OPTIMAL EXPERIMENTAL DESIGNS FOR HYPERPARAMETER ESTIMATION IN HIERARCHICAL LINEAR MODELS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Qing Liu, B.S., M.S.

* * * * *

The Ohio State University

2006

Dissertation Committee:

Angela M. Dean, Adviser

Greg M. Allenby

William I. Notz

Thomas J. Santner

Approved by

_____

Adviser

Graduate Program in Statistics

© Copyright by

Qing Liu

2006

ABSTRACT

Hierarchical models have received wide attention in a variety of fields; for example, social and behavioral sciences, agriculture, education, medical and healthcare research, and marketing. A general theory of optimal design for efficient estimation of hyperparameters in hierarchical linear models is developed. A statistical design optimality criterion is proposed for this setting, and optimal design structures are investigated under various forms of the variance-covariance matrix of the random effects. It is shown that, when the variance-covariance matrix is assumed to be diagonal, that is, when the random effects are assumed to be independent, level-balanced orthogonal designs are optimal if they exist. When the random effects are not independent, however, orthogonal designs cease to be optimal. The structure of the optimal continuous design in each scenario is derived.

Hierarchical models have been proved powerful in addressing a broad class of problems that involve the learning of effect sizes and the drivers of maximal or minimal effect sizes. The modeling of the "level effect" proposed in this dissertation is such an example. The "level effect" refers to the phenomenon in market research studies that consumer preference sensitivity to a product attribute is affected by the set of attribute levels displayed. Standard conjoint analysis models in marketing research do not adjust for the variety of attribute levels offered and, consequently, can not predict

consumer buying behavior well in a different context. A hierarchical Bayes approach is proposed that models the individual consumer behavior and, by incorporating a number of ideas from the Psychology literature, models the level effect through hyperparameters. To evaluate the effectiveness of the proposed model and the effect of different survey designs, web-based survey studies are conducted on credit card products. The optimal design theory developed for efficient estimation of hyperparameters is applied to select the alternative survey designs. The proposed model offers a good fit to the data and predicts consumer preference well in a different context. Furthermore, designs with higher efficiency lead to better estimation and prediction accuracy in general, confirming the optimal design theory.

To Feng, my parents and grandparents

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere appreciation and gratitude to my advisers, Angela Dean and Greg Allenby, for their continuous encouragement, guidance and support. Angela's class spurred my interest on experimental designs. Greg's paper on Bayesian approach in conjoint analysis inspired my research in marketing. I have had the dream to complete a Ph.D. since I was a little girl, and it was Angela and Greg's belief and support in me that have made it possible for me to return and pursue my academic dream after several years at the industry. For that, I cannot thank them enough.

I would also like to thank Professors Bill Notz and Tom Santner, who have kindly agreed to be on my committee to help me with their expertise on experimental designs. I am also indebted to many professors whose lectures have benefited me greatly. Professor Michael Browne's class on multivariate analysis provided me with the skills for the analytical derivations in this dissertation. His course notes are the ones that I constantly refer to during the completion of this dissertation. Professors Tom Santner and Doug Wolfe's classes on statistical inference, Professor Elizabeth Stasny's class on categorical data analysis, Professors Bill Notz and Mark Berliner's classes on regression models, Professor Steve Maceachern's class on decision theories, and Professor Joe Verducci's classes on point estimation and hypothesis testing laid the

foundation for me to conduct my dissertation research. Professors Doug Crithlow and Omer Ozturk introduced me to nonparametric statistical analysis, Professor Jason Hsu introduced me to research on multiple comparisons, and Professor Haikady Nagaraja introduced me to theories on large samples. In the marketing department, Professor Bob Leone's class offered me a broad review of research topics in marketing strategy, Professors Rao Unnava and Patricia West showed me extant research in consumer behavior, and Professor Thomas Otter helped me greatly in the understanding of different modeling approaches to marketing problems.

During the course of this work, I was supported by the Statistics Department teaching assistantship and in part by the grant from National Science Foundation SES-0437251.

I am also grateful to David Bakken at Harris Interactive, who offered his generous help in implementing the web-based survey to collect data for the empirical analysis section of my dissertation.

A special thanks goes to my dear parents and grandparents, who have loved me unconditionally and always believed in me. Also to my brother, with whom I have shared wonderful childhood memories, and to my little niece, who can always make me laugh. I am also indebted to my parents-in-law, as well as my aunts and uncles, for their understanding and support of my academic pursuit.

Finally, I would like to thank my husband Feng, who has loved and supported me all these years. He has always been there for me during difficult times, and encouraged me to hold on to my dreams. He has made my life so beautiful and for that he has my whole-hearted thanks and everlasting love.

VITA

October 25, 1972	Born – P.R. China.
1993	B.S. Science and English, University of Science and Technology of China.
1995	M.S. Statistics, The Ohio State University.
1993-1995	Graduate Fellow, Graduate Teaching Assistant, The Ohio State University.
1996-2001	Statistician, Statistician manager, Senior Statistician Manager, Capital One Financial Services, Inc.
2001-2003	Manager, RCI, Cendant Corporation.
2003-2004	Vice President, JPMorgan Chase.
2004-Present	Lecturer, Graduate Research Assistant, Graduate Teaching Assistant, The Ohio State University.

FIELDS OF STUDY

Major Field: Statistics

TABLE OF CONTENTS

Page

Abstract ii

Dedication iv

Acknowledgments v

Vita viii

List of Figures xi

List of Tables xii

Chapters:

1. Introduction 1

    1.1 Model Specification 3
    1.2 Experimental Designs 5

2. Optimal Designs for Estimation of $\boldsymbol{\theta}$ 9

    2.1 Design Criterion 9
    2.2 Optimal Designs When $\mathbf{Z}_i = \mathbf{I}$, $\mathbf{X}_i = \mathbf{X}$, and $\sigma_i^2 = \sigma^2$ 11
        2.2.1 The Main-Effects Model 21
        2.2.2 Scenario I: Random Effects Are Independently Distributed With Equal Within-factor Variances 26
        2.2.3 Scenario II: Random Effects Are Equi-correlated and with Equal Variances 27
        2.2.4 Design Examples for Scenario II 31
        2.2.5 Scenario III: Random Effects are Equi-correlated Within-groups but Independent Between-groups 40
    2.3 Optimal Designs When Z_i Can Be Controlled 41

3. Optimal Designs for Estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\Lambda}$ 45

3.1 Scenario I: Random Effects are Independent and homoscedastic . . 46
3.2 Scenario II: Random Effects are Equi-correlated and With Equal Variances 52

4. A Practical Application in Market Research 54

4.1 The "Level Effect" in Conjoint Studies 54
4.2 The Psychological Process 57
4.3 Modeling the Level Effect 61
4.3.1 Single-attribute Case 61
4.3.2 General Expression in Matrix Forms 62
4.4 Survey Designs 70
4.5 Model Fitting and Validation 77

5. Summary and Future Research 86

Appendices:

A. Computer Search Design Algorithm 89

A.1 Algorithm Used in Example 2.2 and 2.3 89
A.2 Algorithm Used for Survey Designs in Section 4.4 92

B. MCMC Algorithm For Model Fitting 95

LIST OF FIGURES

Figure **Page**

B.1 Time Series Plots of Posterior Draws 98

LIST OF TABLES

Table		Page
2.1	10-Run Designs Under the Main Effects Model of Example 2.2	34
2.2	8-Run Designs Under the Main Effects Model of Example 2.3	38
4.1	The 12-profile Credit Card Survey Studies in Example 4.1	65
4.2	The Credit Card Profiles for Survey Evaluations	66
4.3	Web-based 12-profile Credit Card Survey Studies	71
4.4	Designs Used in Study 2	75
4.5	Designs Used in Study 3	76
4.6	Posterior Estimates of $\boldsymbol{\theta}^*$ From Study 1 and Each Design in Study 2	78
4.7	Predictions of Mean Profile Ratings in Study 3 Version 1	82
4.8	Predictions of Mean Profile Ratings in Study 3 Version 2	83
4.9	Predictions of Mean Profile Ratings in Study 3 Version 3	84
4.10	Mean and Std. Dev. of Squared Error Loss of Mean Rating Predictions	85
4.11	Comparison of the Four 12-Run Designs in Study 2	85
4.12	Comparison of Six 8-Run Designs in Study 2	85

CHAPTER 1

Introduction

Hierarchical models have received wide attention and rapid development in a variety of fields during the past two decades. Various forms of hierarchical models have been used under the terminology "multi-level linear models", "mixed-effects models", "random-effects models", "population models", "random-coefficient regression models" and "covariance components models" (see Raudenbush and Bryk, 2002). Areas in which hierarchical models are applied include the social and behavioral sciences, agriculture, education, medicine, healthcare studies, and marketing. For example, in educational research, data usually contain repeated measurements over time for each individual within a number of different institutions, and hierarchical models have been used to analyze individual's learning curve over time, how the learning rate is affected by individual characteristics, and how the effects are influenced by institutional characteristics (see, for example, Huttenlocher et al., 1991; Raudenbush, 1993; Goldstein, 1995). In marketing research, hierarchical models are often the models of choice for marketing studies in which the learning of effect sizes and the determination of conditions for maximization or minimization of effect sizes are of importance; see, for example, Lenk and Rao (1990) on new product diffusion, Allenby and Lenk (1994,

1995) on brand choice, Manchanda et al. (1999)on multiple category purchase decisions, Bradlow and Rao (2000) on assortment choice, and Montgomery et al. (2004) on online shopping path. In pharmacokinetics, toxicokinetics and pharmacodynamics, hierarchical models, often called "mixed-effects models" or "population models", are used to describe the characteristics of a whole population while taking into consideration the heterogeneity among subjects (see Yuh et al., 1994, for a bibliography).

Among all the above references, two-level hierarchical models are the most commonly used. As we will illustrate in Section 1.1 through an example of a hierarchical linear model, parameters in the first-level of the hierarchy reflect individual-level effects, which are assumed to be random effects and distributed according to a probability distribution characterized by the "hyperparameters" in the second-level of the hierarchy. Examples of hyperparameters include parameters that characterize population characteristics, such as the effect of a new drug on a population, the vocabulary growth rate of children between age 1 and age 2, and the mean consumer sensitivity to a product feature change in a target consumer population; or parameters that reflect the effect of various drivers of the individual-level effects, such as the effect of patient age on individual patient's responsiveness to the new drug, the effect of the sex of the child and the exposure to language on the vocabulary growth of the child, and the effect of household income and other contextual factors on individual consumer sensitivity to the product feature change. In the modeling of the "level effect" in market research studies presented in Chapter 4, the hyperparameters reflect how the number of attribute levels affect consumer preference sensitivity to the attribute.

1.1 Model Specification

This paper focuses on hierarchical linear models. The following example is used to introduce the model. Consider a consumer survey in which respondent i is given a set of m_i questions ($i = 1, \ldots, n$). The questions contain information on various levels of marketing variables (treatment factors), such as price, product attributes or possibly aspects of advertisements. Suppose treatment factor k contains l_k levels ($k = 1, \ldots, t$). The design matrix $\mathbf{X}_i$ includes a column of ones that corresponds to the general mean and $l_k - 1$ dummy-coded columns for treatment factor k, ($k = 1, \ldots, t$) (see Draper and Smith, 1998, chapter 14, for the dummy coding of discrete variables). Let $p = 1 + \sum_{k=1}^{t}(l_k - 1)$, the design matrix $\mathbf{X}_i$ is of size $m_i \times p$. The responses of subject i to the set of questions are represented by the vector $\mathbf{y}_i$ of length m_i. The effects of the variables on respondent i are captured by the p elements in vector $\boldsymbol{\beta}_i$, which are assumed to be random effects that are distributed according to a multivariate normal distribution with $p \times p$ variance-covariance matrix $\boldsymbol{\Lambda}$, and mean $\mathbf{Z}_i\boldsymbol{\theta}$ where $\mathbf{Z}_i$ is a matrix ($p \times q$) of covariates, such as household income, age or other contextual variables that characterize the purchasing environment, and $\boldsymbol{\theta}$ is a parameter vector of length q. Thus, the hierarchical linear model is of the following form:

$$\mathbf{y}_i|\boldsymbol{\beta}_i, \sigma_i^2 = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i \tag{1.1}$$

$$\boldsymbol{\beta}_i|\boldsymbol{\theta}, \boldsymbol{\Lambda} = \mathbf{Z}_i\boldsymbol{\theta} + \boldsymbol{\delta}_i \tag{1.2}$$

The error vector $\boldsymbol{\epsilon}_i$ of length m_i in the first-level hierarchy captures consumer i's response variability to the set of questions, and it is generally assumed to have a Multivariate Normal distribution with mean vector $\mathbf{0}$ of size m_i and variance-covariance

matrix $\sigma_i^2\mathbf{I}_{m_i}$. The parameter vector $\boldsymbol{\beta}_i$ of size p captures the effects of the marketing variables on individual respondent i. The error vector $\boldsymbol{\delta}_i$ of length p in the second-level hierarchy captures the dispersion of individual-level effects $\boldsymbol{\beta}_i$ and is assumed to be Multivariate Normal with mean vector $\mathbf{0}$ of size p and variance-covariance matrix $\boldsymbol{\Lambda}$ of size $p \times p$. As the prior knowledge at the second stage is likely to be weak (see Chaloner and Verdinelli, 1995, page 283; also Lindley and Smith, 1972, page 7), proper, but diffuse, priors are usually assumed for $\boldsymbol{\theta}$, $\boldsymbol{\Lambda}$ and σ_i^2. For example, the prior distribution of $\boldsymbol{\theta}$ is assumed to be Multivariate Normal($\boldsymbol{\theta}_0 = \mathbf{0}_q$, $\mathbf{D}_0^{-1} = 100\mathbf{I}_q$), the prior distribution of $\boldsymbol{\Lambda}$ is Inverted Wishart($\nu_0 = p + 3$, $\mathbf{V}_0 = \nu_0\mathbf{I}_p$), and the prior distribution of σ_i^2 is Inverse Gamma($r_0/2, s_{0i}/2$) where $r_0 = 3$ and $s_{0i} = r_0 * \frac{\sum_{j=1}^{m_i}(y_{ij}-\bar{y}_{i\cdot})^2}{m_i-1}$. (See, for example, Gamerman, 1997; Rossi, Allenby and McCulloch, 2005).

Example 1.1: Suppose every consumer in the survey is asked to rate the three levels (9.99%, 15.99%, 17.99%) of credit card interest rate. The response vector $\mathbf{y}_i$ contains the three ratings $(y_{1i}, y_{2i}, y_{3i})'$ from consumer i on the three interest rate levels from lowest to highest. Using the dummy coding of discrete variables (see Draper and Smith, 1998, chapter 14), matrix $\mathbf{X}_i$ of size 3×3 is coded as follows: the first column is a vector with all elements being 1, the second column and the third column together represent the three levels of interest rate in such a way that the three combinations (0, 0), (1, 0), (0, 1) reflect the reference level 9.99%, the second level 15.99%, and the third level 17.99% respectively. Since every consumer in the survey is asked to rate on the same three interest rate levels, the model matrix $\mathbf{X}_i$ is the same for each consumer. Correspondingly, the vector $\boldsymbol{\beta}_i = (\mu_i, b_i, \gamma_i)'$ contains three parameters: the general mean μ_i which measures consumer i's preference on the

reference level 9.99% of interest rate, the parameter b_i which measures the contrast of consumer i's preference between the 15.99% interest rate and the reference level 9.99%, and the parameter γ_i which measures the contrast of consumer i's preference between the 17.99% interest rate and the reference level 9.99%. The first-level of the hierarchy (1.1) for this example is then of the form:

$$\begin{pmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_i \\ b_i \\ \gamma_i \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \epsilon_{3i} \end{pmatrix},$$

where $(\epsilon_{1i},\ \epsilon_{2i},\ \epsilon_{3i})' \sim N_3(\mathbf{0}, \sigma_i^2 \mathbf{I}_3)$. Assuming in this example that no covariates are measured, the matrix $\mathbf{Z}_i$ simplifies to the identity matrix of size 3 for every consumer i, and the vector $\boldsymbol{\theta}$ of size 3 then represents the mean of the $\boldsymbol{\beta}_i$'s across the population of consumers, and $\boldsymbol{\Lambda}$ (3×3) represents the variation of the individual-level effects $\boldsymbol{\beta}_i$ due to the difference in individual consumer preference. The second-level of the hierarchy (1.2) for this example is then of the form:

$$\begin{pmatrix} \mu_i \\ b_i \\ \gamma_i \end{pmatrix} = \begin{pmatrix} \mu \\ b \\ \gamma \end{pmatrix} + \begin{pmatrix} \delta_{\mu_i} \\ \delta_{b_i} \\ \delta_{\gamma_i} \end{pmatrix},$$

where $(\delta_{\mu_i},\ \delta_{b_i},\ \delta_{\gamma_i})' \sim N_3(\mathbf{0}, \boldsymbol{\Lambda})$. The hyperparameters in this model are the population parameters (μ, b, γ) and the elements in the variance-covariance matrix $\boldsymbol{\Lambda}$.

1.2 Experimental Designs

While it is important to have accurate information on individual-level effects $\boldsymbol{\beta}_i$ in situations such as direct marketing, which focuses on individual customization of products, in other situations, such as those in pharmacokinetics where population

parameters are of interest, or in situations where predictions of consumer preferences in a new target population are required, accurate estimation of hyperparameters $\boldsymbol{\theta}$, $\boldsymbol{\Lambda}$ is important because the hyperparameters capture the population characteristics, suggest conditions in which individual-level effects $\boldsymbol{\beta}_i$ are big, and enable predictions to new contexts with information on covariates Z_i. In this paper, experimental designs that provide efficient estimation of the hyperparameters are explored.

A vast amount of research on experimental designs has focused on the estimation of parameters in the traditional fixed-effects (non-hierarchical) linear models (see, Wu and Hamada, 2000, for an extensive review). Orthogonal designs are efficient in this setting (see Hedayat, Sloane and Stufken, 1999, for the construction and use of orthogonal arrays;). Bayesian designs that take into consideration the uncertainty of parameters through prior distributions have also been investigated (see Chaloner and Verdinelli, 1995, for a review). Relatively fewer papers, however, exist in the literature on experimental designs within the context of a hierarchical model.

In the set-up of hierarchical nonlinear models, which is outside the scope of this dissertation, such as choice models in marketing, Arora and Huber (2001), and Sándor and Wedel (2001, 2005) examined a heuristic approach to the construction of efficient factorial designs in the survey setting by relabeling, swapping and cycling the factor levels of an original orthogonal design, focusing on the estimation of individual-level parameters. On hyperparameter estimation, Sándor and Wedel (2002) applied the same heuristic approach to finding efficient factorial designs for the estimation of the mean and dispersion of individual-level random effects which are assumed to be

independently distributed. In the field of pharmacokinetics, toxicokinetics and pharmacodynamics, researchers have investigated optimal designs for the estimation of the same type of hyperparameters, also called the "population parameters" in the literature. For example, Mentré et al. (1997) proposed an algorithm to search for optimal sampling schedules for each group of subjects among a finite set of possible sampling times given a fixed cost constraint. Tod et al. (1998) extended the results of Mentré et al. (1997) by taking into consideration the uncertainty of the population parameters to be estimated. Han and Chaloner (2004) took a Bayesian-theoretic approach and explored efficient Bayesian designs where the prior distribution used at the design stage may differ from the prior distribution used at the analysis stage. They used MCMC to compare the candidate designs in order to find the optimal sampling schedules and number of subjects.

Within the context of a hierarchical linear model, as considered in this dissertation, Smith and Verdinelli (1980) investigated optimal Bayesian designs for the estimation of individual-level parameters under the one-way analysis of variance model in which there is one single predictor variable. For the estimation of hyperparameters, Lenk, Desarbo, Green and Young (1996) investigated, in the survey setting, the tradeoff between the number of subjects and the number of questions per subject under a cost constraint, given the design structure being orthogonal and the assumption that the random effects are independent and homoscedastic (i,e., variance-covariance matrix $\mathbf{\Lambda} = \lambda_0^2 \mathbf{I}_p$). In this paper, efficient factorial designs under some general forms of the variance-covariance matrix $\mathbf{\Lambda}$ of the random effects are investigated, and the optimal design structure under each form of $\mathbf{\Lambda}$ is analytically derived. Optimal experimental

designs investigated in this dissertation are under the constraint of a fixed number of observations per subject, as seen in the survey studies where survey questions are designed given a fixed length of the questionnaire.

In Chapter 2, optimal designs for the estimation of hyperparameter $\boldsymbol{\theta}$ in a hierarchical linear model are studied while the variance-covariance parameters in $\boldsymbol{\Lambda}$ are considered as nuisance parameters. Optimal factorial design structures are analytically derived under some general forms of the variance-covariance matrix $\boldsymbol{\Lambda}$ for two specific design settings. In the first setting, every respondent receives the same design (i.e., $\mathbf{X}_i = \mathbf{X}$, i=1, ..., n), and the covariate matrix Z_i is the identity matrix for each respondent so that the hyperparameters in vector $\boldsymbol{\theta}$ capture the characteristics of the population. In the second setting, $\mathbf{Z}_i$ contains information on covariates such as demographic variables or contextual variables, and optimal design structures are considered for both the design matrix $\mathbf{X}$ and the covariate matrix $\mathbf{Z}_i$. In Chapter 3, optimal experimental designs are investigated for the joint estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\Lambda}$ in the situation when the random effects are independent and homoscedastic (i.e., $\boldsymbol{\Lambda} = \lambda_0^2 \mathbf{I}_p$). A practical application is illustrated in Chapter 4 where the "level effect" in market research studies is examined. By incorporating ideas in psychology, a hierarchical linear model is proposed that models the "level effect" through hyperparameters. Survey studies are conducted to test the model and the effect of different designs. The optimal design theory developed in Chapter 2 is used as guidance to select the various survey designs. Data from the surveys offer support for the model and the optimal design theory. Summary and future research are given in Chapter 5.

CHAPTER 2

Optimal Designs for Estimation of $\boldsymbol{\theta}$

2.1 Design Criterion

Consider the hierarchical linear model introduced in (1.1) and (1.2) of Section 1.1,

$$\mathbf{y}_i|\boldsymbol{\beta}_i, \sigma_i^2 = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i,$$

$$\boldsymbol{\beta}_i|\boldsymbol{\theta}, \boldsymbol{\Lambda} = \mathbf{Z}_i\boldsymbol{\theta} + \boldsymbol{\delta}_i,$$

$$\text{where } \boldsymbol{\epsilon}_i \sim N_{m_i}(\mathbf{0}, \sigma_i^2\mathbf{I}), \text{ and } \boldsymbol{\delta}_i \sim N_p(\mathbf{0}, \boldsymbol{\Lambda}).$$

When the interest is in the estimation of hyperparameters $\boldsymbol{\theta}$ and $\boldsymbol{\Lambda}$, the above equations can be combined (see Lenk et al., 1996, pg 187) to obtain

$$\mathbf{y}_i|\boldsymbol{\theta}, \boldsymbol{\Lambda}, \sigma_i^2 = \mathbf{X}_i\mathbf{Z}_i\boldsymbol{\theta} + \mathbf{X}_i\boldsymbol{\delta}_i + \boldsymbol{\epsilon}_i,$$

$$\text{where } \mathbf{X}_i\boldsymbol{\delta}_i + \boldsymbol{\epsilon}_i \sim N_{m_i}(\mathbf{0}, \mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i' + \sigma_i^2\mathbf{I}_{m_i}).$$

Which is equivalently,

$$\mathbf{y}_i|\boldsymbol{\theta}, \boldsymbol{\Lambda}, \sigma^2 \sim N_{m_i}(\mathbf{X}_i\mathbf{Z}_i\boldsymbol{\theta},\ \boldsymbol{\Sigma}_i = \sigma_i^2\mathbf{I}_{m_i} + \mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i') \tag{2.1}$$

As aforementioned in Section 1.1, proper but diffuse priors are usually assumed for $\boldsymbol{\theta}$, $\boldsymbol{\Lambda}$, and σ_i^2. For example,

$$\begin{aligned}
\boldsymbol{\theta} &\sim N_q(\boldsymbol{\theta}_0 = \mathbf{0}, \mathbf{D}_0^{-1} = 100\mathbf{I}), \\
\boldsymbol{\Lambda} &\sim \text{ IW } (\nu_0 = p + 3, \mathbf{V}_0 = \nu_0 \mathbf{I}_p), \\
\sigma_i^2 &\sim \text{ IG } (r_0/2, s_{0i}/2)
\end{aligned}$$

where $r_0 = 3$ and $s_{0i} = r_0 * \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{i.})^2/(m_i - 1)$. (see, for example, Gamerman, 1997; Rossi, Allenby and McCulloch, 2005)

Let us first look at the simple situation where the estimation of $\boldsymbol{\theta}$ in (2.1) is of primary interest. For example, a medical researcher is interested in finding out the effect of a new drug in a target population of patients, an education specialist is interested in the effect of parents' involvement on children's vocabulary growth, or a marketer is interested to learn how household income affects consumer price sensitivity, while not interested in the dispersion of the effects at the level of individual patient, child or consumer. In this situation, the focus is on the accurate estimation of $\boldsymbol{\theta}$ while elements in $\boldsymbol{\Lambda}$ and the response error variance parameters σ_i^2 are treated as nuisance parameters. To find which designs are more efficient for the estimation of $\boldsymbol{\theta}$, we can first assume $\boldsymbol{\Lambda}$ and σ_i^2 are given and seek the designs that minimize the determinant of the posterior variance of $\boldsymbol{\theta}$ given $\boldsymbol{\Lambda}$ and σ_i^2, which corresponds to maximizing the expected gain in Shannon information (cf. Chaloner and Verdinelli, 1995, page 277).

It is shown, for example, by Rossi et al. (2005, pg. 66) that under model (2.1), the posterior distribution of $\boldsymbol{\theta}$ given $\boldsymbol{\Lambda}$ and $\{\sigma_i^2\}$ is normal with mean vector

$$(\sum_{i=1}^{n} \mathbf{Z}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\mathbf{Z}_i + \mathbf{D}_0)^{-1}(\sum_{i=1}^{n} \mathbf{Z}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{y}_i + \mathbf{D}_0\boldsymbol{\theta}_0)$$

and variance-covariance matrix

$$(\sum_{i=1}^{n} \mathbf{Z}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\mathbf{Z}_i + \mathbf{D}_0)^{-1}, \text{ where } \boldsymbol{\Sigma}_i = \sigma_i^2\mathbf{I}_{m_i} + \mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i'.$$

Therefore, we seek to minimize

$$\left|(\sum_{i=1}^{n} \mathbf{Z}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\mathbf{Z}_i + \mathbf{D}_0)^{-1}\right|$$

or, equivalently, to maximize

$$\Big|\sum_{i=1}^{n}(\mathbf{Z}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\mathbf{Z}_i) + \mathbf{D}_0\Big|. \tag{2.2}$$

When the diffuse prior distribution $\mathbf{D}_0 \approx \mathbf{0}$ is assumed for $\boldsymbol{\theta}$, or when the sample size n becomes large and the influence of the prior information becomes negligible, (2.2) can be maximized by maximizing

$$\Big|\sum_{i=1}^{n}(\mathbf{Z}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\mathbf{Z}_i)\Big|, \quad \text{where} \quad \boldsymbol{\Sigma}_i = \sigma_i^2\mathbf{I}_{m_i} + \mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i', \tag{2.3}$$

which is the same expression as that derived from the non-Bayesian framework through the Fisher Information approach as used by Lenk et al. (1996), Mentré et al. (1997) and Tod et al. (1998).

2.2 Optimal Designs When $\mathbf{Z}_i = \mathbf{I}$, $\mathbf{X}_i = \mathbf{X}$, and $\sigma_i^2 = \sigma^2$

Let us begin with the simplest design setting that involves the following assumptions:

(i) Every subject receives the same design, i.e., $\mathbf{X}_i = \mathbf{X},\ m_i = m$.

(ii) Response errors are homoscedastic, i.e., $\sigma_i^2 = \sigma^2$.

(iii) $\mathbf{Z}_i = \mathbf{I}_p$ so that the hyperparameter vector $\boldsymbol{\theta}$ captures the population characteristics.

This design setting is seen in "population designs" (see Entholzner et al., 2006) that examine population characteristics in fields such as medical research, social sciences and agriculture, and in market research survey designs as shown in Lenk et al (1996). With the above assumptions, the maximization of (2.3) becomes the maximization of

$$\mathcal{Q}_1 = |\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}| \quad \text{where} \quad \boldsymbol{\Sigma} = \sigma^2\mathbf{I}_m + \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}' \tag{2.4}$$

In the following lemma, the criterion $\mathcal{Q}_1$ in (2.4) is written in an alternative form.

Lemma 1. *The maximization of $\mathcal{Q}_1$ is equivalent to*

(i) the maximization of

$$\mathcal{Q}_2 = |\mathbf{X}'\mathbf{X}(\sigma^2\mathbf{I}_p + \boldsymbol{\Lambda}\mathbf{X}'\mathbf{X})^{-1}| = \frac{|\mathbf{X}'\mathbf{X}|}{|\sigma^2\mathbf{I}_p + \boldsymbol{\Lambda}\mathbf{X}'\mathbf{X}|}, \tag{2.5}$$

(ii) the minimization of

$$\left|(\mathbf{X}'\mathbf{X})^{-1} + (\boldsymbol{\Lambda}/\sigma^2)\mathbf{I}_p\right|, \text{ for nonsingular } \mathbf{X}'\mathbf{X},$$

(iii) the maximization of

$$\psi = \log\frac{|\mathbf{M}(\eta)|}{|\frac{\sigma^2}{m}\mathbf{I}_p + \boldsymbol{\Lambda}\mathbf{M}(\eta)|}, \quad \text{where} \quad \mathbf{M}(\eta) = \frac{1}{m}\mathbf{X}'\mathbf{X}. \tag{2.6}$$

For $\mathbf{X}'\mathbf{X}$ singular, ψ is defined to be $-\infty$.

