Statistica Sinica **22** (2012), 393-417
doi:http://dx.doi.org/10.5705/ss.2009.319

BAYESIAN DESIGNS FOR HIERARCHICAL LINEAR MODELS

Qing Liu[1], Angela M. Dean[2] and Greg M. Allenby[2]

[1] *University of Wisconsin-Madison and* [2] *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 $\boldsymbol{\theta}$. We focus on a specific case in which all subjects receive the same set of treatments and in which the covariates are independent of treatments. We obtain the explicit structure of ψ_β- and ψ_θ- optimal continuous (approximate) designs for the case of independent random effects, and for some special cases of correlated random effects. Through examples and simulations, we 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, but sensitive to the specifications of the signs of the correlations of the random effects.

Key words and phrases: Bayesian design, D-optimality, design robustness, hierarchical linear model, hyperparameter, random effects model.

1. Introduction

Over the past two decades, various forms of hierarchical models have been used in such fields as the social and behavioral sciences, agriculture, education, medicine, healthcare studies, and marketing. These models have been termed "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 commonly-used two-level hierarchical models, 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 hyperparameters in the second-level of the hierarchy (see Section 2). Hyperparameters may reflect population characteristics, 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 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 a product feature change (Allenby and Ginter (1995)). In situations such as direct marketing, 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 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.

Most research on efficient designs under hierarchical linear models has focused on non-Bayesian designs that assume fixed values for the variance and covariance parameters (local designs). For example, Giovagnoli and Sebastiani (1989) used a local design criterion for the estimation of hyperparameters in the one-way random effects model with a 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, assuming independent, identically distributed random effects. Fedorov and Hackl (1997, p.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, along with some optimal two-factor designs in the uncorrelated setting.

Bayesian designs for hierarchical linear models, on the other hand, take into account the uncertainty of the model parameters. Smith and Verdinelli (1980) investigated Bayesian designs 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 elements of 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 – the ψ_β criterion for the estimation of the individual-level effects for each respondent, and the ψ_θ criterion 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 when the random effects are correlated (see Sections 6 and 7).

The paper is organized as follows. In Section 2, we describe the hierarchical linear model, and in Section 3 we specify the two Bayesian design criteria investigated. 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 in which 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 the case of independent random effects, and for some specific cases 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 discussion in Section 9.

2. The Model

We take a hierarchical linear model of the 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 ($i = 1, \ldots n$) 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 with information on 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 (see, for example, Rossi, Allenby, and McCulloch (2005)):

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

They are replaced by more informative priors when information is available. There has been discussion on the appropriate diffuse prior to use for $\boldsymbol{\Lambda}$. The standard Jefferey's prior is generally not recommended in this context due to its inadequacies in higher dimensions (see Yang and Berger (1994, Sec. 2.2)). The Inverted Wishart prior has also been criticized 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 decompositions of the covariance matrix, by, Yang and Berger (1994), Pinheiro and Bates (1996), and Pourahmadi (1999), for example. We follow the formulation of Barnard, McCulloch, and Meng (2000) and break the covariance matrix down to components of variances v_{ii} and correlations $r_{ij} = r_{ji}$, with covariances $v_{ij} = r_{ij}\sqrt{v_{ii}v_{jj}}$ for $i < j, i, j \in \{1, 2, \ldots, p\}$. We assume Inverse Gamma distributions on the variance components and allow for different degrees of freedom. Correlation components 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 $\mathbf{\Lambda}$. Let $R = \{r_{ij}\}$ and let $f(R)$ denote the probability density function of R. Our priors for the components of $\mathbf{\Lambda}$ 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 our Bayesian D-criterion for the estimation of individual-level effects $\boldsymbol{\beta}_i$ for subject i. In Section 3.2, we define our Bayesian D-criterion for the estimation of the hyperparameter vector $\boldsymbol{\theta}$, where the covariates $\mathbf{Z}_i$ can be controlled by the experimenter. We call these the ψ_β and ψ_θ criteria, respectively.

Following Chaloner (1984, p.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.

