

As we know, in the inference statistics we face to unknown parameters, which be estimated (if those are identifiable). In this science subject we try to find the best estimator for parameters. One of criteria that we can distinguish which kind of estimator (every function of random sample) is better that the other is the variance of the estimator. In the optimal situation, this criterion is proportional to the inverse of the information matrix. Therefore, and basis of the Fisher information matrix some alphabet optimal criteria have been introduced to select the best elements of population in sample. One of these criteria which have been concentrated in this thesis is D-optimal criterion that is a function of the determinant of the information matrix. This criterion has some properties, which have been introduced in Chapter 3.

As stated previously were told selecting an alternative with the highest utility is the basic assumption to obtain the logit models. Then if we want to have the best selection, we must have the optimal combination of the levels in attributes for creating alternatives and the optimal combination of alternatives in choice sets. Thus, to define a design the choice sets are considered as support points. Based on this philosophy a MNL model which includes two attributes each with two levels have been considered (Chapter 3) and the design region for two-point and three-point design have been achieved.

We know that the MNL model is useful to analyze data when the property IIA be upheld. But, if there are some alternatives more similar than the others this property will not be hold. Thus, it must be considered the others models like Nested logit model to analyze data.

In Chapter 4 have been discussed about the NMNL models which consist M nests and to obtain optimal design we have considered Bayesian D-optimal criterion. But, since there is no closed form for this quantity, we have used local D-optimality criterion for optimal design. Opposite the MNL models, in this kind of logit models (NMNL) there are some extra parameters (dissimilarity parameters) in addition part-worth parameters. In this situation, we encounter more complex work than the MNL models to calculate optimal design.

To obtain optimal designs we could define designs with equal or not equal support points, but in the some of them some parameters were not identifiable, thus we had to combine them for creating a suitable new design with together. Of course, with respect to the total number of parameters which were much more, we had to fixed some of them and discussed about the remain of them.

According to similarity which exist between alternatives in the some nests, it may be used the other models of the logit family to analyze data that one of them in Chapter 5 have been considered.

As have been already told, the respondents choose just one alternative with the highest utility. But, in most times, the alternatives with the lower utility may be considered, also. In this situation, a model of the logit family have been applied, which is called Rank-Order (RO) logit models. In the last Chapter 6, we have first introduced the choice probabilities for this model and after defining optimal criterion the locally D-optimal design have been calculated.

A NOMENCLATURE

MNL	Multinomial Logit	9
NMNL	Nested Multinomial Logit	9
$\mathcal{C}$	Complete set of all choice sets	14
$\mathbb{C}_c$	a choice set with J_c alternatives	14
$\mathcal{J}$	number of all alternatives	14
$\boldsymbol{\beta}$	part-worth parameters vector	14
$\mathbf{I}_n$	n-dimensional identity matrix	15
ε	error term	15
Cov	covariance between two random variables	15
Σ	variance-covariance matrix	15
∂	partial derivative	16
$\Re$	set of real numbers	16
ϕ	empty set	23
ρ	Correlation coefficient	23
$i.i.d$	independent identically distribution	23
ξ	design	37
$\mathcal{X}$	design region	37
Ξ	set of designs	37
$\mathbf{M}(\xi)$	information matrix of design ξ	38

Ψ	optimality criterion	40
$\det(\mathbf{A})$	the determinate of matrix $\mathbf{A}$	40
$\lambda(x)$	efficiency function	41
$\sup(f(x))$	the spermium of function $f(x)$	41
$tr(\mathbf{A})$	the trace of matrix $\mathbf{A}$	41
$E(.)$	the expectation of a random variable	46
$\pi(.)$	prior distribution	46
$\mathcal{S}$	a set of choice sets with the same alternatives	49
$\inf(f(x))$	the infimium of function $f(x)$	70
$dim(\mathbf{A})$	dimension of matrix $\mathbf{A}$	71
$\mathbf{1}_r$	a r-dimension vector with elements one	75