Proof. First, from Morrison (1990) page 69, Equation 8, for nonsingular matrices $\mathbf{A}$ and $\mathbf{C}$ of sizes $(m \times m)$ and $(p \times p)$ respectively, and rectangular $(m \times p)$ matrix $\mathbf{B}$, the following holds:

$$(\mathbf{A} + \mathbf{BCB}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}^{-1}. \tag{2.7}$$

Now, if we postmultiply both sides of Equation (2.7) by $\mathbf{B}$, we get

$$\begin{aligned}
(\mathbf{A} + \mathbf{BCB}')^{-1}\mathbf{B} &= \mathbf{A}^{-1}\mathbf{B} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}^{-1}\mathbf{B} \\
&= \mathbf{A}^{-1}\mathbf{B}[\mathbf{I} - (\mathbf{C}^{-1} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}] \\
&= \mathbf{A}^{-1}\mathbf{B}[\mathbf{I} - (\mathbf{C}^{-1} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}(\mathbf{C}^{-1} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B} - \mathbf{C}^{-1})] \\
&= \mathbf{A}^{-1}\mathbf{B}[\mathbf{I} - \mathbf{I} + (\mathbf{C}^{-1} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}^{-1}] \\
&= \mathbf{A}^{-1}\mathbf{B}[(\mathbf{C}^{-1} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}^{-1}] \\
&= \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{CB}'\mathbf{A}^{-1}\mathbf{B})^{-1}.
\end{aligned}$$

Applying this result to $\mathbf{\Sigma}^{-1}\mathbf{X}$ in the optimality criterion $\mathcal{Q}_1$ in (2.4), we let $\mathbf{A} = \sigma^2\mathbf{I}_m,\ \mathbf{B} = \mathbf{X}$ and $\mathbf{C} = \mathbf{\Lambda}$, then

$$(\sigma^2\mathbf{I}_m + \mathbf{X\Lambda X}')^{-1}\mathbf{X} = \sigma^{-2}\mathbf{X}(\mathbf{I}_p + \sigma^{-2}\mathbf{\Lambda X}'\mathbf{X})^{-1} = \mathbf{X}(\sigma^2\mathbf{I}_p + \mathbf{\Lambda X}'\mathbf{X})^{-1}. \tag{2.8}$$

Therefore, $\mathcal{Q}_1$ becomes

$$\left|\mathbf{X}'(\sigma^2\mathbf{I}_m + \mathbf{X\Lambda X}')^{-1}\mathbf{X}\right| = |\mathbf{X}'\mathbf{X}(\sigma^2\mathbf{I}_p + \mathbf{\Lambda X}'\mathbf{X})^{-1}| = \frac{|\mathbf{X}'\mathbf{X}|}{|\sigma^2\mathbf{I}_p + \mathbf{\Lambda X}'\mathbf{X}|},$$

since for two square matrices $\mathbf{A}$ and $\mathbf{C}$, $|\mathbf{A} * \mathbf{C}| = |\mathbf{A}| * |\mathbf{C}|$ (Graybill, 1983, Theorem 1.5.8).

Note that for $\boldsymbol{\theta}$ to be estimable, $\mathbf{X}'\mathbf{X}$ must be non-singular, and the maximization of $\mathcal{Q}_2$ in (2.5) over the set of designs such that $\mathbf{X}'\mathbf{X}$ is non-singular is equivalent to

the maximization of

$$\frac{1}{|\sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{\Lambda}\mathbf{I}_p|},$$

and equivalent to the minimization of

$$\left|(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{\Lambda}/\sigma^2)\mathbf{I}_p\right|,$$

which is the same expression as the "mixed-effects model D-criterion" examined in Entholzner et al. (2006).

Now if we substitute $\mathbf{X}'\mathbf{X}$ by $m\mathbf{M}(\eta)$ in $\mathcal{Q}_2$ and take the natural log function of $\mathcal{Q}_2$, we get

$$\log \frac{|m\mathbf{M}(\eta)|}{|\sigma^2\mathbf{I}_p + m\mathbf{\Lambda}\mathbf{M}(\eta)|}$$

Since the natural log function is an increasing function, the maximization of $\mathcal{Q}_2$ in (2.5) over the set of designs such that $\mathbf{X}'\mathbf{X}$ is non-singular is equivalent to the maximization of

$$\psi = \log \frac{|\mathbf{M}(\eta)|}{|\frac{\sigma^2}{m}\mathbf{I}_p + \mathbf{\Lambda}\mathbf{M}(\eta)|}$$

□

As described by Silvey (1980), a typical m-observation design contains a number of (say, h) distinct design points $\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(h)}$ replicated respectively $r_1, \ldots, r_h$ times, with $\sum r_i = m$ number of observations in $\mathbf{y}$. We can associate this design with a probability distribution η_m which assigns probability $p_i = r_i/m$ at $\mathbf{x}_{(i)}, i = 1, \ldots, h$. Now let $\tilde{\mathbf{x}}$ be a random vector of size p with distribution η_m and define the $p \times p$ symmetric matrix $\mathbf{M}(\eta_m) = E(\tilde{\mathbf{x}}\tilde{\mathbf{x}}') = \sum p_i \mathbf{x}_{(i)}\mathbf{x}'_{(i)} = m^{-1}\mathbf{X}'\mathbf{X} = m^{-1}\mathbf{M}(\mathbf{X})$. For a concave and increasing design criterion function ϕ, the problem of finding an optimal

m-observation design can be interpreted as that of finding η_m^*, a probability distribution that maximizes $\phi\{\mathbf{M}(\eta_m)\}$. Furthermore, we can extend the definition of $\mathbf{M}(\eta)$ to the set H of all probability distributions η on the Borel sets of $\mathcal{X}$, a compact subset of Euclidean p-space that contains all possible design points (vectors x). Let $\eta \in H$ and define

$$\mathbf{M}(\eta) = E(\tilde{x}\tilde{x}'), \quad \mathcal{M} = \{\mathbf{M}(\eta) : \eta \in H\}. \tag{2.9}$$

Now the problem of finding optimal designs under the criterion ϕ becomes one of finding the measure η^* that maximizes $\phi\{\mathbf{M}(\eta)\}$ over the set H. Note that as remarked by Silvey (1980) on page 15 and 16, the set $\mathcal{M}$ is a closed convex hull of $\{\mathbf{x}\mathbf{x}' : \mathbf{x} \in \mathcal{X}\}$, and each $p \times p$ non-negative definite symmetric matrix in $\mathcal{M}$ can be considered as a vector of size $p(p+1)/2$ and represented by a point in the Euclidean space $\mathcal{R}^{p(p+1)/2}$.

Use the same definitions described by Silvey (1980), let us define the design space $\mathcal{X}$ as a compact subset of Euclidean p-space, i.e., $\mathcal{X} \subseteq \mathcal{R}^p$ and define $\mathcal{M}$ as in (2.9). We now consider the criterion function ψ in (2.6) and show that ψ is concave and increasing in $\mathcal{M}$.

Lemma 2. *The criterion function in (2.6)*

$$\psi = \begin{cases} \log \frac{|\mathbf{M}(\eta)|}{|\frac{\sigma^2}{m}\mathbf{I}_p + \mathbf{\Lambda}\mathbf{M}(\eta)|}, & \text{if } \mathbf{M}(\eta) \text{ is nonsingular} \\ -\infty & \text{if } \mathbf{M}(\eta) \text{ is singular} \end{cases}$$

is concave and strictly increasing in $\mathcal{M}$ where $\mathcal{M}$ is defined in (2.9).

Proof. When $\mathbf{M}(\eta)$ is nonsingular, pre-multiply both the numerator and the denominator of $\frac{|\mathbf{M}(\eta)|}{|\frac{\sigma^2}{m}\mathbf{I}_p + \mathbf{\Lambda}\mathbf{M}(\eta)|}$ with $|\mathbf{M}(\eta)^{-1}|$, the criterion function ψ becomes equivalently

$$\psi = \log \frac{1}{|\frac{\sigma^2}{m}\mathbf{M}(\eta)^{-1} + \mathbf{\Lambda}|} = -\log |\frac{\sigma^2}{m}\mathbf{M}(\eta)^{-1} + \mathbf{\Lambda}|$$

Let us first show that ψ is strictly increasing in $\mathcal{M}$. According to the Minkowski's determinant theorem (Theorem 28, Magnus and Neudecker, 1999, page 227),

$$|\mathbf{A}+\mathbf{B}|^{1/n} \geq |\mathbf{A}|^{1/n} + |\mathbf{B}|^{1/n},$$

if $\mathbf{A}$ and $\mathbf{B}$ are non-negative definite and symmetric (as are all information matrices). The equality holds if and only if $|\mathbf{A}+\mathbf{B}| = 0$ or $\mathbf{A} = \mu\mathbf{B}$ for some $\mu > 0$. If we set $n = 1$, and $\mathbf{B} = \mathbf{C} - \mathbf{A}$, then

$$|\mathbf{C}| \geq |\mathbf{A}| + |\mathbf{C}-\mathbf{A}|.$$

Since $|\mathbf{C}-\mathbf{A}| = |\mathbf{B}| \geq 0$, we have $|\mathbf{C}| \geq |\mathbf{A}|$. If $\mathbf{B}$ is positive definite, then $|\mathbf{C}| > |\mathbf{A}|$. This proves that $|\mathbf{M}(\eta)|$ is strictly increasing in $\mathcal{M}$, i.e., if $\mathbf{M}_1 - \mathbf{M}_2$ is positive definite, where $\mathbf{M}_1$ denotes $\mathbf{M}(\eta_1)$ and $\mathbf{M}_2$ denotes $\mathbf{M}(\eta_2)$, then $|\mathbf{M}_1| > |\mathbf{M}_2|$. Now according to Theorem 12.2.14 (2) in Graybill (1983), if $\mathbf{M}_1$ and $\mathbf{M}_2$ are nonsingular and $\mathbf{M}_1 - \mathbf{M}_2$ is positive definite, then $\mathbf{M}_2^{-1} - \mathbf{M}_1^{-1}$ is positive definite. Therefore $(\frac{\sigma^2}{m}\mathbf{M}_2^{-1} + \mathbf{\Lambda}) - (\frac{\sigma^2}{m}\mathbf{M}_1^{-1} + \mathbf{\Lambda})$ is also positive definite for positive σ^2 and m. Use the result that $|\mathbf{M}(\eta)|$ is strictly increasing, we get

$$\left|\tfrac{\sigma^2}{m}\mathbf{M}_2^{-1} + \mathbf{\Lambda}\right| > \left|\tfrac{\sigma^2}{m}\mathbf{M}_1^{-1} + \mathbf{\Lambda}\right|. \tag{2.10}$$

Also, since the variance-covariance matrix $\mathbf{\Lambda} = [\frac{\sigma^2}{m}\mathbf{M}(\eta)^{-1} + \mathbf{\Lambda}] - \frac{\sigma^2}{m}\mathbf{M}(\eta)^{-1}$ is positive definite, we have

$$\left|\tfrac{\sigma^2}{m}\mathbf{M}_1^{-1} + \mathbf{\Lambda}\right| > \left|\tfrac{\sigma^2}{m}\mathbf{M}_1^{-1}\right| = \frac{(\sigma^2/m)^p}{|\mathbf{M}_1|} > 0,$$
$$\left|\tfrac{\sigma^2}{m}\mathbf{M}_2^{-1} + \mathbf{\Lambda}\right| > \left|\tfrac{\sigma^2}{m}\mathbf{M}_2^{-1}\right| = \frac{(\sigma^2/m)^p}{|\mathbf{M}_2|} > 0.$$

Therefore, we can take log function on both sides of (2.10) and since the log function is strictly increasing, we obtain

$$\log\left|\tfrac{\sigma^2}{m}\mathbf{M}_2^{-1} + \mathbf{\Lambda}\right| > \log\left|\tfrac{\sigma^2}{m}\mathbf{M}_1^{-1} + \mathbf{\Lambda}\right|,$$

which is equivalent to

$$-\log\left|\tfrac{\sigma^2}{m}\mathbf{M}_2^{-1}+\mathbf{\Lambda}\right| < -\log\left|\tfrac{\sigma^2}{m}\mathbf{M}_1^{-1}+\mathbf{\Lambda}\right|,$$

$$\text{i.e. } \psi(\mathbf{M}_1) > \psi(\mathbf{M}_2) \text{ for positive definite } (\mathbf{M}_1-\mathbf{M}_2).$$

Therefore $\psi(\mathbf{M}(\eta))$ is strictly increasing.

To prove the concavity of ψ, we use the result in Silvey (1980) Appendix 1, which shows that $\mathbf{M}(\eta)^{-1}$ is convex. The convexity of $\mathbf{M}(\eta)^{-1}$ suggests

$$[\lambda\mathbf{M}_1+(1-\lambda)\mathbf{M}_2]^{-1} \le \lambda\mathbf{M}_1^{-1}+(1-\lambda)\mathbf{M}_2^{-1},$$

for $0 \le \lambda \le 1$, $\mathbf{M}_1, \mathbf{M}_2 \in \mathcal{M}$ and $\mathbf{M}_1, \mathbf{M}_2$ nonsingular. Multiply both sides of the inequality with the positive scalar σ^2/m and add $\mathbf{\Lambda}$ to both sides of the inequality, we obtain,

$$\begin{aligned}\tfrac{\sigma^2}{m}[\lambda\mathbf{M}_1+(1-\lambda)\mathbf{M}_2]^{-1}+\mathbf{\Lambda} &\le \lambda\tfrac{\sigma^2}{m}\mathbf{M}_1^{-1}+(1-\lambda)\tfrac{\sigma^2}{m}\mathbf{M}_2^{-1}+\mathbf{\Lambda}\\ &= \lambda(\tfrac{\sigma^2}{m}\mathbf{M}_1^{-1}+\mathbf{\Lambda})+(1-\lambda)(\tfrac{\sigma^2}{m}\mathbf{M}_2^{-1}+\mathbf{\Lambda}),\end{aligned}$$

which shows that $\frac{\sigma^2}{m}\mathbf{M}(\eta)^{-1}+\mathbf{\Lambda}$ is convex, and by A.2 in Silvey (1980), Appendix 1, $(\frac{\sigma^2}{m}\mathbf{M}(\eta)^{-1}+\mathbf{\Lambda})^{-1}$ is concave. According to Theorem 25 on page 222 in Magnus and Neudecker (1999), the function "$\log|\mathbf{A}|$" on the set of non-negative definite matrices $\mathbf{A}$ is concave. Also, "$\log|\mathbf{A}|$" is increasing since $|\mathbf{A}|$ is increasing and the "log" function is increasing. Therefore, by Theorem 15 on page 76 in Magnus and Neudecker (1999), taking an increasing concave function "$\log|\mathbf{A}|$" of a concave function $\mathbf{A} = (\frac{\sigma^2}{m}\mathbf{M}(\eta)^{-1}+\mathbf{\Lambda})^{-1}$ results in $\psi = -\log|\mathbf{M}(\eta)^{-1}+\mathbf{\Lambda}|$ being concave. □

To identify optimal designs under the concave and strictly increasing criterion function ψ in (2.6), two theorems in Silvey (1980) are used, as descried below.

Theorem 3.6 (Silvey 1980): *When ϕ is concave on $\mathcal{M}$, η^* is ϕ-optimal if and only if $F_\phi\{\mathbf{M}(\eta^*), \mathbf{M}(\eta)\} \leq 0$ for all $\eta \in H$.*

The function $F_\phi\{\mathbf{M}(\eta^*), \mathbf{M}(\eta)\}$ is the Fréchet derivative of ϕ at $\mathbf{M}(\eta^*)$ in the direction of $\mathbf{M}(\eta)$, defined as:

$$F_\phi\{\mathbf{M}(\eta^*), \mathbf{M}(\eta)\} = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \left\{\phi[(1-\epsilon)\mathbf{M}(\eta^*) + \epsilon \mathbf{M}(\eta)] - \phi(\mathbf{M}(\eta^*))\right\}.$$

The above theorem says that $\phi(M(\eta^*))$ is the maximum of a concave function, and there is no direction in which we can look upward to a higher value of the function. Considering that there are many possible directions to look in, however, the above theorem is of much less practical use than *Theorem 3.7* in Silvey (1980) which basically says that if the concave function is smooth at the summit then it is not necessary to look around in every direction. Instead we only need to look toward the extreme points of $\mathcal{M}$, the closed convex hull of $\{\mathbf{x}\mathbf{x}' : \mathbf{x} \in \mathcal{X}\}$. The theorem is quoted below

Theorem 3.7 (Silvey 1980): *If ϕ is concave on $\mathcal{M}$ and differentiable at $\mathbf{M}(\eta_*)$, then η_* is ϕ-optimal if and only if $F_\phi\{\mathbf{M}(\eta_*), \mathbf{x}\mathbf{x}'\} \leq 0$ for all $\mathbf{x} \in \mathcal{X}$.*

To check if ϕ is Fréchet differentiable at $\mathbf{M}_1$ when ϕ is concave and $\phi(\mathbf{M}_1)$ is finite, the sufficient and necessary condition stated in Silvey (1980), Section 4.1 on page 75, can be used. That is, one can check whether the Gâteaux derivative is linear in its second argument, i.e.,

$$G_\phi(\mathbf{M}_1, \sum a_i \mathbf{M}_2) = \sum a_i G_\phi(\mathbf{M}_1, \mathbf{M}_2), \tag{2.11}$$

where a_i is a real number, and

$$G_\phi\{\mathbf{M}_1, \mathbf{M}_2\} = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon}[\phi\{\mathbf{M}_1 + \epsilon\mathbf{M}_2\} - \phi(\mathbf{M}_1)].$$

When ϕ is concave and $\phi(\mathbf{M}_1)$ is finite, if (2.11) holds, then ϕ is Fréchet differentiable at $\mathbf{M}_1$.

Based on the above theorems from Silvey (1980), let us now derive the sufficient and necessary condition for the optimal design η^* that maximizes the criterion function ψ.

Theorem 1. *Let η be a design measure in the class of probability distributions H on the Borel sets of a compact design space $\mathcal{X} \subseteq \mathcal{R}^p$. Under the criterion of maximizing ψ defined in (2.6), a design η^* is optimal if and only if*

$$\mathbf{x}'\mathbf{M}(\eta^*)^{-1}[\tfrac{\sigma^2}{m}\mathbf{I}_p + \mathbf{M}(\eta^*)\mathbf{\Lambda}]^{-1}\mathbf{x} \leq Tr[(\tfrac{\sigma^2}{m}\mathbf{I}_p + \mathbf{M}(\eta^*)\mathbf{\Lambda})^{-1}] \quad \textit{for all} \quad \mathbf{x} \in \mathcal{X}. \qquad (2.12)$$

Proof. The Gâteaux derivative of criterion function ψ defined in (2.6) is:

$$\begin{aligned}
G_\psi\{\mathbf{M}_1, \mathbf{M}_2\} &= \lim_{\{\epsilon \to 0^+\}} \frac{1}{\epsilon} \log \frac{|\frac{\sigma^2}{m}\mathbf{I} + \Lambda\mathbf{M}_1||\mathbf{M}_1 + \epsilon\mathbf{M}_2|}{|\mathbf{M}_1||\frac{\sigma^2}{m}\mathbf{I} + \Lambda\mathbf{M}_1 + \epsilon\Lambda\mathbf{M}_2|} \\
&= \lim_{\{\epsilon \to 0^+\}} \frac{1}{\epsilon} \log \frac{|\mathbf{M}_1 + \epsilon\mathbf{M}_2|}{|\mathbf{M}_1|} - \lim_{\{\epsilon \to 0^+\}} \frac{1}{\epsilon} \log \frac{|\frac{\sigma^2}{m}\mathbf{I} + \Lambda\mathbf{M}_1 + \epsilon\Lambda\mathbf{M}_2|}{|\frac{\sigma^2}{m}\mathbf{I} + \Lambda\mathbf{M}_1|} \\
&= \lim_{\{\epsilon \to 0^+\}} \frac{1}{\epsilon}[1 + \epsilon Tr(\mathbf{M}_2\mathbf{M}_1^{-1}) + O(\epsilon^2)] \\
&\quad - \lim_{\{\epsilon \to 0^+\}} \frac{1}{\epsilon}[1 + \epsilon Tr(\mathbf{\Lambda}\mathbf{M}_2(\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{\Lambda}\mathbf{M}_1)^{-1}) + O(\epsilon^2)]
\end{aligned}$$

since $\frac{1}{\epsilon}O(\epsilon^2) \to 0$ as $\epsilon \to 0$,

$$\begin{aligned}
G_\psi\{\mathbf{M}_1, \mathbf{M}_2\} &= \lim_{\{\epsilon \to 0^+\}} \frac{1}{\epsilon}[\epsilon Tr(\mathbf{M}_2\mathbf{M}_1^{-1}) - \epsilon Tr(\mathbf{\Lambda}\mathbf{M}_2(\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{\Lambda}\mathbf{M}_1)^{-1}) + O(\epsilon^2)] \\
&= Tr(\mathbf{M}_2\mathbf{M}_1^{-1}) - Tr(\mathbf{M}_2(\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{\Lambda}\mathbf{M}_1)^{-1}\mathbf{\Lambda}) \\
&= Tr[\mathbf{M}_2\mathbf{M}_1^{-1}(\mathbf{I} - (\tfrac{\sigma^2}{m}\mathbf{M}_1^{-1} + \mathbf{\Lambda})^{-1}\mathbf{\Lambda})] \\
&= Tr[\mathbf{M}_2\mathbf{M}_1^{-1}(\mathbf{I} - (\tfrac{\sigma^2}{m}\mathbf{\Lambda}^{-1}\mathbf{M}_1^{-1} + \mathbf{I})^{-1})] \\
&= Tr[\mathbf{M}_2\mathbf{M}_1^{-1}(\tfrac{\sigma^2}{m}\mathbf{\Lambda}^{-1}\mathbf{M}_1^{-1} + \mathbf{I})^{-1}(\tfrac{\sigma^2}{m}\mathbf{\Lambda}^{-1}\mathbf{M}_1^{-1})] \\
&= Tr[\mathbf{M}_2\mathbf{M}_1^{-1}(\tfrac{\sigma^2}{m}\mathbf{M}_1^{-1} + \mathbf{\Lambda})^{-1}(\tfrac{\sigma^2}{m}\mathbf{M}_1^{-1})] \\
&= \tfrac{\sigma^2}{m}Tr[\mathbf{M}_2\mathbf{M}_1^{-1}(\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}_1\mathbf{\Lambda})^{-1}]
\end{aligned}$$

The Gâteaux derivative is linear in $\mathbf{M}_2$, and since only η for which $\mathbf{M}(\eta)$ is non-singular can be optimal in this case, ψ is Fréchet differentiable at $\mathbf{M}_1$. Therefore, we can apply Theorem 3.7 in Silvey (1980) to get the sufficient and necessary condition for the optimal design $\mathbf{M}(\eta^*)$:

$$\begin{aligned}
F_\psi\{\mathbf{M}(\eta^*), \mathbf{x}\mathbf{x}'\} &= G_\psi\{\mathbf{M}(\eta^*), \mathbf{x}\mathbf{x}' - \mathbf{M}(\eta^*)\} \\
&= \tfrac{\sigma^2}{m}\Big\{Tr[\mathbf{x}\mathbf{x}'\mathbf{M}_1^{-1}(\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}(\eta^*)\mathbf{\Lambda})^{-1}] - Tr[(\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}(\eta^*)\mathbf{\Lambda})^{-1}]\Big\}
\end{aligned}$$

The sufficient and necessary condition $F_\psi\{\mathbf{M}(\eta^*), \mathbf{x}\mathbf{x}'\} \leq 0, \quad \text{for all} \quad \mathbf{x} \in \mathcal{X}$, is therefore equivalent to

$$\mathbf{x}'\mathbf{M}(\eta^*)^{-1}[\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}(\eta^*)\mathbf{\Lambda}]^{-1}\mathbf{x} \leq Tr[(\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}(\eta^*)\mathbf{\Lambda})^{-1}], \quad \text{for all} \quad \mathbf{x} \in \mathcal{X}.$$

□

Theorem 1 agrees with the equivalence theorem for the random-coefficients regression model stated in Fedorov and Hackl (1997, page 75). It applies to a general compact design space $\mathcal{X} \subseteq \mathcal{R}^p$.

2.2.1 The Main-Effects Model

Let us now look at the situation where the predictor variables in model matrix $\mathbf{X}$ are discrete such as the interest rate with three levels in Example 1.1. We focus on the main-effects model, which assumes that no interactions between treatment factors are present in the parameter vector $\boldsymbol{\beta}_i$ in (1.1). Suppose there are t treatment factors where each treatment factor k $(k = 1, \ldots, t)$ contains l_k levels, then the parameter vector $\boldsymbol{\beta}_i$ for each respondent i in (1.1) includes only the mean parameter and the $l_k - 1$ contrasts for each factor k. The size of $\boldsymbol{\beta}_i$ is $1 + \sum_{k=1}^{t}(l_k - 1) = p$. With the assumptions (i) to (iii) in Section 2.2, the hyperparameter vector $\boldsymbol{\theta}$ of size p captures the mean of the individual-level effects $\boldsymbol{\beta}_i$. The model illustrated in Example 1.1 is an example of such a main-effects model.

To find optimal designs for the estimation of the hyperparameter vector $\boldsymbol{\theta}$ under the main-effects model, let us first define the design space $\mathcal{X}$. Suppose the contrasts within each treatment factor in the parameter vector $\boldsymbol{\beta}_i$ are orthogonal contrasts scaled in such a way that the corresponding model matrix $\mathbf{X}$ is coded as described by Kuhfeld in SAS TS-722C, 2005, page 64, under the name of the "standardized orthogonal effects coding". In such a coding scheme, the first column of the model matrix $\mathbf{X}$ is a column of ones corresponding to the mean parameter, then there are $(l_k - 1)$ columns corresponding to each treatment factor k, $(k = 1, \ldots, t)$. It is shown next in Lemma 3 that these $(l_k - 1)$ columns satisfy the following condition: the sum of squares of the elements in every row is equal to $(l_k - 1)$, $k = 1, \ldots, t$.

Lemma 3. *Suppose there are t treatment factors where each treatment factor k $(k = 1, \ldots, t)$ contains l_k levels. Let x_{k_s} denote the element in the model matrix*

X *corresponding to the* s*th* ($s = 1, \ldots, l_k - 1$) *of the* $l_k - 1$ *contrasts of treatment factor* k ($k = 1, \ldots, t$)*, then under the standardized orthogonal effects coding,*

$$x_{k_1}^2 + x_{k_2}^2 + \ldots + x_{k_{(l_k-1)}}^2 = l_k - 1.$$

Proof. Without loss of generality, let us examine the first treatment factor with l_1 levels. The coefficients of the unscaled $l_1 - 1$ orthogonal contrasts of the factor levels are:

$$\begin{pmatrix} l_1 - 1 & 0 & \cdots & 0 & 0 \\ -1 & l_1 - 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & -1 & \cdots & 2 & 0 \\ -1 & -1 & \cdots & -1 & 1 \\ -1 & -1 & \cdots & -1 & -1 \end{pmatrix}, \tag{2.13}$$

where each of the l_1 rows corresponds to each of the l_1 levels of the factor. Note that for the sth column of the $(l_1 - 1)$ columns, there are $(l_1 - s)$ number of -1s, one $(l_1 - s)$, and $(s - 1)$ zeros, $(s = 1, \ldots, l_1 - 1)$. In the standardized orthogonal effects coding, the elements in each column are scaled so that the column sum of squares equals l_1 for every column. For example, in the first column, every element is multiplied by $\sqrt{\frac{l_1}{l_1-1+(l_1-1)^2}}$, such that the sum of squares of the elements in the first column equals $(l_1-1)^2 * \frac{l_1}{l_1-1+(l_1-1)^2} + (l_1-1) * (-1)^2 * \frac{l_1}{l_1-1+(l_1-1)^2} = l_1$. Similarly, for the sth column, every element is multiplied by $\sqrt{\frac{l_1}{l_1-s+(l_1-s)^2}}$, $(s = 1, \ldots, l_k - 1)$.