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

In this section, we suppose 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. 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 d_i that minimizes the pre-posterior risk

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

Here $\mathcal{I}_{GFIM}$ is the generalized Fisher Information matrix (see Ferreira (1981)) which, under the assumption of normality, is

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

where $\boldsymbol{\beta}_i'$ is the transpose of $\boldsymbol{\beta}_i$. Since the posterior $f(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\theta}, \mathbf{\Lambda}, \sigma_i^2)$ is a normal density function with covariance matrix $(\sigma_i^{-2}\mathbf{X}_i'\mathbf{X}_i + \mathbf{\Lambda}^{-1})^{-1}$ 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 of Rossi, Allenby, and McCulloch (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 $\boldsymbol{\beta}_i$ or $\boldsymbol{\theta}$. 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. In this paper, since the number of parameters p in vector $\boldsymbol{\beta}_i$ is fixed for any given experiment, a ψ_β-optimal design is one 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 (3.4) involves only the model matrix $\mathbf{X}_i$ and does not depend on the covariate matrix $\mathbf{Z}_i$.

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

In this section, we suppose 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 write

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

(see Lenk et al. (1996, p.187)).

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 $\tilde{d} = (d_1, \ldots, d_n)$, let $\widetilde{\mathbf{X}}' = (\mathbf{X}_1', \ldots, \mathbf{X}_n')$ 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 $\mathbf{\Lambda}$ and $\boldsymbol{\varsigma}$ is normal with mean vector and covariance matrix

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

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

(see Chapter 2 of Rossi, Allenby, and McCulloch (2005)), we have

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

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

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

When 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, a ψ_θ-optimal design maximizes

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

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, and an approximation to (3.8) is

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

A ψ_θ-optimal design is taken to be 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 $\{\mathbf{Z}_i\}, i = 1, \ldots, n$.

Frequently, as in many survey studies, experimenters have control over the sampling of subjects on the basis of covariates such as gender and age, as well as independent control over the treatment allocation which determines the model matrix $\mathbf{X}_i$. In some circumstances, such as the study of the "level effect" (see Liu et al. (2009)). the covariate matrix $\mathbf{Z}_i$ can be controlled but not determined independently of $\mathbf{X}_i$. In other settings, for example when subjects are pre-designated or scarce, the sampling of the subjects cannot be controlled and the covariate matrix $\mathbf{Z}_i$ is treated as fixed in (3.8) and (3.9).

Here we consider that both treatment allocation and selection of covariates can be controlled and 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 are fixed for all survey responses from respondent i. 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 $\mathbf{Z}_i = \mathbf{z}_i' \otimes \mathbf{I}_p$ 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

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

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

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 $\mathbf{X}$.

$\psi_{\boldsymbol{\beta}}$ criterion The ψ_β criterion involves the search of an m-point design in $\mathcal{D}(m)$, with model matrix $\mathbf{X}$, that maximizes (3.4) which becomes

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

$\psi_{\boldsymbol{\theta}}$ criterion The covariate matrix $\widetilde{\mathbf{Z}}' = (\mathbf{Z}'_1, \ldots, \mathbf{Z}'_n)$ is determined independently of $\mathbf{X}$ and, by Theorem 8.8.10 of Graybill (1983), (4.2) simplifies to

$$\log\left\{\left|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right|^{q/p}\Big|\sum_{i=1}^{n}(\mathbf{z}_i\mathbf{z}_i')\Big|^{p}\right\} = \tfrac{q}{p}\log\left|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right| + p\log\Big|\sum_{i=1}^{n}(\mathbf{z}_i\mathbf{z}_i')\Big|. \quad (5.2)$$

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, Chapters 10 and 11 of Atkinson and Donev (1992)). We 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, with $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_m + \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}'$, the ψ_θ criterion involves the search of a design in $\mathcal{D}(m)$, with model matrix $\mathbf{X}$, that maximizes

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

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) that is more convenient in the design search. The proof follows from Morrison (1990, p.69) by letting $\mathbf{A} = \sigma^2\mathbf{I}_m$, $\mathbf{B} = \mathbf{X}$, $\mathbf{C} = \boldsymbol{\Lambda}$, and noting that $\mathbf{I}_p = (\boldsymbol{\Lambda}^{-1} + \sigma^{-2}\mathbf{X}'\mathbf{X})^{-1}(\boldsymbol{\Lambda}^{-1} + \sigma^{-2}\mathbf{X}'\mathbf{X})$.