Bibliography

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

Bartelts, K., Y. Boztug, and M. Muller (1999). Testing the multinomial logit model. *Sonderforschungbereich 373, Humboldt-University at Berlin, Germany (Preprint).*

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

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

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

Bernardo, J.M. (1979). Expected information as expected utility. *Ann. Statist., 7, 686-690.*

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

Box, G.E.P. and H.L. Lucas (1959). Design of experiments in non-linear situations. *Biometrika, 46, 77-90.*

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

Bradley, R.A. and M.E. Terry, (1952). Rank analysis of incomplete block design, I. The method of paired comparisons. *Biometrika 39, 324-345.*

Braess, D. and H. Dette (2004). On the number of suport point of maximin and Bayesian D-optimal designs in non-linear regression models. *Ruhr-Universität Bochum (Preprint).*

Brownstone, D. and K. Small (1989). Efficient estimation of nested logit model. *Journal of Business and Economic Statistics, 7, 67-74.*

Burgess, L. and D. Street (2003). Optimal design for 2^k choice experiments. *Communication in Statistics, Theory and Methods, 32, 2185-2206.*

Chaloner, K. (1993). A note on optimal Bayesian design for nonlinear problems. *Journal of Statistical Planning and Inference 37, 229-235.*

Chaloner, K. and K. Larntz (1989). Optimal Bayesian designs applied to logistic regresion experiments. *Journal of Statistical and Planning and Inference, 21, 191-208.*

Chaloner, K. and I. Verdinelli (1995). Bayesian experimental design: A review. *Statistical Science, 10, 273-304.*

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

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Statistics, 24, 586-602.*

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

Cosslett, S. (1981). Efficient estimation of discrete choice models, in C.Manski and D. McFadden, eds. *Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, MA, pp. 51-111.*

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

DeGroot, M.H. (1962). Uncertainty, information and sequential experiments. *Ann. Statist. 33, 404-419.*

DeGroot, M.H. (1986). Concepts of information based on utility. *In Recent Developments in the Foundations of Utility and Risk Theory. L. Daboni, A. Montesano, and M. Lines eds. 265-275. Reidel, Dordrecht, Holland.*

Dette, H. and H.M. Neugebauer (1996). Bayesian optimal one point designs for one parametric nonlinear regression models. *Journal of Statistcal Planning and Inference, 52, 17-31.*

Dette, H. and H.M. Neugebauer (1997). Bayesian optimal design for exponential regression models. *Journal of Statistical Planning and Inference, 60, 331-349.*

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

Ford, I., B. Torsney and C.F. Wu (1992). The use of a canonical form in the construction of locally optimal design for nonlinear problems. *Journal of the Royal Statistical Society. Series B, 54, 569-583.*

Forinash, C. and F. Koppelman (1993). Application and interpretation of nested logit models of intencity model choice. *Transportation Research Record 1413, 98-106.*

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

Goos, P. (2002). The optimal design of blocked and split-plot experiments. *New York: Springer.*

Greene, H. (1997). Econometric analysis. *3.d ed. Englewood Cliffs, NJ: Prentice Hall.*

Graßhoff U. and R. Schwabe (2008). Optimal design for the Bradley-Terry paired comparison model. *Statistical Methods and Application, 17, 275-289.*

Großmann, H., H. Holling and R. Schwabe (2002). Advances in optimum experimental design for conjoint analysis and discrete choice models. In *P. H. Franses and A. L. Montgomery (Hrsg.), Advances in Econometrics, 16. Econometric Models in Marketing. JAI Press, Amsterdam (2002) 93-117.*

Guadagni, P. M. and J.D.C. Little (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science, 2, 203-238.*

Haines, L. (1995). A geometric approach to optimal design for one parameter non linear models. *Journal of Royal Statistical Society Series B, 57, 575-598.*

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

Hausman, A. and D. Wise (1978). A conditional Probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica, 46, 403-426.*

Hensher, D. and G. Greene (2002). Specification and estimation of nested logit model. *Transportation Research B, 36, 1-17.*

