BAYESIAN DESIGNS FOR HIERARCHICAL LINEAR MODELS

Qing Liu[*], Angela M. Dean[†] and Greg M. Allenby[‡]

[*]*Department of Marketing, University of Wisconsin-Madison*

[†]*Department of Statistics,The Ohio State University*

[‡]*Department of Marketing, The Ohio State University*

Abstract: Two Bayesian optimal design criteria for hierarchical linear models are discussed – the ψ_β criterion for the estimation of individual-level parameters $\boldsymbol{\beta}$, and the ψ_θ criterion for the estimation of hyperparameters $\boldsymbol{\theta}$. While the ψ_β criterion involves only the specification of the treatments, the ψ_θ criterion involves the specification of both the treatments and the covariates. We focus on a specific case in which all subjects receive the same set of treatments and the covariates are independent of treatments. We obtain the explicit structure of ψ_β- and ψ_θ- optimal continuous (approximate) designs for both the situation of independent random effects and some special situations of correlated random effects. Through examples and simulations we then compare ψ_β- and ψ_θ-optimal designs under more general scenarios of correlated random effects. While orthogonal designs are often ψ_β-optimal even when the random effects are correlated, ψ_θ-optimal designs tend to be nonorthogonal and unbalanced. In our study of the robustness of ψ_β- and ψ_θ-optimal designs, both types of designs are found to be insensitive to various specifications of the response errors and the variances of the random effects. However, they are sensitive to the specifications of the signs of the correlations of the random effects, especially the ψ_θ-optimal designs. Resulting implications for practical applications are discussed.

Key words and phrases: Bayesian Design, D-optimality, Design Robustness, Random Effects Model, Hierarchical Linear Model, Hyperparameter.

1 Introduction

Over the past two decades, various forms of hierarchical models have been used in a wide variety of fields such as the social and behavioral sciences, agriculture, education, medicine, healthcare studies, and marketing. These models have been used under the terminology of "multi-level models", "mixed-effects models", "random-effects models", "population models", "random-coefficient regression models" and "covariance components models" (see a review by Raudenbush and Bryk, 2002).

Hierarchical models consist of at least two levels by definition. In a two-level hierarchical model commonly seen in the literature, parameters in the first level of the hierarchy capture individual-level effects, which are assumed to be random and the probability distribution of the random effects is characterized by the hyperparameters in the second-level of the hierarchy (see Section 2). Hyperparameters may reflect population characteristics ("population parameters"), for example, the mean and dispersion of the effects of a new drug on patients in a certain population (see Yuh et al., 1994); or may reflect the effects of various covariates which drive the individual-level effects, such as the effect of exposure to language on vocabulary growth of a child (Huttenlocher et al., 1991), and the effects of consumer demographics on consumer sensitivity to the product feature change (Allenby and Ginter, 1995).

In situations such as direct marketing, which focuses on individual customization of products, it is important to have accurate information on individual-level effects. 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 is important as these capture the population characteristics, and enable predictions to new contexts.

Extant research on efficient designs under hierarchical linear models has focused mainly on non-Bayesian (local) designs that assume fixed values for the variance and covariance parameters. For example, for the estimation of hyperparameters, Giovagnoli and Sebastiani (1989) used a local design criterion that allows for different emphasis on the estimation of the mean and variance for a special case of the hierarchical linear model – the one-way random effects model with one single factor or predictor variable. Lenk et al. (1996) investigated the tradeoff between the number of subjects in a survey setting and the number of questions per subject under a cost constraint and an orthogonal design structure, assuming independent, identically distributed random effects. Fedorov and Hackl (1997, pg. 78) derived a necessary and sufficient condition for a design to be optimal under a hierarchical linear model with random effects that may be correlated. Some examples of optimal one-factor designs were given by Entholzner et al. (2005) in the correlated setting and some optimal two-factor designs in the uncorrelated setting.

Bayesian designs for hierarchical linear models that take into account of the uncertainty of the model parameters were investigated by Smith and Verdinelli (1980) for the estimation of individual-level effects under the one-way random effects model. Using the same model, Lohr (1995) derived optimal Bayesian designs for the estimation of the ratio of the variance components. Liu, Dean and Allenby (2007) investigated Bayesian designs for the joint estimation of the mean and covariance matrix of the random effects for the general form of the hierarchical linear model with multiple predictor variables.

In this paper, we treat the covariance matrix of the random effects as nuisance parameters and focus our attention on two types of Bayesian designs under the general form of the hierarchical linear model – one for the estimation of the individual-level effects for each respondent, and the other for the estimation of the effects of the covariates. When there are no covariates, the latter criterion becomes the criterion for estimating the mean of the random effects (i.e., population mean). Comparisons between the two types of Bayesian designs suggest that they are quite different from each other when the random effects are correlated (see Sections 6 and 7). For the designs investigated in this paper, a fixed number of observations per subject is assumed, as seen in survey studies where survey questions are designed given a fixed length of the questionnaire.

The paper is organized as follows. In Section 2, we describe the hierarchical linear model used in this paper. In Section 3, we specify the two Bayesian design criteria investigated — the ψ_β criterion for the estimation of individual-level parameters $\boldsymbol{\beta}_i$ for respondent i, and the ψ_θ criterion for the estimation of the hyperparameter vector $\boldsymbol{\theta}$. We discuss the issue of experimenter-controlled covariates briefly in Section 4. Then, in Section 5 and later, we focus our attention on the special scenario when all subjects receive the same treatments and the covariates are independent of the treatments. In Section 6, we derive forms of optimal continuous (approximate) designs under the ψ_β and the ψ_θ criteria for both the situation of independent random effects and some specific situations of correlated random effects. For more general situations, ψ_β- and ψ_θ-optimal exact designs are examined through examples in Section 7. Design robustness is investigated in Section 8 under different specifications of the response errors and of the covariance matrix of the random effects. We end the paper with conclusion and discussion in Section 9.

2 The Model

We take a hierarchical linear model of the following form:

$$\mathbf{y}_i|\boldsymbol{\beta}_i, \sigma_i^2 \sim N_{m_i}(\mathbf{X}_i\boldsymbol{\beta}_i, \sigma_i^2\mathbf{I}_{m_i}), \tag{2.1}$$

$$\boldsymbol{\beta}_i|\boldsymbol{\theta}, \boldsymbol{\Lambda} \sim N_p(\mathbf{Z}_i\boldsymbol{\theta}, \boldsymbol{\Lambda}), \tag{2.2}$$

where responses of subject i ($i = 1, \dots n$) are represented by the vector $\mathbf{y}_i$ of length m_i, corresponding to the $m_i \times p$ model matrix $\mathbf{X}_i$, which depends upon the treatments (or stimuli) allocated to the subjects. The effects of the stimuli 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 $p \times p$ covariance matrix $\boldsymbol{\Lambda}$ and mean $\mathbf{Z}_i\boldsymbol{\theta}$ where $\mathbf{Z}_i$ is a $p \times q$ matrix of covariates, such as household income and age, and $\boldsymbol{\theta}$ is the corresponding parameter vector of length q.

The following diffuse conjugate priors are often assumed for $\boldsymbol{\theta}$ and σ_i^2, corresponding to weak prior knowledge in data analysis (see, for example, Rossi et al., 2005):

$$\boldsymbol{\theta} \sim \text{Normal}(\mathbf{0}_q, 100\mathbf{I}_q), \quad \sigma_i^2 \sim \text{Inverse Gamma}(1.5, 0.5). \tag{2.3}$$

These are replaced by more informative priors when information is available. There have been some discussions on the appropriate diffuse prior to use for $\boldsymbol{\Lambda}$. The standard Jeferey's prior is not recommended due to the inadequacies of the Jefferey's prior in higher dimensions (see Yang and Berger, Section 2.2, 1994, for more details and references). The Inverted Wishart prior has also been criticized for being inadequate because it only allows one degree-of-freedom or shape parameter for all components of the covariance matrix (see, for example, Daniels and Kass, 1999). More flexible priors have been proposed based on various decompositions of the covariance matrix, such as the reference prior by Yang and Berger (1994) based on the spectral decomposition of the covariance matrix, or priors based on the cholesky decomposition of the covariance matrix (Pinheiro and Bates, 1996) or of the inverse of the covariance matrix (Pourahmadi, 1999, 2000). In this paper, we follow Barnard, McCulloch and Meng (2000) and break the covariance matrix down to components of variances v_{ii} and correlations $r_{ij} = r_{ji}$, corresponding to covariances $v_{ij} = r_{ij}\sqrt{v_{ii}v_{jj}}$ for $i < j, i, j \in \{1, 2, \dots, p\}$. We assume Inverse Gamma distributions on the variance components v_{ii} and allow for different degrees of freedom. Correlation components r_{ij} are assumed to follow a jointly Uniform prior, where the support regions of the components are sequentially determined to ensure a positive-definite covariance matrix $\boldsymbol{\Lambda}$ (see Barnard et al., 2000, for details). Let $R = \{r_{ij}\}$ and let $f(R)$ denote the probability density function of R. The priors we are using for components of $\boldsymbol{\Lambda}$ can then be expressed as

$$v_{ii} \sim \text{Inverse Gamma}(a_i, b_i), \quad f(R) \propto 1. \tag{2.4}$$

3 Bayesian Design Criteria

We consider two Bayesian design criteria for the hierarchical linear model specified in (2.1) and (2.2). In Section 3.1, we define the Bayesian D-criterion for the estimation of individual-level effects $\boldsymbol{\beta}_i$ for subject i. In Section 3.2, we define the Bayesian D-criterion for the estimation of the hyperparameter vector $\boldsymbol{\theta}$ where the covariates $\mathbf{Z}_i$ can be controlled by the experimenter. To distinguish between the two criteria and from the traditional non-Bayesian D-criteria, we call the criterion in Section 3.1 the ψ_β criterion, and call the criterion in Section 3.2 the ψ_θ criterion.

Following Chaloner (1984, page 284), we define each design criterion as the minimization of the pre-posterior risk (see Berger, 1985) where the posterior loss is defined based on the posterior conditional distribution of the corresponding parameter of interest given nuisance parameters, as shown in (3.1) and (3.6), respectively. While the ψ_β criterion only involves the specification of the treatments which determines the model matrix X_i, the ψ_θ criterion involves the specification of both the treatments and the covariates which may or may not be independent of each other.

3.1 ψ_β criterion for estimation of $\boldsymbol{\beta}_i$

In this section, we consider the situation where interest is in the accurate estimation of individual-level effects $\boldsymbol{\beta}_i$ for subject i, while all other parameters are considered to be nuisance parameters. This occurs when there is a focus on the individual customization of products, such as direct marketing. We define the ψ_β criterion as follows. Let d_i be the design allocated to subject i, with corresponding $m_i \times p$ model matrix $\mathbf{X}_i$. We seek a design for subject i that minimizes the pre-posterior risk

$$E_{\boldsymbol{\theta},\boldsymbol{\Lambda},\sigma_i^2} E_{\boldsymbol{\beta}_i|\boldsymbol{\theta},\boldsymbol{\Lambda},\sigma_i^2} \left\{ \log \left| \mathcal{I}_{GFIM}(\boldsymbol{\beta}_i|\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \sigma_i^2) \right|^{-1/p} \right\}, \tag{3.1}$$

where the expectation is taken over the prior distributions of $\boldsymbol{\theta}$, $\boldsymbol{\Lambda}$, and σ_i^2. The $\mathcal{I}_{GFIM}$ is the generalized Fisher Information matrix (see Ferreira, 1981) which, under the assumptions of normality, is obtained by taking the negative expectation of the second derivative of the logarithm of the posterior density function, that is,

$$\mathcal{I}_{GFIM}(\boldsymbol{\beta}_i|\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \sigma_i^2) = -E\left[\frac{\partial^2 \log f(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \sigma_i^2)}{\partial \boldsymbol{\beta}_i \boldsymbol{\beta}_i'}\right], \tag{3.2}$$

where $\boldsymbol{\beta}_i'$ is the transpose of $\boldsymbol{\beta}_i$. Since the posterior $f(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \sigma_i^2)$ is a normal density function with covariance matrix $(\sigma_i^{-2}\mathbf{X}_i'\mathbf{X}_i + \boldsymbol{\Lambda}^{-1})^{-1}$ and mean $(\sigma_i^{-2}\mathbf{X}_i'\mathbf{X}_i +$

$\mathbf{\Lambda}^{-1})^{-1}(\sigma_i^{-2}\mathbf{X}_i'\mathbf{y}_i + \mathbf{\Lambda}^{-1}\mathbf{Z}_i\boldsymbol{\theta})$ (see, for example, Chapter 2, Rossi et al., 2005), we have

$$\mathcal{I}_{GFIM}(\boldsymbol{\beta}_i|\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\theta}, \mathbf{\Lambda}, \sigma_i^2) = \sigma_i^{-2}\mathbf{X}_i'\mathbf{X}_i + \mathbf{\Lambda}^{-1},$$

which does not depend on $\boldsymbol{\theta}$. Therefore, (3.1) simplifies to

$$\int \left\{\log \left|\sigma_i^{-2}\mathbf{X}_i'\mathbf{X}_i + \mathbf{\Lambda}^{-1}\right|^{-1/p}\right\} f(\mathbf{\Lambda})f(\sigma_i^2)d\mathbf{\Lambda}d\sigma_i^2, \tag{3.3}$$

where $f(\Lambda)$ and $f(\sigma_i^2)$ are the prior probability density functions of $\mathbf{\Lambda}$ and σ_i^2, respectively. An optimal ψ_β design is a design that minimizes (3.3). Equivalently, since the number of parameters p in vector $\boldsymbol{\beta}_i$ is fixed for any given experiment in this paper, an optimal ψ_β design is a design that maximizes

$$\int \log \left|\sigma_i^{-2}\mathbf{X}_i'\mathbf{X}_i + \mathbf{\Lambda}^{-1}\right| f(\mathbf{\Lambda})f(\sigma_i^2)d\mathbf{\Lambda}d\sigma_i^2, \tag{3.4}$$

for each $i = 1, \ldots, n$. Note that the ψ_β criterion function (3.4) only involves the model matrix $\mathbf{X}_i$ and does not depend on the covariates matrix $\mathbf{Z}_i$.