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

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

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 $\boldsymbol{\Lambda}$ and σ^2 are degenerate. For example, when $\boldsymbol{\Lambda}$ and σ^2 are fixed, the ψ_θ criterion is equivalent to the minimization of $\left|(\mathbf{X}'\mathbf{X})^{-1} + (\boldsymbol{\Lambda}/\sigma^2)\mathbf{I}_p\right|$, which is the non-Bayesian criterion used by Fedorov and Hackl (1997, Eq. 5.2.6), and the "mixed-effects model D-criterion" used by Entholzner et al. (2005).

6. Theoretical Results on $\psi_{\boldsymbol{\beta}}$- and $\psi_{\boldsymbol{\theta}}$-Optimal Designs

In this section we identify ψ_β-optimal designs and ψ_θ-optimal designs for both the case of independent random effects and some special cases of correlated random effects under assumptions (i), (ii), (iii) in Section 5. We employ the standard continuous design framework (see, for example, Silvey (1980) and Pukelsheim (1993, p.26)).

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 $\mathbf{x}$. Here, $\mathbf{X}'\mathbf{X} = m\int \mathbf{x}\mathbf{x}'d\eta(\mathbf{x}), \eta \in \mathcal{H}, \mathbf{x} \in \mathcal{X}$, and we take

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

Following (Silvey, 1980, p.16), 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^* maximizes

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

Similarly, a ψ_{θ}-optimal continuous design $\eta^{\diamond}$ maximizes

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

Lemma 2. *The functions of* (6.2) *and* (6.3) *are concave and monotone in* $\mathcal{M}$.

The proof of Lemma 2 follows since the integrands in (6.2) and (6.3) are monotone and concave (see Chaloner (1984); Fedorov and Hackl (1997, p.31), and integration is a linear operation.

Theorem 1. *A design* η^* *is* ψ_{β}*-optimal if and only if*

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

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

Theorem 2. *A design* $\eta^{\diamond}$ *is* ψ_{θ}*-optimal if and only if*

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

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, using (2.6.11) of Fedorov and Hackl (1997), the proofs of Theorems 1 and 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, and Theorem 3.7 in Silvey (1980) or Theorem 2.3.2 in Fedorov and Hackl (1997).

6.2. The design space

For the remainder of the paper, the design space $\mathcal{X}$ is taken to be

$$\mathcal{X} = \Big\{ \mathbf{x} = [x_0, x_1, \ldots, x_{p-1}]' \quad \text{such that} \quad x_0 = 1 \quad \text{and} \quad \sum_{k=0}^{p-1} x_k^2 \leq p \Big\}, \tag{6.6}$$

where p is the number of parameters or, equivelently, the number of columns of the model matrix $\mathbf{X}$. Thus, when there are two treatment factors having two levels each, the rows $\mathbf{x}$ of $\mathbf{X}$ can be $[1,-1,-1]'$, $[1,-1,\ 1]'$, $[1,\ 1,-1]'$, $[1,\ 1,\ 1]'$. Similarly, for a factor with three levels where the objective is to measure the linear and quadratic contrasts, one might take the rows of $\mathbf{X}$, corresponding to the three different levels of the treatment factor as

$$\left[1, \frac{-\sqrt{3}}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right]', \quad \left[1,\ 0, \frac{2}{\sqrt{2}}\right]', \quad \left[1, \frac{\sqrt{3}}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right]'$$

(see the "standardized orthogonal effects coding" of Kuhfeld (2005)).

6.3. Independent random effects

Here we identify a 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 identical prior knowledge on the variances of the random effects. The proof of Theorem 3 is provided in the Appendix.