Herriges, J.A. and C.L. Kling (1996). Testing the consistency of nested logt modles with utiliy maximizarion. *Economics Letters, 50, 33-39.*

Hoel, P.G. (1958). Efficiency problems in polynomial estimation. *Annals of Mathematical Statistics, 29, 1134-1145.*

Horowitz, J.L., D. Bolduc, S. Divakar, J. Geweke, F. Gonul, V. Hajivassiliou, F.S. Koppelman, M. Keane, R. Matzkin, P. Rossi, and P. Ruud(1994). Advances in random utility models. *Marketting Letters 5, 311-322.*

Huber, J. and K. Zwerina (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research, 33, 307-317.*

Kessels, R., B. Jones, P. Goos and M. Vandebroek (2006). An efficient algorithm for constructing Bayesian optimal choice designs. *Katholieke Universiteit LEUVEN, KBI 0616 (Preprint).*

Kessels, R., P. Goos and M. Vandebroek (2006). A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research 43, 409-419.*

Kiefer, J. (1961). Optimum designs in regression problems (II). *Annals of Mathematical Statistics, 32, 295-325.*

Kiefer, J. and J. Wolfowitz (1959). Optimum designs in regression problems (I). *Annals of Mathematical Statistics, 30, 271-294.*

Kiefer, J. and J. Wolfowitz (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics, 12, 363-366.*

Kitsos, C.P., D.M. Titterington and B. Torsney (1988). An optimal design problem in rhythometry. *Biometrics, 44, 647-671.*

Koning, R. and G. Ridder(2002). Discrete choice and stochastic utility maximization. *USC Center for Law, Economics & Organization Research Paper No. C02-16.*

Koppelman, F. and C. Wen (1998). Alternative nested logit models: Structure, properties and estimation. *Transportation Research B, 32, 289-298.*

Kullbach, S. and R.A. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics 22, 79-86.*

Lee, B. (1999). Calling patterns and usage of residential toll service under self-selecting tariffs. *Journal of Regulatory Economics 16, 45-82.*

Lindley, D.V. (1956). On the measure of information provided by an experiment. *Ann. Statist. 27, 986-1005.*

Louviere, J., D. Hensher and J. Swait (2000). Stated choice methods: Analysis and applications. *Combridge University Press, New York.*

Louviere, J.J. and G. Woodworth (1983). Design and analysis of simulated consumer choice or allocation experiments: An approach based on homogeneous data. *Journal of Marketing Research, 20, 350-367.*

Luce, R.D. (1959). Individual choice behavior. *Wiley: New York.*

Majumdar, D. (1988). Optimal block designs for comparing new treatments with a standard treatment. *In Optimal Design and Analysis of Experiments, Dodge, Y., V. Fedorov and H. Wynn, North-Holland, Amsterdam, 15-27.*

Majumdar, D. (1992). Optimal block designs for comparing test treatment with a control utilizing prior information. *Annals of Statistics, 20, 216-237.*

Manski, C. and D. McFadden (1981). Alternative estimatores and sample designs for discrete choice analysis, in C. Manski and D. McFadden, eds. *Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, MA, 305-319.*

Marschak, J. (1960). Binary choice constrains on random utility indicators,in *K. Arrow, ed., Stanford Symposium on Mathematical Methods in the Social Sciences, Standford University Press: Stanford, pp. 312-329.*

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

McFadden, D. (1978a). Quantitative methods for analyzing travel behaviour of individuals: Some recent developments, in: D. Hensher and P. Stopher, eds., Behavioral Travel Modelling. *Croom Helm, London, 279-318.*

McFadden, D. (1978b). Modeling the choice of residential location, in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (eds.). *Spatial Interaction Theory and Planning Models , 75-96, North Holland: Amsterdam, Reprinted in J. Quigley (ed.), THE Economics of Housing, Vol. I, 531-552, Edward Elgar: London, 1997.*