3.2 ψ_θ criterion for estimation of hyperparameter $\boldsymbol{\theta}$

In this section, we consider the situation where interest is in the effects $\boldsymbol{\theta}$, of covariates, with the dispersion parameters $\mathbf{\Lambda}$ and σ_i^2 regarded as nuisance parameters. When there are no covariates, i.e., when $\mathbf{Z}_i = \mathbf{I}$, $\boldsymbol{\theta}$ simply captures the mean of the random effects $\boldsymbol{\beta}_i$. In this setting, the two layers (2.1) and (2.2) of the hierarchical model can be combined to obtain

$$\mathbf{y}_i|\boldsymbol{\theta}, \mathbf{\Lambda}, \sigma_i^2 \sim N_{m_i}(\mathbf{X}_i\mathbf{Z}_i\boldsymbol{\theta},\ \mathbf{\Sigma}_i = \sigma_i^2\mathbf{I}_{m_i} + \mathbf{X}_i\mathbf{\Lambda}\mathbf{X}_i'), \tag{3.5}$$

(see Lenk et al., 1996, pg 187). Diffuse priors (2.3) and (2.4) are used in this paper for $\boldsymbol{\theta}$, σ_i^2, and $\mathbf{\Lambda}$, $i = 1, \ldots, n$.

Let $\mathcal{D}(m_1, \ldots, m_n)$ be a class of designs $\tilde{d} = (d_1, \ldots, d_n)$, where d_i is the m_i-point sub-design allocated to subject i. When $m_1 = m_2 = \ldots = m_n$, we write $\mathcal{D}(m)$. For a given $\tilde{d} = (d_1, \ldots, d_n)$, define $\widetilde{\mathbf{X}}' = (\mathbf{X}_1', \ldots, \mathbf{X}_n')$ and $\widetilde{\mathbf{Z}}' = (\mathbf{Z}_1', \ldots, \mathbf{Z}_n')$, where $\mathbf{X}_i$ is the $m_i \times p$ model matrix corresponding to d_i and $\mathbf{Z}_i$ is the corresponding $p \times q$ matrix of covariates. Under the ψ_θ criterion, we seek a design $\tilde{d}^*$ in $\mathcal{D}(m_1, \ldots, m_n)$ that minimizes the pre-posterior risk

$$E_{\boldsymbol{\theta},\mathbf{\Lambda},\boldsymbol{\varsigma}} \left\{\log \left|\mathcal{I}_{GFIM}(\boldsymbol{\theta}|\widetilde{\mathbf{X}}, \widetilde{\mathbf{Z}}, \mathbf{\Lambda}, \boldsymbol{\varsigma})\right|^{-1/q}\right\}, \tag{3.6}$$

where $\boldsymbol{\varsigma} = (\sigma_1^2, \ldots, \sigma_n^2)'$. Since the posterior distribution of $\boldsymbol{\theta}$ given $\boldsymbol{\Lambda}$ and $\boldsymbol{\varsigma}$ is normal with mean vector

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

and covariance matrix

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

(see Chapter 2, Rossi et al., 2005), we have

$$\mathcal{I}_{GFIM}(\boldsymbol{\theta}|\widetilde{\mathbf{X}}, \widetilde{\mathbf{Z}}, \boldsymbol{\Lambda}, \boldsymbol{\varsigma}) = \sum_{i=1}^{n} \mathbf{Z}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\mathbf{Z}_i + \mathbf{D}_0^{-1},$$

where $\boldsymbol{\theta}_0$ and $\mathbf{D}_0$ are the prior mean and covariance matrix of the hyperparameter vector $\boldsymbol{\theta}$. Therefore, (3.6) simplifies to

$$\int \left\{ \log \left| \sum_{i=1}^{n} \mathbf{Z}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\mathbf{Z}_i + \mathbf{D}_0^{-1} \right|^{-1/q} \right\} f(\boldsymbol{\Lambda}) f(\boldsymbol{\varsigma}) d\boldsymbol{\Lambda} d\boldsymbol{\varsigma}. \tag{3.7}$$

Since for a given experiment we assume that the number of parameters q in vector $\boldsymbol{\theta}$ is fixed so that the dimension of the covariates $\mathbf{Z}_i$ is fixed for each subject i, an optimal ψ_θ design is a design that maximizes

$$\int \log \left| \sum_{i=1}^{n} \mathbf{Z}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\mathbf{Z}_i + \mathbf{D}_0^{-1} \right| f(\boldsymbol{\Lambda}) f(\boldsymbol{\varsigma}) d\boldsymbol{\Lambda} d\boldsymbol{\varsigma}. \tag{3.8}$$

In this paper, we assume the diffuse prior in (2.3) for $\boldsymbol{\theta}$ where $\mathbf{D}_0^{-1} = (100\mathbf{I}_q)^{-1}$. When the diffuse prior is used, or when the number of subjects n is large, the influence of the prior information becomes negligible, an approximation to (3.8) is

$$\int \log \left| \sum_{i=1}^{n} \mathbf{Z}_i'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\mathbf{Z}_i \right| f(\boldsymbol{\Lambda}) f(\boldsymbol{\varsigma}) d\boldsymbol{\Lambda} d\boldsymbol{\varsigma}. \tag{3.9}$$

An optimal ψ_θ design is a design that maximizes (3.8) or (3.9) depending on how informative and influential the prior is. In this paper, we use (3.9).

4 Controlled Covariates $\mathbf{Z}_i$

The ψ_β criterion in (3.4) requires the search for optimal design d_i which involves the specification of the model matrix $\mathbf{X}_i$. This search is done separately for each $i = 1, \ldots, n$. However, the ψ_θ criterion in (3.9) requires the search for optimal design $\tilde{d} = (d_1, \ldots, d_n)$ which involves the specifications of both the set of model matrices $\{\mathbf{X}_i\}$ and the set of matrices of covariates $\{\mathbf{Z}_i\}, i = 1, \ldots, n$.

When experimenters do not have control over the sampling of the subjects which determines the set of matrices of covariates, such as when subjects are pre-designated or scarce, the covariates need to be taken as given in the ψ_θ criterion. However, as often in survey studies, experimenters do have control over the sampling of the subjects on the basis of the covariates such as gender and age, in addition to the control over the treatment allocation which determines the model matrices.

In this section, we consider this situation in which both the treatment allocation and the selection of covariates can be controlled. First, consider the situation in which $\mathbf{X}_i$ and $\mathbf{Z}_i$ can be determined independently of each other. For example, if the covariates consist of the age a_i and household income h_i for each respondent i, these values remain constant for all responses from respondent i and do not depend on the survey questions asked. Therefore, $\mathbf{Z}_i$ can be expressed as $\mathbf{z}_i' \otimes \mathbf{I}_p$, where $\otimes$ denotes Kronecker product, and $\mathbf{z}_i' = [1, \;\; a_i, \;\; h_i]$. So,

$$\mathbf{Z}_i = [\mathbf{I}_p, \;\; a_i\mathbf{I}_p, \;\;\; h_i\mathbf{I}_p] = \mathbf{z}_i' \otimes \mathbf{I}_p. \tag{4.1}$$

The hyperparameter vector $\boldsymbol{\theta}$ in (2.2) is then of length $q = 3p$. Here, the first set of p hyperparameters corresponds to the first p columns of $\mathbf{Z}_i$ in (4.1), that is, the p columns of $\mathbf{I}_p$, and captures the general mean of the random effects $\boldsymbol{\beta}_i$; the second set of p hyperparameters corresponds to the p columns of $a_i\mathbf{I}_p$ in (4.1) and captures the influence of respondent age a_i on $\boldsymbol{\beta}_i$; the third set of p hyperparameters in $\boldsymbol{\theta}$ corresponds to the p columns of $h_i\mathbf{I}_p$ and captures the influence of household income h_i on $\boldsymbol{\beta}_i$. Using (4.1), and noting that $\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i$ is a $p \times p$ matrix, the integrand of (3.9) becomes

$$\log\left|\sum_{i=1}^{n}\left[(\mathbf{z}_i' \otimes \mathbf{I}_p)'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i(\mathbf{z}_i' \otimes \mathbf{I}_p)\right]\right| = \log\left|\sum_{i=1}^{n}\left[(\mathbf{z}_i\mathbf{z}_i') \otimes (\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i)\right]\right|, \tag{4.2}$$

where $\boldsymbol{\Sigma}_i = \sigma_i^2\mathbf{I}_{m_i} + \mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i'$.

When $\mathbf{X}_i$ and $\mathbf{Z}_i$ cannot be determined independently but are linked, $\mathbf{Z}_i$ cannot be written in the form of (4.1) and computer search algorithms need to be used to find the optimal combination of $\mathbf{X}_i$ and $\mathbf{Z}_i$. An example of this situation is studied in Liu et al. (2009) in the modeling of the level-effect in marketing research.

5 Special Case of $\mathbf{X}_i = \mathbf{X}$, $\sigma_i^2 = \sigma^2$ and $\mathbf{Z}_i$ Independent of $\mathbf{X}$

In the remainder of this paper, we focus on the special case where
(i) every subject receives the same design so that $\mathbf{X}_i = \mathbf{X}$, and $m_i = m$
(ii) the response errors are homoscedastic so that $\sigma_i^2 = \sigma^2$, and
(iii) $\mathbf{Z}_i$ independent of $\mathbf{X}$.
Note that when $p = q$ and $\mathbf{Z}_i = \mathbf{I}_p$, the hyperparameter vector $\boldsymbol{\theta}$ captures the population characteristics.

