Submitted to the Annals of Statistics

OPTIMAL EXPERIMENTAL DESIGNS FOR HYPERPARAMETER ESTIMATION IN HIERARCHICAL LINEAR MODELS *

By Qing Liu, Angela M. Dean, and Greg M. Allenby

The Ohio State University

Optimal design for the efficient estimation of hyperparameters in hierarchical linear models is discussed. A criterion is derived under a Bayesian formulation for both the situation of independent random effects and that of correlated random effects. It is shown by example that designs obtained by fixing the error variance and the random effects covariance matrix to the means of their respective prior distributions can be as efficient or almost as efficient as optimal designs obtained by integrating over the prior distributions. We obtain explicit forms of the structure of such optimal designs and study the efficiency of exact designs when the optimal structure cannot be achieved. Design robustness is studied under various prior mean specifications of the covariance matrix, and resulting implications for practical applications are discussed.

1. Introduction. Over the past two decades, various forms of hierarchical models have been used to model a wide variety of data in 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

*Supported in part by NSF Grant SES-0437251

AMS 2000 subject classifications: Primary 62K05; secondary 62K15

Keywords and phrases: Bayesian Design, D-optimality, Design Robustness, Hierarchical Linear Model, Hyperparameter, Optimal Design, Random Effects Model

models", "population models", "random-coefficient regression models" and "covariance components models" (see Raudenbush and Bryk, 2002). In marketing research, for example, hierarchical models have been used by Lenk and Rao (1990) for studying new product diffusion, Allenby and Lenk (1994) for brand choice, Manchanda et al. (1999) for multiple category purchase decisions, Bradlow and Rao (2000) for assortment choice, and Montgomery et al. (2004) for online shopping path. In the social sciences, hierarchical models have been used by Huttenlocher et al. (1991) for studying children's vocabulary growth, Rumberger (1995) for students' dropping out of school, Horney et al. (1995) for criminal behavior, Raudenbush and Sampson (1999) for physical and social disorder. In pharmaceutical research, Yuh et al (1994) compiled a bibliography listing the numerous applications of hierarchical models in pharmacokinetic and pharmacodynamics.

Hierarchical models, linear or nonlinear, often consist of two levels, where parameters in the first level of the hierarchy reflect individual-level effects, which are assumed to be random and distributed according to a probability distribution characterized by the hyperparameters in the second-level of the hierarchy. Hyperparameters may characterize population characteristics ("population parameters"), such as the mean effect of a new drug in a patient population (see Yuh et al., 1994), and the mean consumer sensitivity to a product feature change in a target consumer population (e.g., Allenby and Lenk, 1994); or may reflect the effect of various covariates which drive the individual-level effects, such as the effect of the exposure to language on the 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 cus-

tomization of products, it is important to have accurate information on individual-level effects. However, 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 they capture the population characteristics, and enable predictions to new contexts.

A few researchers have discussed experimental designs that provide efficient estimation of hyperparameters under hierarchical nonlinear models. For example, in marketing research, Sándor and Wedel (2002) used relabeling, swapping and cycling procedures to find efficient factorial designs for the estimation of population parameters where the random effects are assumed to be independent and identically distributed. In pharmaceutical research, Mentré et al. (1997) proposed first-order linearization of a nonlinear model and used Fisher information for the linear model parameters as the basis to compare a given finite set of candidate designs, assuming the random effects to be independent. Tod et al. (1998) extended the work of Mentré et al. (1997) by accounting for the uncertainty of population parameters and employing a stochastic gradient search algorithm to find the optimal design, assuming the random effects to be independent and identically distributed. Han and Chaloner (2004) took a Bayesian-theoretic approach and used MCMC nested within Monte Carlo simulation to compare a predetermined set of candidate designs. A critical comparison of these different approaches is given by Han and Chaloner (2005).

Under a hierarchical linear model with independent, identically distributed random effects, 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. 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.

In this paper, we obtain explicit forms of the structure of optimal designs under hierarchical Bayesian linear models for both the situation of independent random effects and that of correlated random effects. We study the efficiency of exact designs when the optimal structure cannot be achieved. 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.