Theorem 3. *Let the random individual-level effects $\boldsymbol{\beta}_i$ in* (2.1) *be independent such that*

$$\mathbf{\Lambda} = Diag\left(\lambda_0^2,\ \lambda_1^2,\ \ldots,\ \lambda_{p-1}^2\right), \tag{6.7}$$

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

Note that a design with an equal number of occurrences of every treatment combination, called a level-balanced orthogonal design, has $\mathbf{X}'\mathbf{X} = m\mathbf{I}$ under the standardized orthogonal effects coding of the model matrix $\mathbf{X}$, and so $\mathbf{M}(\eta) = \mathbf{I}$; such a design 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, one must have $\tilde{a} > 0$ and $\tilde{a} + p\tilde{d} > 0$. The prior distribution of $\boldsymbol{\Lambda}$ in this case reduces to the prior distributions of $\tilde{a}$ and $\tilde{d}$.

Information matrices of ψ_β- and ψ_θ-optimal designs 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 $\eta^\diamond$, for any prior distributions (diffuse or more informative) of σ^2, $\tilde{a}$, and $\tilde{d}$, as long as σ^2 is positive and $\boldsymbol{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$ is positive definite. These optimal continuous designs provide efficiency bounds for exact designs (see Section 7). Proofs of Theorems 4 and 5 are provided in the Appendix.

Theorem 4. *Given $\boldsymbol{\Lambda}$ 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*

$$E_{\tilde{a},\tilde{d},\sigma^2}\left\{\left(\frac{\tilde{a}\sigma^2}{m\tilde{a}(1+\kappa)+\sigma^2}\right)\left(\frac{\kappa m\tilde{a}(\tilde{a}+p\tilde{d})+\tilde{d}\sigma^2}{m\tilde{a}(\tilde{a}+p\tilde{d})[1-(p-1)\kappa]+\tilde{a}\sigma^2}\right)\right\} = 0, \quad (6.8)$$

and $\kappa \in (-1, 1/(p-1))$.

Theorem 5. *Given $\boldsymbol{\Lambda}$ 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*

$$E_{\tilde{a},\tilde{d},\sigma^2}\left\{\frac{\epsilon^2[m(\tilde{a}+p\tilde{d})(p-2)+m\tilde{d}]-\epsilon[2m(\tilde{a}+p\tilde{d})+\sigma^2-2m\tilde{d}]+m\tilde{d}}{[m\tilde{a}(1+\epsilon)+\sigma^2][\sigma^2+m(\tilde{a}+p\tilde{d})(1-(p-1)\epsilon)]}\right\} = 0, \quad (6.9)$$

and $\epsilon \in (-1, \frac{1}{p-1})$.

Note that, although there are no closed-form solutions for κ and ϵ in (6.8) and (6.9), we can use a grid search within the interval $(-1, 1/(p-1))$ to find approximate solutions. Theorems 4 and 5 can easily be extended to more general settings where the random effects are interchangeable within groups and independent between groups.

Corollary 1. *Given $\boldsymbol{\Lambda}$ of the block diagonal form $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_0, \boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \ldots, \boldsymbol{\Lambda}_G\}$, where $\lambda_0 > 0$, $\boldsymbol{\Lambda}_g = \tilde{a}_g\mathbf{I}_{p_g} + \tilde{d}_g\mathbf{J}_{p_g}$ 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 $g = 1, \ldots, G$, a design η ($\eta = \eta^*$ or $\eta^\diamond$) with $\mathbf{M}(\eta) = \mathrm{diag}\{1, \mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_G\}$ satisfies the following:*

(i) *It 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 $\kappa_g \in (-1, 1/(p_g-1))$ *for* $g = 1,\ldots,G$.

(ii) *It 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 $\epsilon_g \in (-1, 1/(p_g-1))$ *for* $g = 1,\ldots,G$.

Theorems 3, 4, and 5, together with Corollary 1, suggest that the matrix $\mathbf{M}(\eta^*)$ $(\mathbf{M}(\eta^\diamond))$ of a ψ_β-optimal (ψ_θ-optimal) design often has a structure similar to the covariance matrix, $\mathbf{\Lambda}$, of the random effects.