ψ_β criterion In the special case of (i), (ii), (iii), the ψ_β criterion involves the search of an m-point design in $\mathcal{D}(m)$ with model matrix $\mathbf{X}$ that maximizes the integral (3.4) which becomes

$$\int \log \left|\sigma^{-2}\mathbf{X}'\mathbf{X} + \boldsymbol{\Lambda}^{-1}\right| f(\boldsymbol{\Lambda}) f(\sigma^2) d\boldsymbol{\Lambda} d\sigma^2. \tag{5.1}$$

ψ_θ criterion In the special case of (i), (ii), (iii), the covariates matrix $\widetilde{\mathbf{Z}}' = (\mathbf{Z}_1', \ldots, \mathbf{Z}_n')$ is determined independently of $\mathbf{X}$. Equation (4.2) simplifies to

$$\begin{aligned} \log \left| \left[\sum_{i=1}^{n} (\mathbf{z}_i \mathbf{z}_i') \otimes (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}) \right] \right| &= \log \left\{ \left| \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} \right|^{q/p} \left| \sum_{i=1}^{n} (\mathbf{z}_i \mathbf{z}_i') \right|^{p} \right\} \\ &= \tfrac{q}{p} \log \left| \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} \right| + p \log \left| \sum_{i=1}^{n} (\mathbf{z}_i \mathbf{z}_i') \right|, \end{aligned} \tag{5.2}$$

where the second equality follows from Graybill (1983, Theorem 8.8.10). By (5.2), with the independence of $\mathbf{z}_i$ and $\mathbf{X}$, and given the number of parameters p and q, the maximization of the ψ_θ design criterion function in (3.9) is achieved through the individual maximization of $\log|\sum_{i=1}^{n}(\mathbf{z}_i\mathbf{z}_i')|$ and $\int \log|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}| f(\boldsymbol{\Lambda}) f(\boldsymbol{\varsigma}) d\boldsymbol{\Lambda} d\boldsymbol{\varsigma}$. For the maximization of $\log|\sum_{i=1}^{n}(\mathbf{z}_i\mathbf{z}_i')|$, the classical fixed-effects D-optimal design theory applies (see, for example, Chapter 10 and 11, Atkinson and Donev, 1992). We will therefore focus on the maximization of $\int \log|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}| f(\boldsymbol{\Lambda}) f(\boldsymbol{\varsigma}) d\boldsymbol{\Lambda} d\boldsymbol{\varsigma}$ for the ψ_θ criterion. So, the ψ_θ criterion in this special case involves the search of a design in $\mathcal{D}(m)$ with model matrix $\mathbf{X}$ that maximizes

$$\int \left\{ \log \left| \mathbf{X}'(\sigma^2 \mathbf{I}_m + \mathbf{X}\mathbf{\Lambda}\mathbf{X}')^{-1}\mathbf{X} \right| \right\} f(\mathbf{\Lambda}) f(\sigma^2) d\mathbf{\Lambda} d\sigma^2, \tag{5.3}$$

due to (4.2), (5.2), and $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_m + \mathbf{X}\mathbf{\Lambda}\mathbf{X}'$. We restrict our search to those designs with nonsingular $\mathbf{X}'\mathbf{X}$ so that the data inform the entire posterior distribution of $\boldsymbol{\theta}$ under vague prior assumptions. Lemma 1 gives an alternative form of (5.3) which is more convenient to use in a search for a ψ_θ optimal design. The proof follows from Morrison (1990, page 69) by letting $\mathbf{A} = \sigma^2 \mathbf{I}_m,\ \mathbf{B} = \mathbf{X}$, $\mathbf{C} = \mathbf{\Lambda}$ and noting that $\mathbf{I}_p = (\mathbf{\Lambda}^{-1} + \sigma^{-2}\mathbf{X}'\mathbf{X})^{-1}(\mathbf{\Lambda}^{-1} + \sigma^{-2}\mathbf{X}'\mathbf{X})$.

Lemma 1. *Under (i), (ii), (iii), the ψ_θ optimal design maximizes*

$$\int \log \left(\frac{1}{|\sigma^2 (\mathbf{X}'\mathbf{X})^{-1} + \mathbf{\Lambda}|} \right) f(\mathbf{\Lambda}) f(\sigma^2) d\mathbf{\Lambda} d\sigma^2. \tag{5.4}$$

Note that our Bayesian design criteria ψ_β and ψ_θ nest the corresponding non-Bayesian criteria which can be considered as special cases when the prior distributions of $\mathbf{\Lambda}$ and σ^2 are degenerate. For example, when $\mathbf{\Lambda}$ and σ^2 are fixed, our ψ_θ criterion is equivalent to the minimization of $\left|(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{\Lambda}/\sigma^2)\mathbf{I}_p\right|$, which is the non-Bayesian criterion used by Fedorov and Hackl (1997, Equation 5.2.6), who obtained the necessary and sufficient condition for optimal designs under this criterion. It is also the "mixed-effects model D-criterion" used by Entholzner et al. (2005) who examined optimal designs for the special cases of one or two-factor designs.

6 Theoretical Results on ψ_β- and ψ_θ-Optimal Designs

In this section we identify ψ_β-optimal designs and ψ_θ-optimal designs, respectively, for both the case of independent random effects and some special cases of correlated random effects under assumptions (i) to (iii) in Section 5. It is difficult to obtain theoretical results on optimal designs in a space of discrete design points. Consequently, we follow the strategy of Silvey (1980), Pukelsheim (1993, page 26), and others and first obtain results in a continuous space where fractions of an observation are allowed at any given design point.

6.1 ψ_β and ψ_θ criteria

Let η be a continuous design measure in the class of probability distributions $\mathcal{H}$ on the Borel sets of $\mathcal{X}$, a compact subset of Euclidean p-space ($\mathcal{R}^p$) that contains all possible design points. In this continuous space, we have $\mathbf{X}'\mathbf{X} = m\int \mathbf{x}\mathbf{x}'d\eta(\mathbf{x}), \eta \in \mathcal{H}, \mathbf{x} \in \mathcal{X}$. Define the set $\mathcal{M}$ to be

$$\mathcal{M} = \{\mathbf{M}(\eta) : \mathbf{M}(\eta) = \int \mathbf{x}\mathbf{x}'d\eta(\mathbf{x}), \eta \in \mathcal{H}, \mathbf{x} \in \mathcal{X}\}. \tag{6.1}$$

By Silvey (1980), the set $\mathcal{M}$ is a closed convex hull of $\{\mathbf{x}\mathbf{x}' : \mathbf{x} \in \mathcal{X}\}$, and a ψ_β-optimal continuous design η^* under assumptions (i), (ii) and (iii) is such that $\eta^* \in \mathcal{H}$ maximizes the continuous analog of (5.1), namely,

$$\psi_\beta(\mathbf{M}(\eta)) = \begin{cases} \int \left\{\log|\frac{m}{\sigma^2}\mathbf{M}(\eta) + \mathbf{\Lambda}^{-1}|\right\} f(\mathbf{\Lambda})f(\sigma^2)d\mathbf{\Lambda}d\sigma^2 & \text{for } \mathbf{M}(\eta) \text{ nonsingular,} \\ -\infty & \text{for } \mathbf{M}(\eta) \text{ singular.} \end{cases} \tag{6.2}$$

Similarly, a ψ_θ-optimal continuous design $\eta^\diamond$ under assumptions (i), (ii) and (iii) is such that $\eta^\diamond \in \mathcal{H}$ maximizes the continuous analog of (5.4), namely,

$$\psi_\theta(\mathbf{M}(\eta)) = \begin{cases} \int \left\{-\log|\frac{\sigma^2}{m}\mathbf{M}(\eta)^{-1} + \mathbf{\Lambda}|\right\} f(\mathbf{\Lambda})f(\sigma^2)d\mathbf{\Lambda}d\sigma^2 & \text{for } \mathbf{M}(\eta) \text{ nonsingular,} \\ -\infty & \text{for } \mathbf{M}(\eta) \text{ singular.} \end{cases} \tag{6.3}$$

Lemma 2. *Functions $\psi_\beta(M(\eta))$ in (6.2) and $\psi_\theta(M(\eta)$ in (6.3) are each concave and monotone in $\mathcal{M}$ where $\mathcal{M}$ is defined in (6.1).*

The proof of Lemma 2 follows from the fact that both integrands in (6.2) and (6.3) are monotone and concave (see Chaloner 1984; Fedorov and Hackl 1997, page 31), and that integration is a linear operation. The following two theorems give necessary and sufficient conditions, respectively, for a ψ_β-optimal continuous design η^* and a ψ_θ-optimal continuous design $\eta^\diamond$.

Theorem 1. *Let η be a design measure in the class of probability distributions $\mathcal{H}$ on the Borel sets of a compact design space $\mathcal{X} \subseteq \mathcal{R}^p$. A design η^* is ψ_β-optimal if and only if*

$$\int \left\{\mathbf{x}'\left[\tfrac{m}{\sigma^2}\mathbf{M}(\eta^*) + \mathbf{\Lambda}^{-1}\right]^{-1}\mathbf{x}\right\} f(\mathbf{\Lambda})f(\sigma^2)d\mathbf{\Lambda}d\sigma^2 \leq \int \left\{Tr\left[\tfrac{m}{\sigma^2}\mathbf{I}_p + \mathbf{M}(\eta^*)^{-1}\mathbf{\Lambda}^{-1}\right]^{-1}\right\} f(\mathbf{\Lambda})f(\sigma^2)d\mathbf{\Lambda}d\sigma^2, \tag{6.4}$$

for all $\mathbf{x} \in \mathcal{X}$.

Theorem 2. *Let η be a design measure in the class of probability distributions $\mathcal{H}$ on the Borel sets of a compact design space $\mathcal{X} \subseteq \mathcal{R}^p$. A design $\eta^\diamond$ is ψ_θ-optimal if and only if*

$$\int \left\{ \mathbf{x}'\mathbf{M}(\eta^\diamond)^{-1}[\tfrac{\sigma^2}{m}\mathbf{I}_p + \mathbf{M}(\eta^\diamond)\mathbf{\Lambda}]^{-1}\mathbf{x} \right\} f(\mathbf{\Lambda})f(\sigma^2)d\mathbf{\Lambda}d\sigma^2$$
$$\leq \int \left\{ Tr[(\tfrac{\sigma^2}{m}\mathbf{I}_p + \mathbf{M}(\eta^\diamond)\mathbf{\Lambda})^{-1}] \right\} f(\mathbf{\Lambda})f(\sigma^2)d\mathbf{\Lambda}d\sigma^2, \quad (6.5)$$

for all $\mathbf{x} \in \mathcal{X}$.

Since the integration (over $\mathbf{\Lambda}$ and σ^2) in the ψ_β and ψ_θ criteria is a linear operation, by (2.6.11), Fedorov and Hackl (1997), the proofs of Theorem 1 and Theorem 2 can be obtained by taking integrals (over $\mathbf{\Lambda}$ and σ^2) of the necessary and sufficient conditions for optimal ψ_β and ψ_θ designs when $\mathbf{\Lambda}$ and σ^2 are known, the latter of which follow from Lemma 2 above, and Theorem 3.7 in Silvey (1980) or Theorem 2.3.2 in Fedorov and Hackl (1997).

6.2 Model matrix X

In the model matrix $\mathbf{X}$ we employ the "standardized orthogonal effects coding" (see Kuhfeld, 2005) so that for a treatment factor, each column of the standardized coefficients has squared length equal to the number of levels of the corresponding factor. For example, for a three-level factor let

$$\mathbf{H} = \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ -1 & -1 \end{pmatrix}\mathbf{S},$$

where the two columns correspond to two orthogonal contrasts in the three levels of the factor, and $\mathbf{S}$ is a diagonal matrix with standardization coefficients $\sqrt{3}/\sqrt{6}$ and $\sqrt{3}/\sqrt{2}$ on the diagonal so that the sum of squares in each column of $\mathbf{H}$ equals the number of levels of the factor which is 3 in this example. With this coding, the pair of values $(\frac{2\sqrt{3}}{\sqrt{6}}, 0)$ in the first row represents the lowest level, $(\frac{-\sqrt{3}}{\sqrt{6}}, \frac{\sqrt{3}}{\sqrt{2}})$ represents the middle level, and $(\frac{-\sqrt{3}}{\sqrt{6}}, \frac{-\sqrt{3}}{\sqrt{2}})$ represents the highest level of the three-level factor.

For an arbitrary treatment factor with h levels, the matrix $\mathbf{H}_h$ of the standardized coefficients is:

$$\mathbf{H}_h = \begin{pmatrix} h-1 & 0 & \cdots & 0 & 0 \\ -1 & h-2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & -1 & \cdots & 2 & 0 \\ -1 & -1 & \cdots & -1 & 1 \\ -1 & -1 & \cdots & -1 & -1 \end{pmatrix}\mathbf{S}_h, \quad (6.6)$$

where $\mathbf{S}_h = Diag(s_1, s_2, \ldots, s_{h-1})$ where $s_j = \sqrt{h/[(h-j)^2 + h - j]}$ for $j = 1, \ldots, h-1$.

In addition, this standardization ensures that the sum of squares for each row of $\mathbf{H}_h$ equals the number of levels of the factor minus one. This can be seen by first noting that the matrix $\mathbf{H}_h^* = [\mathbf{1}_h \;\; \mathbf{H}_h]$ is orthogonal, and so $\mathbf{H}_h^{*\prime}\mathbf{H}_h^* = \mathbf{H}_h^*\mathbf{H}_h^{*\prime} = h\mathbf{I}_h$, and

$$\mathbf{H}_h^*\mathbf{H}_h^{*\prime} = [\mathbf{1}_h \;\; \mathbf{H}_h] \begin{bmatrix} \mathbf{1}_h' \\ \mathbf{H}_h' \end{bmatrix} = \mathbf{J}_h + \mathbf{H}_h\mathbf{H}_h', \tag{6.7}$$

where $\mathbf{1}_h$ is an $h \times 1$ vector of all 1's and $\mathbf{J}_h = \mathbf{1}_h\mathbf{1}_h'$. Consequently, $\mathbf{H}_h\mathbf{H}_h' = h\mathbf{I}_h - \mathbf{J}_h$ which has each diagonal element equal to $h-1$, and each diagonal element of $\mathbf{H}_h\mathbf{H}_h'$ equals the sum of the squares of the elements in each row of $\mathbf{H}_h$.