Now to check the row sum of squares, note that the rth row in the first (l_1-1) rows (i.e., $1 \leq r \leq l_1-1$) contains $(r-1)$ number of -1s, one (l_1-r), and (l_1-1-r) number of zeros, where the sth element of the row is scaled by $\sqrt{\frac{l_1}{l_1-s+(l_1-s)^2}}$, $(s = 1, \ldots, l_k-1)$. The l_1th row, i.e., the last row, however, contains $(l_1 - 1)$ number of -1s, scaled by $\sqrt{\frac{l_1}{l_1-s+(l_1-s)^2}}$ for the sth element, $(s = 1, \ldots, l_k - 1)$. The sum of squares of the

elements in the rth row $(1 \leq r \leq l_1 - 1)$ is therefore

$$\begin{aligned}
&\sum_{s=1}^{r-1}\left(\frac{-1\sqrt{l_1}}{\sqrt{l_1-s+(l_1-s)^2}}\right)^2+\left(\frac{(l_1-r)\sqrt{l_1}}{\sqrt{l_1-(r-1)+(l_1-(r-1))^2}}\right)^2\\
=&l_1\left(\sum_{s=1}^{r-1}\frac{1}{l_1-s+(l_1-s)^2}+\frac{(l_1-r)^2}{l_1-r+(l_1-r)^2}\right)\\
=&l_1\left(\sum_{s=1}^{r-1}\frac{1}{(l_1-s)(l_1-s+1)}+\frac{(l_1-r)^2}{(l_1-r)(l_1-r+1)}\right)\\
=&l_1\left(\sum_{s=1}^{r-1}(\frac{1}{l_1-s}-\frac{1}{l_1-s+1})+\frac{l_1-r}{l_1-r+1}\right)\\
=&l_1\left(-\frac{1}{l_1}+\frac{1}{l_1-r+1}+\frac{(l_1-r)}{(l_1-r+1}\right)\\
=&l_1\left(-\frac{1}{l_1}+1\right)=l_1-1.
\end{aligned}$$

The sum of squares of the elements in the l_1th row is:

$$\begin{aligned}
\sum_{s=1}^{l_1-1}\left(\frac{-1\sqrt{l_1}}{\sqrt{l_1-s+(l_1-s)^2}}\right)^2&=l_1\left(\sum_{s=1}^{l_1-1}\frac{1}{l_1-s+(l_1-s)^2}\right)\\
&=l_1\left(\sum_{s=1}^{l_1-1}\frac{1}{(l_1-s)(l_1-s+1)}\right)\\
&=l_1\left(\sum_{s=1}^{l_1-1}(\frac{1}{l_1-s}-\frac{1}{l_1-s+1})\right)\\
&=l_1\left(-\frac{1}{l_1}+1\right)=l_1-1.
\end{aligned}$$

Therefore, the row sum of squares equals (l_1-1) for every row of the elements in the (l_1-1) columns of $\mathbf{X}$ corresponding to the (l_1-1) standardized orthogonal contrasts of the first factor. The same can be generalized to any of the treatment factor k, $(k = 1, \ldots, t)$. □

Let $\mathbf{x}$ denote a row vector in model matrix $\mathbf{X}$ under the standard orthogonal effects coding, we can then use such points $\mathbf{x}$ as the border points to form the compact continuous design space $\mathcal{X}$, a subset of $\mathcal{R}^p$ in which the first coordinate of $\mathbf{x} \in \mathcal{X}$

is constrained to be 1 (corresponding to the mean parameter) and, corresponding to each treatment factor k, $x_{k_1}^2 + x_{k_2}^2 + \ldots + x_{k_{(l_k-1)}}^2 \leq l_k - 1$. That is, the induced compact continuous design space $\mathcal{X}$ is

$$\mathcal{X} = \Big\{ \mathbf{x}' = [1, \ldots, x_{k_1}, \ldots, x_{k_{(l_k-1)}}, \ldots, x_{t_1}, \ldots, x_{t_{(l_t-1)}}] \Big\}$$
$$\text{s.t.} \quad \sum_{s=1}^{l_k-1} x_{k_s}^2 \leq l_k - 1, k = 1, \ldots, t. \quad (2.14)$$

Note that for every design point $\mathbf{x}$ in $\mathcal{X}$,

$$\mathbf{x}'\mathbf{x} = 1 + \sum_{k=1}^{t} \sum_{s=1}^{l_k-1} x_{k_s}^2 \leq 1 + \sum_{k=1}^{t} (l_k - 1) = p, \ \forall \ \mathbf{x} \in \mathcal{X}. \quad (2.15)$$

A special example is when all factors are binary, the induced continuous design space $\mathcal{X}$ then is

$$\mathcal{X} = \Big\{ \mathbf{x}' = [1, -1 \leq x_1 \leq 1, -1 \leq x_2 \leq 1, \ldots, -1 \leq x_{p-1} \leq 1] \Big\},$$

which is a compact subspace of $\mathcal{R}^p$ that includes design points used for the estimation of the intercept and the main effects of the $(p-1)$ predictor variables in the multiple regression model.

Remark: The main effects may be captured by contrasts other than the standardized orthogonal contrasts, such as the contrasts of the treatment levels relative to a reference level illustrated in Example 1.1 When such contrasts are used, the corresponding model matrix $\mathbf{X}$ uses the dummy coding as seen in Example 1.1 The relationship of $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Lambda}}, \tilde{\mathbf{X}})$ under one coding scheme and $(\boldsymbol{\theta}, \boldsymbol{\Lambda}, \mathbf{X})$ under another coding scheme is as follows:

$$\mathbf{X} = \tilde{\mathbf{X}}\mathbf{T}, \ \boldsymbol{\theta} = \mathbf{T}^{-1}\tilde{\boldsymbol{\theta}}, \ \ \boldsymbol{\Lambda} = \mathbf{T}^{-1}\tilde{\boldsymbol{\Lambda}}\mathbf{T}^{-1\,\prime},$$

where $\mathbf{T}$ is the $p \times p$ non-singular transformation matrix. The above relationship simply reflects the change of variables from one coding scheme to another coding scheme, as seen through

$$\tilde{\mathbf{X}}\tilde{\boldsymbol{\theta}} = \tilde{\mathbf{X}}\mathbf{T}\mathbf{T}^{-1}\tilde{\boldsymbol{\theta}} = \mathbf{X}\boldsymbol{\theta},$$

$$\tilde{\mathbf{X}}\tilde{\boldsymbol{\Lambda}}\tilde{\mathbf{X}}' = \tilde{\mathbf{X}}\mathbf{T}\mathbf{T}^{-1}\tilde{\boldsymbol{\Lambda}}\mathbf{T}^{-1\,\prime}\mathbf{T}'\tilde{\mathbf{X}}' = \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}'.$$

Note that since the original form of the ψ criterion $\mathcal{Q}_1$ in (2.4) is

$$\begin{aligned}
\mathcal{Q}_1 &= |\mathbf{X}'(\sigma^2\mathbf{I}_m + \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}')^{-1}\mathbf{X}| \\
&= |\mathbf{T}'\tilde{\mathbf{X}}'(\sigma^2\mathbf{I}_m + \tilde{\mathbf{X}}\mathbf{T}\mathbf{T}^{-1}\tilde{\boldsymbol{\Lambda}}\mathbf{T}^{-1\prime}\mathbf{T}'\tilde{\mathbf{X}}')^{-1}\tilde{\mathbf{X}}\mathbf{T}| \\
&= |\mathbf{T}'\mathbf{T}||\tilde{\mathbf{X}}'(\sigma^2\mathbf{I}_m + \tilde{\mathbf{X}}\tilde{\boldsymbol{\Lambda}}\tilde{\mathbf{X}}')^{-1}\tilde{\mathbf{X}}| \quad \text{as } \mathbf{T} \text{ is a square matrix,}
\end{aligned}$$

and that the transformation matrix $\mathbf{T}$ is the same for every design, it shows that the same design that is optimal for estimating $\tilde{\boldsymbol{\theta}}$ given $\tilde{\boldsymbol{\Lambda}}$ is also optimal for estimating $\boldsymbol{\theta}$ given $\boldsymbol{\Lambda}$.

Example 2.1: The coding used in the model matrix $\mathbf{X}$ in Example 1.1 is the dummy-coding such that the three levels of interest rate are represented by $(0,\ 0)$, $(1,\ 0)$, $(0,\ 1)$ in the second and third columns of the model matrix. Alternatively, the standardized orthogonal effects coding (Kuhfeld, 2005) can be used, in which the three levels of interest rate are represented by $(-1.22, -0.71)$, $(0,\ 1.41)$, $(1.22, -0.71)$ in the second and third columns of the model matrix $\tilde{\mathbf{X}}$, while the first column remains unchanged. Note that $(-1.22)^2 + (-0.71)^2 = 0^2 + 1.41^2 = 1.22^2 + (-0.71)^2 = 2$. The transformation matrix $\mathbf{T}$ can be obtained by solving the system of equations $\mathbf{X} = \tilde{\mathbf{X}}\mathbf{T}$. The hyperparameters $\tilde{\boldsymbol{\theta}}$ to be estimated in the standardized orthogonal effects coding are the transformed contrasts $\tilde{\boldsymbol{\theta}} = \mathbf{T}\boldsymbol{\theta}$, i.e., $(\tilde{\mu}, \tilde{b}, \tilde{\gamma})' = (\mu + 0.33b +$

$0.33\gamma,\ 0.41\gamma,\ 0.47b - 0.24\gamma)'$. If an experimental design is optimal for the estimation of $(\tilde{\mu}, \tilde{b}, \tilde{\gamma})'$ under the standardized orthogonal effects coding of the model matrix, then the same design is also optimal for the estimation of $(\mu, b, \gamma)'$ under the dummy coding of the model matrix.

In the next three subsections let us examine three scenarios for $\mathbf{\Lambda}$, and derive the optimal design η^* under each scenario for the estimation of hyperparameter vector $\boldsymbol{\theta}$ in a main-effects hierarchical linear model.

2.2.2 Scenario I: Random Effects Are Independently Distributed With Equal Within-factor Variances

Theorem 2. *Let η be a design measure in the class of probability distributions H on the Borel sets of $\mathcal{X}$ where $\mathcal{X}$ is a compact subspace of $\mathcal{R}^p$ defined in (2.14). When the random individual-level main effects $\boldsymbol{\beta}_i$ in (1.1) are independently distributed with equal within-factor variances such that $\mathbf{\Lambda}$ is a diagonal matrix*

$$\mathbf{\Lambda} = Diag(\lambda_0^2,\ \lambda_1^2\mathbf{I}_{(l_1-1)},\ \ \ldots,\ \lambda_k^2\mathbf{I}_{(l_k-1)},\ \ \ldots,\ \lambda_t^2\mathbf{I}_{(l_t-1)}), \tag{2.16}$$

where the k^{th} factor has l_k levels, for the estimation of $\boldsymbol{\theta}$ in (1.2) given such a form of $\mathbf{\Lambda}$, any design that satisfies $\mathbf{M}(\eta) = \mathbf{I}$ is optimal under the ψ criterion defined in (2.6).

Proof. To prove that $\mathbf{M}(\eta) = \mathbf{I}$ is optimal in this scenario, we need to show that if $\mathbf{M}(\eta) = \mathbf{I}$ satisfies the sufficient and necessary condition of Theorem 1.

The left hand side of Equation (2.12) is now

$$
\begin{aligned}
&(\tfrac{\sigma^2}{m}+\lambda_0^2)^{-1}+\sum_{k=1}^{t}\big[(\tfrac{\sigma^2}{m}+\lambda_k^2)^{-1}\sum_{j=1}^{l_k-1}x_{k_j}^2\big] \\
\leq&(\tfrac{\sigma^2}{m}+\lambda_0^2)^{-1}+\sum_{k=1}^{t}\big[(\tfrac{\sigma^2}{m}+\lambda_k^2)^{-1}(l_k-1)\big] \quad \text{by (2.14)} \\
=&Tr[(\tfrac{\sigma^2}{m}I+\boldsymbol{\Lambda})^{-1}], \quad \text{the right hand side.}
\end{aligned}
$$

Therefore, the condition is satisfied and this proves the theorem. □

Remark: Note that for an m-observation design, the design measure η may not yield integer number of observations on the design points to form an exact design. However, for certain combination of m, the number of treatment factors and factor levels, in the standardized orthogonal coding, level-balanced orthogonal designs $\mathbf{X}$ satisfies $\mathbf{X}'\mathbf{X} = m\mathbf{I}$ for the main-effects model, and equivalently $\mathbf{M}(\eta) = \mathbf{I}$. Therefore, level-balanced orthogonal designs are optimal, if they exist, for the estimation of hyperparameters $\boldsymbol{\theta}$ in the above scenario.

2.2.3 Scenario II: Random Effects Are Equi-correlated and with Equal Variances

In this scenario, the random individual-level main effects $\boldsymbol{\beta}_i$ in (1.1) are assumed to be equally correlated and with equal variances. Let a^* and d^* be scalars, $\mathbf{J}_p$ be the $p \times p$ matrix with all elements equal to 1, and let $\boldsymbol{\Lambda} = a^*\mathbf{I}_p + d^*\mathbf{J}_p$. We first show that a design η satisfying $\mathbf{M}(\eta) = \mathbf{I}_p$ is not optimal for the estimation of $\boldsymbol{\theta}$ in (1.2) given the form of $\boldsymbol{\Lambda}$ in this scenario. For display clarity in subsequent derivations, the subscript p is omitted in matrices $\mathbf{I}_p$ and $\mathbf{J}_p$.

With $\mathbf{\Lambda} = a^*\mathbf{I} + d^*\mathbf{J}$ and $\mathbf{M}(\eta) = \mathbf{I}$, the left hand side of the sufficient and necessary condition (2.12) for optimal designs now becomes

$$\frac{m}{\sigma^2}\mathbf{x}'[(a+1)\mathbf{I} + d\mathbf{J}]^{-1}\mathbf{x},$$

and the right hand side is

$$\frac{m}{\sigma^2}Tr[(a+1)\mathbf{I} + d\mathbf{J}]^{-1},$$

where $a = ma^*/\sigma^2, \quad d = md^*/\sigma^2$. From Theorem 8.3.4 in Graybill (1983),

$$[(a+1)\mathbf{I} + d\mathbf{J}]^{-1} = \frac{1}{a+1}\left(\mathbf{I} - \frac{d}{a+1+pd}\mathbf{J}\right). \tag{2.17}$$

The right hand side is now equal to

$$\frac{m}{\sigma^2}Tr\left[\tfrac{1}{a+1}\left(\mathbf{I} - \tfrac{d}{a+1+pd}\mathbf{J}\right)\right] = \frac{m}{\sigma^2}\frac{p}{a+1}\left(1 - \frac{d}{a+1+pd}\right).$$

On the left hand side, $\dfrac{\mathbf{x}'\mathbf{V}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \leq e_{max}(\mathbf{V})$ holds for any vector $\mathbf{x}$, where $e_{max}(\mathbf{V})$ represents the maximum eigenvalue of a non-negative $p \times p$ matrix $\mathbf{V}$ (see Graybill, 1983, Theorem 12.2.14). According to Theorem 8.5.3 in Graybill (1983), one eigenvalue of $[(a+1)\mathbf{I} + d\mathbf{J}]^{-1}$ is $\frac{1}{a+1}(1 - \frac{pd}{a+1+pd})$ and the other $(p-1)$ eigenvalues are equal to $\frac{1}{a+1}$. From (2.15), $\mathbf{x}'\mathbf{x} \leq p$. Since $d \neq 0$ in this scenario, the value of $\frac{d}{a+1+pd}$ can only be greater than 0 or less than 0. Let us now examine each of these two possibilities. If $\frac{d}{a+1+pd} > 0$, then $\frac{pd}{a+1+pd} > 0$. Thus the maximal eigenvalue of $[(a+1)\mathbf{I} + d\mathbf{J}]^{-1}$ is $\frac{1}{a+1}$ since $1 - \frac{pd}{a+1+pd} < 1$, and the maximum of the left hand side is

$$\begin{aligned} Max_{\{\mathbf{x}\in\mathcal{X}\}}\frac{m}{\sigma^2}\mathbf{x}'[(a+1)\mathbf{I} + d\mathbf{J}]^{-1}\mathbf{x} &= \left(\frac{m}{\sigma^2}\right) * e_{max}[(a+1)\mathbf{I} + d\mathbf{J}] * Max(\mathbf{x}'\mathbf{x}) \\ &= \frac{m}{\sigma^2}\frac{p}{a+1} > \frac{m}{\sigma^2}\frac{p}{a+1}\left(1 - \frac{d}{a+1+pd}\right), \\ &\quad \text{since } \ \tfrac{d}{a+1+pd} > 0. \end{aligned}$$

If $\frac{d}{a+1+pd} < 0$, then $\frac{pd}{a+1+pd} < 0$, and $\frac{pd}{a+1+pd} < \frac{d}{a+1+pd}$ since $p > 1$. Thus the maximal eigenvalue of $[(a+1)\mathbf{I} + d\mathbf{J}]^{-1}$ is $\frac{1}{a+1}(1 - \frac{pd}{a+1+pd})$ since $1 - \frac{pd}{a+1+pd} > 1$, and the maximum of the left hand side is

$$\begin{aligned} Max_{\{\mathbf{x}\in\mathcal{X}\}} \frac{m}{\sigma^2}\mathbf{x}'[(a+1)\mathbf{I} + d\mathbf{J}]^{-1}\mathbf{x} &= \left(\frac{m}{\sigma^2}\right) * e_{max}[(a+1)\mathbf{I} + d\mathbf{J}] * Max(\mathbf{x}'\mathbf{x}) \\ &= \frac{m}{\sigma^2}\frac{p}{a+1}(1 - \frac{pd}{a+1+pd}) \\ &> \frac{m}{\sigma^2}\frac{p}{a+1}\left(1 - \frac{d}{a+1+pd}\right), \\ &\text{since } \tfrac{pd}{a+1+pd} < \tfrac{d}{a+1+pd}. \end{aligned}$$

This shows that level-balanced orthogonal designs do not satisfy the necessary and sufficient condition and therefore are not optimal in this scenario.

Theorem 3. *Let η be a design measure in the class of probability distributions H on the Borel sets of $\mathcal{X}$ where $\mathcal{X}$ is a compact subspace of $\mathcal{R}^p$ defined in (2.14). When $\mathbf{\Lambda} = a^*I + d^*J$, for the estimation of $\boldsymbol{\theta}$ in (1.2) given such a form of $\mathbf{\Lambda}$, designs of structure $\mathbf{M}(\eta) = (1+\epsilon)\mathbf{I} - \epsilon\mathbf{J}$ are optimal under the ψ criterion in (2.6), where*

$$\epsilon = \frac{2(a+pd) + 1 - 2d - \sqrt{4(a+1)(a+pd) - 4d + 1}}{2(a+pd)(p-2) + 2d}, \tag{2.18}$$

$a = ma^*/\sigma^2$, and $d = md^*/\sigma^2$.

Proof. To prove that such a design is optimal we need to check that it satisfies the sufficient and necessary condition in Equation (2.12), which in this scenario, can be rewritten as:

$$\mathbf{x}'\mathbf{M}(\eta^*)^{-1}[\mathbf{I} + \mathbf{M}(\eta^*)(a\mathbf{I} + d\mathbf{J})]^{-1}\mathbf{x} \leq Tr[\mathbf{I} + \mathbf{M}(\eta^*)(a\mathbf{I} + d\mathbf{J})]^{-1}, \forall \mathbf{x} \in \mathcal{X}, \tag{2.19}$$

where $\mathbf{\Lambda} = a^*\mathbf{I} + d^*\mathbf{J}$, and $a = ma^*/\sigma^2, \quad d = md^*/\sigma^2$.

Now with $\mathbf{M}(\eta^*) = (1+\epsilon)\mathbf{I} - \epsilon\mathbf{J}$, use (2.17) to get the inverse of $\mathbf{M}(\eta^*)$ such that

$$\mathbf{M}(\eta^*)^{-1} = \frac{1}{1+\epsilon}\left(\mathbf{I} + \frac{\epsilon}{1-(p-1)\epsilon}\mathbf{J}\right).$$

In addition, we have

$$\mathbf{I} + \mathbf{M}(\eta^*)(a\mathbf{I} + d\mathbf{J}) = \big(1 + a(1+\epsilon)\big)\mathbf{I} + \big(d(1+\epsilon) - dp\epsilon - a\epsilon\big)\mathbf{J},$$

$$\big[\mathbf{I} + \mathbf{M}(\eta^*)(a\mathbf{I} + d\mathbf{J})\big]^{-1} = \frac{1}{1+a(1+\epsilon)} \left(\mathbf{I} - \frac{d(1+\epsilon) - dp\epsilon - a\epsilon}{1 + a(1+\epsilon) + pd(1+\epsilon) - p^2 d\epsilon - pa\epsilon}\mathbf{J}\right),$$

$$\mathbf{M}(\eta^*)^{-1}\big[\mathbf{I} + \mathbf{M}(\eta^*)(a\mathbf{I} + d\mathbf{J})\big]^{-1} = \frac{1}{(1+\epsilon)[1+a(1+\epsilon)]} \left[\mathbf{I} - \frac{\epsilon^2[(a+pd)(p-2)+d] - \epsilon[2(a+pd)+1-2d]+d}{[1-(p-1)\epsilon][1+(a+pd)(1-(p-1)\epsilon)]}\mathbf{J}\right].$$

Take the expression of ϵ in equation (2.18) into the above equation, the last item in the last equation becomes $\mathbf{0}$ and we have

$$\mathbf{M}(\eta^*)^{-1}[\mathbf{I} + \mathbf{M}(\eta^*)(a\mathbf{I} + d\mathbf{J})]^{-1} = \frac{1}{(1+\epsilon)[1+a(1+\epsilon)]}\mathbf{I}.$$

On the right hand side of (2.19), we have

$$Tr[\mathbf{I} + \mathbf{M}(\eta^*)(a\mathbf{I} + d\mathbf{J})]^{-1} = \frac{p}{(1+\epsilon)[1+a(1+\epsilon)]} \left(1 - \frac{\epsilon^2[(a+pd)(p-2)+d] - \epsilon[2(a+pd)+1-2d]+d}{1+(a+pd)[1-(p-1)\epsilon]}\right).$$

Again, take the expression of ϵ in equation (2.18) into the above equation, the last item in the last equation becomes 0 and we have

$$Tr[\mathbf{I} + \mathbf{M}(\eta^*)(a\mathbf{I} + d\mathbf{J})]^{-1} = \frac{p}{(1+\epsilon)[1+a(1+\epsilon)]}.$$

Since $\mathbf{x}'\mathbf{x} \leq p$ from (2.15), we have

$$\mathbf{x}'\mathbf{M}(\eta^*)^{-1}[\mathbf{I} + \mathbf{M}(\eta^*)(a\mathbf{I} + d\mathbf{J})]^{-1}\mathbf{x} \leq Tr[\mathbf{I} + \mathbf{M}(\eta^*)(a\mathbf{I} + d\mathbf{J})]^{-1}, \forall \mathbf{x} \in \mathcal{X}.$$

The condition holds and therefore the design η^* is optimal. □

Remark: Theorem 3 suggests that when the random effects are equi-correlated, optimal designs are nonorthogonal and unbalanced (i.e., the off-diagonal elements of $\mathbf{M}(\eta^*)$ are non-zero). This is quite different from the designs considered optimal under the classical fixed-effects D-criterion. In Section 2.2.4 let us look at some examples of exact designs and compare their design efficiencies under the ψ-criterion in (2.6) and under the classical fixed-effects D-criterion. The efficiency of the design η under the ψ criterion in (2.6) is calculated as

$$\psi\text{-eff}(\eta) = \left(\frac{|\mathbf{M}(\eta)| \left| \frac{\sigma^2}{m}\mathbf{I} + \mathbf{\Lambda}\mathbf{M}_\psi(\eta^*) \right|}{|\mathbf{M}_\psi(\eta^*)| \left| \frac{\sigma^2}{m}\mathbf{I} + \mathbf{\Lambda}\mathbf{M}(\eta) \right|} \right)^{1/p}, \tag{2.20}$$

where η^* denotes the optimal design under the ψ criterion, whose information matrix is given in Theorem 3 as $\mathbf{M}_\psi(\eta^*) = (1+\epsilon)\mathbf{I} - \epsilon\mathbf{J}$ where the value of ϵ is given in (2.18). The efficiency of design η under the classical fixed-effects D-criterion is calculated as

$$D\text{-eff}(\eta) = \left(\frac{|\mathbf{M}(\eta)|}{|\mathbf{M}_D(\eta^+)|} \right)^{1/p} = |\mathbf{M}(\eta)|^{1/p} \text{ since } \mathbf{M}_D(\eta^+) = \mathbf{I}, \tag{2.21}$$

where η^+ is the optimal design under the classical fixed-effects D-criterion whose information matrix is $\mathbf{M}_D(\eta^+) = \mathbf{I}$ (see, for example, Kuhfeld, Tobias, and Garratt, 1994).

2.2.4 Design Examples for Scenario II

According to Theorem 3 of Section 2.2, when $\mathbf{\Lambda} = a^*\mathbf{I} + d^*\mathbf{J}$, an optimal continuous designs η^* satisfies $M(\eta^*) = (1+\epsilon)\mathbf{I} - \epsilon\mathbf{J}$ where ϵ is defined as a function of m,

a, d, and σ^2 in (2.18). As this $M(\eta^*)$ may not render a design with an integer number of observations at the design points (exact designs), a computer-search algorithm is used to search for designs that have high efficiency according to the ψ criterion in (2.6). The algorithm is developed using SAS PROC IML and employs a simple exchange rule as suggested by Mitchell and Miller (1970) (see Atkinson and Donev, 1992, page 172 for an explanation). The steps of the algorithm are outlined in in Appendix A.1. The SAS code of the algorithm can be obtained through the website: `http://www.stat.ohio-state.edu/~amd/dissertations.html`.

For the search of the designs, let us assume without loss of generality that $\sigma^2 = 1$, and examine different ratios of $d/a = d^*/a^*$ where $a = ma^*/\sigma^2$, $\quad d = md^*/\sigma^2$, and $\mathbf{\Lambda} = a^*\mathbf{I} + d^*\mathbf{J}$. Let $r = d/a$, if we divide both the numerator and the denominator of ϵ in (2.18) by a, then we get

$$\epsilon = \frac{2(1+pr) + \frac{1}{a} - 2r - \sqrt{4(1+\frac{1}{a})(1+pr) - 4(\frac{r}{a}) + \frac{1}{a^2}}}{2(1+pr)(p-2) + 2r}$$

As the number of runs m gets large while a^* and σ^2 remain fixed, $a = ma^*/\sigma^2$ gets large and $1/a$ becomes small. For example, when $\sigma^2 = 1$, $a^* = 1$ and $m = 10$, then $1/a = 0.1$ and $1/a^2 = 0.01$, so small that ϵ is almost equal to

$$\frac{2(1+pr) - 2r - \sqrt{4(1+pr)}}{2(1+pr)(p-2) + 2r},$$

which suggests that given the number of parameters p, the ϵ in the optimal design $\mathbf{M}(\eta^*)$ is mainly determined by the ratio of $r = d/a$.

Example 2.2: Consider a simple example with one treatment factor having 2 levels in a 10-run experiment (i.e., $m = 10$, and $p = 2$), and let $\mathbf{\Lambda} = a^*\mathbf{I} + d^*\mathbf{J}$. Table 2.1 reports the searching results for different values of d/a. In the second column of the table, the value of ϵ in (2.18) is reported. In the third column, the design from the search algorithm is expressed as (m_1,m_2), where $m_1(m_2)$ is the number of times level 1(2) occurs in the design. The resulting $\mathbf{X}'\mathbf{X}$ is also reported in the same column. The ψ-efficiency for each design is calculated from (2.20). The D-efficiency of the design is also reported based on the classical D-criterion for fixed effects, as shown in (2.21). Both efficiencies are reported in Column 4.

In addition, we check the performance of the design under alternative d/a ratios, as well as an alternative form of $\mathbf{\Lambda}$ where $\mathbf{\Lambda}_e = \begin{pmatrix} 1 & 0.825 \\ 0.825 & 0.6833 \end{pmatrix}$ for positively correlated random effects that are not equally correlated nor with equal variances, and $\mathbf{\Lambda}_{e-} = \begin{pmatrix} 1 & -0.825 \\ -0.825 & 0.6833 \end{pmatrix}$ for negatively correlated random effects that are not equally correlated nor with equal variances. The ψ-efficiencies of the design under alternative d/a ratios are reported in Columns 5 to 8. Since we do not know what the optimal design is under the general form of $\mathbf{\Lambda}_e$ and $\mathbf{\Lambda}_{e-}$, the orthogonal design is set as a reference point, and efficiencies relative to the orthogonal design is reported in the last column of the table. There are two parts of the table where the first part is for positively correlated random effects characterized by positive d/a ratios and $\mathbf{\Lambda}_e$, and the second part is for negatively correlated random effects with negative d/a ratios and $\mathbf{\Lambda}_{e-}$. Note that since $\mathbf{\Lambda}$ needs to be positive definite, d/a has to be greater than -0.5. When d/a is equal to, or greater than, 50, the random effects are almost perfectly correlated.