Examples of ψ_β- and ψ_θ-optimal continuous designs Consider an experiment with two treatment factors each with two levels under a hierarchical linear model. With (i), (ii), and (iii) from Section 5, 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 $p = 3$.

The prior distributions 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 $\mathbf{\Lambda} = \tilde{a}\mathbf{I}+\tilde{d}\mathbf{J}$. We examine the situation when the correlation $\tilde{d}/(\tilde{a}+\tilde{d})$ is Uniform $(-0.5, 1)$, as well as three separate situations where the correlation is assumed to be negative, low positive, and high positive, with Uniform $(-0.5, 0)$, $(0, 0.5)$, and $(0.5, 1)$ priors, respectively. Note that through variable transformation, we can obtain the corresponding priors for $\tilde{a}$ and $\tilde{d}$. For these four prior distributions, and $m = 12$ observations per subject, Table 1 shows the κ and ϵ values corresponding to the ψ_β- and ψ_θ-optimal continuous designs of Theorems 4 and 5, respectively. We used a grid search to find the κ and ϵ values to satisfy equations (6.8) and (6.9). Under given values of κ and ϵ, the Monte Carlo method was used to obtain the expectation over $(\tilde{a}, \tilde{d}, \sigma^2)$. Fixing the support points of the continuous designs to correspond to $\mathbf{x}_1 = [1,-1,-1]'$, $\mathbf{x}_2 = [1,-1,\ 1]'$, $\mathbf{x}_3 = [1,\ 1,-1]'$ and $\mathbf{x}_4 = [1,\ 1,\ 1]'$, we also report weights on these four support points, denoted m_{11}, m_{12}, m_{21}, and m_{22}, respectively, that give rise to the optimal designs.

The results in Table 1 show that ψ_β- and ψ_θ-optimal continuous designs have opposite signs on the off-diagonal for the special scenarios considered in

Table 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.

$m = 12,\ \ p = 3,\ \ \sigma^2 \sim IG(1.5, 0.5)$		
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)

Theorems 4 and 5, with $\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$. In addition, the magnitudes of ϵ are larger 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.

7. ψ_β- and ψ_θ- Optimal Exact Designs

In this section we obtain, through computer search, optimal exact designs that have integer numbers of observations at the design points. We examine ψ_β-optimal and ψ_θ-optimal exact designs for the special forms of the random effects covariance matrix $\mathbf{\Lambda}$ seen in Table 1, and some more general forms. The design setting used in the examples above is repeated here, and the assumptions (i), (ii), and (iii) made in Section 5. Simple-exchange algorithms (see Atkinson and Donev (1992, Chap. 15)) are used to obtain the ψ_β- and ψ_θ-optimal exact designs. The 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. From Theorems 3, 4, 5, and Corollary 1, we know explicit forms of ψ_β- and ψ_θ-optimal continuous designs. These 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 (which has $\mathbf{M}(\eta) = \mathbf{I}$) as the base design and calculate the relative ψ_β-efficiency of a design with model matrix $\mathbf{X}$ as

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

and the relative ψ_θ-efficiency 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 2 provides the ψ_β-optimal and ψ_θ-optimal exact designs as found through computer search for independent random effects and for equally correlated random effects with equal variances. Designs are given 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 $\mathbf{X}'\mathbf{X}$ are also reported. The last column of Table 2 shows that all optimal exact designs obtained through computer search have efficiencies over 99% relative to their continuous counterparts.

Exact designs obtained from rounding the weights of the continuous designs in Table 1 tend to be the same as or very close to the exact optimal designs found through computer search. For example, the rounding procedure in Pukelsheim (1993, p.309) leads to the exact design (2,2,2,6) under the ψ_θ criterion for the second case in Table 1. This design has efficiency 99.71% relative to the optimal continuous design, and is almost as efficient as the the optimal exact design (3,2,2,5) obtained through computer search and listed in Table 2 with 99.96% efficiency. Similarly, the rounding procedure leads to six possible designs (3,3,2,4), (3,2,3,4), (2,3,3,4), (2,3,2,5), (2,2,3,5), and (3,2,2,5) for the last case in Table 1 under the ψ_β criterion. The last three of these have efficiency 99.47% relative to the optimal continuous design, while the second and the third have the same efficiency as the first design that was found through computer search, with efficiency 99.66%.