The model matrix $\mathbf{X}$ contains m rows selected from

$$[\mathbf{1}_m,\; \mathbf{H}_{h_1} \otimes \mathbf{1}_{h_2} \otimes \cdots \otimes \mathbf{1}_{h_f}, \quad \ldots, \quad \mathbf{1}_{h_1} \otimes \cdots \otimes \mathbf{1}_{h_{f-1}} \otimes \mathbf{H}_{h_f}],$$

corresponding to the factor combinations in the design plus a column of ones in the first column for the general mean, where h_i is the number of levels of factor i $(i = 1, \ldots, f)$. For example, if there are two factors with 2 and 3 levels respectively and each treatment combination occurs once in the design, then we have

$$\mathbf{X} = \begin{pmatrix} 1 & -1 & 2 & 0 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix} \mathbf{S}, \tag{6.8}$$

where $\mathbf{S}$ is the diagonal standardization matrix such that $\mathbf{S} = Diag\{1, \frac{\sqrt{2}}{\sqrt{2}}, \frac{\sqrt{3}}{\sqrt{6}}, \frac{\sqrt{3}}{\sqrt{2}}\}$. If some of the six combinations are not presented, then the corresponding rows of $\mathbf{X}$ are removed. The sum of squares for each row of $\mathbf{X}$ then equals $1 + (2-1) + (3-1) = 4$ which equals the total number of parameters or the number of columns of $\mathbf{X}$. Thus for every row $\mathbf{x}' = [x_0, x_1, \ldots, x_{p-1}]$ of $\mathbf{X}$, we have $\sum_{k=0}^{p-1} x_k^2 = p$. With these points as the border points, we create a continuous induced design space $\mathcal{X}$, expressed as

$$\mathcal{X} = \Big\{ \mathbf{x} = [x_0, x_1, \ldots, x_{p-1}]' \quad \text{such that} \quad x_0 = 1 \quad \text{and} \quad \sum_{k=0}^{p-1} x_k^2 \le p \Big\}. \tag{6.9}$$

Note that with the above definition, a design with an equal number of occurrences of every treatment combination, called a level-balanced orthogonal design, such as the design with the model matrix (6.8), has $\mathbf{X}'\mathbf{X} = m\mathbf{I}$, and so $\mathbf{M}(\eta) = \mathbf{I}$.

6.3 Independent random effects

Our first result identifies a continuous design that is both ψ_β-optimal and ψ_θ-optimal when the random effects $\boldsymbol{\beta}_i$ in (2.1) are independent and there is weak prior knowledge on the variances of the random effects. The proof of Theorem 3 is provided in the appendix.

Theorem 3. *Let η be a design measure in the class of probability distributions $\mathcal{H}$ on the Borel sets of $\mathcal{X}$ where $\mathcal{X}$ is a compact subspace of $\mathcal{R}^p$ defined in (6.9). Let the random individual-level effects $\boldsymbol{\beta}_i$ in (2.1) be independent such that $\boldsymbol{\Lambda}$ is*

$$\boldsymbol{\Lambda} = Diag\left(\lambda_0^2,\ \lambda_2^2,\ \ldots,\ \lambda_{p-1}^2\right), \tag{6.10}$$

where the prior distributions of $\lambda_k^2, k = 0, \ldots, p-1$ are identical. Given such a situation, any design η^* that satisfies $\mathbf{M}(\eta^*) = \mathbf{I}$ is both ψ_β- and ψ_θ-optimal.

The following corollary follows directly from Theorem 3 by noting that, for a level-balanced orthogonal design, the corresponding model matrix satisfies $\mathbf{M}(\eta) = \mathbf{I}$.

Corollary 1. *Under the conditions of Theorem 3, if a level-balanced orthogonal design exists, it is both ψ_β- and ψ_θ-optimal.*

6.4 Special cases of correlated random effects

When the random individual-level effects $\boldsymbol{\beta}_i$ in (2.1) are equally correlated and with equal variances, $\boldsymbol{\Lambda}$ is of the form $\tilde{a}\mathbf{I}+\tilde{d}\mathbf{J}$, where $\tilde{a}$ and $\tilde{d}$ are scalars, and $\mathbf{J}$ is the $p\times p$ matrix with all elements equal to 1. Note that for $\boldsymbol{\Lambda}$ to be positive definite, $\tilde{a}$ and $\tilde{d}$ are constrained such that $\tilde{a} > 0$ and $\tilde{a}+p\tilde{d} > 0$. The prior distribution of $\boldsymbol{\Lambda}$ in this special case reduces to the prior distributions of $\tilde{a}$ and $\tilde{d}$. Specifically, as in (2.4), the diffuse prior of the variance term $(\tilde{a}+\tilde{d})$ is assumed to be Inverse Gamma (1.5,0.5), and the prior of the correlation term $\tilde{d}/(\tilde{a}+\tilde{d})$ is uniform $(1/(1-p), 1)$ to ensure that $\boldsymbol{\Lambda}$ is positive definite. When prior knowledge is available, more informative priors can be used on the variance term $\tilde{a}+\tilde{d}$ and the correlation term $\tilde{d}/(\tilde{a}+\tilde{d})$. For example, if there is knowledge that the correlation is low and positive the prior distribution of $\tilde{d}/(\tilde{a}+\tilde{d})$ may be taken as a Uniform (0, 0.2) distribution. Note that through variable transformation, we can obtain the corresponding priors for $\tilde{a}$ and $\tilde{d}$.

Theorems 4 and 5 give the forms of the matrix $\mathbf{M}(\eta^*)$ and $\mathbf{M}(\eta^\diamond)$, respectively, of a ψ_β-optimal design η^* and a ψ_θ-optimal design $\eta^\diamond$. These optimal continuous designs provide efficiency bounds, respectively, for ψ_β-optimal and ψ_θ-optimal exact designs (see Section 7). Proofs of Theorems 4 and 5 are provided in the appendix.

Theorem 4. *Let η be a design measure in the class of probability distributions $\mathcal{H}$ on the Borel sets of $\mathcal{X}$ where $\mathcal{X}$ is a compact subspace of $\mathcal{R}^p$ defined in (6.9). Given $\mathbf{\Lambda}$ of the form $\tilde{a}\mathbf{I}+\tilde{d}\mathbf{J}$ such that $\tilde{a}>0$ and $\tilde{a}+p\tilde{d}>0$, a design η^* with $\mathbf{M}(\eta^*)=(1+\kappa)\mathbf{I}-\kappa\mathbf{J}$ is ψ_β-optimal, if*

$$E_{\tilde{a},\tilde{d},\sigma^2}\left\{\left(\frac{\tilde{a}\sigma^2}{m\tilde{a}(1+\kappa)+\sigma^2}\right)\left(\frac{\kappa m\tilde{a}(\tilde{a}+p\tilde{d})+\tilde{d}\sigma^2}{m\tilde{a}(\tilde{a}+p\tilde{d})[1-(p-1)\kappa]+\tilde{a}\sigma^2}\right)\right\}=0, \tag{6.11}$$

and the condition $\kappa\in(-1,\frac{1}{p-1})$ is satisfied.

Theorem 5. *Let η be a design measure in the class of probability distributions $\mathcal{H}$ on the Borel sets of $\mathcal{X}$ where $\mathcal{X}$ is a compact subspace of $\mathcal{R}^p$ defined in (6.9). Given $\mathbf{\Lambda}$ of the form $\tilde{a}\mathbf{I}+\tilde{d}\mathbf{J}$ such that $\tilde{a}>0$ and $\tilde{a}+p\tilde{d}>0$, a design $\eta^\diamond$ with $\mathbf{M}(\eta^\diamond)=(1+\epsilon)\mathbf{I}-\epsilon\mathbf{J}$ is ψ_θ-optimal, if*

$$E_{\tilde{a},\tilde{d},\sigma^2}\left\{\frac{\epsilon^2[m(\tilde{a}+p\tilde{d})(p-2)+m\tilde{d}]-\epsilon[2m(\tilde{a}+p\tilde{d})+\sigma^2-2m\tilde{d}]+m\tilde{d}}{[m\tilde{a}(1+\epsilon)+\sigma^2][\sigma^2+m(\tilde{a}+p\tilde{d})(1-(p-1)\epsilon)]}\right\}=0, \tag{6.12}$$

and the condition $\epsilon\in(-1,\frac{1}{p-1})$ is satisfied.

Note that although there are no closed-form solutions for κ and ϵ in Equations (6.11) and (6.12), we can use a grid search within the interval $(-1,\frac{1}{p-1})$ to find the approximate solutions, as in our examples in Section 6.5.

Theorems 4 and 5 can easily be extended to more general settings where the random effects are interchangeable within groups and independent between groups, as described in the following corollary.

Corollary 2. *Given $\mathbf{\Lambda}$ of the block diagonal form*

$$\mathbf{\Lambda}=\begin{pmatrix}\lambda_0 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0}\\ \mathbf{0} & \mathbf{\Lambda}_1 & \mathbf{0} & \dots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{\Lambda}_2 & \dots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{\Lambda}_G\end{pmatrix}, \tag{6.13}$$

where $\lambda_0>0$, $\mathbf{\Lambda}_g=\tilde{a}_g\mathbf{I}_{p_g}+\tilde{d}_g\mathbf{J}_{p_g}$ such that $\tilde{a_g}>0$, $\tilde{a_g}+p_g\tilde{d_g}>0$, and the prior distributions for $(\tilde{a_g},\tilde{d_g},\sigma^2)$ are identical for $g=1,\dots,G$. The following conclusions hold *for a design η ($\eta = \eta^*$ or $\eta^\diamond$) with*

$$\mathbf{M}(\eta) = \begin{pmatrix} 1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{M}_2 & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{M}_G \end{pmatrix}. \tag{6.14}$$

(i)The design is ψ_β-optimal if $\mathbf{M}_g = (1+\kappa_g)\mathbf{I}_{p_g} - \kappa_g \mathbf{J}_{p_g}$, where

$$E_{\tilde{a}_g,\tilde{d}_g,\sigma^2}\left\{\left(\frac{\tilde{a}_g\sigma^2}{m\tilde{a}_g(1+\kappa_g)+\sigma^2}\right)\left(\frac{\kappa_g m\tilde{a}_g(\tilde{a}_g+p_g\tilde{d}_g)+\tilde{d}_g\sigma^2}{m\tilde{a}_g(\tilde{a}_g+p_g\tilde{d}_g)[1-(p_g-1)\kappa_g]+\tilde{a}_g\sigma^2}\right)\right\} = 0,$$

and the condition $\kappa_g \in (-1, \frac{1}{p_g-1})$ is satisfied for $g = 1, \dots, G$.

(ii)The design is ψ_θ-optimal if $\mathbf{M}_g = (1+\epsilon_g)\mathbf{I}_{p_g} - \epsilon_g \mathbf{J}_{p_g}$, where

$$E_{\tilde{a}_g,\tilde{d}_g,\sigma^2}\left\{\frac{\epsilon_g^2[m(\tilde{a}_g+p_g\tilde{d}_g)(p_g-2)+m\tilde{d}_g]-\epsilon_g[2m(\tilde{a}_g+p_g\tilde{d}_g)+\sigma^2-2m\tilde{d}_g]+m\tilde{d}_g}{[m\tilde{a}_g(1+\epsilon_g)+\sigma^2][\sigma^2+m(\tilde{a}_g+p_g\tilde{d}_g)(1-(p_g-1)\epsilon_g)]}\right\} = 0,$$

and the condition $\epsilon_g \in (-1, \frac{1}{p_g-1})$ is satisfied for $g = 1, \dots, G$.

Theorems 3, 4 and 5, together with Corollary 2, suggest that the matrix $\mathbf{M}(\eta^*)$ ($\mathbf{M}(\eta^\diamond)$) of a ψ_β-optimal (ψ_θ-optimal) design will often have a structure similar to the covariance matrix, $\mathbf{\Lambda}$, of the random effects. Some examples of ψ_β- and ψ_θ- optimal continuous designs are provided in Section 6.5 according to Theorems 4 and 5, respectively, for the special case of equi-correlated random effects with $\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$.