$\mathbf{\Lambda}$	M^*	Design $\mathbf{X}$	Efficiency	Design Performance Under alt. $\mathbf{\Lambda}$				
d/a	ϵ	(m_1, m_2)	ψ, D	ψ-eff. under alt. d/a				$\mathbf{\Lambda}_e$
		$\mathbf{X'X}$		0	1.25	5	50	rel. ψ-eff
0	0	(5, 5) 10I	$\psi = 1$ $D = 1$	1	0.9953	0.9891	0.9815	1
0.25	0.1	(5, 5) 10I	$\psi = 0.9993$ $D = 1$	1	0.9953	0.9891	0.9815	1
0.5 *	0.17	(4, 6) 12I - 2J	$\psi = 0.9999$ $D = 0.9798$	0.9964	0.9995	0.9956	0.9889	1.084 (=1/0.923)
1.25 *	0.30	(4, 6) 12I - 2J	$\psi = 0.9995$ $D = 0.9798$	0.9964	0.9995	0.9956	0.9889	1.084 (=1/0.923)
2 *	0.37	(3, 7) 14I - 4J	$\psi = 1$ $D = 0.9165$	0.9839	0.9993	0.9992	0.9942	1.152 (=1/0.868)
5	0.53	(2, 8) 16I - 6J	$\psi = 0.9997$ $D = 0.8$	0.9545	0.9923	0.9997	0.9979	1.196 (=1/0.836)
10	0.63	(2, 8) 16I - 6J	$\psi = 0.9999$ $D = 0.8$	0.9545	0.9923	0.9997	0.9979	1.196 (=1/0.836)
15	0.69	(2, 8) 16I - 6J	$\psi = 0.9996$ $D = 0.8$	0.9545	0.9923	0.9997	0.9979	1.196 (=1/0.836)
50	0.82	(1, 9) 18I - 8J	$\psi = 1$ $D = 0.6$	0.8742	0.9639	0.992	1	1.174 (=1/0.852)
d/a	ϵ	(m_1, m_2)	ψ, D	ψ-eff. under alt. d/a				$\mathbf{\Lambda}_{e-}$
		$\mathbf{X'X}$		0	1.25	5	50	rel. ψ-eff
-0.49	-0.61	(8, 2) 4I + 6J	$\psi = 0.9999$ $D = 0.8$	0.9545	0.958	0.9659	0.9817	1.196 (=1/0.836)
-0.4 *	-0.34	(7, 3) 6I + 4J	$\psi = 0.9993$ $D = 0.9165$	0.9839	0.9865	0.992	0.9993	1.152 (=1/0.868)
-0.25 *	-0.16	(6, 4) 8I + 2J	$\psi = 0.9998$ $D = 0.9798$	0.9964	0.9978	0.9998	0.996	1.084 (=1/0.923)
-0.1	-0.05	(5, 5) 10I	$\psi = 0.9997$ $D = 1$	1	0.9997	0.997	0.9789	1

*: recommended setting for the search of optimal exact designs when random effects are correlated

Table 2.1: 10-Run Designs Under the Main Effects Model of Example 2.2

From the first part of the table, we can see that when $d/a = 0$, i.e., the random effects are independent, the orthogonal design (equal allocation of m_1 and m_2) is D- and ψ- optimal. As the random effects become more and more correlated (as d/a increases), the ψ-optimal design becomes more unbalanced in the allocation of m_1 and m_2. As the random effects get more correlated, the D-efficiencies of the ψ-optimal designs drop significantly, for example, when $d/a = 5$, the D-efficiency (in the fourth column) of the ψ-optimal design is only 80%, suggesting that the optimal designs for the estimation of the hyperpameter $\boldsymbol{\theta}$ under the ψ criterion are far from what are considered optimal under the classical D-criterion for the estimation of fixed effects.

Since the actual $\boldsymbol{\Lambda}$ may very well be different from the $\boldsymbol{\Lambda}$ that we used to search for the ψ-optimal design, the performance of the ψ-optimal design searched under one value of d/a is examined under alternative values of d/a, as well as under the $\boldsymbol{\Lambda}_e$ for highly positively correlated random effects that are not equally correlated nor with equal variances, and $\boldsymbol{\Lambda}_{e-}$ for highly negatively correlated random effects that are not equally correlated nor with equal variances. The data suggest that the orthogonal design that is ψ-optimal under $d/a = 0$ performs well if the actual d/a takes on alternative values. In particular, the orthogonal design is 99.53% ψ-efficient if the actual $d/a = 1.25$, 98.91% ψ-efficient if $d/a = 5$, and 98.15% ψ-efficient if $d/a = 50$ (see the 1st row corresponding to $d/a = 0$, and the 6th to the 8th column of Table 2.1). However, if the actual $\boldsymbol{\Lambda}$ is $\boldsymbol{\Lambda}_e$, then the orthogonal design has a much lower ψ-efficiency. For example, it is only 92.3% as efficient as the unbalanced design $(4,\ 6)$ (see the 3rd row corresponding to $d/a = 0.5$ and the last column of Table 2.1), and 83.6% as efficient as the unbalanced design $(2,\ 8)$ (see the 6th row corresponding

to $d/a = 5$ and the last column of Table 2.1). On the other hand, the unbalanced design (4, 6) that is ψ-optimal under $d/a = 0.5$ and $d/a = 1.25$, performs well across all alternative forms of $\mathbf{\Lambda}$. In particular, it is 99.64% ψ-efficient when the random effects are actually independent (i.e., $d/a = 0$), and it is 99.95%, 99.56%, 99.89% ψ-efficient when $d/a = 1.25$, 5, and 50, respectively (see the 3rd and the 4th rows corresponding to $d/a = 0.5$ and $d/a = 1.25$, and the 5th to the 8th column of Table 2.1). When the actual $\mathbf{\Lambda} = \mathbf{\Lambda}_e$, the unbalanced design (4, 6) is 1.084 (=1/92.3%) times ψ-efficient in comparison to the orthogonal design (see the 3rd and the 4th rows corresponding to $d/a = 0.5$ and $d/a = 1.25$, and the 9th column of Table 2.1). Similarly, the unbalanced design (3, 7) which is optimal under ($d/a = 2$) also performs with high efficiencies across all scenarios of $\mathbf{\Lambda}$. The results in the second part of the table with negative ratios of d/a are similar and therefore are not discussed here.

This example suggests that when the individual-level random effects are correlated, designs with unbalanced allocation of the two factor levels are more efficient than the orthogonal design with equal allocation under the ψ criterion in (2.6), especially when the correlation is strong (such as that shown in $\mathbf{\Lambda}_e$ and $\mathbf{\Lambda}_{e-}$), which agrees with the finding in Example 1 of Entholzner et al. (2006).

Example 2.3: Consider an 8-run experiment with 2 treatment factors each having 2 levels for the estimation of hyperparameters $\boldsymbol{\theta}$ in a main-effects model as discussed in Section 2.2.1 such that $m = 8$ and $p = 3$. Table 2.2 reports the searching results for different values of d/a. The structure of the table is the same as that in Table 2.1, with three exceptions. First, in the third column, the design from the search algorithm is

expressed as $(m_{11},\ m_{12}, m_{21},\ m_{22})$, where m_{ij} is the number of times level i of factor 1 and level j of factor 2 occur together in the design. Second, two general forms of $\mathbf{\Lambda}$ are examined, with

$$\mathbf{\Lambda}_{e1} = \begin{pmatrix} 1 & 0.825 & 0.825 \\ 0.825 & 0.6833 & 0.68 \\ 0.825 & 0.68 & 1 \end{pmatrix}, \quad \text{and} \quad \mathbf{\Lambda}_{e2} = \begin{pmatrix} 1 & 0.825 & 0.825 \\ 0.825 & 1 & 0.68 \\ 0.825 & 0.68 & 0.6833 \end{pmatrix},$$

which correspond to the scenarios of highly correlated random effects which are not equally correlated nor with equal variances. Specifically, the correlation between the first two elements of $\boldsymbol{\theta}$ is particularly high in $\mathbf{\Lambda}_{e_1}$ and the correlation between the first element and the third element of $\boldsymbol{\theta}$ is particularly high in $\mathbf{\Lambda}_{e_2}$. Third, negative d/a ratios are not examined here since Example 2.2 suggest that the optimal designs in the case of negative d/a ratios are the mirror images of the optimal designs in the case of positive d/a ratios (i.e., the designs are the same with an exchange of the treatment labels).

Similar to the findings in Example 2.2, from the first part of the table, we can see that when $d/a = 0$, i.e., the random effects are independent, the orthogonal design (equal allocation of m_{11}, m_{12}, m_{21} and m_{22}) is D- and ψ- optimal. As the random effects become more and more correlated (as d/a increases), the optimal ψ-design becomes nonorthogonal and unbalanced. For example, when $d/a \geq 5$, $m_{11} = 0$, i.e., the ψ-optimal design does not contain the combination of the first level of the two attributes. As the random effects get more correlated, the D-efficiencies of the ψ-optimal designs drop significantly, for example, when $d/a = 5$, the D-efficiency of the ψ-optimal design is only 82.6%, suggesting that the optimal designs for the estimation

of the hyperpameter $\boldsymbol{\theta}$ under the ψ criterion are far from what are considered optimal under the classical D-criterion for the estimation of fixed effects.

$\boldsymbol{\Lambda}$	M^*	Design $\mathbf{X}$	Efficiency	Design Performance Under alt. $\boldsymbol{\Lambda}$				
d/a	ϵ	$(m_{11}, m_{12}, m_{21}, m_{22})$	ψ, D	ψ-eff. under alt. d/a			$\boldsymbol{\Lambda}_{e1}$	$\boldsymbol{\Lambda}_{e2}$
		$\mathbf{X}'\mathbf{X}$		0	5	50	rel. ψ-eff	rel. ψ-eff
0	0	(2,2,2,2) 8I	$\psi = 1$ $D = 1$	1	0.986	0.979	1	1
0.25	0.08	(2,2,2,2) 8I	$\psi = 0.999$ $D = 1$	1	0.986	0.979	1	1
0.5 *	0.13	(1,3,2,2) $\begin{pmatrix} 8 & 0 & -2 \\ 0 & 8 & -2 \\ -2 & -2 & 8 \end{pmatrix}$	$\psi = 0.998$ $D = 0.957$	0.990	0.994	0.988	1.022 (1/0.978)	1.061 (1/0.943)
1.25 *	0.21	(1,2,3,2) $\begin{pmatrix} 8 & -2 & 0 \\ -2 & 8 & -2 \\ 0 & -2 & 8 \end{pmatrix}$	$\psi = 0.998$ $D = 0.957$	0.990	0.994	0.988	1.061 (1/0.943)	1.022 (1/0.978)
2 *	0.26	(1,3,2,2) $\begin{pmatrix} 8 & 0 & -2 \\ 0 & 8 & -2 \\ -2 & -2 & 8 \end{pmatrix}$	$\psi = 0.997$ $D = 0.957$	0.990	0.994	0.988	1.022 (1/0.978)	1.061 (1/0.943)
5	0.33	(0,3,3,2) $\begin{pmatrix} 8 & -2 & -2 \\ -2 & 8 & -4 \\ -2 & -4 & 8 \end{pmatrix}$	$\psi = 0.998$ $D = 0.826$	0.948	0.998	0.995	1.052 (1/0.95)	1.052 (1/0.95)
10	0.37	(0,3,3,2) $\begin{pmatrix} 8 & -2 & -2 \\ -2 & 8 & -4 \\ -2 & -4 & 8 \end{pmatrix}$	$\psi = 0.998$ $D = 0.826$	0.948	0.998	0.995	1.052 (1/0.95)	1.052 (1/0.95)
15	0.39	(0,3,2,3) $\begin{pmatrix} 8 & -2 & -4 \\ -2 & 8 & -2 \\ -4 & -2 & 8 \end{pmatrix}$	$\psi = 0.997$ $D = 0.826$	0.948	0.998	0.995	1.077 (1/0.929)	1.113 (1/0.898)
50	0.44	(0,3,2,3) $\begin{pmatrix} 8 & -2 & -4 \\ -2 & 8 & -2 \\ -4 & -2 & 8 \end{pmatrix}$	$\psi = 0.995$ $D = 0.8256$	0.948	0.998	0.995	1.077 (1/0.929)	1.113 (1/0.898)

*: recommended setting for the search of optimal exact designs when random effects are correlated

Table 2.2: 8-Run Designs Under the Main Effects Model of Example 2.3

Also similar to the findings in Example 2.2, the orthogonal design (ψ-optimal under $d/a = 0$) performs well when the actual d/a takes on different values. In particular, the orthogonal design is 98.6% ψ-efficient if the actual $d/a = 5$, and 97.9% ψ-efficient if $d/a = 50$ (see the 1st row corresponding to $d/a = 0$, and the 6th to the 7th column of Table 2.2). However, if $\mathbf{\Lambda} = \mathbf{\Lambda}_{e1}$ or $\mathbf{\Lambda} = \mathbf{\Lambda}_{e2}$, the orthogonal design has a much lower ψ-efficiency relative to the nonorthogonal unbalanced designs. For example, it is only 94.3% as efficient as the nonorthogonal unbalanced design (1, 2, 3, 2) (ψ-optimal under $d/a = 1.25$) under $\mathbf{\Lambda}_{e1}$. On the other hand, the nonorthogonal unbalanced design (1, 2, 3, 2) that is $\psi-$optimal under $d/a = 1.25$ performs well across all alternative forms of $\mathbf{\Lambda}$. In particular, it is 99% ψ-efficient when the random effects are actually independent (i.e., $d/a = 0$), and it is 99.4%, 98.8% ψ-efficient when $d/a = 5$ and 50 respectively (see the 4th row corresponding to $d/a = 1.25$, and the 5th to the 7th column of Table 2.2). When the actual $\mathbf{\Lambda} = \mathbf{\Lambda}_{e1}$, the nonorthogonal unbalanced design (1, 2, 3, 2) is 1.061 (=1/94.3%) times ψ-efficient in comparison to the orthogonal design, and is 1.022 (=1/97.8%) times ψ-efficient in comparison to the orthogonal design (see the 4th row corresponding to $d/a = 1.25$, and the last two columns of Table 2.2).

Examples 2.2 and 2.3 suggest that, in practical applications, when the experimenter believes that the random effects are correlated, orthogonal designs are not good choices of designs. Instead, nonorthogonal designs searched under moderate ratios of d/a such as $d/a = 0.5, \ , 1.25$, or 2 are more efficient for the estimation of hyperparameters $\boldsymbol{\theta}$ in a hierarchical linear main-effects model than the orthogonal

designs, and moreover have high efficiencies across a variety of possible forms of positively correlated $\mathbf{\Lambda}$. This is very useful since random effects are seen to be correlated. The experimenter may have some idea on the direction of the correlation (positive or negative) but does not know the specific correlation structure in $\mathbf{\Lambda}$. From the results in the two examples it is recommended to use $d/a = 0.5,\ 1.25$ or 2 to search for the optimal exact designs in the situation of positively correlated random effects, and to use $d/a = -0.25$,or-0.4 in the situation of negatively correlated random effects.

2.2.5 Scenario III: Random Effects are Equi-correlated Within-groups but Independent Between-groups

This scenario is an extension of Scenario I and Scenario II. Let $g = 1, \ldots, G$ denote the groups of random effects, and let p_g denote the size of the gth group. Suppose that

$$\mathbf{\Lambda} = Diag(\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \ldots, \mathbf{\Lambda}_G), \quad \mathbf{\Lambda}_g = a_g^* \mathbf{I}_{p_g} + d_g^* \mathbf{J}_{p_g}.$$

For example, in the market research study of consumer preference of credit card products, different levels of credit card attributes are tested. Some attributes are money related such as credit limit and annual fee, while some other attributes are non-money related such as the logo on the card and the co-brand of the card. It is reasonable to assume that attributes are independent across groups, but are correlated within the same group. The following theorem gives the optimal factorial design structure in this scenario.

Theorem 4. *When* $\mathbf{\Lambda} = Diag(\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \ldots, \mathbf{\Lambda}_G)$, *where* $\mathbf{\Lambda}_g = a_g^* \mathbf{I}_{p_g} + d_g^* \mathbf{J}_{p_g}$, *design* η^* *with information matrix* $\mathbf{M}(\eta^*) = Diag(\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_G)$, *where* $\mathbf{M}_g = (1 + \epsilon_g)\mathbf{I} -$

$\epsilon_g \mathbf{J}$, *is optimal under the* ψ *criterion, with*

$$\epsilon_g = \frac{2(a_g + p_g d_g) + 1 - 2d_g - \sqrt{4(a_g+1)(a_g + p_g d_g) - 4d_g + 1}}{2(a_g + p_g d_g)(p_g - 2) + 2d_g}, \tag{2.22}$$

$a_g = ma_g^*/\sigma^2$, *and* $d_g = md_g^*/\sigma^2$, $g = 1, \ldots, G$.

Proof. Express the vector $\mathbf{x}$ as $\mathbf{x} = (\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_G')'$, where vector $\mathbf{x}_g$ is of dimension p_g, $g = 1, \ldots, G$. Due to the block diagonal nature of $\mathbf{\Lambda}$ and $\mathbf{M}$, the left hand side of (2.12) becomes

$$\sum_{g=1}^{G} \mathbf{x}_g' \mathbf{M}_g^{-1} [\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}_g \mathbf{\Lambda}_g]^{-1} \mathbf{x}_g,$$

and the right hand side becomes

$$\sum_{g=1}^{G} Tr[\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}_g \mathbf{\Lambda}_g]^{-1}.$$

Since we proved in Scenario II that, for a given g,

$$\mathbf{x}_g' \mathbf{M}_g^{-1} [\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}_g \mathbf{\Lambda}_g]^{-1} \mathbf{x}_g \leq Tr[\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}_g \mathbf{\Lambda}_g]^{-1},$$

we have

$$\sum_{g=1}^{G} \mathbf{x}_g' \mathbf{M}_g^{-1} [\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}_g \mathbf{\Lambda}_g]^{-1} \mathbf{x}_g \leq \sum_{g=1}^{G} Tr[\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}_g \mathbf{\Lambda}_g]^{-1}$$

The condition (2.12) is satisfied and this proves the optimality of the design η^*. □

2.3 Optimal Designs When Z_i Can Be Controlled

In Section 2.2, optimal design structures were investigated when the covariate matrix $\mathbf{Z}_i$ for subject i is equal to the identity matrix. In this section, the design problem is investigated in which both the matrix $\mathbf{X}$ ($\mathbf{X}_i = \mathbf{X}$ for all i), and the matrix $\mathbf{Z}_i$ can be controlled by the experimenter. For example, in a survey study,

the experimenter can control the questions to be administered to respondents (which determines $\mathbf{X}$) as well as the sampling of the respondents on the basis of certain demographic information such as age or income (which determines $\mathbf{Z}_i$). Consider the situation in which $\mathbf{X}$ and $\mathbf{Z}_i$ can be determined independently of each other, and suppose that the length of the hyperparameter vector $\boldsymbol{\theta}$ (which is q) is a multiple of the length of the individual-level effect $\boldsymbol{\beta}_i$ (which is p). For example, when the $\mathbf{Z}_i$ includes respondent-specific age and household income, the hyperparameter vector $\boldsymbol{\theta}$ in (1.2) includes the means (over the respondents) of the p individual-level parameters in $\boldsymbol{\beta}_i$ (which corresponds to the first set of p hyperparameters in $\boldsymbol{\theta}$), the effects of respondent age on the p individual-level parameters in $\boldsymbol{\beta}_i$ (which corresponds to the second set of p hyperparameters in $\boldsymbol{\theta}$), as well as the effects of respondent household income on the p individual-level parameters in $\boldsymbol{\beta}_i$ (which corresponds to the third set of p hyperparameters in $\boldsymbol{\theta}$). The size of the hyperparameter vector $\boldsymbol{\theta}$ is therefore $q = p + p + p = 3p$. In this situation, the $p \times q$ covariate matrix $\mathbf{Z}_i$ is

$$\mathbf{Z}_i = [\mathbf{I}_p,\ a_i\mathbf{I}_p,\quad h_i\mathbf{I}_p] = \mathbf{I}_p \otimes \mathbf{z}_i',$$

where $\mathbf{z}_i' = [1,\quad a_i,\quad h_i]$. a_i represents the age of respondent i, and h_i represents respondent i's household income. The size of the vector $\mathbf{z}_i$ is q/p. Expression (2.3) becomes

$$\Big|\sum_{i=1}^{n}\left[(\mathbf{I}_p \otimes \mathbf{z}_i')'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i(\mathbf{I}_p \otimes \mathbf{z}_i')\right]\Big|. \tag{2.23}$$

According to Theorem 8.8.4 in Graybill (1983), for any matrices $\mathbf{A}$ and $\mathbf{B}$, $(\mathbf{A}\otimes\mathbf{B})' = \mathbf{A}'\otimes\mathbf{B}'$. In addition, according to Theorem 8.8.6 in Graybill (1983), for rectangular matrices $\mathbf{C}$ $\mathbf{D}$, $\mathbf{F}$, $\mathbf{G}$ of sizes $m_1 \times n_1$, $m_2 \times n_2$, $n_1 \times k_1$ and $n_2 \times k_2$ respectively,

$(\mathbf{C} \otimes \mathbf{D})(\mathbf{F} \otimes \mathbf{G}) = (\mathbf{CF}) \otimes (\mathbf{DG})$. Expression (2.23) is therefore equivalent to

$$\begin{aligned}
&\Big| \sum_{i=1}^{n} \big[(\mathbf{I}_p \otimes \mathbf{z}_i) \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i (\mathbf{I}_p \otimes \mathbf{z}_i') \big] \Big| \\
=&\Big| \sum_{i=1}^{n} \big[(\mathbf{I}_p \otimes \mathbf{z}_i) (\mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \otimes 1) (\mathbf{I}_p \otimes \mathbf{z}_i') \big] \Big| \\
&\quad \text{since} \;\; \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \otimes 1 = \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \\
=&\Big| \sum_{i=1}^{n} \big[(\mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \otimes \mathbf{z}_i) (\mathbf{I}_p \otimes \mathbf{z}_i') \big] \Big| \\
=&\Big| \sum_{i=1}^{n} \big[(\mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i) \otimes (\mathbf{z}_i \mathbf{z}_i') \big] \Big| \qquad (2.24)
\end{aligned}$$

When all subjects receive the same design, i.e., $\mathbf{X}_i = \mathbf{X}$, and when the response errors are homoscedastic, i.e., $\sigma_i^2 = \sigma^2$, as in assumption (i) and (ii) of Section 2.2, equation (2.24) simplifies to

$$\begin{aligned}
&\Big| \sum_{i=1}^{n} \big[(\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X}) \otimes (\mathbf{z}_i \mathbf{z}_i') \big] \Big| \\
=&\Big| (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X}) \otimes \big[\sum_{i=1}^{n} (\mathbf{z}_i \mathbf{z}_i') \big] \Big|, \;\; \text{by Theorem 8.8.11 in Graybill (1983)} \\
=&\Big| \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X}) \Big|^{q/p} \Big| \sum_{i=1}^{n} (\mathbf{z}_i \mathbf{z}_i') \Big|^{p}, \;\; \text{by Theorem 8.8.10 in Graybill (1983)}
\end{aligned}$$

where q/p is the size of vector $\mathbf{z}_i$. With the independence of $\mathbf{z}_i$ and $\mathbf{X}$, and given the number of parameters p and q, the maximization of the determinant in the above equation is achieved through the individual maximization of $|\sum_{i=1}^{n} (\mathbf{z}_i \mathbf{z}_i')|$ and $|\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X}|$, the latter of which has been discussed in Section 2.2. For the maximization of $|\sum_{i=1}^{n} (\mathbf{z}_i \mathbf{z}_i')|$, the classical fixed-effects D-optimal design theory applies which suggests that, for example, if the effect of household income on consumer price sensitivity $\boldsymbol{\beta}_i$ is quadratic, then $|\sum_{i=1}^{n} (\mathbf{z}_i \mathbf{z}_i')|$ is maximized by selecting the consumer sample so that there are equal number of households at the low, middle and the high income

levels. On the other hand, if the effect is assumed to be linear, then $|\sum_{i=1}^{n}(\mathbf{z}_i\mathbf{z}_i')|$ is maximized by selecting the consumer sample so that there are equal number of households at the two extreme income levels.

Note that in other circumstances when $\mathbf{Z}_i = \mathbf{I}_p \otimes \mathbf{z}_i'$ does not hold, for example, when $\mathbf{X}_i$ and $\mathbf{Z}_i$ are linked, such as in the modeling of the "level effect" in Chapter 4, when $\mathbf{Z}_i$ consists of relative ranges and frequencies of the attribute levels present in $\mathbf{X}_i$, computer search algorithms need to be used to find the optimal combination of $\mathbf{X}_i$ and $\mathbf{Z}_i$.

CHAPTER 3

Optimal Designs for Estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\Lambda}$

In this chapter, consider the situation in which estimation of both $\boldsymbol{\theta}$ and $\boldsymbol{\Lambda}$ model (2.1) is of interest to the experimenter. For example, a retailer is not only interested in knowing the mean consumer preference but also the dispersion of individual consumer preferences. The response error variance σ^2 is considered to be a nuisance parameter. In this situation, we seek to minimize the determinant of the posterior variance of $(\boldsymbol{\theta}, \boldsymbol{\Lambda})|\sigma^2$. However, since the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\Lambda})|\mathbf{y}, \sigma^2$ cannot be expressed in closed form, the Fisher Information approach is taken in which the posterior variance of $(\boldsymbol{\theta}, \boldsymbol{\Lambda})|\mathbf{y}, \sigma^2$ is approximated by the Fisher Information matrix $FI_{\boldsymbol{\theta},\, \boldsymbol{\Lambda}}$ (see Chaloner and Verdinelli, 1995, page 286), and the maximization of the determinant of $FI_{\boldsymbol{\theta},\, \boldsymbol{\Lambda}}$ is used as the optimal experimental design criterion, that is, maximize

$$\mathcal{Q} = |FI_{\boldsymbol{\theta},\, \boldsymbol{\Lambda}}| = \begin{vmatrix} FI(\boldsymbol{\theta}, \boldsymbol{\theta}) & FI(\boldsymbol{\theta}, \boldsymbol{\Lambda}) \\ FI(\boldsymbol{\theta}, \boldsymbol{\Lambda}) & FI(\boldsymbol{\Lambda}, \boldsymbol{\Lambda}) \end{vmatrix}$$

Let $[\boldsymbol{\Lambda}]_{uv} = \lambda_{u,v}$ denote the $(u, v)^{th}$ element of $\boldsymbol{\Lambda}$, then it can be shown (see Lenk et al. 1996) that under model (2.1),

$$FI(\boldsymbol{\theta}, \boldsymbol{\Lambda}) = \mathbf{0},$$

$$\mathcal{Q} = |FI_{\boldsymbol{\theta},\,\boldsymbol{\Lambda}}| = |FI(\boldsymbol{\theta},\boldsymbol{\theta})||FI(\boldsymbol{\Lambda},\boldsymbol{\Lambda})|, \quad \text{since} \quad FI(\boldsymbol{\theta},\boldsymbol{\Lambda}) = \mathbf{0}. \tag{3.1}$$

$$FI(\boldsymbol{\theta},\boldsymbol{\theta}) = \sum_{i=1}^{n}(\mathbf{Z}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\mathbf{Z}_i) \quad \text{where} \quad \boldsymbol{\Sigma}_i = \sigma_i^2\mathbf{I}_{m_i} + \mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i',$$

$$FI(\lambda_{u,v},\lambda_{r,s}) = \frac{1}{2}\sum_{i=1}^{n} Tr\left(\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\lambda_{u,v}}\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\lambda_{r,s}}\right). \tag{3.2}$$

These will be used in Section 3.1 and 3.2.

3.1 Scenario I: Random Effects are Independent and homoscedastic

As in Section 2.2, we first take $\mathbf{X}_i = \mathbf{X}$, $\sigma_i^2 = \sigma^2$, $\mathbf{Z}_i = \mathbf{I}_p$ and look at the simplified situation in which the random effects are independently distributed with equal variances, i.e., $\boldsymbol{\Lambda}$ is assumed to be diagonal with equal diagonal elements such that $\boldsymbol{\Lambda} = \lambda_0^2\mathbf{I}_p$, and $\boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}_m + \lambda_0^2\mathbf{X}\mathbf{X}'$ for all i. So from (3.2),

$$\begin{aligned} FI(\lambda_0,\lambda_0) &= \frac{n}{2}Tr\left[(\sigma^2\mathbf{I}_m + \lambda_0^2\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\mathbf{X}'(\sigma^2\mathbf{I}_m + \lambda_0^2\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\mathbf{X}'\right] \\ &= \frac{n}{2}Tr\left\{[\mathbf{X}'\mathbf{X}(\sigma^2\mathbf{I}_p + \lambda_0^2\mathbf{X}'\mathbf{X})^{-1}]^2\right\} \\ &\quad \text{since } (\sigma^2\mathbf{I}_m + \lambda_0^2\mathbf{X}\mathbf{X}')^{-1}\mathbf{X} = \mathbf{X}(\sigma^2\mathbf{I}_p + \lambda_0^2\mathbf{X}'\mathbf{X})^{-1} \text{ by (2.8)} \end{aligned}$$

By Lemma 1 (i) in Section 2.2,

$$|FI(\boldsymbol{\theta},\boldsymbol{\theta})| = |\mathbf{X}'\mathbf{X}(\sigma^2\mathbf{I}_p + \boldsymbol{\Lambda}\mathbf{X}'\mathbf{X})^{-1}| = \frac{|\mathbf{X}'\mathbf{X}|}{|\sigma^2\mathbf{I}_p + \boldsymbol{\Lambda}\mathbf{X}'\mathbf{X}|}$$

The maximization of the criterion $\mathcal{Q}$ in (3.1) now simplifies to the maximization of

$$\mathcal{Q}_3 = \frac{|\mathbf{X}'\mathbf{X}|}{|\sigma^2\mathbf{I}_p + \lambda_0^2\mathbf{X}'\mathbf{X}|}Tr\left\{[\mathbf{X}'\mathbf{X}(\sigma^2\mathbf{I}_p + \lambda_0^2\mathbf{X}'\mathbf{X})^{-1}]^2\right\}, \tag{3.3}$$

Since $\mathbf{M}(\eta) = \frac{1}{m}\mathbf{X}'\mathbf{X}$ where m is fixed, the maximization of (3.3) is equivalent to the maximization of

$$\frac{|\mathbf{M}(\eta)|}{|\frac{\sigma^2}{m}\mathbf{I}_p + \lambda_0^2\mathbf{M}(\eta)|}Tr\left\{[\mathbf{M}(\eta)(\tfrac{\sigma^2}{m}\mathbf{I}_p + \lambda_0^2\mathbf{M}(\eta))^{-1}]^2\right\},$$

which is also equivalent to the maximization of

$$\frac{|\mathbf{M}(\eta)|}{|\mathbf{I}_p + c\mathbf{M}(\eta)|} Tr\left\{[\mathbf{M}(\eta)(\mathbf{I}_p + c\mathbf{M}(\eta))^{-1}]^2\right\}, \tag{3.4}$$

for a given σ^2 and a given λ_0^2. The scalar $c = m\lambda_0^2/\sigma^2$.