Table 2 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 Uniform (0.5, 1) (last section of Table 2), the ψ_θ-optimal design has $m_{22} = 0$. Nevertheless, the main effects of the two factors are still estimable and observations on the combination 22 adds little or no information on the main effect parameters due to the high correlation. The ψ_β-optimal design in this case is also nonorthogonal and unbalanced, but to a lesser degree. 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 in Table 1

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 $\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$. Instead, the three variances in the $\mathbf{\Lambda}$ of our example are assumed to be independently distributed, each as Inverse Gamma (1.5, 0.5). For the three correlations, we examine three scenarios: in the first, there is no restriction on the signs of the three correlations; in the second, all three correlations are positive; in the third, 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 3 show that, when there is prior knowledge of 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 2, ψ_β-optimal designs are very different from ψ_θ-optimal designs, with opposite signs on the off-diagonals of the $\mathbf{X}'\mathbf{X}$ matrix.

8. Design Robustness

Here we examine 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, we take the ψ_β and the ψ_θ-optimal exact designs for each of

Table 2. ψ_β- and ψ_θ-optimal 12-run exact designs for specified random effects covariance matrix $\mathbf{\Lambda}$. 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)

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%

eight assumed prior distributions (these are summarized in Section 8.1), evaluate these designs under the different true σ^2 and $\mathbf{\Lambda}$ distributions listed in Section 8.2, and report our simulation results in Section 8.3.

8.1. Designs constructed for assumed σ^2 and $\mathbf{\Lambda}$ distributions

Table 4 lists the ψ_β and the ψ_θ-optimal designs for eight assumed σ^2 and $\mathbf{\Lambda}$ distributions. For each design criterion, D_1 is an orthogonal design obtained without search; D_2–D_4 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 2; D_5 and D_6 are the optimal exact designs obtained for the general forms of $\mathbf{\Lambda}$ in Table 3; D_7

Table 3. 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, McCulloch, and Meng (2000). Efficiencies are calculated through (7.3) and (7.4)

Signs of random effects correlations	ψ_β- and ψ_θ-optimal design $(m_{11}, m_{12}, m_{21}, m_{22})$	Matrix $\mathbf{X}'\mathbf{X}$	Eff. Rel. to orthogonal design ψ_β	Eff. Rel. to orthogonal design 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 4. Summary of the designs and assumed prior distributions used in the robustness study.

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

and D_8 are local optimal exact designs obtained under the fixed values of σ^2 and $\mathbf{\Lambda}$ specified in Table 4.

Since the local ψ_β-optimal exact designs $D_{7\beta}$ and $D_{8\beta}$, obtained under the assumed fixed values of σ^2 and $\mathbf{\Lambda}$, and the designs $D_{2\beta}$ and $D_{3\beta}$, obtained under special forms of $\mathbf{\Lambda}$ with assumed prior correlation distributions, are orthogonal, they are listed with the orthogonal design $D_{1\beta}$ in our simulation study, in the first row of Table 6. Similarly, $D_{3\theta}$ and $D_{7\theta}$ have the same form and are listed together in Table 7.

Table 5. Summary of the eight true σ^2 and $\mathbf{\Lambda}$ distribution scenarios in the robustness study.

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 6. 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	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$	$\mathcal{S}_5$	$\mathcal{S}_6$	$\mathcal{S}_7$	$\mathcal{S}_8$
$D_{1\beta}, D_{2\beta}$, $D_{3\beta}, D_{7\beta}, D_{8\beta}$	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

8.2. True σ^2 and $\mathbf{\Lambda}$ distributions

We evaluate the designs $D_{1\beta}$–$D_{8\beta}$ and $D_{1\theta}$–$D_{8\theta}$ under eight true σ^2 and $\mathbf{\Lambda}$ distributions, as listed in Table 5. For σ^2, we use the Inverse Gamma (1.5, 0.5) and the Inverse Gamma (3, 1). In the first distribution, the mean is 1 and the variance is undefined. In the second, the mean is 0.5 and the variance is 0.25. Two distributions are used for the variance components of $\mathbf{\Lambda}$: all three variances are Inverse Gamma (1.5, 0.5); the three variances are Inverse Gamma (1.5, 0.5), (3, 1), and (2.5, 1.5). Two distribution assumptions are made for the correlation components of $\mathbf{\Lambda}$: "all positive" and "all negative".