6.5 ψ_β- and ψ_θ- optimal continuous design examples

Consider an experiment with two treatment factors each with two levels under a hierarchical linear model. Under assumptions (i), (ii) and (iii) from Section 5, each subject i (i=1, …, n) receives the same treatment allocation ($\mathbf{X}_i = \mathbf{X}$), response errors are homoscedastic ($\sigma_i^2 = \sigma^2$), and the covariates $\mathbf{Z}_i$ are independent of $\mathbf{X}$. For ease of exposition, we let $\mathbf{Z}_i = \mathbf{I}$. For subject i, the individual-level random effects $\boldsymbol{\beta}_i$ in (2.1) include the general mean, the main effects of factors 1 and 2, and thus $p = 3$. $\boldsymbol{\beta}_i$ is assumed to be randomly distributed according to a multivariate normal distribution with mean $\boldsymbol{\theta}$ and covariance matrix $\mathbf{\Lambda}$ as in (2.2).

The prior distribution for σ^2 and the equal variances $\tilde{a} + \tilde{d}$ of the random effects are assumed to be Inverse Gamma (1.5,0.5). The correlation of the random effects $\tilde{d}/(\tilde{a} + \tilde{d})$

is constrained to be in $(-0.5, 1)$ to ensure a positive definite $\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$. In addition to the situation when the correlation $\tilde{d}/(\tilde{a} + \tilde{d})$ is assumed to be from Uniform $(-0.5, 1)$, we also examine three separate situations when the correlation is negative, low positive, and high positive, with Uniform $(-0.5, 0)$, $(0, 0.5)$ and $(0.5, 1)$ priors, respectively.

Let the number of observations per subject be $m = 12$. Table 6.1 shows for different scenarios the κ and ϵ values corresponding to the ψ_β- and ψ_θ- optimal continuous designs by Theorems 4 and 5, respectively. We used a grid search to find the κ and ϵ values that satisfy equations (6.11) and (6.12). Under given values of κ and ϵ, Monte Carlo method was used to obtain the expectation over $(\tilde{a}, \tilde{d}, \sigma^2)$. Fixing the support points to be the four level combinations of the two treatment factors, that is, $\mathbf{x}_1 = (1, -1, -1)$, $\mathbf{x}_2 = (1, -1, 1)$, $\mathbf{x}_3 = (1, 1, -1)$ and $\mathbf{x}_4 = (1, 1, 1)$, we also report in Table 6.1 the weights on these four support points that give rise to the corresponding optimal continuous designs, denoted as m_{11}, m_{12}, m_{21}, and m_{22}, respectively.

The results in Table 6.1 show that ψ_β- and ψ_θ- optimal continuous designs, $\mathbf{M}(\eta^*)$ and $\mathbf{M}(\eta^\diamond)$, have opposite signs on the off-diagonal for the special scenarios considered in Theorems 4 and 5, with $\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$. In addition, the magnitudes of ϵ are higher than the corresponding κ's. The weights on the support points also suggest that ψ_θ- optimal continuous designs tend to be less balanced than ψ_β- optimal continuous designs.

$m = 12, \quad p = 3, \quad \sigma^2 \sim IG(1.5, 0.5)$		
Covariance $\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$ Prior	ψ_β-optimal $\mathbf{M}(\eta^*) = (1+\kappa)\mathbf{I} - \kappa\mathbf{J}$ $(m_{11}, m_{12}, m_{21}, m_{22})$	ψ_θ-optimal $\mathbf{M}(\eta^\diamond) = (1+\epsilon)\mathbf{I} - \epsilon\mathbf{J}$ $(m_{11}, m_{12}, m_{21}, m_{22})$
$(\tilde{a} + \tilde{d}) \sim IG(1.5, 0.5)$ $\tilde{d}/(\tilde{a} + \tilde{d}) \sim U(-0.5, 1)$	$\kappa = -0.04$ (2.88, 2.88, 2.88, 3.36)	$\epsilon = 0.09$ (3.27, 3.27, 3.27, 2.19)
$(\tilde{a} + \tilde{d}) \sim IG(1.5, 0.5)$ $\tilde{d}/(\tilde{a} + \tilde{d}) \sim U(-0.5, 0)$	$\kappa = 0.08$ (3.24, 3.24, 3.24, 2.28)	$\epsilon = -0.28$ (2.16, 2.16, 2.16, 5.52)
$(\tilde{a} + \tilde{d}) \sim IG(1.5, 0.5)$ $\tilde{d}/(\tilde{a} + \tilde{d}) \sim U(0, 0.5)$	$\kappa = -0.04$ (2.88, 2.88, 2.88, 3.36)	$\epsilon = 0.10$ (3.30, 3.30, 3.30, 2.10)
$(\tilde{a} + \tilde{d}) \sim IG(1.5, 0.5)$ $\tilde{d}/(\tilde{a} + \tilde{d}) \sim U(0.5, 1)$	$\kappa = -0.20$ (2.40, 2.40, 2.40, 4.80)	$\epsilon = 0.30$ (3.90, 3.90, 3.90, 0.30)

Table 6.1: κ and ϵ values as in Theorems 4 and 5, respectively, for ψ_β- and ψ_θ-optimal designs under different prior assumptions, together with corresponding weights on the four level combinations of the two treatment factors.

7 ψ_β- and ψ_θ- Optimal Exact Designs

The theoretical results on optimal continuous designs of Theorems 3, 4, 5, and Corollary 2 of Section 6, hold for some special forms of the random effects covariance matrix $\mathbf{\Lambda}$. In general, optimal continuous designs may not give rise to integer numbers of observations at the design points in the design space $\mathcal{X}$, as seen in Table 6.1. In this section, we obtain through computer search optimal exact designs that have integer numbers of observations at the design points. We examine through examples ψ_β-optimal and ψ_θ-optimal exact designs respectively, for both the special forms and some general forms of the random effects covariance matrix $\mathbf{\Lambda}$.

The same design setting as described in Section 6.5 is used here, and this incorporates assumptions (i), (ii) and (iii) from Section 5. Simple-exchange algorithms (see Atkinson and Donev, 1992, Chapter 15) are used to obtain the ψ_β- and ψ_θ-optimal exact designs. Monte Carlo method is used for the integration over the prior distributions of $\mathbf{\Lambda}$ and σ^2 in the evaluation of the design criteria.

7.1 Efficiencies

Efficiency relative to an optimal continuous design For the special cases in Theorems 3, 4, 5 and Corollary 2, we know the explicit forms of $\mathbf{M}(\eta^*)$ and $\mathbf{M}(\eta^\diamond)$ for ψ_β- and ψ_θ-optimal continuous designs η^* and $\eta^\diamond$ respectively. These optimal continuous designs can be used to provide efficiency bounds for exact designs when the optimal exact design is unknown. For an exact design with model matrix $\mathbf{X}$, we define its efficiency relative to a ψ_β-optimal continuous design η^* as

$$\int \left(\frac{\left|\sigma^{-2}\mathbf{X}'\mathbf{X} + \mathbf{\Lambda}^{-1}\right|}{\left|\sigma^{-2}m\mathbf{M}(\eta^*) + \mathbf{\Lambda}^{-1}\right|} \right)^{1/p} f(\mathbf{\Lambda})f(\sigma^2)d\mathbf{\Lambda}d\sigma^2. \tag{7.1}$$

Similarly, we define its efficiency relative to a ψ_θ-optimal continuous design $\eta^\diamond$ as

$$\int \left(\frac{\left|\sigma^{2}(m\mathbf{M}(\eta^\diamond))^{-1} + \mathbf{\Lambda}\right|}{\left|\sigma^{2}(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{\Lambda}\right|} \right)^{1/p} f(\mathbf{\Lambda})f(\sigma^2)d\mathbf{\Lambda}d\sigma^2. \tag{7.2}$$

Efficiency relative to an orthogonal design For the general cases when the structures of the ψ_β and ψ_θ-optimal continuous designs are unknown, we use an orthogonal design η^* as the base design and calculate the relative ψ_β-efficiency of a design with model matrix

$\mathbf{X}$ as

$$\text{rel. } \psi_\beta\text{-eff} = \int \left(\frac{\left|\sigma^{-2}\mathbf{X}'\mathbf{X} + \mathbf{\Lambda}^{-1}\right|}{\left|\sigma^{-2}m\mathbf{I} + \mathbf{\Lambda}^{-1}\right|} \right)^{1/p} f(\mathbf{\Lambda})f(\sigma^2)d\mathbf{\Lambda}d\sigma^2, \tag{7.3}$$

where (7.3) is obtained by replacing $\mathbf{M}(\eta^*)$ of the orthogonal design with $\mathbf{I}$. Similarly, we define the relative ψ_θ-efficiency with model matrix $\mathbf{X}$ relative to the orthogonal design as

$$\text{rel. } \psi_\theta\text{-eff} = \int \left(\frac{\left|\sigma^2 m^{-1}\mathbf{I} + \mathbf{\Lambda}\right|}{\left|\sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{\Lambda}\right|} \right)^{1/p} f(\mathbf{\Lambda})f(\sigma^2)d\mathbf{\Lambda}d\sigma^2. \tag{7.4}$$

7.2 Special forms of $\mathbf{\Lambda}$

Table 7.1 provides the ψ_β-optimal and ψ_θ-optimal exact designs as found through computer search under both the situation of independent random effects (as seen in Theorem 3) and the situation of equally correlated random effects with equal variances (as seen in Theorems 4 and 5). The designs are 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 same design. The resulting matrices $\mathbf{X}'\mathbf{X}$ of the designs are also reported. The last column of Table 7.1 shows that all the optimal exact designs obtained through computer search have efficiencies over 99% relative to their continuous counterparts. Note that these optimal exact designs obtained through computer search are the same as or very close to designs obtained by rounding the weights of the continuous designs in Table 6.1 to the nearest integers.

Table 7.1 shows that the ψ_β-optimal designs are often orthogonal even when the random effects are correlated. However, ψ_θ-optimal designs tend to be nonorthogonal and unbalanced, and the degree of imbalance increases as the random effects become more highly correlated. For example, when the correlation of the random effects is high and follows a uniform (0.5, 1) distribution (last section of Table 7.1), the ψ_θ-optimal design does not contain observations on the combination of the second levels of the two factors, that is, $m_{22} = 0$. Nevertheless, the main effects of the two factors are still estimable and observations on the combination 22 will add little or no information on the main effect parameters due to their high correlation. The ψ_β-optimal design in this case is also nonorthogonal and unbalanced, but to a lesser degree than the corresponding ψ_θ-optimal design. In addition, the signs of the off-diagonal elements of the $\mathbf{X}'\mathbf{X}$ matrix of the ψ_β-optimal design are the opposite of the signs of the off-diagonal elements of the $\mathbf{X}'\mathbf{X}$ matrix of the corresponding ψ_θ-optimal design. This is consistent with our findings on ψ_β- and ψ_θ-optimal continuous designs in Table 6.1.