The maximization of the first term of (3.4) is a special case of Scenario I investigated in Section 2.2.2. Theorem 2 shows that the maximization is achieved by the design η^* that satisfies $\mathbf{M}(\eta^*) = \mathbf{I}_p$. Let us now focus on the second term of (3.4) and first prove that it is an increasing and concave function in $\mathcal{M}$ as defined in (2.9).

Lemma 4. *The function*

$$\xi = \begin{cases} Tr\left\{[\mathbf{M}(\eta)(\mathbf{I}_p + c\mathbf{M}(\eta))^{-1}]^2\right\}, & \text{if } \mathbf{M}(\eta) \text{ is nonsingular} \\ -\infty & \text{if } \mathbf{M}(\eta) \text{ is singular} \end{cases}$$

is concave and increasing in $\mathcal{M}$ where $\mathcal{M}$ is defined in (2.9).

Proof. For display clarity, let us omit the subscript and use $\mathbf{M}$ to represent $\mathbf{M}(\eta)$. When $\mathbf{M}$ is nonsingular, $\xi = Tr([\mathbf{M}(\mathbf{I} + c\mathbf{M})^{-1}]^2) = Tr(c\mathbf{I} + \mathbf{M}^{-1})^{-2}$. Let $\tilde{\mathbf{M}} = (c\mathbf{I} + \mathbf{M}^{-1})^{-1}$, then for $\mathbf{M}_1 > \mathbf{M}_2$ (i.e., $\mathbf{M}_1 - \mathbf{M}_2$ is positive definite),

$$c\mathbf{I} + \mathbf{M}_1^{-1} < c\mathbf{I} + \mathbf{M}_2^{-1},$$

and since $c\mathbf{I} + \mathbf{M}^{-1}$ is nonsingular,

$$(c\mathbf{I} + \mathbf{M}_1^{-1})^{-1} > (c\mathbf{I} + \mathbf{M}_2^{-1})^{-1}, \text{ i.e. } \tilde{\mathbf{M}}_1 > \tilde{\mathbf{M}}_2$$

(see proof in Lemma 2 and Theorem 12.2.14, Graybill, 1983). Now,

$$\begin{aligned} \xi(\mathbf{M}_1) - \xi(\mathbf{M}_2) &= Tr(\tilde{\mathbf{M}}_1^2) - Tr(\tilde{\mathbf{M}}_2^2) \\ &= Tr\left[(\tilde{\mathbf{M}}_1 - \tilde{\mathbf{M}}_2)(\tilde{\mathbf{M}}_1 + \tilde{\mathbf{M}}_2)\right], \text{ since } Tr(\tilde{\mathbf{M}}_1\tilde{\mathbf{M}}_2) = Tr(\tilde{\mathbf{M}}_2\tilde{\mathbf{M}}_1) \\ &= Tr\left[(\tilde{\mathbf{M}}_1 - \tilde{\mathbf{M}}_2)\tilde{\mathbf{M}}_1\right] + Tr\left[(\tilde{\mathbf{M}}_1 - \tilde{\mathbf{M}}_2)\tilde{\mathbf{M}}_2\right] \end{aligned} \tag{3.5}$$

Since $\mathbf{M}$ is positive definite, its eigenvalues $e_i,\ (i = 1, \ldots p)$ are all positive, and the eigenvalues of the symmetric matrix $\tilde{\mathbf{M}}$ which are $(c + 1/e_i)^{-1}$ are also all positive. Therefore, $\tilde{\mathbf{M}}$ is positive definite. According to Theorem 12.2.3 in Graybill (1983), for two positive definite matrices $\mathbf{A}$ and $\mathbf{B}$ of size $p \times p$, $Tr(\mathbf{AB}) > 0$. If we let $\mathbf{A} = \tilde{\mathbf{M}}_1 - \tilde{\mathbf{M}}_2$, and let $\mathbf{B} = \tilde{\mathbf{M}}_1$ and $\mathbf{B} = \tilde{\mathbf{M}}_2$ respectively for the first term and the second term of (3.5), we get

$$\xi(\mathbf{M}_1) - \xi(\mathbf{M}_2) > 0, \text{ for } \mathbf{M}_1 > \mathbf{M}_2.$$

Therefore, the function ξ is strictly increasing.

To prove that ξ is concave, write ξ as

$$\xi = Tr(\tilde{\tilde{\mathbf{M}}}^{-1}), \quad \text{where } \tilde{\tilde{\mathbf{M}}} = (c\mathbf{I} + \mathbf{M}^{-1})^2.$$

Note that since $\mathbf{M}$ is symmetric,

$$\tilde{\tilde{\mathbf{M}}} = (c\mathbf{I} + \mathbf{M}^{-1})^2 = (c\mathbf{I} + \mathbf{M}^{-1})'(c\mathbf{I} + \mathbf{M}^{-1}).$$

Therefore, $\tilde{\tilde{\mathbf{M}}}$ is positive definite and $\tilde{\tilde{\mathbf{M}}} \in \mathcal{M}$. According to A.2 in Silvey (1980) Appendix 1, $\tilde{\tilde{\mathbf{M}}}^{-1}$ is concave on $\mathcal{M}$. Then since the trace function is a linear increasing function, $\xi = Tr(\tilde{\tilde{\mathbf{M}}}^{-1})$ is also concave.

□

Theorem 5. *Let η be a design measure in the class of probability distributions H on the Borel sets of a compact design space $\mathcal{X} \subseteq \mathcal{R}^p$. A sufficient and necessary condition for a design η^* to maximize ξ is as follows:*

$$\mathbf{x}'(c\mathbf{M} + \mathbf{I})^{-1}\mathbf{M}(c\mathbf{M} + \mathbf{I})^{-2}\mathbf{x} \leq Tr\left(\mathbf{M}(c\mathbf{M} + \mathbf{I})^{-1}\mathbf{M}(c\mathbf{M} + \mathbf{I})^{-2}\right), \forall \mathbf{x} \in \mathcal{X}.$$

Proof. The Gâteaux derivative of function ξ is:

$$
\begin{aligned}
G_\xi\{\mathbf{M}_1,\mathbf{M}_2\} &= \lim_{\{\epsilon\to 0^+\}} \tfrac{1}{\epsilon}\Big\{Tr\big(c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}\big)^{-2} - Tr(c\mathbf{I}+\mathbf{M}_1^{-1})^{-2}\Big\} \\
&= \lim_{\{\epsilon\to 0^+\}} \tfrac{1}{\epsilon}\Big\{Tr\Big[\ \big[(c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1})^{-1}+(c\mathbf{I}+\mathbf{M}_1^{-1})^{-1}\big] \\
&\quad \big[(c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1})^{-1}-(cI+\mathbf{M}_1^{-1})^{-1}\big]\Big]\Big\} \\
&= \lim_{\{\epsilon\to 0^+\}} \tfrac{1}{\epsilon}\Big\{Tr\Big[(c\mathbf{I}+\mathbf{M}_1^{-1})^{-1}\big[(c\mathbf{I}+\mathbf{M}_1^{-1})(c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1})^{-1}+\mathbf{I}\big] \\
&\quad (c\mathbf{I}+\mathbf{M}_1^{-1})^{-1}\big[(c\mathbf{I}+\mathbf{M}_1^{-1})(c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1})^{-1}-\mathbf{I}\big]\Big]\Big\} \\
&= \lim_{\{\epsilon\to 0^+\}} \tfrac{1}{\epsilon}\Big\{Tr\Big[(c\mathbf{I}+\mathbf{M}_1^{-1})^{-1}(c\mathbf{I}+\mathbf{M}_1^{-1}+c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}) \\
&\quad \big[c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}\big]^{-1}(c\mathbf{I}+\mathbf{M}_1^{-1})^{-1}(c\mathbf{I}+\mathbf{M}_1^{-1}-c\mathbf{I}-(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}) \\
&\quad \big[c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}\big]^{-1}\Big]\Big\}, \\
&\qquad \text{since}\ \ \mathbf{I}=[c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}][c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}]^{-1}. \\
&= \lim_{\{\epsilon\to 0^+\}} \tfrac{1}{\epsilon}\Big\{Tr\Big[(c\mathbf{I}+\mathbf{M}_1^{-1})^{-1}(c\mathbf{I}+\mathbf{M}_1^{-1}+c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}) \\
&\quad \big[c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}\big]^{-1}(c\mathbf{I}+\mathbf{M}_1^{-1})^{-1}\mathbf{M}_1^{-1}[\mathbf{I}-(\mathbf{I}+\epsilon\mathbf{M}_2\mathbf{M}_1^{-1})^{-1}] \\
&\quad \big[c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}\big]^{-1}\Big]\Big\} \\
&= \lim_{\{\epsilon\to 0^+\}} \tfrac{1}{\epsilon}\Big\{Tr\Big[(c\mathbf{I}+\mathbf{M}_1^{-1})^{-1}(c\mathbf{I}+\mathbf{M}_1^{-1}+c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}) \\
&\quad \big[c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}\big]^{-1}(c\mathbf{I}+\mathbf{M}_1^{-1})^{-1}\mathbf{M}_1^{-1}(\mathbf{I}+\epsilon\mathbf{M}_2\mathbf{M}_1^{-1})^{-1} \\
&\quad (\mathbf{I}+\epsilon\mathbf{M}_2\mathbf{M}_1^{-1}-\mathbf{I})\big[c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}\big]^{-1}\Big]\Big\} \\
&= \lim_{\{\epsilon\to 0^+\}} \Big\{Tr\Big[\mathbf{M}_2\mathbf{M}_1^{-1}\big[c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}\big]^{-1}(c\mathbf{I}+\mathbf{M}_1^{-1})^{-1} \\
&\quad (c\mathbf{I}+\mathbf{M}_1^{-1}+c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1})\big[c\mathbf{I}+(\mathbf{M}_1+\epsilon\mathbf{M}_2)^{-1}\big]^{-1} \\
&\quad (c\mathbf{I}+\mathbf{M}_1^{-1})^{-1}\mathbf{M}_1^{-1}(\mathbf{I}+\epsilon\mathbf{M}_2\mathbf{M}_1^{-1})^{-1}\Big]\Big\}
\end{aligned}
$$

By Equation 8, page 69, from Morrison (1990) as quoted in Equation (2.7),

$$
\begin{aligned}
(\mathbf{M}_1 + \epsilon\mathbf{M}_2)^{-1} &= \mathbf{M}_1^{-1} - \mathbf{M}_1^{-1}\left(\epsilon^{-1}\mathbf{M}_2^{-1} + \mathbf{M}_1^{-1}\right)^{-1}\mathbf{M}_1^{-1} = \mathbf{M}_1^{-1} + O(\epsilon) \\
[c\mathbf{I} + (\mathbf{M}_1 + \epsilon\mathbf{M}_2)^{-1}]^{-1} &= c^{-1}\mathbf{I} - c^{-1}\left(\mathbf{M}_1 + \epsilon\mathbf{M}_2 + c^{-1}\mathbf{I}\right)^{-1}c^{-1} \\
&= c^{-1}\mathbf{I} - c^{-2}\left[(\mathbf{M}_1 + c^{-1}\mathbf{I})^{-1} + O(\epsilon)\right] \\
&= c^{-1}\mathbf{I} - c^{-1}(c\mathbf{M}_1 + \mathbf{I})^{-1} + O(\epsilon) \\
&= c^{-1}(c\mathbf{M}_1 + \mathbf{I})^{-1}(c\mathbf{M}_1 + \mathbf{I}) - c^{-1}(c\mathbf{M}_1 + \mathbf{I})^{-1} + O(\epsilon) \\
&= c^{-1}(c\mathbf{M}_1 + \mathbf{I})^{-1}c\mathbf{M}_1 + O(\epsilon) \\
&= (c\mathbf{I} + \mathbf{M}_1^{-1})^{-1} + O(\epsilon)
\end{aligned}
$$

Therefore, The Gâteaux derivative of function ξ is

$$
\begin{aligned}
G_\xi\{\mathbf{M}_1, \mathbf{M}_2\} = \lim_{\{\epsilon\to 0^+\}} \Big\{ & Tr\Big[\mathbf{M}_2\mathbf{M}_1^{-1}\big[c\mathbf{I} + (\mathbf{M}_1 + \epsilon\mathbf{M}_2)^{-1}\big]^{-1}(c\mathbf{I} + \mathbf{M}_1^{-1})^{-1} \\
& (c\mathbf{I} + \mathbf{M}_1^{-1} + c\mathbf{I} + (\mathbf{M}_1 + \epsilon\mathbf{M}_2)^{-1})\big[c\mathbf{I} + (\mathbf{M}_1 + \epsilon\mathbf{M}_2)^{-1}\big]^{-1} \\
& (c\mathbf{I} + \mathbf{M}_1^{-1})^{-1}\mathbf{M}_1^{-1}(\mathbf{I} + \epsilon\mathbf{M}_2\mathbf{M}_1^{-1})^{-1}\Big]\Big\} \\
= \lim_{\{\epsilon\to 0^+\}} \Big\{ & Tr\Big[\mathbf{M}_2\mathbf{M}_1^{-1}\big[(c\mathbf{I} + \mathbf{M}_1^{-1})^{-1} + O(\epsilon)\big](c\mathbf{I} + \mathbf{M}_1^{-1})^{-1} \\
& (c\mathbf{I} + \mathbf{M}_1^{-1} + c\mathbf{I} + \mathbf{M}_1^{-1} + O(\epsilon))\big[(c\mathbf{I} + \mathbf{M}_1^{-1})^{-1} + O(\epsilon)\big] \\
& (c\mathbf{I} + \mathbf{M}_1^{-1})^{-1}(\mathbf{M}_1^{-1} + O(\epsilon))\Big]\Big\} \\
= \lim_{\{\epsilon\to 0^+\}} \Big\{ & Tr\big[\mathbf{M}_2\mathbf{M}_1^{-1}(c\mathbf{I} + \mathbf{M}_1^{-1})^{-1}(c\mathbf{I} + \mathbf{M}_1^{-1})^{-1}2(c\mathbf{I} + \mathbf{M}_1^{-1}) \\
& (c\mathbf{I} + \mathbf{M}_1^{-1})^{-1}(c\mathbf{I} + \mathbf{M}_1^{-1})^{-1}\mathbf{M}_1^{-1}\big] + O(\epsilon)\Big\} \\
= & 2Tr\left[\mathbf{M}_2(c\mathbf{M}_1 + \mathbf{I})^{-1}\mathbf{M}_1(c\mathbf{M}_1 + \mathbf{I})^{-2}\right]
\end{aligned}
$$

$$
\begin{aligned}
F_\xi\{\mathbf{M}_1, \mathbf{M}_2\} = & 2Tr\left[\mathbf{M}_2(c\mathbf{M}_1 + \mathbf{I})^{-1}\mathbf{M}_1(c\mathbf{M}_1 + \mathbf{I})^{-2}\right] \\
& - 2Tr\left[\mathbf{M}_1(c\mathbf{M}_1 + \mathbf{I})^{-1}\mathbf{M}_1(c\mathbf{M}_1 + \mathbf{I})^{-2}\right],
\end{aligned}
$$

where the $F_\xi\{\mathbf{M}_1, \mathbf{M}_2\}$ is the Fréchet derivative such that

$$F_\xi\{\mathbf{M}_1, \mathbf{M}_2\} = G_\xi\{\mathbf{M}_1, \mathbf{M}_2 - \mathbf{M}_1\}.$$

The Gâteaux derivative is linear in $\mathbf{M}_2$, and since only η for which $\mathbf{M}(\eta)$ is non-singular can be optimal in this case, the Fréchet derivative is differentiable at $\mathbf{M}_1$. Therefore, we can apply Theorem 3.7 in Silvey (1980) to get the sufficient and necessary condition for the maximization of ξ:

$$\mathbf{x}'(c\mathbf{M} + \mathbf{I})^{-1}\mathbf{M}(c\mathbf{M} + \mathbf{I})^{-2}x \leq Tr\left[\mathbf{M}(c\mathbf{M} + \mathbf{I})^{-1}\mathbf{M}(c\mathbf{M} + \mathbf{I})^{-2}\right], \forall \mathbf{x} \in \mathcal{X}. \quad (3.6)$$

□

Theorem 6. *Let η be a design measure in the class of probability distributions H on the Borel sets of $\mathcal{X}$ where $\mathcal{X}$ is a compact subspace of $\mathcal{R}^p$ defined in (2.14). When $\mathbf{\Lambda} = \lambda_0^2\mathbf{I}$, any design η that satisfies $\mathbf{M}(\eta) = \mathbf{I}$ is optimal under the criterion in (3.4) for the estimation of $(\boldsymbol{\theta},\ \mathbf{\Lambda})$.*

Proof. The maximization of the first term of (3.4) is a special case of Scenario I investigated in Section 2.2.2. Theorem 2 shows that the maximization is achieved by the design η^* that satisfies $\mathbf{M}(\eta^*) = \mathbf{I}_p$. Now if we can prove that such a design also maximizes the second term of (3.4) which is the function ξ, then the theorem is proved. To prove this, we examine the sufficient and necessary condition (3.6) in Theorem 5. With $\mathbf{M} = \mathbf{I}$, the left side of (3.6) becomes $\frac{1}{(c+1)^3}\mathbf{x}'\mathbf{x}$, and the right hand side is $\frac{p}{(c+1)^3}$. Since $\mathbf{x}'\mathbf{x} \leq p$ from (2.15), the condition (3.6) holds and this proves the theorem. □

Remark: Note that for an m-observation design, the design measure η may not yield integer number of observations on the design points to form an exact design.

However, for certain combination of m, the number of treatment factors and factor levels, in the standardized orthogonal coding, level-balanced orthogonal designs $\mathbf{X}$ satisfies $\mathbf{X}'\mathbf{X} = m\mathbf{I}$ for the main-effects model, and equivalently $\mathbf{M}(\eta) = \mathbf{I}$. Therefore, level-balanced orthogonal designs are optimal, if they exist, for the estimation of hyperparameters $\boldsymbol{\theta}$ in the above scenario.

3.2 Scenario II: Random Effects are Equi-correlated and With Equal Variances

In this scenario, the random individual-level main effects $\boldsymbol{\beta}_i$ in (1.1) are assumed to be equally correlated and with equal variances, i.e., $\boldsymbol{\Lambda} = a^*\mathbf{I}_p + d^*\mathbf{J}_p$. The Fisher Information on $\boldsymbol{\Lambda}$ is then

$$FI(\boldsymbol{\Lambda},\ \boldsymbol{\Lambda}) = \begin{pmatrix} FI(a^*, a^*) & FI(a^*, d^*) \\ FI(a^*, d^*) & FI(d^*, d^*) \end{pmatrix},$$

where, by (3.2),

$$\begin{aligned}
FI(a,a) &= \frac{n}{2} Tr\left[(\sigma^2\mathbf{I}_m + \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}')^{-1}\mathbf{X}\mathbf{X}'(\sigma^2\mathbf{I}_m + \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}')^{-1}\mathbf{X}\mathbf{X}'\right] \\
&= \frac{n}{2} Tr\left\{[\mathbf{X}'(\sigma^2\mathbf{I}_m + \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}')^{-1}\mathbf{X}]^2\right\} \\
FI(a,b) &= \frac{n}{2} Tr\left[(\sigma^2\mathbf{I}_m + \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}')^{-1}\mathbf{X}\mathbf{X}'(\sigma^2\mathbf{I}_m + \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}')^{-1}\mathbf{X}\mathbf{J}\mathbf{X}'\right] \\
&= \frac{n}{2}\left\{\mathbf{1}'[\mathbf{X}'(\sigma^2\mathbf{I}_m + \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}')^{-1}\mathbf{X}]^2\mathbf{1}\right\} \\
FI(b,b) &= \frac{n}{2} Tr\left[(\sigma^2\mathbf{I}_m + \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}')^{-1}\mathbf{X}\mathbf{J}\mathbf{X}'(\sigma^2\mathbf{I}_m + \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}')^{-1}\mathbf{X}\mathbf{J}\mathbf{X}'\right] \\
&= \frac{n}{2}[\mathbf{1}'\mathbf{X}'(\sigma^2\mathbf{I}_m + \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}')^{-1}\mathbf{X}\mathbf{1}]^2
\end{aligned}$$

since

$$|FI(\boldsymbol{\Lambda}, \boldsymbol{\Lambda})| = FI(a,a)FI(b,b) - FI(a,b)^2$$

To maximize the design criterion $\mathcal{Q}$ in (3.1) is to maximize

$$\frac{|\mathbf{X}'\mathbf{X}|}{|\sigma^2\mathbf{I}_p + \mathbf{\Lambda}\mathbf{X}'\mathbf{X}|}\left\{Tr(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^2(\mathbf{1}'\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{1})^2 - [\mathbf{1}'(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^2\mathbf{1}]^2\right\}, \tag{3.7}$$

where $\mathbf{\Sigma} = \sigma^2\mathbf{I}_m + \mathbf{X}\mathbf{\Lambda}\mathbf{X}'$, and $\mathbf{\Lambda} = a^*\mathbf{I}_p + d^*\mathbf{J}_p$.

The investigation of optimal designs under the criterion (3.7) is one of the future research topics which will be discussed in Chapter 5.

CHAPTER 4

A Practical Application in Market Research

4.1 The "Level Effect" in Conjoint Studies

A practical use of the optimal design theory developed in Chapter 2 is in the learning of the "level effect" in conjoint studies. Introduced in the early 1970s, Conjoint Studies are widely used in market research to study consumers' preferences of products or services (also called "Conjoint Analysis"). Features that characterize a product are called product attributes; for example, credit limit, annual fee and annual interest rate are the attributes of a credit card product. Each attribute has various manifestations which are called levels; for example, $8.99\%, 9.99\%, 17.99\%$ are three possible levels of the annual interest rate of a credit card product.

In conjoint studies, respondents are usually asked to evaluate a number of product concepts (also called "profiles"). These product concepts are created by factorial combinations of attribute levels. The responses are used to estimate the "part-worths" of the attribute levels, which explain the contributions of the attribute levels to the overall profile evaluation. By setting one level within each attribute as the reference level (usually the lowest level), the part-worth of a certain level of a certain attribute

measures the influence of that attribute at that specific level relative to the reference level. The estimated difference between the highest-level part-worth and the lowest-level part-worth of an attribute is defined as the derived attribute importance (see, for example, Wittink et al., 1982, 1990, 1997). In other words, the part-worths are the dummy-coded contrasts of the attribute levels as discussed in Example 2.1 in Section 2.2.1.

There are three kinds of measures typically used in conjoint studies: Rating (scores), Ranking and Choice. When the rating scale is used, respondents are asked to express their preferences on a scale that indicates the intensity of the preference. For example, 10 points for the most preferred, and 1 point for the least preferred. Ranking, on the other hand, only asks for the preference order of products and does not allow a statement on preference intensity. When the Choice measure is used, respondents are asked to select their most preferred product out of all the alternatives. For the model discussed in this paper, we focus on the rating measure.

It was observed first by Currim, Weinberg and Wittink (1981) that the derived attribute importance increases with the addition of intermediate attribute levels in monotonic attributes (such as the interest rate of a credit card product). This intriguing phenomenon was named the number-of-attribute levels (NOL) effect, also known as the "level effect". Since then the level effect has been observed across different measurement scales, various data collection methods and estimation techniques (see Wittink et al., 1982, 1990, 1997; Steenkamp and Wittink, 1994; and Creyer and Ross, 1988). In a review article of conjoint analysis, Green and Srinivasan (1990) include

the level effect as one of the topics that call for further investigation. Note that the level effect examined in the above references is on monotonic attributes only. It does not apply to an attribute in which an ideal-point preference exists, such as the temperature of a room. On nominal attributes such as the color of a car, it is not clear if the level effect exists, i.e., if increasing choices of colors on cars will make people more sensitive to the color of the car when making purchasing decisions.

Possible explanations of the level effect have been proposed in the literature. Currim, Weinberg and Wittink (1981) find that the ordinal properties of ranking and rating measures contribute to the occurrence of the level effect. Wittink et al. (1990), Steenkamp and Wittink (1994), Verlegh, Schifferstein and Wittink (2002) find that certain experimental design methods such as the utility-balance approach in ACA (Adaptive Conjoint Analysis), and small sampling errors seem to reduce the magnitude of the level effect. Steenkamp and Wittink (1994) also investigate the attention-based explanation which states that the respondents' attention to an attribute may increase with the addition of attribute levels. However, only minimal support of the attention-based explanation has been found in the empirical studies. Most recently, Verlegh, Schifferstein and Wittink (2002) suggest that the level effect is most likely due to respondents' tendencies to uniformly distribute their responses over the corresponding continuum of the measurement scale.

The existence of the level effect suggests that consumer preference varies with context. In particular, it implies that, in the market place, consumer preference sensitivity to a monotonic product attribute is affected by the variety of attribute

levels displayed, which varies from store to store. Such a finding is consistent with a large body of consumer behavior literature (e.g., Huber, Payne and Puto 1982, Lynch, Chakravarti and Mitra 1991, Simonson and Tversky 1992) describing the influence of contextual effects in judgment and choice. Standard conjoint analysis models in marketing research do not adjust for the varieties offered and, consequently, cannot predict consumer buying behavior well in different contexts. In this chapter the level effect is modeled by incorporating a number of ideas from the Psychology literature. A summary of these ideas is given in Section 4.2, and then, based on these ideas, a hierarchical linear model is proposed in Section 4.3. To evaluate the effectiveness of the proposed model and the effect of different survey designs, a national web-based survey study was conducted on credit card products. The optimal experimental design theory developed in Chapter 2 is applied to survey designs, see Section 4.4. The data were collected through Harris Interactive, a global market research firm. In Section 4.5 the proposed model is fit to the data and consumer preference in a different context is predicted using validation data. The effect of different survey designs on prediction accuracy is examined. MCMC is used to derive the posterior estimates of the model parameters. Details of the algorithm is included in Appendix B.

4.2 The Psychological Process

Mellers and Cooke (1994) describe three stages of the psychological process involved in preference judgment of products: first, the *Perception Stage* where physical values of product attributes are converted to their corresponding subjective values (perceptual values); second, the *Integration Stage* where perceptual values of attributes are combined by means of a composition rule to form an overall impression

of the multi-attribute product; last, the *Decision Stage* where the overall impression is converted through a judgment function (usually a linear function) to the final response, i.e., a rating, a choice or a rank.

To illustrate this with an example from Mellers and Cooke (1994), assume that a customer is evaluating cars based on price and gas mileage. At the perception stage, the physical values ϕ_{price} and ϕ_{gas} of a car are transformed to the corresponding perceptual values ψ_{price} and ψ_{gas} through a function H. Then, at the integration stage, ψ_{price} and ψ_{gas} are combined by a function C to form an overall impression Ψ of the car. Last, the overall impression Ψ is transformed to the final response A via a function G.

$$\begin{array}{ccccc} \phi_{price} & \stackrel{H}{\rightarrow} & \psi_{price} & \stackrel{C}{\searrow} & \\ \phi_{gas} & \rightarrow & \psi_{gas} & \nearrow & \end{array} \Psi \stackrel{G}{\rightarrow} A$$

Suppose that a customer is presented with a car priced at $\$\ 20,000$ and gas efficiency of 23 miles per gallon. At the first stage, the physical values of $\$\ 20,000$ and 23 mpg are transformed into perceptual values. The consumer may perceive the price of $\$\ 20,000$ as a "good deal" and the gas efficiency of 23 mpg as "average". So on a preference scale of 1 to 10 with 1 being the least attractive and 10 being the most attractive, the price of the $\$\ 20,000$ may be perceived by this consumer as a 9, and the gas efficiency of 23 mpg as a 5. Then the consumer weighs how important the two features (price and gas efficiency) are and makes a decision. If price is given a much greater weight than the gas efficiency then the car will have a quite high rating on the scale of attractiveness and he or she may end up buying the car. On the other hand, if the consumer thinks gas efficiency is far more important, then the car may

not be appealing and he or she may want to look elsewhere.

When modeling the process with rating data, the function G can, in practice, be set as the identity function, i.e., $\Psi = A$, because the rating, A, is closely related to the overall impression, Ψ. Choosing G to be the identity function may not be appropriate in other response contexts, e.g., choice. In addition, if we consider it appropriate to assume the additive structure of function C, as is present in a main-effects model, then $\Psi = \psi_{price} + \psi_{gas}$ and the source of the level-effect is assumed to be from the function H at the first stage – the Perception Stage.