McFadden, D. (1981). Econometric models of probabilisticchoice, in C. Manski and D. McFadden (eds). *Structural Analysis of Discrete Data, MIT Press: Cambridge, 198-272.*

McFadden, D. (1984). Econometric analysis of qualitative response models, in Z. Griliches and M. Intriligator (eds). *Handbook of Econometrics, 2, Elsevier Science: Amsterdam, 1395-1457.*

McFadden, D. (1989). A method of simulation moments for estimation of discrete response models without numerical integration. *Econometrica, 57, 665-1026.*

McFadden, D. (2001). Observational studies: Choice-based sampling. *International Encyclopedia of Social and Behavioral Sciences, 2.1, Article 92, Elsevier Science: Amsterdam.*

McFadden, D. and J. Hausman (1984). Specification tests for the multinomial logit model. *Econometrica, Econometric Society, 52, 1219-1240.*

Melas, V.B. (2006). Functional approach to optimal experimental design. *New York; Springer.*

Mukhopadhyay, S. and L. Haines (1995). Bayesian D-optimal designs for the exponential growth model. *Journal of Statistical Planning and Inference, 44, 385-397.*

Pukelsheim, F.(1993). Optimal design of experiments. *Wiley, New York.*

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

Sandor, Z. and M. Wedel (2002). Profile construction in experimental choice designs for mixed logit models. *Marketing Science, 21, 55-75.*

Sandor, Z. and M. Wedel (2005). Heterogeneous conjoint choice designs. *Journal of Marketing Research, 42, 210-218.*

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

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Tech. J., 27, 376-423.*

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

Slutsky, E.E. (1952). On the theory of the budget of the consumer. *In: G.I. Stigler and K.E. Boulding, Editors, Readings in price theory, Irwin, Homewood, IL , 2756.*

Stone, M. (1959a). Application of a measure of information to the design and comparison of regression experiment. *Annals of Statistics, 30, 55-70.*

Tekle, F.B., E.S. Frans and P.F. Berger (2008). Maximin *D*-optimal designs for binary longitudinal responses. *Computational Statistics and Data Analysis, 52, 5253-5262.*

Thurston, L. (1927). A law of comparative judgment. *Psychological Review, 34, 273-286.*

Toman, B. (1994). Bayes optimal designs for two and three level fractorial experiment. *JASA, 89, 937-946.*

Train, K.(1986). Qualitative choice analysis. *MIT Press, Cambridge, MA.*

Train, K. (2003). Discrete choice methods with simulation. *Cambridge University Press.*

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

Vermeulen, B., P. Goos, M. Vandebroek (2007). Rank-Order conjoint experiments: Efficiency and design. *Katholieke Universiteit LEUVEN, KBI 0734 (Preprint).*

White, L.V. (1973). An extension of general equivalence theorem to nonlinear models. *Biometrika, 60, 345-348.*

White, L.V. (1975). The optimal design of experiments for estimation in nonlinear models. *PhD. Dissertation, University of London.*

Williams, H. (1977). On the formation of travel demand models and economic evaluation measures of user benefits. *Environment and Planning A, 9, 285-344.*

Wu, C.F. (1988). Optimal design for percentile estimation of a quantal response curve. In *Optimal Design and Anaylsis of Experiments, Y.Dodge, V.V.Fedorov and H.P. Wynn eds. North-Holland, Amsterdam, 213-223.*

Yu, J., P. Goos and M. Vandebroek. The importance of attribute interactions in conjoint choice design and modeling. *Katholieke Universiteit LEUVEN, KBI 0601 (Preprint).*

Yatchew, A. and Z. Griliches(1985). Specification error in Probit model. *Review of Economics and Statistics 67, 134-139.*

Zacks, S. (1977). Problems and approaches in design of experiments for estimation and testing in non-linear models. In *Multivariate Analysis, IV, 209-223, P.R. Krishnaiah ed., North-Holland, Amsterdam, 209-223.*

Zeng, L. (2000). A heteroscedastic generalized exterem value discrete choice model. *Sociological methods and research, 29, 118-144.*