Covariance matrix $\mathbf{\Lambda}$	ψ_β- and ψ_θ-optimal design $(m_{11}, m_{12}, m_{21}, m_{22})$	Matrix $\mathbf{X}'\mathbf{X}$	Eff. rel. to optimal continuous
$\mathbf{\Lambda} = Diag(\lambda_1^2, \ldots, \lambda_p^2)$ $\lambda_k^2 \sim IG(1.5, 0.5)$, $k = 1, \ldots, p$	(3,3,3,3)	$12\mathbf{I}_3$	100%
$\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$ $(\tilde{a} + \tilde{d}) \sim IG(1.5, 0.5)$ $\tilde{d}/(\tilde{a} + \tilde{d}) \sim U(-0.5, 1)$	ψ_β-optimal: (3,3,3,3)	$12\mathbf{I}_3$	99.94%
	ψ_θ-optimal: $(3, 3, 3, 3)$	$12\mathbf{I}_3$	99.83%
$\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$ $(\tilde{a} + \tilde{d}) \sim IG(1.5, 0.5)$ $\tilde{d}/(\tilde{a} + \tilde{d}) \sim U(-0.5, 0)$	ψ_β-optimal: (3,3,3,3)	$12\mathbf{I}_3$	99.66%
	ψ_θ-optimal: (3,2,2,5)	$\begin{pmatrix} 12 & 2 & 2 \\ 2 & 12 & 4 \\ 2 & 4 & 12 \end{pmatrix}$	99.96%
$\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$ $(\tilde{a} + \tilde{d}) \sim IG(1.5, 0.5)$ $\tilde{d}/(\tilde{a} + \tilde{d}) \sim U(0, 0.5)$	ψ_β-optimal: (3,3,3,3)	$12\mathbf{I}_3$	99.96%
	ψ_θ-optimal: (4,3,3,2)	$\begin{pmatrix} 12 & -2 & -2 \\ -2 & 12 & 0 \\ -2 & 0 & 12 \end{pmatrix}$	99.86%
$\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$ $(\tilde{a} + \tilde{d}) \sim IG(1.5, 0.5)$ $\tilde{d}/(\tilde{a} + \tilde{d}) \sim U(0.5, 1)$	ψ_β-optimal: (3,3,2,4)	$\begin{pmatrix} 12 & 0 & 2 \\ 0 & 12 & 2 \\ 2 & 2 & 12 \end{pmatrix}$	99.66%
	ψ_θ-optimal: (4,4,4,0)	$\begin{pmatrix} 12 & -4 & -4 \\ -4 & 12 & -4 \\ -4 & -4 & 12 \end{pmatrix}$	99.70%

Table 7.1: ψ_β- and ψ_θ-optimal 12-run exact designs when the random effects covariance matrix $\mathbf{\Lambda}$ is diagonal or has equal diagonal and off-diagonal elements. The prior distributions of σ^2 and the variance component $\tilde{a} + \tilde{d}$ of $\mathbf{\Lambda}$ are assumed to be Inverse Gamma (1.5,0.5) and the prior distribution of the correlation component $\tilde{d}/(\tilde{a} + \tilde{d})$ is assumed to be Uniform (-0.5, 1), (-0.5, 0), (0, 0.5), and (0.5, 1), respectively. Efficiencies are calculated through (7.1) and (7.2)

7.3 General forms of $\mathbf{\Lambda}$

Now we consider the general case when the covariance matrix $\mathbf{\Lambda}$ is not restricted to be of the form $\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$ with equal variances and covariances of the random effects. Instead, the variance and correlation components of $\mathbf{\Lambda}$ follow the prior distributions as specified in (2.4) with $a_i = 1.5$ and $b_i = 0.5$; that is, each of the three variances in $\mathbf{\Lambda}$ is assumed to be independently distributed according to an Inverse Gamma (1.5, 0.5) distribution. For the three correlation components, we examine three scenarios. In the first scenario, there is no restriction on the signs of the three correlation components so that the correlations can be positive or negative. In the second scenario, all three correlations are positive, and in the third scenario, all three correlations are negative. We do not know the forms of optimal continuous designs for these general forms of $\mathbf{\Lambda}$, and therefore the optimal exact designs obtained through computer search are compared with an orthogonal design by using the relative efficiencies (7.3) and (7.4). Results in Table 7.2 show that, when there is prior knowledge on the signs of the correlations of the random effects (e.g., all three correlations positive, or all negative), both ψ_β- and ψ_θ-optimal designs are nonorthogonal and unbalanced. Consistent with our findings in Table 7.1, ψ_β-optimal designs are very different from ψ_θ-optimal designs with opposite signs on the off-diagonals of the $\mathbf{X'X}$ matrix.

Signs of random effects correlations	ψ_β- and ψ_θ-optimal design $(m_{11}, m_{12}, m_{21}, m_{22})$	Matrix $\mathbf{X'X}$	Eff. Rel. to orthogonal design ψ_β	ψ_θ
No restrictions	(3,3,3,3)	$12\mathbf{I}_3$	1.000	1.000
All positive	ψ_β-optimal: (3,2,3,4)	$\begin{pmatrix} 12 & 2 & 0 \\ 2 & 12 & 2 \\ 0 & 2 & 12 \end{pmatrix}$	1.002	0.979
	ψ_θ-optimal: (3,4,3,2)	$\begin{pmatrix} 12 & -2 & 0 \\ -2 & 12 & -2 \\ 0 & -2 & 12 \end{pmatrix}$	0.973	1.009
All negative	ψ_β-optimal: (4,4,3,1)	$\begin{pmatrix} 12 & -4 & -2 \\ -4 & 12 & -2 \\ -2 & -2 & 12 \end{pmatrix}$	1.023	0.915
	ψ_θ-optimal: (2,2,3,5)	$\begin{pmatrix} 12 & 4 & 2 \\ 4 & 12 & 2 \\ 2 & 2 & 12 \end{pmatrix}$	0.923	1.027

Table 7.2: ψ_β- and ψ_θ-optimal 12-run exact designs under general forms of the random effects covariance matrix $\mathbf{\Lambda}$. The prior distributions of σ^2 and the three variance components of $\mathbf{\Lambda}$ are assumed to be independently distributed Inverse Gamma (1.5,0.5) and the prior distributions of the three correlation components are assumed to be Uniform of which the support regions are sequentially determined according to Barnard et al. (2000). Three scenarios of the support region are considered: no restrictions on the signs of the three correlations, all three correlations are positive and all three correlations are negative. Efficiencies are calculated through (7.3) and (7.4)

8 Design Robustness

We examine in this section the robustness of designs when the true distributions of σ^2 and $\mathbf{\Lambda}$ deviate from the assumed prior distributions used in the design construction. Specifically, for each design criterion, we take the ψ_β and the ψ_θ-optimal exact designs for each of eight assumed prior distributions, evaluate these designs under different true σ^2 and $\mathbf{\Lambda}$ distributions. The designs and the assumed prior distributions of σ^2 and $\mathbf{\Lambda}$ are summarized in Section 8.1, the true distributions of σ^2 and $\mathbf{\Lambda}$ are summarized in Section 8.2, and our simulation results are reported in Section 8.3.

8.1 The assumed and the true distribution scenarios

Table 8.1 lists the ψ_β and the ψ_θ-optimal designs for the eight assumed prior distributions. For each design criterion, the first listed design is an orthogonal design obtained without search, the next three designs are the optimal exact designs obtained for the special forms of $\mathbf{\Lambda}$ where $\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$ under the assumed priors in Table 7.1; the next two designs are the optimal exact designs obtained for the general forms of $\mathbf{\Lambda}$ in Table 7.2; the last two designs are local optimal exact designs obtained under fixed values of σ^2 and $\mathbf{\Lambda}$.

The local ψ_β-optimal exact designs $D_{7\beta}$ and $D_{8\beta}$, obtained under assumed fixed values of σ^2 and $\mathbf{\Lambda}$, are orthogonal, as are designs $D_{2\beta}$ and $D_{3\beta}$, obtained under special forms of $\mathbf{\Lambda}$ with assumed prior correlation distributions (see Table 7.1). These designs will all be listed together with the orthogonal design in our simulation study in Section 8.3 (as in the first row of Table 8.3). Similarly, $D_{3\theta}$ and $D_{7\theta}$ have the same form and are listed together in Table 8.4 of our simulation study.

Design Type	Designs		Prior for σ^2	Prior for $\mathbf{\Lambda}$
Orthogonal	$D_{1\beta}$	$D_{1\theta}$	–	–
Optimal under $\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$	$D_{2\beta}$	$D_{2\theta}$	IG(1.5, 0.5)	Var. $\sim IG(1.5, 0.5)$, Cor.$\sim U(-0.5, 0)$
	$D_{3\beta}$	$D_{3\theta}$	IG(1.5, 0.5)	Var. $\sim IG(1.5, 0.5)$, Cor.$\sim U(0, 0.5)$
	$D_{4\beta}$	$D_{4\theta}$	IG(1.5, 0.5)	Var. $\sim IG(1.5, 0.5)$, Cor.$\sim U(0.5, 1)$
Optimal under general $\mathbf{\Lambda}$	$D_{5\beta}$	$D_{5\theta}$	IG(1.5, 0.5)	Var. $\sim IG(1.5, 0.5)$, Cor. all +
	$D_{6\beta}$	$D_{6\theta}$	IG(1.5, 0.5)	Var. $\sim IG(1.5, 0.5)$, Cor. all −
Local optimal under fixed $\sigma^2, \mathbf{\Lambda}$	$D_{7\beta}$	$D_{7\theta}$	$\sigma^2 = 1$	$\mathbf{\Lambda} = \mathbf{I} + 0.5\mathbf{J}$
	$D_{8\beta}$	$D_{8\theta}$	$\sigma^2 = 1$	$\mathbf{\Lambda} = \mathbf{I} - 0.2\mathbf{J}$

Table 8.1: Summary of the designs and assumed prior distributions used in the robustness study.

8.2 True σ^2 and $\mathbf{\Lambda}$ distribution Scenarios

We evaluated the designs $D_{1\beta}$-$D_{8\beta}$ and $D_{1\theta}$-$D_{8\theta}$ under eight different true σ^2 and $\mathbf{\Lambda}$ distribution scenarios $\mathcal{S}_1$-$\mathcal{S}_8$ listed in Table 8.2. There are two distribution scenarios for σ^2, namely the Inverse Gamma (1.5, 0.5) and the Inverse Gamma (3, 1). While the former distribution has a mean equal to 1 and an undefined variance, the latter distribution has a mean equal to 0.5 and variance 0.25. Therefore the σ^2's from the former distribution are much more dispersed than those from the latter distribution and with a different central location. Two distribution scenarios are taken for the variance components of $\mathbf{\Lambda}$, namely (i) all three variances come from Inverse Gamma (1.5, 0.5); (ii) the three variances respectively come from Inverse Gamma (1.5, 0.5), (3, 1) and (2.5, 1.5) with different means and dispersions. Finally, two distribution assumptions are taken for the correlation components of $\mathbf{\Lambda}$, namely "all positive" and "all negative". This gives a total of eight scenarios of true σ^2 and $\mathbf{\Lambda}$ distributions.

Scenario	True dist. for σ^2	True dist. for $\mathbf{\Lambda}$
$\mathcal{S}_1$	$\sigma^2 \sim IG(1.5, 0.5)$	Var. $\sim IG(1.5, 0.5)$, Cor. all +
$\mathcal{S}_2$	$\sigma^2 \sim IG(1.5, 0.5)$	Var. $\sim IG(1.5, 0.5)$, Cor. all −
$\mathcal{S}_3$	$\sigma^2 \sim IG(1.5, 0.5)$	Var. $\sim IG(1.5, 0.5), IG(3, 1), IG(2.5, 1.5)$, Cor. all +
$\mathcal{S}_4$	$\sigma^2 \sim IG(1.5, 0.5)$	Var.$\sim IG(1.5, 0.5), IG(3, 1), IG(2.5, 1.5)$, Cor. all −
$\mathcal{S}_5$	$\sigma^2 \sim IG(3, 1)$	Var. $\sim IG(1.5, 0.5)$, Cor. all +
$\mathcal{S}_6$	$\sigma^2 \sim IG(3, 1)$	Var. $\sim IG(1.5, 0.5)$, Cor. all −
$\mathcal{S}_7$	$\sigma^2 \sim IG(3, 1)$	Var. $\sim IG(1.5, 0.5), IG(3, 1), IG(2.5, 1.5)$, Cor. all +
$\mathcal{S}_8$	$\sigma^2 \sim IG(3, 1)$	Var. $\sim IG(1.5, 0.5), IG(3, 1), IG(2.5, 1.5)$, Cor. all −

Table 8.2: Summary of the eight true σ^2 and $\mathbf{\Lambda}$ distribution scenarios in the robustness study.

8.3 Simulation results

Tables 8.3 and 8.4 report the performances of $D_{1\beta}$-$D_{8\beta}$ and $D_{1\theta}$-$D_{8\theta}$ under the eight true distribution scenarios $\mathcal{S}_1$-$\mathcal{S}_8$. The ψ_β-efficiency of each of $D_{1\beta}$-$D_{8\beta}$ under each true distribution scenario $\mathcal{S}_1$-$\mathcal{S}_8$ is obtained by replacing the $mM(\eta^*)$ in the denominator of (7.1) with $\mathbf{X}^{*\prime}\mathbf{X}^*$ where $\mathbf{X}^*$ is the model matrix of the ψ_β-optimal exact design obtained through computer search under the true distribution. Similarly, ψ_θ-efficiency of each of $D_{1\theta}$-$D_{8\theta}$ under each true distribution scenario $\mathcal{S}_1$-$\mathcal{S}_8$ is obtained by replacing the $mM(\eta^\diamond)$ with $\mathbf{X}^{\diamond\prime}\mathbf{X}^\diamond$ where $\mathbf{X}^\diamond$ is the model matrix of the ψ_θ-optimal exact design obtained through computer search under the true distribution.