It has been recognized in the psychology literature that judgment is relative to the context (see, for example, Parducci, 1965, 1974, 1982; Krumhansl, 1978; Mellers and Birnbaum, 1982; Mellers and Cooke 1994, 1996). These contextual factors include the range of an attribute, the number of attribute levels, and the order of the profile presentation. Parducci (1965) proposes a theory on how the range of the attribute and the number of levels affect respondents' evaluation of the attribute levels. The theory posits that the evaluation is a compromise between the range principle and the frequency principle. The range principle describes the tendency of a respondent to map the range of the attribute linearly onto the range of the response scale. The frequency principle describes the tendency to use the categories along the response scale with equal frequency. To illustrate the principles with an example, suppose respondents are presented with five possible car prices:

$$\$20,000 \quad \$21,000 \quad \$26,000 \quad \$28,000 \quad \$30,000$$

On a scale of $0 - 10$, under the Range Principle, perceptual attractiveness values of the prices are:

$$10 \quad 9 \quad 4 \quad 2 \quad 0$$

and under the Frequency Principle, perceptual values are :

$$10 \quad 7.5 \quad 5 \quad 2.5 \quad 0$$

Cooke et al. (2004) formalize Parducci's range-frequency theory into the following model (in the case of one respondent and a single attribute):

$$A_g = b + mJ_g \tag{4.1}$$

$$J_g = \omega R_g + (1 - \omega)F_g \tag{4.2}$$

$$R_g = (S_g - S_{min})/(S_{max} - S_{min}) \tag{4.3}$$

$$F_g = (Rank(g) - 1)/(N - 1) \tag{4.4}$$

where N is the total number of levels in the attribute, and $Rank(g)$ is the rank of the attribute level g which is 1 if g is the lowest level of the attribute and N if g is the highest level of the attribute. F_g is the relative rank of the level g of the attribute $(0 \leq F_g \leq 1)$. S_g, S_{min}, S_{max} are the physical numerical value of the attribute at level g, the minimal level and the maximal level respectively, and R_g is the relative location of attribute level g compared with the whole range $(0 \leq R_g \leq 1)$. ω is the weight $(0 \leq \omega \leq 1)$ that designates the relative contribution of the range and the frequency information to the normalized perceptual value J_g of the attribute level g, and A_g is the perceptual value of the attribute at level g. The constant b captures the perceptual value of the lowest level of the attribute (i.e., when $J_g = 0$). When the part-worth of the lowest level is set to 0, the constant $b = 0$. The coefficient m

captures the perceived distance between the highest level and the lowest level of the attribute.

4.3 Modeling the Level Effect

4.3.1 Single-attribute Case

The deterministic model proposed by Cooke et al. (2004), (4.1) to (4.4), is extended by adding a random error term reflecting possible deviations from the expected perceptual value $(b + mJ_g)$. Thus a respondent's perceptual attractiveness of an attribute at level g is modeled as:

$$A_g = b + mJ_g + \epsilon_g \tag{4.5}$$

Combining (4.5) and (4.2), we obtain:

$$A_g = b + mF_g + m\omega(R_g - F_g) + \epsilon_g, \tag{4.6}$$

where ϵ_g is an error term distributed according to a Normal$(0, \sigma^2)$ distribution.

The perceived distance m between the lowest level and the highest levels is modeled according to the Distance-Density model in Krumhansl (1978) which states that "...two points in a relatively dense region of a stimulus space would have a smaller similarity measure than two points of equal interpoint distance but located in a less dense region of the space...". Specifically, m is assumed to be a linear function of the number of attribute levels L:

$$m = \kappa + \nu L, \tag{4.7}$$

and the final form of the model is:

$$A_g = b + \kappa F_g + \nu L F_g + \kappa\omega(R_g - F_g) + \nu\omega L(R_g - F_g) + \epsilon_g. \tag{4.8}$$

Parameter ν measures the level effect. For ν to be estimable, we need to have multiple studies with different number of attribute levels L. The difference $R_g - F_g$ in equation (4.6) represents how far away the relative range is from the relative rank and provides a measure of the skewness of the distribution of the attribute levels. If an attribute has levels packed on the lower end, for example, $(10, 15, 20, 30, 35, 40, 70, 90, 100)$, then we will observe prevailingly $R_g - F_g < 0$. Similarly, if an attribute has levels that are packed up on the higher end, we will have prevailingly $R_g - F_g > 0$. If an attribute has equally-spaced levels or when an attribute has only 2 levels, then $R_g - F_g = 0$ for all g. Thus, for ω in equation (4.8) to be estimable, we need to have a study in which the attribute has more than two levels and the levels are not equally spaced.

4.3.2 General Expression in Matrix Forms

While the model is illustrated in the previous section for the case of a single respondent rating levels of a single attribute, let us now express the model in matrix form for the general case of multiple respondents rating products of multiple attributes in multiple conjoint studies. The model takes the form of a hierarchical linear model. It is based on the model for the single-attribute case in the previous section. An additive structure is assumed on the combination of the attributes, i.e., the model contains main effects only and no interaction terms. The first-level of the hierarchical model has the standard main-effects conjoint model structure:

$$\mathbf{y}_{i,s}|\boldsymbol{\beta}_{i,s}, \sigma^2 \sim \text{Normal}(\mathbf{X}_{i,s}\boldsymbol{\beta}_{i,s}, \sigma^2\mathbf{I}), \tag{4.9}$$

where s indexes the study ($s = 1, 2, \cdots, S$) where different studies have different numbers of attribute levels. $\mathbf{y}_{i,s}$ is the vector of profile ratings from respondent i in study s ($i = 1, \ldots, n$), $\boldsymbol{\beta}_{i,s}$ is the vector of attribute level part-worths, and $\mathbf{X}_{i,s}$ is

the model matrix. If all respondents in study s are asked to rate on the same set of profiles then the model matrix is the same across all respondents in that study, i.e. $\mathbf{X}_{i,s} = \mathbf{X}_s$. Note that (4.9) for a given study s is the same as (1.1), the first level of the hierarchical model introduced in Chapter 1.

In the second-level of the hierarchy, the part-worths $\boldsymbol{\beta}_{i,s}$ are modeled according to the Range-Frequency/Distance-Density theory as formalized in (4.8). Let the hyperparameter vector $\boldsymbol{\theta}$ be

$$\boldsymbol{\theta} = (\mu, \kappa_1, \nu_1, \kappa_1\omega_1, \nu_1\omega_1, \cdots, \kappa_\tau, \nu_\tau, \kappa_\tau\omega_\tau, \nu_\tau\omega_\tau \cdots, \kappa_T, \nu_T, \kappa_T\omega_T, \nu_T\omega_T)', \quad (4.10)$$

where τ indicates the τth attribute ($\tau = 1, \cdots, T$), and the set of $(\kappa_\tau, \nu_\tau, \omega_\tau)$ corresponds to the (κ, ν, ω) in (4.8) for the attribute τ. As mentioned under (4.8), ν_τ is only estimable when there are multiple studies with different numbers of attribute levels L_τ on attribute τ. When L_τ remains the same across the studies, ν_τ is not estimable and is omitted from the $\boldsymbol{\theta}$ vector together with $\nu_\tau\omega_\tau$ in (4.10). Similarly, if the attribute levels in an attribute τ are equally-spaced or only contain the two extreme levels in all of the studies s ($s = 1, \ldots, S$), then $R - F = 0$ in (4.8) for all levels of the attribute in all the studies, and ω_τ is not estimable and therefore the elements $\kappa_\tau\omega_\tau$ and $\nu_\tau\omega_\tau$ are taken out of the $\boldsymbol{\theta}$ vector in (4.10). The expression of the second-level of the hierarchy is:

$$\boldsymbol{\beta}_{i,s}|\boldsymbol{\theta}, \boldsymbol{\Lambda}_s \sim \text{Normal}(\mathbf{Z}_s\boldsymbol{\theta}, \boldsymbol{\Lambda}_s), \quad (4.11)$$

where the covariate matrix $\mathbf{Z}_s$ for study s is of a block diagonal structure

$$Z_s = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & Z_{1,s} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & Z_{\tau,s} & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & Z_{T,s} \end{pmatrix}. \tag{4.12}$$

When there are different numbers of attribute levels $L_{\tau,s}$ on attribute τ for $\tau = 1, \ldots, T$ and $s = 1, \ldots, S$, the block component $Z_{\tau,s}$ in matrix Z_s consists of F, $L_{\tau,s}F$, $R - F$ and $L_{\tau,s}(R - F)$ on the various levels of attribute τ, as required by Equation (4.8). When $L_{\tau,s} = L_\tau$ for every study s for an attribute τ, the level effect parameter ν_τ in (4.10) is not estimable and is taken out of the $\boldsymbol{\theta}$ vector together with $\nu_\tau\omega_\tau$ in (4.10), and therefore the block component $Z_{\tau,s}$ in matrix Z_s only consists of F and $R - F$ on the various levels of attribute τ, corresponding to the parameters κ_τ and ω_τ, as shown in (4.8) for the given attribute τ. In addition, when the levels of attribute τ are equally spaced or only contain the two extreme levels across all the studies s ($s = 1, \ldots, S$), then ω_τ is not estimable and therefore the block component $Z_{\tau,s}$ in matrix Z_s only consists of F on various levels of the attribute τ.

Note that the expression for $\boldsymbol{\theta}$ in (4.10) is for the situation when all of the T attributes are monotonic. When there are non-monotonic attributes such as nominal attributes, the level effect does not apply and the hyperparameters in vector $\boldsymbol{\theta}$ corresponding to these attributes are simply the mean of the individual-level effects $\boldsymbol{\beta}_{i,s}$ over the population of the respondents and the studies. Example 4.1 illustrates the model matrices $\mathbf{X}_{i,s}$, the hyperparameter vector $\boldsymbol{\theta}$ and covariate matrices $\mathbf{Z}_s$ in (4.9)

and (4.11) for an investigation of the level effect on credit card interest rate through two survey studies.

Example 4.1 Consider two studies on credit card products characterized by three product attributes: APR, card provider, and rewards. The levels of the attributes in the two studies are as follows: Respondents are randomly divided into each of the

Study 1	APR:	(8.99%, 17.99%)
	Provider:	(Capital One, Citibank)
	Reward:	(None, Cash, Travel)
Study 2	APR:	(8.99%, 9.99%, 15.99%, 17.99%)
	Provider:	(Capital One, Citibank)
	Reward:	(None, Cash, Travel)

Table 4.1: The 12-profile Credit Card Survey Studies in Example 4.1

studies. All respondents in Study 1 are asked to rate the 12 products on the left side of Table 4.2, and all respondents in Study 2 are asked to rate the 12 products on the right side of Table 4.2. The products appear in random order to each respondent. The model matrix $\mathbf{X}_{i,s}$ in (4.9) is the same for every respondent i ($i = 1, \ldots, n$), i.e., $\mathbf{X}_{i,s} = \mathbf{X}_s$ since every respondent in the same study s ($s = 1, 2$) is asked to rate the same 12 products. Corresponding to the parameter vector $\boldsymbol{\beta}_{i,s}$ of attribute level part-worths which are the dummy-coded main effects discussed in Example 2.1, the matrix $\mathbf{X}_s$ for each study s ($s = 1,\ 2$) is shown in (4.13).

The 12 Profiles in Study 1			The 12 Profiles in Study 2		
APR	Provider	Rewards	APR	Provider	Rewards
17.99%	Capital One	Travel	17.99%	Capital One	Cash
17.99%	Capital One	Cash	17.99%	Citibank	Travel
17.99%	Capital One	None	17.99%	Citibank	None
17.99%	Citibank	Travel	15.99%	Capital One	Travel
17.99%	Citibank	Cash	15.99%	Capital One	None
17.99%	Citibank	None	15.99%	Citibank	Cash
8.99%	Capital One	Travel	9.99%	Capital One	Travel
8.99%	Capital One	Cash	9.99%	Capital One	Cash
8.99%	Capital One	None	9.99%	Citibank	None
8.99%	Citibank	Travel	8.99%	Capital One	None
8.99%	Citibank	Cash	8.99%	Citibank	Travel
8.99%	Citibank	None	8.99%	Citibank	Cash

Table 4.2: The Credit Card Profiles for Survey Evaluations

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{X}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}. \tag{4.13}$$

The first column of the 12×5 (i.e., $m = 12$, $p_1 = 5$) matrix $\mathbf{X}_1$ in (4.13) corresponds to the mean parameter in $\boldsymbol{\beta}_{i,1}$ in (4.9). The second column corresponds to the

contrast between the 8.99% APR and the reference level 17.99%. The third column corresponds to the contrast between the two providers: Capital One vs. Citibank. The fourth and the fifth columns correspond to the effect of the cash reward in contrast to no reward, and the effect of the travel reward in contrast to no reward, respectively. Similarly, the first column of the 12×7 (i.e., $m = 12$, $p_2 = 7$) matrix $\mathbf{X}_2$ in (4.13) corresponds to the mean parameter in $\boldsymbol{\beta}_{i,2}$ in (4.9). The second, third and fourth columns correspond respectively to the contrast between the 15.99% APR and the reference level 17.99%, the contrast between the 9.99% and the reference level 17.99%, and the contrast between the 8.99% and the reference level 17.99%. The fifth column corresponds to the contrast between the two providers: Capital One vs. Citibank . The sixth and the seventh columns correspond to the effect of the cash reward in contrast to no reward, and the effect of the travel reward in contrast to no reward, respectively.

Now at the second-level of the hierarchy in (4.11), since the card provider and the rewards are both nominal attributes, the elements in the hyperparameter vector $\boldsymbol{\theta}$ corresponding to these attributes are simply the means of the individual-level effects $\boldsymbol{\beta}_{i,s}$ over the respondents and the studies. Therefore, the length of $\boldsymbol{\theta}$ in (4.11) is 8 (i.e., $q = 8$) with

$$\boldsymbol{\theta}' = (\mu, \kappa_A, \nu_A, \kappa_A\omega_A, \nu_A\omega_A, \theta_P, \theta_C, \theta_T), \tag{4.14}$$

where the $(\kappa_A, \nu_A, \omega_A)$ reflect the (κ, ν, ω) in (4.8) on the APR attribute, θ_P is the mean contrast between the two card providers, θ_C is the mean contrast between cash reward and no reward, and θ_T is the mean contrast between travel reward and no reward. The corresponding covariate matrix $\mathbf{Z}_s$ in (4.8) for each study is

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{Z}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.33 & 1.33 & -0.11 & -0.44 & 0 & 0 & 0 \\ 0 & 0.67 & 2.67 & 0.22 & 0.88 & 0 & 0 & 0 \\ 0 & 1 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4.15}$$

The first element of the 5×8 matrix $\mathbf{Z}_1$ corresponds to the mean parameter in $\boldsymbol{\theta}$ in (4.14). The block component $\mathbf{Z}_{APR,1}$ in (4.12) corresponding to the monotonic APR attribute is of size 1×4, which includes the elements from the 2nd column to the 5th column in the second row of $\mathbf{Z}_1$, i.e., $(1,\ 2,\ 0, 0)$. The first element of the block component $\mathbf{Z}_{APR,1}$ is the relative rank F in (4.8) of the 8.99% APR level as defined by (4.4) (in this example, the rank of the reference level 17.99% is set to 1, and it follows that the relative rank of 8.99% is $(2-1)/(2-1) = 1$). The second element is the number of APR levels in this study (i.e., 2) times the relative rank of the 8.99% APR level. Since the difference between the relative range R as defined by (4.3) of the 8.99% APR level is $(8.99 - 17.99)/(8.99 - 17.99) = 1$, the third and the fourth elements of $\mathbf{Z}_{APR,1}$ corresponding to the $R - F$ and $L(R - F)$ in (4.8) of the 8.99% APR are both 0. Then corresponding to the nominal attributes, card provider and rewards, are the three ones in the $(3,\ 6)$th position, the $(4,\ 7)$th, and the $(5,\ 8)$th position of $\mathbf{Z}_1$ for the mean contrast of Capital One vs. Citibank, cash reward vs. no reward, and travel reward vs. no reward.

Similarly, The first column of the 7×8 matrix $\mathbf{Z}_2$ corresponds to the mean parameter in $\boldsymbol{\theta}$ in (4.14). The block component $\mathbf{Z}_{APR,2}$ in (4.12) corresponding to the monotonic APR attribute is of size 3×4, which includes the elements from the 2nd column to the 5th column in the second to the fourth row of $\mathbf{Z}_2$, i.e.,

$$\begin{pmatrix} 0.33 & 1.33 & -0.11 & -0.44 \\ 0.67 & 2.67 & 0.22 & 0.88 \\ 1 & 4 & 0 & 0 \end{pmatrix} \tag{4.16}$$

The first column of the block component $\mathbf{Z}_{APR,2}$ in (4.16) contains the relative ranks F of the three APR levels 15.99%, 9.99%, and 8.99%, as defined in (4.4) (the rank of the reference level 17.99% is set to 1, and it follows that the relative ranks of the 15.99%, 9.99%, and 8.99% APR level are $(2-1)/(4-1) = 0.33$, $(3-1)/(4-1) = 0.67$, $(4-1)/(4-1) = 1$, respectively). The second column of $\mathbf{Z}_{APR,2}$ in (4.16) is obtained by multiplying the first column by the number of APR levels in this study (i.e., 4). Since the relative range of the 15.99%, 9.99%, and 8.99% APR level as defined in (4.3) is $(15.99 - 17.99)/(8.99 - 17.99) = 0.22$, $(9.99 - 17.99)/(8.99 - 17.99) = 0.89$, and $(8.99 - 17.99)/(8.99 - 17.99) = 1$ respectively, the difference between the relative range and the relative rank $R - F$ in (4.8) for the 15.99%, 9.99%, and 8.99% APR level is $0.22 - 0.33 = -0.11$, $0.89 - 0.67 = 0.22$, and $1 - 1 = 0$, respectively, as shown in the 3rd column of the block component $\mathbf{Z}_{APR,2}$ in (4.16). The fourth component of $\mathbf{Z}_{APR,2}$ is obtained by multiplying the third column by the number of APR levels in this study (i.e., 4). Then just as for $\mathbf{Z}_1$, corresponding to the nominal attributes, card provider and rewards, are the three ones in the (5, 6)th position, the (6, 7)th, and the (7, 8)th position of $\mathbf{Z}_2$ for the mean contrast of Capital One vs. Citibank, cash reward vs. no reward, and travel reward vs. no reward.

4.4 Survey Designs

To evaluate the effectiveness of the proposed model, survey studies were conducted through Harris Interactive, a global marketing research firm. Credit card products were used as stimuli in the surveys. Measurement of the level effect in credit-card preferences requires information from multiple surveys where the number of attribute-levels varies. The attributes used in the surveys are the APR (interest rate) of the card, the card provider and the reward program. Study 1 involved two levels of APR, study 2 had four levels of APR, and study 3 APR had three levels. Data from studies 1 and 2 were used to calibrate the model parameters, and study 3 was reserved for hold-out predictive testing. Each respondent in the survey is asked to evaluate on 12 credit card products on a 0 to 10 rating scale with 0 being least likely to apply and 10 being most likely to apply. Table 4.3 provides a list of attributes and attribute-levels. In Example 4.1 the expressions of the model matrices $\mathbf{X}_s$, $\mathbf{Z}_s$ for study 1 and 2, and the hyperparameter vector $\boldsymbol{\theta}$ in (4.11) are illustrated for a particular design of study 1 and study 2 of the survey.

To select the 12 credit card products to be included in each study of the survey, let us recap the optimal design theory developed in Chapter 2. Specifically, within the context of a hierarchical linear model, when the interest is in the estimation of hyperparameter $\boldsymbol{\theta}$, a good experimental design is one that minimizes the posterior variance of $\boldsymbol{\theta}$ for given $\boldsymbol{\Lambda}$ and σ^2, which corresponds to the maximization of (2.3). When there are s studies and n_s respondents in study s who rate on the same set of products within each study, and the response errors are homoscedastic within each study (i.e., $\sigma^2_{i,s} = \sigma^2_s$ for all respondent i in study s), the maximization of (2.3)

Study 1	APR:	(8.99%, 17.99%)
	Provider:	(Capital One, Citibank)
	Reward:	(None, Cash, Travel)
Study 2	APR:	(8.99%, 9.99%, 15.99%, 17.99%)
	Provider:	(Capital One, Citibank)
	Reward:	(None, Cash, Travel)
Study 3	APR:	(8.99%, 9.99%, 17.99%)
	Provider:	(Capital One, Citibank)
	Reward:	(None, Cash, Travel)

Table 4.3: Web-based 12-profile Credit Card Survey Studies

simplifies to the maximization of

$$\Big|\sum_{s=1}^{S}(n_s\mathbf{Z}_s'\mathbf{X}_s'\boldsymbol{\Sigma}_s^{-1}\mathbf{X}_s\mathbf{Z}_s)\Big|, \quad \text{where} \quad \boldsymbol{\Sigma}_s = \sigma_s^2\mathbf{I} + \mathbf{X}_s\boldsymbol{\Lambda}_s\mathbf{X}_s'. \tag{4.17}$$

Furthermore, if there is an equal number of respondents in each study, then the maximization of (4.17) is equivalent to the maximization of

$$\Big|\sum_{s=1}^{S}\mathbf{Z}_s'\mathbf{X}_s'\boldsymbol{\Sigma}_s^{-1}\mathbf{X}_s\mathbf{Z}_s\Big|, \quad \text{where} \quad \boldsymbol{\Sigma}_s = \sigma_s^2\mathbf{I} + \mathbf{X}_s\boldsymbol{\Lambda}_s\mathbf{X}_s'. \tag{4.18}$$

In Section 2.2, optimal designs are investigated under the ψ criterion in (2.6), which corresponds to a special case of the maximization of (4.18) when $\mathbf{Z}_s = \mathbf{I}$ and $\mathbf{X}_s = \mathbf{X}$. Findings in Section 2.2 suggest that when the random effects are correlated, optimal designs are nonorthogonal and unbalanced. A computer search algorithm was developed to find optimal exact designs under the ψ criterion, as shown in Example 2.2 and 2.3. In Section 2.3, optimal designs are investigated under the criterion in (2.23) when both the model matrix $\mathbf{X}$ and the covariate matrix $\mathbf{Z}_i$ can be controlled. A special

case considered in Section 2.3 is when the determination of $\mathbf{X}$ and $\mathbf{Z}_i$ are independent and $\mathbf{Z}_i$ can be expressed as $\mathbf{Z}_i = \mathbf{I}_p \otimes \mathbf{z}_i'$ (see the paragraph above expression (2.23)). In this case, the optimal determination of $\mathbf{X}$ and $\mathbf{Z}_i$ $(i = 1, \ldots, n)$ is achieved by the individual maximization of $|\sum_{i=1}^{n}(\mathbf{z}_i\mathbf{z}_i')|$ and $|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}|$, the latter of which has been discussed in Section 2.2. The maximization of $|\sum_{i=1}^{n}(\mathbf{z}_i\mathbf{z}_i')|$ is achieved through the D-optimal $\tilde{\mathbf{Z}}' = (\mathbf{z}_1,\ \mathbf{z}_2, \ldots, \mathbf{z}_n)$.

Compared to the ψ criterion (2.6) and the criterion in (2.23), the maximization of (4.18) for the estimation of $\boldsymbol{\theta}$ in (4.11) through the survey studies 1 and 2 has three additional levels of complexity: First, the matrices $\mathbf{X}_s$ and $\mathbf{Z}_s$ are different between study 1 and 2 as study 1 involved two levels of APR while study 2 involved 4 levels of APR. Second, the $\mathbf{X}_s$ and $\mathbf{Z}_s$ cannot be determined independently, since the entries in $\mathbf{Z}_s$ related to the APR attribute reflect the relative ranges and relative ranges of the APR levels as illustrated in Example 4.1. Third, the vector of unique parameters to be estimated in $\boldsymbol{\theta}$ in (4.14) is

$$\boldsymbol{\theta}^* = (\mu, \kappa_A, \nu_A, \omega_A, \theta_P, \theta_C, \theta_T)', \tag{4.19}$$

which is a nonlinear function of $\boldsymbol{\theta}$. According to (1.5.1) in Silvey (1980), the optimal design for estimation of $\boldsymbol{\theta}^*$ is the design that minimizes

$$\left| \left(\frac{\partial \boldsymbol{\theta}^*}{\partial \boldsymbol{\theta}} \right)' \left(\sum_{s=1}^{S} \mathbf{Z}_s' \mathbf{X}_s' \boldsymbol{\Sigma}_s^{-1} \mathbf{X}_s \mathbf{Z}_s \right)^{-1} \left(\frac{\partial \boldsymbol{\theta}^*}{\partial \boldsymbol{\theta}} \right) \right|,$$

or equivalently, the design that maximizes

$$\tilde{\Psi} = \frac{1}{\left| \left(\frac{\partial \boldsymbol{\theta}^*}{\partial \boldsymbol{\theta}} \right)' \left(\sum_{s=1}^{S} \mathbf{Z}_s' \mathbf{X}_s' \boldsymbol{\Sigma}_s^{-1} \mathbf{X}_s \mathbf{Z}_s \right)^{-1} \left(\frac{\partial \boldsymbol{\theta}^*}{\partial \boldsymbol{\theta}} \right) \right|}, \tag{4.20}$$

where

$$\frac{\partial \boldsymbol{\theta}^*}{\partial \boldsymbol{\theta}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1/\omega_A & 0 & 1/\kappa_A & 0 & 0 & 0 \\ 0 & 0 & 1/\omega_A & 1/\nu_A & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4.21}$$

To select the $\mathbf{X}_s$ and $\mathbf{Z}_s$ $(s = 1, 2)$ such that $\tilde{\Psi}$ in (4.20) is maximized, the following simplifications are taken:

(i) A full-factorial design as shown in $\mathbf{X}_1$ of (4.13) is used in Study 1 such that $\mathbf{X}_1$ is fixed in (4.20).

(ii) The APR levels in the studies are set such that the 8.99% and 17.99% APR are included in study 1, and 8.99%, 9.99%, 15.99% and 17.99% are included in study 2. Therefore, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are fixed, as shown in (4.15) in Example 4.1.

Let

$$\mathbf{R}_1 = \mathbf{Z}_1'\mathbf{X}_1'\boldsymbol{\Sigma}_1^{-1}\mathbf{X}_1\mathbf{Z}_1, \quad \text{where} \quad \boldsymbol{\Sigma}_1 = \sigma_1^2\mathbf{I}_m + \mathbf{X}_1\boldsymbol{\Lambda}_1\mathbf{X}_1'. \tag{4.22}$$

The problem of maximizing (4.20) now becomes the determination of $\mathbf{X}_2$ that maximizes

$$\tilde{\Psi} = \frac{1}{\left| \left(\frac{\partial \boldsymbol{\theta}^*}{\partial \boldsymbol{\theta}}\right)' (\mathbf{Z}_2'\mathbf{X}_2'\boldsymbol{\Sigma}_2^{-1}\mathbf{X}_2\mathbf{Z}_2 + \mathbf{R}_1)^{-1} \left(\frac{\partial \boldsymbol{\theta}^*}{\partial \boldsymbol{\theta}}\right) \right|} \tag{4.23}$$

for given $\boldsymbol{\Lambda}_s$ and σ_s^2 $(s = 1, 2)$. A computer-search algorithm outlined in Appendix A.2 is used to search for such $\mathbf{X}_2$. This algorithm is based on the simple exchange algorithm in A.1 used to search for optimal exact designs in Example 2.2 and 2.3 in Chapter 2. Following the recommendations in Example 2.2 and 2.3, the following

assumptions are made for the search:

$$\sigma_1^2 = \sigma_2^2 = 1, \quad \mathbf{\Lambda}_1 = \mathbf{I}_5 + 2\mathbf{J}_5, \quad \mathbf{\Lambda}_2 = \mathbf{I}_8 + 2\mathbf{J}_8$$

To evaluate $\partial\boldsymbol{\theta}^*/\partial\boldsymbol{\theta}$ in (4.20), assumptions need to be made on the values of κ_A, ν_A and ω_A. Different values of κ_A, ν_A and ω_A were tested. In particular, numbers generated from a Uniform(1, 5) distribution were used for κ_A, numbers from Uniform(0, 0.5) and Uniform(0, 1) distributions were used for ν_A and ω_A, respectively. It was found that the resulting design was robust to the different specifications of κ_A, ν_A and ω_A. Therefore, in the algorithm outlined in Appendix A.2, the following values are used $\kappa_A = 3$, $\nu_A = 0.2$, and $\omega_A = 0.5$.

Four versions of fractional factorial designs were selected from the algorithm and used in Study 2. As no theoretical result is available on the optimal design structure that maximizes $\tilde{\Psi}$ in (4.20), relative efficiencies (unnormalized over the number of parameters) were reported as defined by $\dfrac{\tilde{\Psi}(\mathbf{X}_{2a})}{\tilde{\Psi}(\mathbf{X}_{2b})}$ for the relative $\tilde{\Psi}$-efficiency between the two designs $\mathbf{X}_{2a}$ and $\mathbf{X}_{2b}$. As a comparison, unnormalized relative D-efficiencies between the two designs $\mathbf{X}_{2a}$ and $\mathbf{X}_{2b}$ were also reported as defined by $\frac{|\mathbf{X}_{2a}'\mathbf{X}_{2a}|}{|\mathbf{X}_{2b}'\mathbf{X}_{2b}|}$. As shown in Table 4.4, the four designs used in Study 2 were comparable to each other under the $\tilde{\Psi}$ criterion. Under the D-criterion, the first three designs were comparable, while the last design had only 93.75% efficiency relative to the first three.

Similarly, three versions of fractional factorial designs were selected from the algorithm and used in Study 3. As shown in Table 4.5, the three designs were comparable

Design I

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Rel. $\tilde{\Psi}$-eff. $= 1$
Rel. D-eff. $= 1$

Design II

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Rel. $\tilde{\Psi}$-eff. $=98.5\%$
Rel. D-eff. $= 1$

Design III

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Rel. $\tilde{\Psi}$-eff. $= 99.6$
Rel. D-eff. $= 1$

Design IV

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Rel. $\tilde{\Psi}$-eff. $=1.004$
Rel. D-eff. $= 93.75\%$

Table 4.4: Designs Used in Study 2

under the $\tilde{\Psi}$ criterion, and under the D-criterion, the first two designs were comparable while the last design had 94.2% efficiency relative to the first two.