We take a hierarchical linear model of the following form:

$$\mathbf{y}_i|\boldsymbol{\beta}_i, \sigma_i^2 = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \qquad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma_i^2), \tag{1}$$

$$\boldsymbol{\beta}_i|\boldsymbol{\theta}, \boldsymbol{\Lambda} = \mathbf{Z}_i\boldsymbol{\theta} + \boldsymbol{\delta}_i, \qquad \boldsymbol{\delta}_i \sim N(\mathbf{0}, \boldsymbol{\Lambda}), \tag{2}$$

where responses of subject i $(i = 1, \ldots n)$ are represented by the vector $\mathbf{y}_i$ of length m_i, and $\mathbf{X}_i$ is the $m_i \times p$ model matrix, 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. We assume the following non-informative priors for $\boldsymbol{\theta}$, $\boldsymbol{\Lambda}$ and σ_i^2 corresponding to weak prior knowledge (see, for example, Rossi et al., 2005). These are replaced by more informative priors when information is available.

$$\boldsymbol{\theta} \sim \text{Normal}(\mathbf{0}_q, 100\mathbf{I}_q), \tag{3}$$

$$\boldsymbol{\Lambda} \sim \text{Inverted Wishart}(\nu_0 = p + 3, \nu_0\mathbf{I}_p), \tag{4}$$

$$\sigma_i^2 \sim \text{Inverse Gamma}(\frac{3}{2}, \frac{s_{0i}}{2}), \quad s_{0i} = \frac{3\sum_{j=1}^{m_i}(y_{ij} - \bar{y}_{i.})^2}{m_i - 1}. \tag{5}$$

In Section 2, we give a criterion for an optimal design for the estimation of the hyperparameter vector $\boldsymbol{\theta}$ while the covariance parameters in $\boldsymbol{\Lambda}$ are considered to be nuisance parameters. Since an optimal design is dependent upon the prior probability distributions of the unknown parameters $\boldsymbol{\Lambda}$ and $\{\sigma_i^2\}$, we investigate the efficiency of designs derived with $\boldsymbol{\Lambda}$ and $\{\sigma_i^2\}$ fixed at the means, $\boldsymbol{\Lambda}_0$ and σ_0^2, of their respective prior distributions. In Section 3, we derive the structure of optimal factorial designs in this setting under main effects models for some special forms of $\boldsymbol{\Lambda}_0$ with the assumptions that every subject receives the same treatment combination (so that $\mathbf{X}_i = \mathbf{X}$), and that no covariates are measured (so that $\mathbf{Z}_i = \mathbf{I}$). This is extended in Section 4, where we discuss optimal factorial designs for the situation when both the treatment combination and the subject covariate selection can be controlled by the experimenter. In Section 5, we present simulated examples which indicate that designs obtained by fixing $\boldsymbol{\Lambda}$ and $\{\sigma_i^2\}$ to their prior means can be as efficient or almost as efficient as the optimal designs obtained through the more computationally intensive procedure of integrating over the prior distributions of $\boldsymbol{\Lambda}$ and $\{\sigma_i^2\}$. In addition, we investigate design robustness to different prior mean specifications of $\boldsymbol{\Lambda}$ and make a recommendation for the specification of the mean of $\boldsymbol{\Lambda}$ in the search for optimal designs. A summary and conclusions are provided in Section 6.

2. Optimal designs for estimation of $\boldsymbol{\theta}$. When interest is in the estimation of hyperparameters $\boldsymbol{\theta}$ and $\boldsymbol{\Lambda}$, the two layers (1) and (2), of the hierarchical model can be combined to obtain

$$\mathbf{y}_i|\boldsymbol{\theta}, \boldsymbol{\Lambda}, \sigma_i^2 \sim N_{m_i}(\mathbf{X}_i\mathbf{Z}_i\boldsymbol{\theta},\ \boldsymbol{\Sigma}_i = \sigma_i^2\mathbf{I}_{m_i} + \mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i')\,, \tag{6}$$

(see Lenk et al., 1996, pg 187) with proper priors (3), (4), and (5) assumed for $\boldsymbol{\theta}$, $\boldsymbol{\Lambda}$, and σ_i^2, $i = 1, \ldots, n$.