Designs	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$	$\mathcal{S}_5$	$\mathcal{S}_6$	$\mathcal{S}_7$	$\mathcal{S}_8$
$D_{1\beta}, D_{2\beta}, D_{3\beta}, D_{7\beta}, D_{8\beta}$	0.998	0.979	1.000	0.986	0.994	0.967	0.998	0.975
$D_{4\beta}$	0.996	0.949	0.993	0.959	0.995	0.936	0.994	0.946
$D_{5\beta}$	1.000	0.946	0.998	0.954	1.000	0.933	1.000	0.941
$D_{6\beta}$	0.925	1.000	0.928	0.999	0.920	1.000	0.925	1.000

Table 8.3: ψ_β-efficiency of the designs $D_{1\beta}$-$D_{8\beta}$ relative to ψ_β-optimal exact designs under the eight true σ^2 and $\mathbf{\Lambda}$ distribution scenarios $\mathcal{S}_1$-$\mathcal{S}_8$.

Designs	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$	$\mathcal{S}_5$	$\mathcal{S}_6$	$\mathcal{S}_7$	$\mathcal{S}_8$
$D_{1\theta}$	0.991	0.974	0.992	0.976	0.990	0.972	0.991	0.973
$D_{2\theta}$	0.952	0.995	0.962	0.995	0.942	0.994	0.953	0.994
$D_{3\theta}, D_{7\theta}$	0.998	0.941	0.998	0.945	0.998	0.933	0.998	0.937
$D_{4\theta}$	0.981	0.815	0.986	0.834	0.975	0.792	0.983	0.810
$D_{5\theta}$	1.000	0.943	0.9996	0.948	1.000	0.989	1.000	0.940
$D_{6\theta}$	0.950	1.000	0.956	1.000	0.939	1.000	0.946	1.000
$D_{8\theta}$	0.975	0.987	0.980	0.985	0.970	0.985	0.976	0.983

Table 8.4: ψ_θ-efficiency of the designs $D_{1\theta}$-$D_{8\theta}$ relative to ψ_θ-optimal exact designs under the eight true σ^2 and $\mathbf{\Lambda}$ distribution scenarios $\mathcal{S}_1$-$\mathcal{S}_8$.

While the orthogonal design tends to be efficient especially under the ψ_β criterion, Tables 8.3 and 8.4 show that more efficient designs can be obtained when the assumed prior distributions used in the design construction correctly reflect the true signs of the correlations of the random effects. For example, in Table 8.4, when all pairs of random effects are negatively correlated as in the true distribution scenarios $\mathcal{S}_2$, $\mathcal{S}_4$, $\mathcal{S}_6$ and $\mathcal{S}_8$, designs $D_{2\theta}$, $D_{6\theta}$, and $D_{8\theta}$, obtained under the prior assumptions of negative correlations are more efficient than the orthogonal design. Note that the designs $D_{2\theta}$ and $D_{8\theta}$ are almost as ψ_θ-efficient in these four true distribution scenarios as the design $D_{6\theta}$. However, the computer search of the designs $D_{2\theta}$ and $D_{8\theta}$, especially the locally-optimal design $D_{8\theta}$, requires much less time than the search of the design $D_{6\theta}$. Similarly, when all pairs of random effects are positively correlated as in scenarios $\mathcal{S}_1$, $\mathcal{S}_3$, $\mathcal{S}_5$ and $\mathcal{S}_7$, designs $D_{3\theta}$, $D_{5\theta}$, and $D_{7\theta}$ obtained under the prior distribution assumptions of positive correlations of moderate size are more efficient than the orthogonal design. In addition, the less computation-intensive designs $D_{3\theta}$ and $D_{7\theta}$ are almost as efficient as the more computation-intensive design $D_{5\theta}$. These findings suggest that while orthogonal design is a good design to use when the random effects are not correlated or there is no knowledge on the possible signs of the correlations, more efficient designs can be obtained when there is knowledge on the signs of the correlations.

9 Conclusion and Discussion

In this paper, we have investigated Bayesian optimal designs for hierarchical linear models. We examined two design criteria – the ψ_β criterion for the estimation of the individual-level parameters $\boldsymbol{\beta}_i$ for subject i, and the ψ_θ criterion for the estimation of the hyperparameter vector $\boldsymbol{\theta}$.We focused on a special case in which (i) all subjects receive the same treatments so that $\mathbf{X}_i = \mathbf{X}$, (ii) the response errors are homoscedastic so that $\sigma_i^2 = \sigma^2$ and, (iii) the covariates $\mathbf{Z}_i$ are independent of $\mathbf{X}$.

Findings from our comparisons between the two types of designs imply that (i) designs that are ψ_β-optimal for the estimation of the individual-level parameters $\boldsymbol{\beta}_i$ are not necessarily ψ_θ-optimal for the estimation of the hyperparameter vector $\boldsymbol{\theta}$; (ii) orthogonal designs may not be a good choice when the interest is in the estimation of the hyperparameter vector $\boldsymbol{\theta}$ and the random effects are expected to be correlated; (iii) for the construction of both ψ_β-optimal and ψ_θ-optimal designs, and especially the ψ_θ-optimal designs, it is important to have the prior of the covariance matrix $\boldsymbol{\Lambda}$ of the random effects reflect the expected algebraic signs of the covariance elements. Designs obtained under moderate sized correlations with the anticipated signs are more likely to be efficient under the corresponding design criterion and also robust to varying distributions of response errors and variances of the random effects; (iv) locally-optimal designs, obtained by fixing $\sigma^2 = 1$ and $\boldsymbol{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$ with moderate sized correlations in accordance with the anticipated signs, can be used as good surrogates, especially when the numbers of runs, factors and factor levels are large.

There are a few limitations in this paper that we discuss next. First, in our definition of each design criterion, we have used the pre-posterior risk based on the posterior conditional distribution of the parameter of interest given all nuisance parameters. For example, the ψ_β criterion is defined based on the posterior conditional distribution of $\boldsymbol{\beta}_i$ given nuisance parameters $\boldsymbol{\theta}, \boldsymbol{\Lambda}$ and σ_i^2. As a result, the ψ_β criterion is independent of the covariates in $\mathbf{Z}_i$. If the full posterior distribution of $\boldsymbol{\beta}_i$ is used where all nuisance parameters $\boldsymbol{\theta}, \boldsymbol{\Lambda}$ and σ_i^2 are integrated out, or if the marginal posterior distribution of $\boldsymbol{\beta}_i$ is used where $\boldsymbol{\theta}$ is integrated out given $\boldsymbol{\Lambda}$ and σ_i^2, then the resulting design criterion would depend on the covariates in

$\mathbf{Z}_i$, with a much higher level of complexity than the ψ_β criterion used in this paper. In particular, the design criteria based on the full posterior or marginal posterior distributions are not in closed forms and can only be approximated through simulations. Although Han and Chaloner (2004) have shown that MCMC within Monte Carlo simulation can be used to compare a pre-determined set of candidate designs, and Clyde, Muller, and Parmigiani (1996) used MCMC within MCMC to find optimal sample size a simple design problem with only one variable, the computation becomes prohibitive for the search of an optimal design over a large design space that involves multiple variables. Nevertheless, with new computational advancements and more computational power, we believe that it will be an interesting future research direction to explore the design criteria based on full or marginal posterior distributions.

Next, related to the first point, it would also be interesting to explore other Bayesian design criteria such as the Bayesian A-criterion where the pre-posterior risk is the expected squared error loss. Chaloner and Verdinelli (1995) provide some excellent examples of utility/loss functions that lead to different Bayesian design criteria with different focus of interest. Designs obtained under these alternative criteria may be quite different from what we have investigated in this paper. However, for the design criterion that involves the maximization of the expected gain in Shannon information (Lindley, 1956) based on the posterior conditional distribution of $\boldsymbol{\beta}_i$ or $\boldsymbol{\theta}$, we note that it can be approximated by the ψ_β criterion as the maximization of (3.4) or the ψ_θ criterion as the maximization of (3.9). This can be seen by following the same approach taken in Liu et al. (2007).

Last, in this paper we restricted to a special case where each respondent gets the same design (homogeneous), response errors are homoscedastic, and the covariates are determined independently of the treatments . A natural extension is to the general case of heterogeneous designs where different respondents with heteroscedastic response errors are given different designs and the specification of the treatment allocation may or may not be independent from that of the covariates. While a jointly optimal allocation of the treatments ($\mathbf{X}_i$) and the covariates ($\mathbf{Z}_i$) may be elusive due to a large number of possible combinations, it is possible to find through computer search optimal treatment allocations for different subjects given the knowledge of the covariates associated with the subjects ($\mathbf{X}_i|\mathbf{Z}_i$) or vice versa ($\mathbf{Z}_i|\mathbf{X}_i$, see an example in Liu et al. 2009).

Acknowledgment This paper is based on the first author's Ph.D. dissertation. The research was supported in part by NSF Grant SES-0437251. The authors would like to thank the reviewers for their helpful comments and suggestions.

Appendix

A: Proof of Theorem 3

Proof. For $\mathbf{\Lambda}$ given in (6.10) and $\mathbf{M}(\eta^*) = \mathbf{I}$, the left hand side of (6.4) is

$$
\begin{aligned}
&\int \left\{ \sum_{k=0}^{p-1} x_k^2 \left(\tfrac{m}{\sigma^2} + \lambda_k^{-2}\right)^{-1} \right\} f(\mathbf{\Lambda}) f(\sigma^2) d\mathbf{\Lambda} d\sigma^2 \\
=& \sum_{k=0}^{p-1} x_k^2 \int \left(\tfrac{m}{\sigma^2} + \lambda_k^{-2}\right)^{-1} f(\mathbf{\Lambda}) f(\sigma^2) d\mathbf{\Lambda} d\sigma^2 \\
=& E\left[\left(\tfrac{m}{\sigma^2} + \lambda_0^{-2}\right)^{-1}\right] \times \sum_{k=0}^{p-1} x_k^2 \quad \text{since } \lambda_0^2, \ldots, \lambda_{p-1}^2 \text{ are identically distributed,} \\
\leq& E\left[\left(\tfrac{m}{\sigma^2} + \lambda_0^{-2}\right)^{-1}\right] \times p \quad \text{by (6.9)} \\
=& E\left\{Tr\left[\left(\tfrac{m}{\sigma^2} + \lambda_0^{-2}\right)^{-1}\right) \mathbf{I}_p\right]\right\} \\
=& \int \left\{Tr\left[\tfrac{m}{\sigma^2} I + M(\eta^*)^{-1} \mathbf{\Lambda}^{-1}\right]^{-1}\right\} f(\mathbf{\Lambda}) f(\sigma^2) d\mathbf{\Lambda} d\sigma^2,
\end{aligned}
$$

which is the right hand side of (6.4). Therefore η^* is ψ_β-optimal from Theorem 1. Similarly, expression (6.5) holds for η^* and therefore it is also ψ_θ-optimal from Theorem 2.