Design I	Design II	Design III
$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$
Rel. $\tilde{\Psi}$-eff. $= 1$	Rel. $\tilde{\Psi}$-eff. $=99.4\%$	Rel. $\tilde{\Psi}$-eff. $=1$
Rel. D-eff. $= 1$	Rel. D-eff. $= 1$	Rel. D-eff. $= 94.2\%$

Table 4.5: Designs Used in Study 3

Studies 1 and 2 were used to estimate the $\boldsymbol{\theta}^*$ in (4.19), and the $\boldsymbol{\theta}^*$ estimates were used to predict the mean preferences (over the respondents) in study 3 (see Section 4.5).

The data were collected via national web-based surveys, through Harris Interactive, a global market research firm. The following screening rules were used during the data collection: a qualified respondent needs to be 18 years of age or older, live in U.S., and have a credit card issued under his/her own name. A total of 1000

respondents participated in the studies. Some respondents gave the same ratings to all the 12 profiles presented to them, some respondents chose one product and rated all the rest eleven products 0, and some respondents used the scale in a reverse order (i.e., 0 was used for most likely to apply, and 10 was used for least likely to apply). These responses were considered invalid and taken out of the data. The final data contained around 760 valid responses, averaging about 95 valid responses per design version of the study.

4.5 Model Fitting and Validation

The MCMC method is used to obtain the estimates of $\boldsymbol{\theta}^*$ in (4.19). A Metropolis-Hastings (M-H) algorithm was used within a Gibbs sampler to generate the posterior draws. Details of the algorithm are provided in Appendix B.

Using the responses from Study 1 and the responses from each design version of study 2, the proposed hierarchical model was fit to the data and the posterior estimates of $\boldsymbol{\theta}^*$ were obtained, as reported in Table 4.6 with the posterior means and the posterior standard deviations in parenthesis. In all four versions of the designs, the posterior probability of ν_A (which is the level effect of APR) greater than 0 is greater than or equal to 95%, confirming the existence of the level effect, i.e., the contrast between the 8.99% and the 17.99% APR increased with the addition of the two intermediate APR levels. Estimation results on other parameters were quite consistent in all occasions. For example, estimates on credit card provider (θ_P) suggest that respondents do not have a preference of one provider (Capital One) over another (Citibank); estimates on rewards (θ_C and θ_T) suggest that respondents prefer cash

$\boldsymbol{\theta}^*$	Design I	Design II	Design III	Design IV
μ	0.325	0.234	0.254	0.202
	(0.117)	(0.115)	(0.112)	(0.112)
κ_A	3.427	3.27	3.495	3.174
	(0.499)	(0.512)	(0.506)	(0.52)
ν_A	0.224	0.289	0.24	0.373
	(0.138)	(0.15)	(0.141)	(0.143)
ω_A	0.925	0.891	0.922	0.943
	(0.07)	(0.092)	(0.07)	(0.054)
θ_P	-0.048	0.061	-0.015	-0.126
	(0.123)	(0.125)	(0.129)	(0.12)
θ_C	1.595	1.705	1.414	1.684
	(0.18)	(0.175)	(0.164)	(0.172)
θ_T	0.529	0.508	0.681	0.631
	(0.163)	(0.163)	(0.151)	(0.153)

Table 4.6: Posterior Estimates of $\boldsymbol{\theta}^*$ From Study 1 and Each Design in Study 2

reward over none and travel over none.

Data from study 3 were used as validation data to check how the model predicts to a new context with a different number of APR levels. Estimates of $\boldsymbol{\theta}^*$ obtained in Table 4.6 from each design k ($k = 1, \ldots, 4$) of study 2 were used to predict mean ratings of the 12 profiles in each version j ($j = 1, 2, 3$) of the three designs used in study 3:

$$\hat{\tilde{\mathbf{y}}}_{3j,k} = \int \mathbf{X}_{3j}\mathbf{Z}_3\boldsymbol{\theta} \times \pi(\boldsymbol{\theta}, \boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \sigma_1^2, \sigma_2^2 | \{\mathbf{y}_{1i}, \mathbf{y}_{2ik}\}, \mathbf{X}_1, \mathbf{X}_{2k}, \mathbf{Z}_1, \mathbf{Z}_2)\ d\boldsymbol{\theta} d\boldsymbol{\Lambda}_1 d\boldsymbol{\Lambda}_2 d\sigma_1^2 d\sigma_2^2,$$

where $\mathbf{X}_{3j}$ is the model matrix in design version j ($j = 1, 2, 3$) of study 3 which involved three APR levels, as shown in Table 4.5. $\mathbf{X}_1$ is the model matrix in study 1, as shown in (4.13), and $\mathbf{X}_{2k}$ is the model matrix in design k ($k = 1, \ldots, 4$) of study

2, as shown in Table 4.4. $\mathbf{Z}_1$, $\mathbf{Z}_2$ and $\mathbf{Z}_3$ is the covariate matrix in study 1, 2 and 3, respectively. $\mathbf{y}_{1i}$ is the vector of responses from respondent i in study 1, and $\mathbf{y}_{2ik}$ is the vector of responses from respondent i in design k of study 2. The prediction results on the three versions of study 3 are provided in Table 4.7 to Table 4.9. For nearly all of the profiles, the actual mean ratings are within the 95% Bayesian credibility intervals of the predictions. Prediction accuracy is measured with expected squared error loss (MSE):

$$\begin{aligned}\mathrm{MSE}_{j,k} = \int & (\bar{\mathbf{y}}_{3j} - \mathbf{X}_{3j}\mathbf{Z}_3\boldsymbol{\theta})'(\bar{\mathbf{y}}_{3j} - \mathbf{X}_{3j}\mathbf{Z}_3\boldsymbol{\theta}) \\ & \times \pi(\boldsymbol{\theta}, \boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \sigma_1^2, \sigma_2^2 | \{\mathbf{y}_{1i}, \mathbf{y}_{2ik}\}, \mathbf{X}_1, \mathbf{X}_{2k})\ d\boldsymbol{\theta} d\boldsymbol{\Lambda}_1 d\boldsymbol{\Lambda}_2 d\sigma_1^2, d\sigma_2^2 \quad (4.24)\end{aligned}$$

where $\bar{\mathbf{y}}_{3j}$ is the actual mean ratings of the 12 profiles in version j ($j = 1, 2, 3$) of the designs used in study 3. Table 4.10 summarizes the prediction accuracy on study 3 from each of the four sets of $\boldsymbol{\theta}^*$ estimates in Table 4.6. The expected squared error losses of the predictions from the four different designs in study 2 are comparable to each other, with the second design being slightly worse than the others and the third design slightly better.

To compare the effect of the four different designs, in addition to the mean and standard error of squared error loss on predictions, two other measures are used: the determinant of the variance-covariance matrix of the posterior estimates of $\boldsymbol{\theta}^*$ normalized by the size of $\boldsymbol{\theta}^*$, and the average variance of the posterior estimates of $\boldsymbol{\theta}^*$ normalized by the size of $\boldsymbol{\theta}^*$, as shown in Table 4.11. On these two measures, the four designs are again comparable. The first and the fourth designs are slightly better than the others, and the second design slightly worse. Note that the fourth design is

comparable to the others under the $\tilde{\Psi}$ criterion but is only 93.75% efficient under the D-criterion, which suggests that the $\tilde{\Psi}$ efficiency of a design might be more consistent with the estimation efficiency of $\boldsymbol{\theta}^*$ and the predictive accuracy on the validation study than the D-efficiency of the design. However, since there is little difference in $\tilde{\Psi}$ efficiency in the four designs, it cannot be verified that the estimation efficiency and the predictive accuracy varies in the same direction of the $\tilde{\Psi}$ efficiency. As an attempt to test the hypothesis, four observations were randomly taken out of the four 12-run designs to produce a set of six 8-observation designs that vary on both the $\tilde{\Psi}$ and the D efficiencies (four out of the six 8-run designs were each from the four 12-run designs, the additional two were from another random deletion of four observations from Design I and Design III of the 12-run designs). Four observations were also taken out of the 12-run design in study 1 to keep the number of runs the same between study 1 and study 2. Given the 8-run design in study 1, the six 8-run designs in study 2 are each used to estimate $\boldsymbol{\theta}^*$ and predict the mean profile ratings in study 3. Measures on estimation efficiency and prediction accuracy are reported in Table 4.12. Although within each individual measure, there is inconsistency between the relative $\tilde{\Psi}$ efficiency and design performance, when all three measures are considered together, designs with higher $\tilde{\Psi}$ efficiency perform better than designs with lower $\tilde{\Psi}$ efficiency in general, while the same does not hold for D-efficiency. For example, if we compare Design III and Design V in Table 4.12 where Design III has a lower $\tilde{\Psi}$ efficiency but a higher D-efficiency than Design V, the performance measures on estimation and prediction accuracy all favor Design V with the higher $\tilde{\Psi}$ efficiency. If we instead compare Design III and Design VI, which has a lower $\tilde{\Psi}$ and a lower D efficiency relative to Design III, the measure on the determinant of the variance-covariance

matrix of the posterior estimates slightly favors Design VI with lower $\tilde{\Psi}$ efficiency, but both the average variance measure and the measure on predictive accuracy favor Design III with higher $\tilde{\Psi}$ efficiency. Therefore, there is evidence that the efficiency of estimation on $\boldsymbol{\theta}^*$ and the predictive accuracy of a design is, in general, consistent with the $\tilde{\Psi}$ efficiency of the design.

Prediction on Version 1 Design of Study 3				
Actual $\bar{\mathbf{y}}_3$	$\hat{\bar{\mathbf{y}}}_3$ from Design I	$\hat{\bar{\mathbf{y}}}_3$ from Design II	$\hat{\bar{\mathbf{y}}}_3$ from Design III	$\hat{\bar{\mathbf{y}}}_3$ from Design IV
1.287	1.885	2.001	1.653	1.756
	(0.224)	(0.219)	(0.197)	(0.196)
0.681	0.281	0.295	0.238	0.071
	(0.146)	(0.135)	(0.139)	(0.138)
1.011	0.863	0.743	0.935	0.828
	(0.177)	(0.171)	(0.156)	(0.163)
1.383	1.934	1.940	1.668	1.885
	(0.193)	(0.190)	(0.172)	(0.181)
4.043	4.326	4.307	4.541	4.437
	(0.227)	(0.240)	(0.219)	(0.209)
3.362	3.793	3.799	3.860	3.810
	(0.197)	(0.208)	(0.190)	(0.182)
5.000	5.447	5.443	5.290	5.624
	(0.220)	(0.228)	(0.207)	(0.197)
3.351	3.843	3.738	3.876	3.939
	(0.192)	(0.212)	(0.187)	(0.174)
5.181	4.898	4.940	5.134	5.004
	(0.214)	(0.235)	(0.204)	(0.209)
5.745	5.969	6.137	5.868	6.062
	(0.214)	(0.216)	(0.197)	(0.196)
5.330	4.948	4.879	5.150	5.133
	(0.218)	(0.236)	(0.207)	(0.206)
4.436	4.415	4.371	4.469	4.506
	(0.187)	(0.209)	(0.184)	(0.178)

Table 4.7: Predictions of Mean Profile Ratings in Study 3 Version 1

Prediction on Version 2 Design of Study 3				
Actual $\bar{\mathbf{y}}_3$	$\hat{\bar{\mathbf{y}}}_3$ from Design I	$\hat{\bar{\mathbf{y}}}_3$ from Design II	$\hat{\bar{\mathbf{y}}}_3$ from Design III	$\hat{\bar{\mathbf{y}}}_3$ from Design IV
0.979	0.814	0.804	0.919	0.698
	(0.193)	(0.190)	(0.180)	(0.180)
0.511	0.281	0.295	0.238	0.071
	(0.146)	(0.135)	(0.139)	(0.138)
1.043	1.934	1.940	1.668	1.885
	(0.193)	(0.190)	(0.172)	(0.181)
0.543	0.330	0.234	0.254	0.201
	(0.115)	(0.115)	(0.112)	(0.113)
5.532	5.397	5.504	5.275	5.495
	(0.235)	(0.240)	(0.216)	(0.203)
3.755	3.793	3.799	3.860	3.810
	(0.197)	(0.208)	(0.190)	(0.182)
4.426	4.376	4.246	4.557	4.566
	(0.227)	(0.240)	(0.213)	(0.204)
5.383	5.447	5.443	5.290	5.624
	(0.220)	(0.228)	(0.207)	(0.197)
5.277	4.898	4.940	5.134	5.004
	(0.214)	(0.235)	(0.204)	(0.209)
6.468	5.969	6.137	5.868	6.062
	(0.214)	(0.216)	(0.197)	(0.196)
5.277	4.948	4.879	5.150	5.133
	(0.218)	(0.236)	(0.207)	(0.206)
4.415	4.415	4.371	4.469	4.506
	(0.187)	(0.209)	(0.184)	(0.178)

Table 4.8: Predictions of Mean Profile Ratings in Study 3 Version 2

Prediction on Version 3 Design of Study 3				
Actual $\bar{\mathbf{y}}_3$	$\hat{\bar{\mathbf{y}}}_3$ from Design I	$\hat{\bar{\mathbf{y}}}_3$ from Design II	$\hat{\bar{\mathbf{y}}}_3$ from Design III	$\hat{\bar{\mathbf{y}}}_3$ from Design IV
1.354	1.885	2.001	1.653	1.756
	(0.224)	(0.219)	(0.197)	(0.196)
0.594	0.281	0.295	0.238	0.071
	(0.146)	(0.135)	(0.139)	(0.138)
1.177	0.863	0.743	0.935	0.828
	(0.177)	(0.171)	(0.156)	(0.163)
1.208	1.934	1.940	1.668	1.885
	(0.193)	(0.190)	(0.172)	(0.181)
0.583	0.330	0.234	0.254	0.201
	(0.115)	(0.115)	(0.112)	(0.113)
4.250	4.326	4.307	4.541	4.437
	(0.227)	(0.240)	(0.219)	(0.209)
3.490	3.793	3.799	3.860	3.810
	(0.197)	(0.208)	(0.190)	(0.182)
5.646	5.447	5.443	5.290	5.624
	(0.220)	(0.228)	(0.207)	(0.197)
5.135	4.898	4.940	5.134	5.004
	(0.214)	(0.235)	(0.204)	(0.209)
6.583	5.969	6.137	5.868	6.062
	(0.214)	(0.216)	(0.197)	(0.196)
5.427	4.948	4.879	5.150	5.133
	(0.218)	(0.236)	(0.207)	(0.206)
5.063	4.415	4.371	4.469	4.506
	(0.187)	(0.209)	(0.184)	(0.178)

Table 4.9: Predictions of Mean Profile Ratings in Study 3 Version 3

	Mean and standard deviation of squared error loss of predictions from			
Predictions on Study 3	Design I Rel. $\tilde{\Psi} = 1$ Rel. $D = 1$	Design II Rel. $\tilde{\Psi} = 98.5\%$ Rel. $D = 1$	Design III Rel. $\tilde{\Psi} = 99.6\%$ Rel. $D = 1$	Design IV Rel. $\tilde{\Psi} = 1.004$ Rel. $D = 93.75\%$
Version 1	0.189 (0.195)	0.210 (0.222)	0.138 (0.158)	0.199 (0.191)
Version 2	0.177 (0.316)	0.166 (0.314)	0.155 (0.243)	0.170 (0.281)
Version 3	0.226 (0.241)	0.251 (0.264)	0.191 (0.207)	0.199 (0.198)
All	0.198 (0.257)	0.209 (0.270)	0.161 (0.207)	0.189 (0.228)

Table 4.10: Mean and Std. Dev. of Squared Error Loss of Mean Rating Predictions

Performance Measure	Design I Rel. $\tilde{\Psi} = 1$ Rel. $D = 1$	Design II Rel. $\tilde{\Psi} = 98.5\%$ Rel. $D = 1$	Design III Rel. $\tilde{\Psi} = 99.6\%$ Rel. $D = 1$	Design IV Rel. $\tilde{\Psi} = 1.004$ Rel. $D = 93.75\%$
$\lvert Var(\boldsymbol{\theta}^*)\rvert^{1/q}$	0.013	0.017	0.014	0.013
$Tr(Var(\boldsymbol{\theta}^*))/q$	0.046	0.054	0.051	0.046
Pred. MSE Std. Dev.	0.198 (0.257)	0.209 (0.270)	0.161 (0.207)	0.189 (0.228)

Table 4.11: Comparison of the Four 12-Run Designs in Study 2

Performance Measure	Design I	Design II	Design III	Design IV	Design V	Design VI
	Relative $\tilde{\Psi} = 1$ $D = 1$	Relative $\tilde{\Psi} = 97\%$ $D = 67\%$	Relative $\tilde{\Psi} = 86\%$ $D = 67\%$	Relative $\tilde{\Psi} = 69\%$ $D = 17\%$	Relative $\tilde{\Psi} = 1.05$ $D = 56\%$	Relative $\tilde{\Psi} = 76\%$ $D = 17\%$
$\lvert Var(\boldsymbol{\theta}^*)\rvert^{1/q}$	0.017	0.019	0.023	0.034	0.021	0.021
$Tr(Var(\boldsymbol{\theta}^*))/q$	0.061	0.057	0.078	0.090	0.062	0.084
Pred. MSE Std. Dev.	0.220 (0.289)	0.186 (0.245)	0.231 (0.293)	0.268 (0.349)	0.21 (0.26)	0.25 (0.34)

Table 4.12: Comparison of Six 8-Run Designs in Study 2

CHAPTER 5

Summary and Future Research

When interest is in the estimation of hyperparameters in hierarchical models, classical fixed-effects optimal design criteria are no longer relevant. Within the context of a linear hierarchical model, an optimal design criterion has been derived for the estimation of hyperparameters in some general scenarios, and the optimal design structure has been analytically derived for each of these scenarios. It is shown that orthogonal designs are optimal only when the random effects are independently distributed. When the random effects are correlated, it is shown that orthogonal designs are not optimal. Interestingly, as proved in Theorem 3 for equi-correlated random-effects, the ψ-optimal design is nonorthogonal and, at the same time, unbalanced. This conclusion has significant implications in experimental studies. Experimenters have long favored the use of orthogonal designs in experimental studies since orthogonal designs produce uncorrelated parameter estimates. However, when the random effects are correlated as is usually seen in practical studies, unconventional nonorthogonal unbalanced designs outperform the conventional orthogonal designs for the estimation of the hyperparameters in hierarchical linear models. Examples have been given to compare the ψ-optimal designs to the traditional orthogonal designs and the efficiency of the orthogonal designs is 83.6% to 99.5% relative to the nonorthogonal

designs when the random effects are correlated, as seen in Example 2.2 and 2.3 .

The computer search algorithms outlined in Appendix A employ a simple exchange algorithm that separates the search for the addition and deletion of design points. For the search of the survey designs that maximize $\tilde{\Psi}$ in (4.20), a simplification approach was taken. Future work is being planned, jointly with Brad Jones from SAS Inc., to develop a computer search algorithm capable of searching for designs that maximize $\tilde{\Psi}$ for the general scenario. An improved exchange algorithm will be employed that considers the addition and the deletion of design points jointly.

An interesting direction for future research is to investigate optimal designs under alternative design criteria. For example, instead of the criterion to minimize the determinant of the posterior variance-covariance matrix of $\boldsymbol{\theta}|(\boldsymbol{\Lambda}, \sigma^2)$, as investigated in Chapter 2, it may be more desirable to minimize the determinant of the marginal posterior variance-covariance matrix of $\boldsymbol{\theta}$. Similarly, instead of the criterion to minimize the determinant of the posterior variance-covariance matrix of $(\boldsymbol{\theta}, \boldsymbol{\Lambda})|\sigma^2$, as investigated in Chapter 3 for a specific form of $\boldsymbol{\Lambda}$ where $\boldsymbol{\Lambda} = \lambda_0^2 \mathbf{I}_p$, it may be more desirable to minimize the determinant of the marginal posterior variance-covariance matrix of $(\boldsymbol{\theta}, \boldsymbol{\Lambda})$. The potential difficulty in the investigation of optimal designs under these two criteria is that these two criteria cannot be expressed in closed forms. Therefore, analytical results on the optimal design structures may not be achievable, and we need to resort to algorithmic approaches such as MCMC to search for optimal designs under these criteria. There are other design criteria that may be of interest for future research, such as a criterion that focuses on future predictions, or a criterion

that aims for efficient model discrimination.

Additional topics for future research include the extension of analytical results on optimal designs in Chapter 3 to the scenario where $\mathbf{\Lambda}$ is of a more general form, and the study of blocked designs in which different groups of subjects receive different designs including incomplete designs. Of special interest is the determination of optimal $\mathbf{X}_i$ and $\mathbf{Z}_i$ where the covariates in $\mathbf{Z}_i$ are block variables such as demographic region or education level. Furthermore, a natural extension of the current research is to the setting of hierarchical non-linear models.

The level effect is an example of a broad class of problems in Marketing and other disciplines that involve the learning of effect sizes and their drivers. Learning the level effect gives us ideas on how consumer sensitivity to a product attribute varies with the range and the number of the attribute levels, and allows us to predict to a new context. A hierarchical linear model has been proposed in this dissertation that models individual consumer behavior and, by incorporating a number of ideas from the Psychology literature, models the level effect through hyperparameters. Currently, the model that is formalized according to the Range-Frequency and Distance-Density theories in Psychology only include Main Effects and no interaction effects. One future research direction is to model the level effect in the presence of first-order interactions. In addition, while we have focused on rating data in this paper, the level effect is also present in Choice data, and another interesting research direction is to model the level effect in choice data.

APPENDIX A

Computer Search Design Algorithm

A.1 Algorithm Used in Example 2.2 and 2.3

The SAS code of the algorithm can be obtained from the website

`http://www.stat.ohio-state.edu/~amd/dissertations.html`

The algorithm contains the following steps:

1. Specify the number of factors, the number of levels within each factor, and the number of runs m for the design (denoted by "n_trials" in the algorithm).

2. Let $\sigma^2 = 1$, $a^* = 1$, and specify the value of d^*, such that the ratio of $d/a = d^*/a^* = d^*$, where $a = ma^*/\sigma^2$, $\quad d = md^*/\sigma^2$, and $\mathbf{\Lambda} = a^*\mathbf{I}_p + d^*\mathbf{J}_p$. (In the algorithm, a^* is denoted by aa where $aa = 1$, and d^* is denoted by bb.)

3. Specify the maximum number of iterations for the exchange.

4. Indicate if you would like to choose the initial design. If so, specify the factor level combinations in the initial design, such as the combination of $(1, 2)$ for the two binary factors. If not, specify a random seed and the number of runs in the

initial design so that an initial design will be randomly generated according to the information specified.

5. Create all candidate design points $\mathbf{x}$ (i.e., all possible factorial combinations) where the elements in $\mathbf{x}$ are the coefficients of the standardized orthogonal contrasts of the factor levels under the main-effects model.

6. Create the matrix $\mathbf{X}$ for the initial design under the standardized orthogonal main-effects coding scheme.

7. If the number of runs of the initial design is less than the number of runs m as specified in Step 1, sequentially add a design point to the initial design until the number of runs reaches m. The design point added each time is the one that adds the most value to $\mathcal{Q}_2 = \frac{|\mathbf{X}'\mathbf{X}|}{|\sigma^2\mathbf{I}_p+\mathbf{\Lambda}\mathbf{X}'\mathbf{X}|}$ in (2.5). Note that when there are n design points in the matrix $\mathbf{X}_n$, adding another design point $\mathbf{x}$ to the matrix leads to

$$\begin{aligned}\mathcal{Q}_2^+ &= \frac{|\mathbf{X}_n'\mathbf{X}_n + \mathbf{x}\mathbf{x}'|}{|\sigma^2\mathbf{I}_p + \mathbf{\Lambda}\mathbf{X}_n'\mathbf{X}_n + \mathbf{\Lambda}\mathbf{x}\mathbf{x}'|} \\ &= \frac{|\mathbf{X}_n'\mathbf{X}_n + \mathbf{x}\mathbf{x}'|}{|\sigma^2\mathbf{I}_p + a^*\mathbf{X}_n'\mathbf{X}_n + d^*\mathbf{J}_p\mathbf{X}_n'\mathbf{X}_n + d^*\mathbf{J}_p\mathbf{x}\mathbf{x}' + a^*\mathbf{x}\mathbf{x}'|} \\ &\quad \text{since} \quad \mathbf{\Lambda} = a^*\mathbf{I}_p + d^*\mathbf{J}_p\end{aligned}$$

According to Theorem 8.9.3 in Graybill (1983),

$$|\mathbf{X}_n'\mathbf{X}_n + \mathbf{x}\mathbf{x}'| = |\mathbf{X}_n'\mathbf{X}_n| \left(1 + \mathbf{x}'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}\right),$$

$$\begin{aligned}|\sigma^2\mathbf{I}_p+a^*\mathbf{X}_n'\mathbf{X}_n+d^*\mathbf{J}_p\mathbf{X}_n'\mathbf{X}_n+d^*\mathbf{J}_p\mathbf{x}\mathbf{x}'+a^*\mathbf{x}\mathbf{x}'| &= |\sigma^2\mathbf{I}_p+a^*\mathbf{X}_n'\mathbf{X}_n+d^*\mathbf{J}_p\mathbf{X}_n'\mathbf{X}_n+d^*\mathbf{J}_p\mathbf{x}\mathbf{x}'| \\ &\quad \times \left(1 + a^*\mathbf{x}'(\sigma^2\mathbf{I}_p + a^*\mathbf{X}_n'\mathbf{X}_n + d^*\mathbf{J}_p\mathbf{X}_n'\mathbf{X}_n + d^*\mathbf{J}_p\mathbf{x}\mathbf{x}')^{-1}\mathbf{x}\right).\end{aligned}$$

Since $|\mathbf{X}_n'\mathbf{X}_n|$ remains the same regardless of which new design point is to be added, the design point that adds the most value to $\mathcal{Q}_2$ (i.e., maximizes $\mathcal{Q}_2^+$) is the design point $\mathbf{x}^+$ that maximizes

$$\frac{1+\mathbf{x}'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}}{1+a^*\mathbf{x}'(\sigma^2\mathbf{I}_p+a^*\mathbf{X}_n'\mathbf{X}_n+d^*\mathbf{J}_p\mathbf{X}_n'\mathbf{X}_n+d^*\mathbf{J}_p\mathbf{x}\mathbf{x}')^{-1}\mathbf{x}} \times \frac{1}{|\sigma^2\mathbf{I}_p+a^*\mathbf{X}_n'\mathbf{X}_n+d^*\mathbf{J}_p\mathbf{X}_n'\mathbf{X}_n+d^*\mathbf{J}_p\mathbf{x}\mathbf{x}'|}. \quad \text{(A.1)}$$

8. Start the exchange process. At each iteration, a new design point $\mathbf{x}_{m+1}^+$ maximizing (A.1) (for $n=m$) from the set of candidate design points is added to the matrix $\mathbf{X}_m$ which adds the most value to $\mathcal{Q}_2$, thus creating the new design matrix $\mathbf{X}_{m+1}$. Then a design point $\mathbf{x}_{m+1}^-$ is taken out of the design matrix $\mathbf{X}_{m+1}$ which subtracts the least value from $\mathcal{Q}_2$, thus creating an updated $\mathbf{X}_m$. After each exchange, the resulting $\mathcal{Q}_2$ is evaluated. The process continues until no further improvement of $\mathcal{Q}_2$ is possible, or when a maximum number of iterations as specified in Step 4 has been reached, whichever occurs first. Note that when the design matrix contains $m+1$ design points, i.e., $n=m+1$, taking a point $\mathbf{x}$ out of the matrix leads to

$$\begin{aligned}\mathcal{Q}_2^- &= \frac{|\mathbf{X}_n'\mathbf{X}_n-\mathbf{x}\mathbf{x}'|}{|\sigma^2\mathbf{I}_p+\mathbf{\Lambda}\mathbf{X}_n'\mathbf{X}_n-\mathbf{\Lambda}\mathbf{x}\mathbf{x}'|} \\ &= \frac{|\mathbf{X}_n'\mathbf{X}_n-\mathbf{x}\mathbf{x}'|}{|\sigma^2\mathbf{I}_p+a^*\mathbf{X}_n'\mathbf{X}_n+d^*\mathbf{J}_p\mathbf{X}_n'\mathbf{X}_n-d^*\mathbf{J}_p\mathbf{x}\mathbf{x}'-a^*\mathbf{x}\mathbf{x}'|} \\ &\quad \text{since} \quad \mathbf{\Lambda}=a^*\mathbf{I}_p+d^*\mathbf{J}_p\end{aligned}$$

Again, by Theorem 8.9.3 in Graybill (1983),

$$|\mathbf{X}_n'\mathbf{X}_n - \mathbf{x}\mathbf{x}'| = |\mathbf{X}_n'\mathbf{X}_n| \left(1 - \mathbf{x}'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}\right),$$

$$|\sigma^2\mathbf{I}_p + a^*\mathbf{X}_n'\mathbf{X}_n + d^*\mathbf{J}_p\mathbf{X}_n'\mathbf{X}_n - d^*\mathbf{J}_p\mathbf{x}\mathbf{x}' - a^*\mathbf{x}\mathbf{x}'| = |\sigma^2\mathbf{I}_p + a^*\mathbf{X}_n'\mathbf{X}_n + d^*\mathbf{J}_p\mathbf{X}_n'\mathbf{X}_n - d^*\mathbf{J}_p\mathbf{x}\mathbf{x}'|$$
$$\times \left(1 - a^*\mathbf{x}'(\sigma^2\mathbf{I}_p + a^*\mathbf{X}_n'\mathbf{X}_n + d^*\mathbf{J}_p\mathbf{X}_n'\mathbf{X}_n - d^*\mathbf{J}_p\mathbf{x}\mathbf{x}')^{-1}\mathbf{x}\right).$$

Since $|\mathbf{X}_n'\mathbf{X}_n|$ is the same regardless of which design point $\mathbf{x}$ is taken out, the design point that subtracts the least value from $\mathcal{Q}_2$ (i.e., maximizes $\mathcal{Q}_2^-$) is the design point $\mathbf{x}_n^-$ that maximizes

$$\frac{1 - \mathbf{x}'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}}{1 - a^*\mathbf{x}'(\sigma^2\mathbf{I}_p + a^*\mathbf{X}_n'\mathbf{X}_n + d^*\mathbf{J}_p\mathbf{X}_n'\mathbf{X}_n - d^*\mathbf{J}_p\mathbf{x}\mathbf{x}')^{-1}\mathbf{x}} \times \frac{1}{|\sigma^2\mathbf{I}_p + a^*\mathbf{X}_n'\mathbf{X}_n + d^*\mathbf{J}_p\mathbf{X}_n'\mathbf{X}_n - d^*\mathbf{J}_p\mathbf{x}\mathbf{x}'|}. \quad \text{(A.2)}$$

A.2 Algorithm Used for Survey Designs in Section 4.4

This algorithm is based on the simple exchange algorithm in A.1. The SAS code of the algorithm can be obtained from the website

`http://www.stat.ohio-state.edu/~amd/dissertations.html`

The algorithm contains the following steps:

1. Specify the number of factors, the number of levels within each factor, and the number of runs m for the design (denoted by "n_trials" in the algorithm).