8.3. Simulation results

Tables 6 and 7 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 (ψ_θ-efficiency) of each design under each true distribution scenario is obtained by replacing $mM(\eta^*)$ in the denominator of (7.1) with $\mathbf{X}^{*\prime}\mathbf{X}^*$ where $\mathbf{X}^*$ is the model matrix of the ψ_β-optimal (ψ_θ-optimal) exact design obtained through computer search under the true distribution.

While the orthogonal design $D_{1\beta}$ tends to be efficient, especially under the ψ_β criterion, Tables 6 and 7 show that more efficient designs can be obtained when

Table 7. 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 $\mathcal{S}_1$-$\mathcal{S}_8$.

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

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 7, when all pairs of random effects are negatively correlated as in 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 $D_{1\theta}$. Note that designs $D_{2\theta}$ and $D_{8\theta}$ are almost as ψ_θ-efficient in these four scenarios as the design $D_{6\theta}$. However, 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 when 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, focusing on the ψ_β criterion for the estimation of the individual-level parameters, and the ψ_θ criterion for the estimation of the hyperparameters. We focused on a special case in which all subjects receive the same treatments, response errors are homoscedastic, and covariates are independent of treatments.

Our results suggest that (i) designs that are ψ_β-optimal for the estimation of the individual-level parameters are not necessarily ψ_θ-optimal for the estimation of the hyperparameters; (ii) orthogonal designs may not be a good choice when interest is in the estimation of hyperparameters and random effects are expected to be correlated; (iii) for the construction of 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 of the random effects reflect the expected algebraic signs of the covariance elements; designs obtained under moderate sized correlations with the anticipated signs are 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, with fixed variances and moderate sized correlations in accordance with the anticipated signs, are less computation-intensive and seem to be good surrogates for the optimal design; (v) for the special case of equally correlated random effects, exact designs obtained from rounding the weights of the continuous designs tend to be either the same as the corresponding optimal exact design found through computer search, or have efficiencies only slightly lower.

We next discuss some possible extensions for future work. 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. With new computational advancements and more computational power, we believe that it is an interesting future research direction to explore design criteria based on full or marginal posterior distributions. 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 focuses. Designs obtained under these alternative criteria may be quite different from those we have investigated in this paper. However, we note that a design criterion that involves the maximization of the expected gain in Shannon information (Lindley (1956)) based on the posterior conditional distribution of individual-level parameters or hyperparameters can be approximated by the ψ_β criterion or the ψ_θ criterion. This can be seen by following the approach taken in Liu, Dean, and Allenby (2007).

A natural extension of our results is to the general case in which 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 and the covariates may be difficult to find 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, or vice versa; see an example in Liu et al. (2009).

Acknowledgement

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 extremely helpful comments and suggestions for future extensions.

Appendix

A: Proof of Theorem 3

For $\mathbf{\Lambda}$ given in (6.7) and $\mathbf{M}(\eta^*) = \mathbf{I}$, the left hand side of (6.4) is