□

B: Proof of Theorem 4

Proof. Let $\mathbf{M}(\eta^*)$ be as defined in the statement of the theorem. For $\mathbf{M}(\eta^*)$ to be positive definite, κ needs to satisfy the condition that $-1 < \kappa < \frac{1}{p-1}$. The inverse of $\mathbf{M}(\eta^*)$ is

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

In addition, with $\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$, we have

$$\begin{aligned}\mathbf{M}(\eta^*)^{-1}\mathbf{\Lambda}^{-1} &= \frac{1}{\tilde{a}(1+\kappa)}\left[\mathbf{I} + \frac{\kappa}{1-(p-1)\kappa}\mathbf{J}\right]\left(\mathbf{I} - \frac{\tilde{d}}{\tilde{a}+p\tilde{d}}\mathbf{J}\right), \\ &= \frac{1}{\tilde{a}(1+\kappa)}\left[\mathbf{I} + \frac{(\tilde{a}+p\tilde{d})\kappa - \tilde{d}(1+\kappa)}{(\tilde{a}+p\tilde{d})\left[1-(p-1)\kappa\right]}\mathbf{J}\right].\end{aligned}$$

$$\begin{aligned}\left[\frac{m}{\sigma^2}\mathbf{I} + \mathbf{M}(\eta^*)^{-1}\mathbf{\Lambda}^{-1}\right]^{-1} &= \frac{\tilde{a}(1+\kappa)\sigma^2}{m\tilde{a}(1+\kappa)+\sigma^2} \\ &\quad \left\{\mathbf{I} + \left(\frac{1}{1+\kappa}\right)\left(\frac{\sigma^2\tilde{d}(1+\kappa) - \sigma^2(\tilde{a}+p\tilde{d})\kappa}{m\tilde{a}(\tilde{a}+p\tilde{d})[1-(p-1)\kappa] + \tilde{a}\sigma^2}\right)\mathbf{J}\right\}.\end{aligned}$$

The right hand side of (6.4) is

$$\begin{aligned}&E_{\tilde{a},\tilde{d},\sigma^2}\left\{Tr\left[\tfrac{m}{\sigma^2}\mathbf{I} + \mathbf{M}(\eta^*)^{-1}\mathbf{\Lambda}^{-1}\right]^{-1}\right\} \\ &= E_{\tilde{a},\tilde{d},\sigma^2}\left\{\frac{p\tilde{a}\sigma^2}{m\tilde{a}(1+\kappa)+\sigma^2}\left[1 + \frac{\kappa^2 m\tilde{a}(\tilde{a}+p\tilde{d})(1-p) + \kappa[m\tilde{a}(\tilde{a}+p\tilde{d}) + \tilde{d}(1-p)\sigma^2] + d\sigma^2}{m\tilde{a}(\tilde{a}+p\tilde{d})[1-(p-1)\kappa] + \tilde{a}\sigma^2}\right]\right\}. \\ &= E_{\tilde{a},\tilde{d},\sigma^2}\left\{\frac{p\tilde{a}\sigma^2}{m\tilde{a}(1+\kappa)+\sigma^2} + \left(\frac{p\left[(1-p)\kappa+1\right]\tilde{a}\sigma^2}{m\tilde{a}(1+\kappa)+\sigma^2}\right)\left(\frac{\kappa m\tilde{a}(\tilde{a}+p\tilde{d}) + \tilde{d}\sigma^2}{m\tilde{a}(\tilde{a}+p\tilde{d})[1-(p-1)\kappa] + \tilde{a}\sigma^2}\right)\right\}.\end{aligned}$$

Using (6.11), the expectation (over $\tilde{a}$, $\tilde{d}$, and σ^2) of the second item inside the curly bracket of the last equation becomes 0 and we obtain

$$E_{\tilde{a},\tilde{d},\sigma^2}\left\{Tr[\tfrac{m}{\sigma^2}\mathbf{I} + \mathbf{M}(\eta^*)^{-1}\mathbf{\Lambda}^{-1}]^{-1}\right\} = E_{\tilde{a},\sigma^2}\left\{\frac{p\tilde{a}\sigma^2}{m\tilde{a}(1+\kappa)+\sigma^2}\right\}.$$

On the left-hand side,

$$\begin{aligned}\left[\tfrac{m}{\sigma^2}\mathbf{M}(\eta^*) + \mathbf{\Lambda}^{-1}\right]^{-1} &= \frac{\tilde{a}\sigma^2}{m\tilde{a}(1+\kappa)+\sigma^2}\mathbf{I} \\ &\quad + \left(\frac{\tilde{a}\sigma^2}{m\tilde{a}(1+\kappa)+\sigma^2}\right)\left(\frac{\kappa m\tilde{a}(\tilde{a}+p\tilde{d}) + \tilde{d}\sigma^2}{m\tilde{a}(\tilde{a}+p\tilde{d})[1-(p-1)\kappa] + \tilde{a}\sigma^2}\right)\mathbf{J}.\end{aligned}$$

Using (6.11), the expectation of the second item in the last equation becomes $\mathbf{0}$ and the left hand side of (6.4) becomes

$$E_{\tilde{a},\tilde{d},\sigma^2}\left\{\mathbf{x}'\left[\tfrac{m}{\sigma^2}\mathbf{M}(\eta^*) + \mathbf{\Lambda}^{-1}\right]^{-1}\mathbf{x}\right\} = E_{\tilde{a},\sigma^2}\left\{\frac{\tilde{a}\sigma^2}{m\tilde{a}(1+\kappa)+\sigma^2}\right\}\mathbf{x}'\mathbf{x}.$$

Since $\mathbf{x}'\mathbf{x} \le p$ from (6.9), the theorem follows from Theorem 1. □

C: Proof of Theorem 5

Proof. Let $\mathbf{M}(\eta^\diamond)$ be as defined in the statement of the theorem. For $\mathbf{M}(\eta^\diamond)$ to be positive definite, ϵ needs to satisfy the condition that $-1 < \epsilon < \frac{1}{p-1}$. The inverse of $\mathbf{M}(\eta^\diamond)$ is

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

In addition, with $\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$, we have

$$\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}(\eta^\diamond)\mathbf{\Lambda} = \frac{m\tilde{a}(1+\epsilon)+\sigma^2}{m}\mathbf{I} + \left[\tilde{d}(1+\epsilon) - (\tilde{a}+p\tilde{d})\epsilon\right]\mathbf{J},$$

$$\left[\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}(\eta^\diamond)\mathbf{\Lambda}\right]^{-1} = \frac{m}{m\tilde{a}(1+\epsilon)+\sigma^2}\left[\mathbf{I} - \frac{m\tilde{d}(1+\epsilon) - m(\tilde{a}+p\tilde{d})\epsilon}{\sigma^2 + m(\tilde{a}+p\tilde{d})(1-(p-1)\epsilon)}\mathbf{J}\right],$$

$$\mathbf{M}(\eta^\diamond)^{-1}\left[\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}(\eta^\diamond)\mathbf{\Lambda}\right]^{-1} = \frac{m}{(1+\epsilon)[m\tilde{a}(1+\epsilon)+\sigma^2]}$$
$$\left\{\mathbf{I} - \frac{\epsilon^2[m(\tilde{a}+p\tilde{d})(p-2) + m\tilde{d}] - \epsilon[2m(\tilde{a}+p\tilde{d}) + \sigma^2 - 2m\tilde{d}] + m\tilde{d}}{[1-(p-1)\epsilon][\sigma^2 + m(\tilde{a}+p\tilde{d})(1-(p-1)\epsilon)]}\mathbf{J}\right\}.$$

Using (6.12), the expectation (over $\tilde{a}$, $\tilde{d}$ and σ^2) of the coefficient of $\mathbf{J}$ in the last equation becomes $\mathbf{0}$ and the left hand side of (6.5) becomes

$$E_{\tilde{a},\tilde{d},\sigma^2}\left\{\mathbf{x}'\mathbf{M}(\eta^\diamond)^{-1}[\mathbf{I} + \mathbf{M}(\eta^\diamond)\mathbf{\Lambda}]^{-1}\mathbf{x}\right\} = E_{\tilde{a},\sigma^2}\left\{\frac{m}{(1+\epsilon)[m\tilde{a}(1+\epsilon)+\sigma^2]}\right\}\mathbf{x}'\mathbf{x}.$$

The right hand side of (6.5) is

$$E_{\tilde{a},\tilde{d},\sigma^2}\left\{Tr[\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}(\eta^\diamond)\mathbf{\Lambda}]^{-1}\right\} = E_{\tilde{a},\tilde{d},\sigma^2}\left\{\frac{pm}{(1+\epsilon)[m\tilde{a}(1+\epsilon)+\sigma^2]}\right\}$$
$$- E_{\tilde{a},\tilde{d},\sigma^2}\left\{\left(\frac{pm}{1+\epsilon}\right)\left(\frac{\epsilon^2[m(\tilde{a}+p\tilde{d})(p-2) + m\tilde{d}] - \epsilon[2m(\tilde{a}+p\tilde{d}) + \sigma^2 - 2m\tilde{d}] + m\tilde{d}}{[m\tilde{a}(1+\epsilon)+\sigma^2]\left[\sigma^2 + m(\tilde{a}+p\tilde{d})(1-(p-1)\epsilon)\right]}\right)\right\}.$$

Using (6.12), the expectation of the second item in the last equation becomes 0 and we obtain

$$E_{\tilde{a},\tilde{d},\sigma^2}\left\{Tr[\tfrac{\sigma^2}{m}\mathbf{I} + \mathbf{M}(\eta^\diamond)\mathbf{\Lambda}]^{-1}\right\} = E_{\tilde{a},\sigma^2}\left\{\frac{pm}{(1+\epsilon)[m\tilde{a}(1+\epsilon)+\sigma^2]}\right\}.$$

Since $\mathbf{x}'\mathbf{x} \le p$ from (6.9), the theorem follows from Theorem 2. □

References

Allenby, G. M. and Ginter, J. L. (1995), 'Using extremes to design products and segment markets', *Journal of Marketing Research* **32 (4)**, 392–403.

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

Barnard, J., McCulloch, R. and Meng, X.-L. (2000), 'Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage', *Statistica Sinica* **20**, 1281–1311.

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, New York: Springer.

Chaloner, K. (1984), 'Optimal bayesian experimental design for linear models', *The Annals of Statistics* **12 (1)**, 283–300.

Chaloner, K. and Verdinelli, I. (1995), 'Bayesian experimental design: A review', *Statistical Science* **10 (3)**, 273–304.

Clyde, M., Muller, P. and Parmigiani, G. (1996), 'Exploring expected utility surfaces by markov chains', Discussion paper 95–39, Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina.

Daniels, M. J. and Kass, R. E. (1999), 'Nonconjugate bayesian estimation of covariance matrices and its use in hierarchical models', *Journal of the American Statistical Association* **94**, 1254–1263.

Entholzner, M., Benda, N., Schmelter, T. and Schwabe, R. (2005), 'A note on designs for estimating population parameters', *Biometrical Letters – Listy Biometryczne* pp. 25–41.

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

Ferreira, P. E. (1981), 'Extending fisher's measure of information', *Biometrika* **68 (3)**, 695–698.

Giovagnoli, A. and Sebastiani, P. (1989), 'Experimental designs for mean and variance estimation in variance components models', *Computational Statistics and Data Analysis* **8**, 21–28.

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

Han, C. and Chaloner, K. (2004), 'Bayesian experimental design for nonlinear mixed-effects models with applications to hiv dynamics', *Biometrics* **60**, 25–33.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. and Lyons, T. (1991), 'Early vocabulary growth: Relation to language input and gender', *Developmental Psychology* **27 (2)**, 236–248.

Kuhfeld, W. F. (2005), Experimental design, efficiency, coding and choice designs, Technical report, TS-722C, SAS Institute, `http://support.sas.com/techsup/tnote/tnote_stat.html`.

Lenk, P. J., Desarbo, W. S., Green, P. E. and Young, M. R. (1996), 'Hierarchical bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs', *Marketing Science* **15 (2)**, 173–191.

Lindley, D. (1956), 'On the measure of information provided by an experiment', *Annals of Statistics* **27**, 986–1005.

Liu, Q., Dean, A. M. and Allenby, G. M. (2007), 'Design for hyperparameter estimation in linear models', *Journal of Statistical Theory and Practice* **1**, 311–328.

Liu, Q., Dean, A. M., Bakken, D. and Allenby, G. M. (2009), 'Studying the level-effect in conjoint analysis: Application of efficient experimental designs for hyperparameter estimation', *Quantitative Marketing and Economics* **7**, 69–93.

Lohr, S. L. (1995), 'Optimal bayesian design of experiments for the one-way random effects model', *Biometrika* **82 (1)**, 175–186.

Morrison, D. F. (1990), *Multivariate Statistical Methods,*, New York: McGraw-Hill, Inc.

Pinheiro, J. C. and Bates, D. M. (1996), 'Unconstrained parameterizations for variance-covariance matrices', *Statistics and Computing* **6**, 289–296.

Pourahmadi, M. (1999), 'Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation', *Biometrika* **86 (3)**, 677–690.

Pourahmadi, M. (2000), 'Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix', *Biometrika* **87 (2)**, 425–435.

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

Rossi, P. E., Allenby, G. M. and McCulloch, R. (2005), *Bayesian Statistics and Marketing*, John Wiley and Sons, Ltd.

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

Smith, A. and Verdinelli, I. (1980), 'A note on bayesian designs for inference using a hierarchical linear model', *Biometrika* **67**, 613–619.

Yang, R. and Berger, J. O. (1994), 'Estimation of a covariance matrix using the reference prior', *The Annals of Statistics* **22 (3)**, 1195–1211.

Yuh, L., Beal, S., Davidian, M., Harrison, F., Hester, A., Kowalski, K., Vonesh, E. and Wolfinger, R. (1994), 'Population pharmacokinetics/pharmacodynamics methodology and applications: A bibliography', *Biometrics* **50**, 566–575.

Department of Marketing, University of Wisconsin-Madison

E-mail: (qliu@bus.wisc.edu) Phone/Fax: 608-263-9298

Department of Statistics, The Ohio State University

E-mail: (amd@stat.osu.edu)

Department of Marketing, The Ohio State University

E-mail: (allenby.1@osu.edu)