2. Let $\sigma^2 = 1$, $a^* = 1$, and specify the value of d^*, such that the ratio of $d/a = d^*/a^* = d^*$, where $a = ma^*/\sigma^2$, $\quad d = md^*/\sigma^2$, and $\mathbf{\Lambda} = a^*\mathbf{I}_p + d^*\mathbf{J}_p$. (In the algorithm, a^* is denoted by aa where $aa = 1$, and d^* is denoted by bb.)

3. Specify the maximum number of iterations for the search.

4. Indicate if you would like to choose the initial design. If so, specify the factor level combinations in the initial design, such as the combinations of $(1, 2, 1)$ and $(4, 1, 3)$ for the 4-level APR, 2-level provider, and 3-level rewards factors. If not, specify a random seed and the number of runs in the initial design so that an initial design will be randomly generated according to the information specified.

5. Specify the $\mathbf{R}_1$ from (4.22), and the covariate matrix $\mathbf{Z}_2$ in Study 2. Note that in the algorithm, the matrices specified are under the standardized orthogonal effects coding scheme. The relationship of $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Lambda}}, \tilde{\mathbf{Z}}, \tilde{\mathbf{X}})$ under the standardized orthogonal effects coding scheme and $(\boldsymbol{\theta}, \boldsymbol{\Lambda}, \mathbf{Z}, \mathbf{X})$ under the dummy coding scheme is as follows:

$$\mathbf{X} = \tilde{\mathbf{X}}\mathbf{T},\ \mathbf{Z} = \mathbf{T}^{-1}\tilde{\mathbf{Z}},\ \boldsymbol{\theta} = \tilde{\boldsymbol{\theta}},\ \ \boldsymbol{\Lambda} = \mathbf{T}^{-1}\tilde{\boldsymbol{\Lambda}}\mathbf{T}^{-1\,\prime},$$

or equivalently,

$$\tilde{\mathbf{X}} = \mathbf{X}\mathbf{T}^{-1},\ \tilde{\mathbf{Z}} = \mathbf{T}\mathbf{Z},\ \tilde{\boldsymbol{\theta}} = \boldsymbol{\theta},\ \ \tilde{\boldsymbol{\Lambda}} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}'$$

where $\mathbf{T}$ is the $p \times p$ non-singular transformation matrix. The above relationship simply reflects the change of variables between the two coding schemes, as seen through

$$\tilde{\mathbf{X}}\tilde{\mathbf{Z}}\tilde{\boldsymbol{\theta}} = \tilde{\mathbf{X}}\mathbf{T}\mathbf{T}^{-1}\tilde{\mathbf{Z}}\tilde{\boldsymbol{\theta}} = \mathbf{X}\mathbf{Z}\boldsymbol{\theta},$$

$$\tilde{\mathbf{X}}\tilde{\boldsymbol{\Lambda}}\tilde{\mathbf{X}}' = \tilde{\mathbf{X}}\mathbf{T}\mathbf{T}^{-1}\tilde{\boldsymbol{\Lambda}}\mathbf{T}^{-1\,\prime}\mathbf{T}'\tilde{\mathbf{X}}' = \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}'.$$

6. Create all candidate design points $\mathbf{x}$ (i.e., all possible factorial combinations) where the elements in $\mathbf{x}$ are the coefficients of the standardized orthogonal contrasts of the factor levels under the main-effects model.

7. Create the matrix $\mathbf{X}$ for the initial design under the standardized orthogonal main-effects coding scheme.

8. If the number of runs of the initial design is less than the number of runs m as specified in Step 1, sequentially add a design point to the initial design until the number of runs reaches m. The design point added each time is the one that adds the most value to $\tilde{\Psi}$ in (4.23).

9. Start the exchange process. At each iteration, a new design point $\mathbf{x}_{m+1}^{+}$ from the set of candidate design points is added to the matrix $\mathbf{X}_m$ which adds the most value to $\tilde{\Psi}$, thus creating the new design matrix $\mathbf{X}_{m+1}$. Then a design point $\mathbf{x}_{m+1}^{-}$ is taken out of the design matrix $\mathbf{X}_{m+1}$ which subtracts the least value from $\tilde{\Psi}$, thus creating an updated $\mathbf{X}_m$. After each exchange, the resulting $\tilde{\Psi}$ is evaluated. The process continues until no further improvement of $\tilde{\Psi}$ is possible, or when a maximum number of iterations as specified in Step 4 has been reached, whichever occurs first.

APPENDIX B

MCMC Algorithm For Model Fitting

The following standard prior distribution assumptions (see, for example, Gamerman, 1997; Rossi, Allenby and McCulloch, 2005) are made on parameters ω_A, $\boldsymbol{\theta}^*_{-\omega_A} = (\mu, \kappa_A, \nu_A, \theta_P, \theta_C, \theta_T)'$, $\{\boldsymbol{\Lambda}_s\}$ and $\{\sigma_s^2\}$ $(s = 1,\ 2)$:

$$\begin{aligned}
\omega_A &\sim \text{Uniform}(0, 1), \\
\boldsymbol{\theta}^*_{-\omega_A} &\sim N_{q-1}(\mathbf{b}_0 = \mathbf{0}, \mathbf{D}_0 = 100\mathbf{I}_{q-1}), \\
\boldsymbol{\Lambda}_s &\sim IW(\eta_{0,s} = p_s + 3, \Delta_{0,s} = \eta_{0,s}\mathbf{I}_{p_s}), \\
\sigma_s^2 &\sim IG(w_0/2, W_{0,s}/2),
\end{aligned}$$

where $q = 7$, $p_1 = 5$, $p_2 = 7$, $w_0 = 3$, and $W_{0,s} = w_0 * \frac{\sum_{i=1}^{n_s}\sum_{j=1}^{m_s}(y_{ij} - \bar{y}_{..})^2}{m_s n_s - 1}$. The m_s represents the number of profiles and n_s represents the number of respondents in the study s, $(s = 1,\ 2)$. In the 12-profile credit card survey study 1 and 2, $m_1 = m_2 = 12$. A Metropolis-Hastings (M-H) algorithm is used within a Gibbs sampler (see Chib and Greenberg, 1995). The estimation algorithm proceeds by generating draws recursively from the following densities:

1. **Independently generate** $\{\boldsymbol{\beta}_{i,s}, i = 1, \cdots, n_s, s = 1, \cdots, S.\}$ from the following multivariate normal distribution:

$$[\boldsymbol{\beta}_{i,s}|\mathbf{y}_{i,s},\mathbf{X}_{i,s},\mathbf{Z}_s,\boldsymbol{\theta},\boldsymbol{\Lambda}_s,\sigma_s^2,] \propto [\mathbf{y}_{i,s}|\mathbf{X}_{i,s},\boldsymbol{\beta}_{i,s},\sigma_s^2][\boldsymbol{\beta}_{i,s}|\mathbf{Z}_s,\boldsymbol{\theta},\boldsymbol{\Lambda}_s] \propto \text{Normal}(\mathbf{b}_{i,s},\mathbf{D}_{i,s})$$

$$\mathbf{D}_{i,s} = (\sigma_s^{-2}\mathbf{X}_{i,s}'\mathbf{X}_{i,s} + \boldsymbol{\Lambda}_s^{-1})^{-1}, \quad \mathbf{b}_{i,s} = \mathbf{D}_{i,s}(\sigma_s^{-2}\mathbf{X}_{i,s}'\mathbf{y}_{i,s} + \boldsymbol{\Lambda}_s^{-1}\mathbf{Z}_s\boldsymbol{\theta})$$

2. Generate $\{\sigma_s^2, s = 1, \cdots, S.\}$ from the following Inverted chi-squared distribution

$$[\sigma_s^2|\{\mathbf{y}_{i,s},\mathbf{X}_{i,s},\boldsymbol{\beta}_{i,s}\}] \propto [\{\mathbf{y}_{i,s}\}|\{\mathbf{X}_{i,s},\boldsymbol{\beta}_{i,s}\},\sigma_s^2][\sigma_s^2] \propto \text{Inverted Chi-squared}(w_s, W_s)$$

$$w_s = w_{0,s} + m_s n_s, \; W_s = [W_{0,s}^{-1} + \sum_{i=1}^{n_s}(\mathbf{y}_{i,s} - \mathbf{X}_{i,s}\boldsymbol{\beta}_{i,s})'(\mathbf{y}_{i,s} - \mathbf{X}_{i,s}\boldsymbol{\beta}_{i,s})]^{-1}$$

3. Generate $\boldsymbol{\theta}^*_{-\omega_A}$

$$[\boldsymbol{\theta}^*_{-\omega_A}|\{\mathbf{Z}_s,\boldsymbol{\beta}_{i,s},\boldsymbol{\Lambda}_s\},\omega_A] \propto [\{\boldsymbol{\beta}_{i,s}\}|\{\mathbf{Z}_s,\boldsymbol{\Lambda}_s\},\boldsymbol{\theta}][\boldsymbol{\theta}^*_{-\omega_A}],$$

where $[\{\boldsymbol{\beta}_{i,s}\}|\{\mathbf{Z}_s,\boldsymbol{\Lambda}_s\},\boldsymbol{\theta}]$ is the product of multivariate normal densities $[\boldsymbol{\beta}_{i,s}|\mathbf{Z}_s,\boldsymbol{\Lambda}_s,\boldsymbol{\theta}]$ for $i = 1,\ldots,n_s$ and $s = 1,\ldots,S$. A random-walk M-H (see Chib and Greenberg, 1995, on a detailed discussion) is used to generate the draws of $\boldsymbol{\theta}^*_{-\omega_A}$. Let

$$\boldsymbol{\theta}^{*(k)}_{-\omega_A} = [\mu^{(k)}, \kappa_A^{(k)}, \nu_A^{(k)}, \theta_P^{(k)}, \theta_C^{(k)}, \theta_T^{(k)}]'$$

be the kth draw, the next draw is given by

$$\boldsymbol{\theta}^{*(k+1)}_{-\omega_A} = \boldsymbol{\theta}^{*(k)}_{-\omega_A} + \boldsymbol{\delta}$$

where $\boldsymbol{\delta}$ is a draw from the candidate generating density Normal$(0, 0.05\mathbf{I})$. Let

$$\boldsymbol{\theta}^{(k)} = [\mu^{(k)}, \kappa_A^{(k)}, \nu_A^{(k)}, \kappa_A^{(k)}\omega_A, \nu_A^{(k)}\omega_A, \theta_P^{(k)}, \theta_C^{(k)}, \theta_T^{(k)}]',$$

the probability of accepting the new draw $\boldsymbol{\theta}^{*(k+1)}_{-\omega_A}$ is given by

$$\min\left[\frac{[\{\boldsymbol{\beta}_{i,s}\}|\{\mathbf{Z}_s,\boldsymbol{\Lambda}_s\},\boldsymbol{\theta}^{(k+1)}][\boldsymbol{\theta}^{*(k+1)}_{-\omega_A}]}{[\{\boldsymbol{\beta}_{i,s}\}|\{\mathbf{Z}_s,\boldsymbol{\Lambda}_s\},\boldsymbol{\theta}^{(k)}][\boldsymbol{\theta}^{*(k)}_{-\omega_A}]},\ 1\right]$$

4. Generate ω_A

$$[\omega_A|\{\mathbf{Z}_s,\boldsymbol{\beta}_{i,s},\boldsymbol{\Lambda}_s\}] \propto [\{\boldsymbol{\beta}_{i,s}\}|\{\mathbf{Z}_s,\boldsymbol{\Lambda}_s\},\boldsymbol{\theta}][\omega_A],$$

where $[\{\boldsymbol{\beta}_{i,s}\}|\{\mathbf{Z}_s,\boldsymbol{\Lambda}_s\},\boldsymbol{\theta}]$ is the product of multivariate normal densities $[\boldsymbol{\beta}_{i,s}|\mathbf{Z}_s,\boldsymbol{\Lambda}_s,\boldsymbol{\theta}]$ for $i = 1,\ldots,n_s$ and $s = 1,\ldots,S$. An independence chain M-H is used to generate the draws of ω_A. Independently generate draws from the Uniform(0,1) distribution. Let $\omega_A^{(k)}$ be the kth draw, and $\omega_A^{(k+1)}$ be the next draw. Let

$$\boldsymbol{\theta}^{(k)} = [\mu,\kappa_A,\nu_A,\kappa_A\omega_A^{(k)},\nu_A\omega_A^{(k)},\theta_P,\theta_C,\theta_T]',$$

the probability of accepting the new draw $\omega_A^{(k+1)}$ is given by

$$\min\left[\frac{[\{\boldsymbol{\beta}_{i,s}\}|\{\mathbf{Z}_s,\boldsymbol{\Lambda}_s\},\boldsymbol{\theta}^{(k+1)}]}{[\{\boldsymbol{\beta}_{i,s}\}|\{\mathbf{Z}_s,\boldsymbol{\Lambda}_s\},\boldsymbol{\theta}^{(k)}]},\ 1\right]$$

5. Generate $\{\boldsymbol{\Lambda}_s, s = 1,\cdots,S.\}$ from Inverted Wishart Distribution

$$[\boldsymbol{\Lambda}_s|\{\boldsymbol{\beta}_{i,s}\},\mathbf{Z}_s,\boldsymbol{\theta}] \propto [\{\boldsymbol{\beta}_{i,s}\}|\mathbf{Z}_s,\boldsymbol{\theta},\boldsymbol{\Lambda}_s][\boldsymbol{\Lambda}_s] \sim \text{Inverted Wishart}(\eta_s,\Delta_s)$$

$$\eta_s = \eta_{0,s} + n_s,\ \Delta_s = [\Delta_{0,s}^{-1} + \sum_{i=1}^{n_s}(\boldsymbol{\beta}_{i,s} - \mathbf{Z}_s\boldsymbol{\theta})(\boldsymbol{\beta}_{i,s} - \mathbf{Z}_s\boldsymbol{\theta})']^{-1}$$

$300,000$ iterations are run where every 30th iteration is kept. Posterior means and variances are calculated using draws after the burn-in period of the first $60,000$ iterations. The convergence of the posterior draws can be seen in the time series plots shown in Figure B.1.

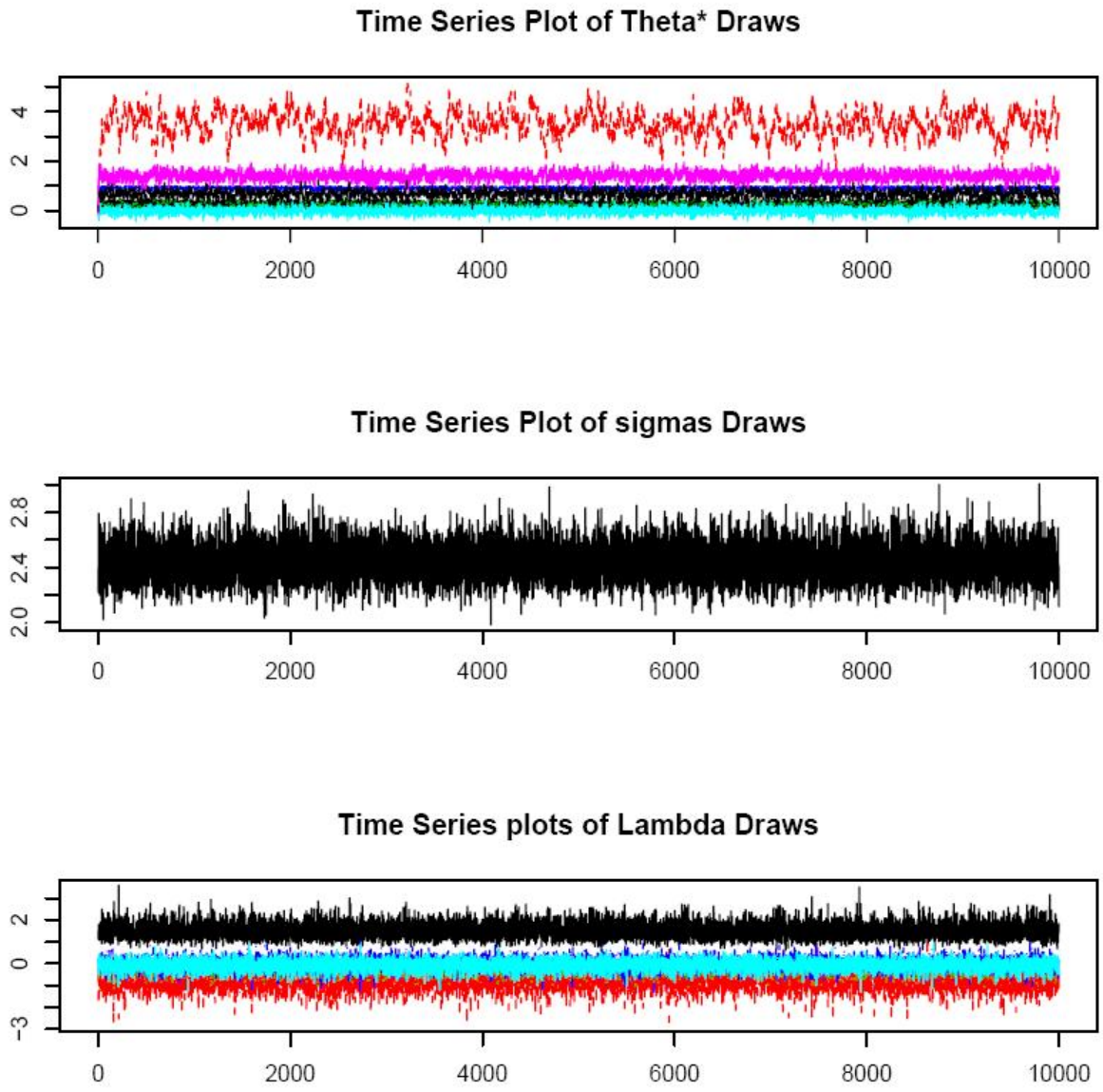


Figure B.1: Time Series Plots of Posterior Draws

BIBLIOGRAPHY

Allenby, Greg M. and Peter J. Lenk (1994), "Modeling Household Purchase Behavior with Logistic Normal Regression," *Journal of American Statistical Association*, 89, 1218–1229.

— (1995), "Reassessing Brand Loyalty, Price Sensitivity, and Merchandising Effects on Consumer Brand Choice," *Journal of Business and Economic Statistics*, 13, 281–298.

Arora, Neeraj and Joel Huber (2001), "Improving Parameter Estimates and Model Prediction by Aggregate Customization in Choice Experiments," *Journal of Consumer Research*, 28 (September), 273–283.

Atkinson, A.C. and A.N. Donev (1992), *Optimum Experimental Designs,*, Clarendon Press, Oxford.

Bradlow, Eric T. and Vithala R. Rao (2000), "A Hierarchical Bayes Model for Assortment Choice," *Journal of Marketing Research*, 37 (2), 259–268.

Chaloner, Kathryn and Isabella Verdinelli (1995), "Bayesian Experimental Design: A Review," *Statistical Science*, 10 (3), 273–304.

Chib, Siddhartha and Edward Greenberg (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49 (4), 327–335.

Cooke, Alan D.J., Chris Janiszewski, Marcus Cunha Jr., Suzanne A. Nasco, and Els De Wilde (2004), "Stimulus Context and the Formation of Consumer Ideals," *Journal of Consumer Research*, 31 (June), 112–124.

Creyer, Elizabeth and William T. Ross (1988), "The Effects of Range-Frequency manipulations on Conjoint Importance Weight Stability," *Advances in Consumer Research*, 15, 505–509.

Currim, Imran S., Charles B. Weinberg, and Dick.R. Wittink (1981), "Design of Subscription Programs for a Performing Arts Series," *Journal of Consumer Research*, 8 (June), 67–75.

Draper, Norman R. and Harry Smith (1998), *Applied Regression Analysis*, New York: Wiley.

Entholzner, Monika, Norbet Benda, Thomas Schmelter, and Rainer Schwabe (2006), "A Note on Designs for Estimating Population parameters," *Working paper*.

Fedorov, Valerii V. and Peter Hackl (1997), *Model-oriented design of experiments*, New York: Springer Verlag.

Gamerman, Dani (1997), *Markov Chain Monte Carlo*, Chapman & Hall.

Goldstein, Harvey (1995), *Multilevel Statistical Models*, New York : Halsted Press.

Graybill, Franklin A. (1983), *Matrices with Applications in Statistics*, Belmont, California: Wadsworth.

Green, Paul E. and V. Srinivasan (1990), "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice," *Journal of Marketing*, 54 (October), 3–19.

Han, Cong and Kathryn Chaloner (2004), "Bayesian Experimental Design for Nonlinear Mixed-effects Models with Applications to HIV Dynamics," *Biometrics*, 60, 25–33.

Hedayat, A.S., Neil J.A. Sloan and John Stufken (1999), *Orthogonal Arrays: Theory and Applications*, Springer-Verlag.

Huber, Joel, John W. Payne, and Christopher Puto (1983), "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis," *Journal of Consumer Research*, 9 (June), 90–98.

Huttenlocher, Janellen, Wendy Haight, Anthony Bryk, Michael Seltzer, and Thomas Lyons (1991), "Early Vocabulary Growth: Relation to Language Input and Gender," *Developmental Psychology*, 27 (2), 236–248.

Krumhansl, Carol L. (1978), "Concerning the Applicability of Geometric Models to Similarity Data: The Interrelationship Between Similarity and Spatial Density," *Psychological Review*, 85 (5), 445–463.

Kuhfeld, Warren F (2005), "Experimental Design, Efficiency, Coding and Choice Designs," Tech. rep., SAS TS-722C, `http://support.sas.com/techsup/tnote/tnote_stat.html`.

Kuhfeld, Warren F., Randall D. Tobias, and Mark Garratt (1994), "Efficient Experimental Design With Marketing Research Applications," *Journal of Marketing Research*, 31 (4), 545–557.

Lenk, Peter J., Wayne S. Desarbo, Paul E. Green, and Martin R. Young (1996), "Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs," *Marketing Science*, 15 (2), 173–191.

Lenk, Peter J. and Ambar G. Rao (1990), "New Models from Old: Forecasting Product Adoption by Hierarchical Bayes Procedures," *Marketing Science*, 9 (1), 42–53.

Lindley, D.V. and A.F.M. Smith (1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society Series B*, 34, 1–41.

Lynch, John G., Dipankar Chakravarti, and Anusree Mitra (1991), "Contrast Effects in Consumer Judgments: Changes in Mental Representations or in the Anchoring of Rating Scales?" *Journal of Consumer Research*, 18 (December), 284–297.

Magnus, Jan R. and Heinz Neudecker (1999), *Matrix differential calculus with applications in statistics and econometrics*, New York: John Wiley.

Manchanda, Puneet, Asim Ansari, and Sunil Gupta (1999), "The "Shopping Basket": A Model for Multicategory Purchase Incidence Decisions," *Marketing Science*, 18 (2), 95–114.

Mellers, Barbara A. and Michael H. Birnbaum (1982), "Loci of contextual effects in judgment," *Journal of Experimental Psychology: Human Perception and Performance*, 8, 582–601.

Mellers, Barbara A. and Alan D.J Cooke (1994), "Trade-offs Depend on Attribute Range," *Journal of Experimental Psychology: Human Perception and Performance*, 20 (5), 1055–1067.

Mellers, Barbara A. and Alan D.J. Cooke (1996), "The Role of Task and Context in Preference Measurement," *Psychological Science*, 7 (2), 76–82.

Mentre, France, Alain Mallet and Doha Baccar (1997), "Optimal Design in Random-Effects Regression Models," *Biometrika*, 84 (2), 429–442.

Merle, Yann and Michel Tod (2001), "Impact of Pharmacokinetic-Pharmacodynamic Model Linearization on the Accuracy of Population Information Matrix and Optimal Design," *Journal of Pharmacokinetics and Pharmacodynamics*, 28 (4), 363–388.

Montgomery, Alan L., Shibo Li, Kannan Srinivasan, and John C. Liechty (2004), "Modeling Online Browsing and Path Analysis Using Clickstream Data," *Marketing Science*, 23 (4), 579–595.

Morrison, Donald F. (1983), *Multivariate Statistical Methods,*, McGraw-Hill, Inc.

Parducci, Allen (1965), "Category Judgment: A range-frequency model," *Psychological Review*, 72, 407–418.

— (1974), "Contextual Effects: A Range-Frequency Analysis," in *Handbook of Perception*, eds. Edward Carterette and Morton Friedman, New York: Academic Press, 127-141.

— (1982), "Category Ratings: Still more contextual effects," in *Social Attitudes and Psychophysical Measurement*, eds. Wegener B., Hillsdale, NJ: Erlbaum, 89-105.

Raudenbush, Stephen W. (1993), "A Crossed Random Effects Model for Unbalanced Data with Applications in Cross-sectional and Longitudinal Research," *Journal of Educational Statistics*, 18, 321–349.

Raudenbush, Stephen W. and Anthony S. Bryk (2002), *Hierarchical Linear Models: Applications and Data Analsis Methods*, Sage Publications.

Rossi, Peter E. and Greg M. Allenby (2000), "Statistics and Marketing," *Journal of American Statistical Association*, 95, 635–638.

Rossi, Peter E., Greg M. Allenby, and Robert McCulloch (2005), *Bayesian Statistics and Marketing*, John Wiley and Sons, Ltd.

Sandor, Zsolt and Michel Wedel (2001), "Designing Conjoint Choice Experiments Using Managers Prior Beliefs," *Journal of Marketing Research*, 38 (4), 430–444.

— (2002), "Profile Construction in Experimental Choice Designs for Mixed Logit Models," *Marketing Science*, 21 (4), 455–475.

— (2005), "Differentiated Bayesian Conjoint Choice Designs," *Journal of Marketing Research*, 55 (2), 210–218.

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

Simonson, Itamar and Amos Tversky (1992), "Choice in Context: Tradeoff Contrast and Extremeness Aversion," *Journal of Marketing Research*, 29 (3), 281–295.

Smith, A.F.M. and Isabella Verdinelli (1980), "A Note on Bayesian Designs for Inference Using a Hierarchical Linear Model," *Biometrika*, 67, 613–619.

Steenkamp, Jan-Benedict E.M. and Dick R. Wittink (1994), "The Metric Quality of Full-Profile Judgments and the Number-of-attribute-levels Effect in Conjoint Analysis," *International Journal of Research in Marketing*, 11, 275–286.

Tod, Michel, France Mentre, Yann Merle, and Alain Mallet (1998), "Robust Optimal Design for the Estimation of Hyperparameters in Population Pharmacokinetics," *Journal of Pharmacokinetics and Biopharmaceuics*, 26, 689–716.

Verlegh, Peter W.J., Hendrik N.J. Schifferstein, and Dick R. Wittink (2002), "Range and Number-of-Levels Effects in Derived and Stated Measures of Attribute Importance," *Marketing Letters*, 13 (1), 41–52.

Wittink, Dick R., Joel Huber, Peter Zandan, and Richard M. Johnson (1992), "The Number of Levels Effect in Conjoint: Where Does It Come From, And Can It Be Eliminated?" in *Sawtooth Software Conference Proceedings*.

Wittink, Dick R., Lakshman Krishnamurthi, and Julia B. Nutter (1982), "Comparing Derived Importance Weights Across Attributes," *Journal of Consumer Research*, 8, 471–474.

Wittink, Dick R., Lakshman Krishnamurthi, and David J. Reibstein (1990), "The Effect of Differences in the Number of Attribute Levels on Conjoint Analysis," *Marketing Letters*, 1 (2), 113–123.

Wittink, Dick R., William G. McLauchlan, and P.B. Seetharaman (1997), "Solving the Number-of-Attribute-Levels Problem in Conjoint Analysis," in *Sawtooth Software Conference proceedings*.

Wu, Jeff C.F. and Michael Hamada (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York: J. Wiley.

Yuh, Lianng, Stuart Beal, Marie Davidian, Ferrin Harrisson, Allen Hester, Kenneth Kowalski, Edward Vonesh, and Russell Wolfinger (1994), "Population pharmacokinetics/Pharmacodynamics Methodology and Applications: A Bibliography," *Biometrics*, 50, 566–575.