$$\begin{aligned}
&\int \left\{ \sum_{k=0}^{p-1} x_k^2 (\tfrac{m}{\sigma^2} + \lambda_k^{-2})^{-1} \right\} f(\mathbf{\Lambda}) f(\sigma^2) d\mathbf{\Lambda} d\sigma^2 \\
&= \sum_{k=0}^{p-1} x_k^2 \int (\tfrac{m}{\sigma^2} + \lambda_k^{-2})^{-1} f(\mathbf{\Lambda}) f(\sigma^2) d\mathbf{\Lambda} d\sigma^2 \\
&= E\left[(\tfrac{m}{\sigma^2} + \lambda_0^{-2})^{-1}\right] \times \sum_{k=0}^{p-1} x_k^2 \quad \text{(since } \lambda_0^2, \ldots, \lambda_{p-1}^2 \text{ are identically distributed)} \\
&\leq E\left[(\tfrac{m}{\sigma^2} + \lambda_0^{-2})^{-1}\right] \times p \quad \text{(by (6.6))} \\
&= E\left\{Tr\left[(\tfrac{m}{\sigma^2} + \lambda_0^{-2})^{-1})\mathbf{I}_p\right]\right\} \\
&= \int \left\{Tr[\tfrac{m}{\sigma^2} I + M(\eta^*)^{-1} \mathbf{\Lambda}^{-1}]^{-1}\right\} f(\mathbf{\Lambda}) f(\sigma^2) d\mathbf{\Lambda} d\sigma^2,
\end{aligned}$$

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

B: Proof of Theorem 4

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 $-1 < \kappa < 1/(p-1)$. With $\mathbf{\Lambda} = \tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$, we have

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

and

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

The right hand side of (6.4) is

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

Using (6.8), the expectation (over $\tilde a$, $\tilde d$, and σ^2) of the second item inside the curly bracket of the last equation is 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,

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

Using (6.8), the expectation of the second item in the last equation is $\mathbf{0}$, and the left hand side of (6.4) is

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

Since $\mathbf{x}'\mathbf{x}\le p$ from (6.6), the theorem follows from Theorem 1.

C: Proof of Theorem 5

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 $-1<\epsilon<\frac{1}{p-1}$. 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 (6.9), the expectation (over $\tilde{a}$, $\tilde{d}$ and σ^2) of the coefficient of $\mathbf{J}$ in the last equation is $\mathbf{0}$, and the left hand side of (6.5) is

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

The right hand side of (6.5) is

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

Using (6.9), the expectation of the second item in the last equation is 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 (6.6), the theorem follows from Theorem 2.

References

Allenby, G. M. and Ginter, J. L. (1995). Using extremes to design products and segment markets. *J. Marketing Res.* **32**, 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. *Statist. Sinica* **20**, 1281-1311.

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

Chaloner, K. (1984). Optimal bayesian experimental design for linear models. *Ann. Statist.* **12**, 283-300.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statist. Sci.* **10**, 273-304.

Daniels, M. J. and Kass, R. E. (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *J. Am. Statist. Assoc.* **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*, 25-41.

Fedorov, V. V. and Hackl, P. (1997). *Model-Oriented Design of Experiments*. Springer Verlag, New York.

Ferreira, P. E. (1981). Extending Fisher's measure of information. *Biometrika* **68**, 695-698.

Giovagnoli, A. and Sebastiani, P. (1989). Experimental designs for mean and variance estimation in variance components models. *Comput. Statist. Data Anal.* **8**, 21-28.

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

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. and Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Dev. Psychol.* **27**, 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 Sci.* **15**, 173-191.

Lindley, D. (1956). On the measure of information provided by an experiment. *Ann. Statist.* **27**, 986-1005.

Liu, Q., Dean, A. M. and Allenby, G. M. (2007). Design for hyperparameter estimation in linear models. *J. Statist. Th. 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**, 175-186.

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

Pinheiro, J. C. and Bates, D. M. (1996). Unconstrained parameterizations for variance-covariance matrices. *Statist. Comput.* **6**, 289-296.

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

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

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

Rossi, P. E., Allenby, G. M. and McCulloch, R. (2005). *Bayesian Statistics and Marketing*. Wiley, Hoboken, NJ.

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. *Ann. Statist.* **22**, 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, Madison, WI 53706 U.S.A.

E-mail: qliu@bus.wisc.edu

Department of Statistics, The Ohio State University, Columbus, OH 43210 U.S.A.

E-mail: amd@stat.osu.edu

Department of Marketing, The Ohio State University, Columbus, OH 43210 U.S.A.

E-mail: allenby.1@osu.edu

(Received December 2009; accepted November 2010)