We first look at the simple situation where the estimation of $\boldsymbol{\theta}$ is of primary interest while elements in $\boldsymbol{\Lambda}$ and the response error variance parameters σ_i^2 are nuisance parameters. This situation may occur, for example, when a medical researcher is interested in finding out the mean effect of a new drug in a target population of patients, or an education specialist is interested in the mean effect of parents' involvement on children's vocabulary growth, or a marketer is interested in learning how household income affects consumer price sensitivity, while the dispersion of the effects at the level of individual patient, child or consumer is not of primary interest.

Let $\mathcal{D}(m_1, \ldots, m_n)$ be a class of designs where m_i design points are allocated to subject i ($i = 1, \ldots, n$), that is, $\tilde{d} = (d_1, \ldots, d_n) \in \mathcal{D}(m_1, \ldots, m_n)$ where d_i is the m_i-point sub-design allocated to subject i. For a given $\tilde{d} = (d_1, \ldots, d_n)$, let $\widetilde{\mathbf{X}}' = (\mathbf{X}_1', \ldots, \mathbf{X}_n')$ where $\mathbf{X}_i$ of size $m_i \times p$ is the corresponding model matrix of d_i. Following Chaloner and Verdinelli (1995, page 277), we seek an optimal design $\tilde{d}^*$ in $\mathcal{D}(m_1, \ldots, m_n)$ that maximizes the expected gain in Shannon Information, that is, we seek a design that gives maximum

$$\int \log p(\boldsymbol{\theta}|\mathbf{y}, \widetilde{\mathbf{X}}) p(\boldsymbol{\theta}|\mathbf{y}, \widetilde{\mathbf{X}}) p(\mathbf{y}|\widetilde{\mathbf{X}}) d\boldsymbol{\theta} d\mathbf{y}, \tag{7}$$

where $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_n')'$ and where $p(\boldsymbol{\theta}|\mathbf{y}, \widetilde{\mathbf{X}})$ is

$$p(\boldsymbol{\theta}|\mathbf{y}, \widetilde{\mathbf{X}}) = \int p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\Lambda}, \{\sigma_i^2\}, \widetilde{\mathbf{X}}) d\boldsymbol{\Lambda} d\sigma_1^2 \ldots d\sigma_n^2\,. \tag{8}$$

Since (8) is not of closed-form, a normal approximation is used as follows. First, let $\boldsymbol{\xi}$ be the vector that includes all the p parameters in $\boldsymbol{\theta}$, the $p(p+1)/2$ parameters in $\boldsymbol{\Lambda}$ and the n error variance parameters $\{\sigma_i^2\}$, then according to Berger (1985, page 224, (iv)), $\boldsymbol{\xi}|\mathbf{y}, \widetilde{\mathbf{X}}$ has the following approximate distribution

$$\boldsymbol{\xi}|\mathbf{y}, \widetilde{\mathbf{X}} \ \sim \ N(\hat{\boldsymbol{\xi}}, \mathbf{I}(\hat{\boldsymbol{\xi}})^{-1}), \tag{9}$$

where $\mathbf{I}(\hat{\boldsymbol{\xi}})$ is the expected Fisher information matrix evaluated at the maximum likelihood estimate $\hat{\boldsymbol{\xi}}$. Now let $\mathbf{A}$ be a square matrix of dimension $[p + p(p+1)/2 + n]$ such that

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

then $\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\xi}$ and, therefore, from (9), the approximate distribution of $\boldsymbol{\theta}|\mathbf{y}, \widetilde{\mathbf{X}}$ is given by

$$\boldsymbol{\theta}|\mathbf{y}, \widetilde{\mathbf{X}} \ \sim \ N(\mathbf{A}\hat{\boldsymbol{\xi}}, \mathbf{A}\mathbf{I}(\hat{\boldsymbol{\xi}})^{-1}\mathbf{A}'). \tag{10}$$

Let $\boldsymbol{\varsigma} = [\sigma_1^2, \ldots, \sigma_n^2]'$, and partition $\mathbf{I}(\boldsymbol{\xi})$ as

$$\mathbf{I}(\boldsymbol{\xi}) = \begin{pmatrix} FI(\boldsymbol{\theta}, \boldsymbol{\theta}) & FI(\boldsymbol{\theta}, \boldsymbol{\Lambda}) & FI(\boldsymbol{\theta}, \boldsymbol{\varsigma}) \\ FI(\boldsymbol{\theta}, \boldsymbol{\Lambda}) & FI(\boldsymbol{\Lambda}, \boldsymbol{\Lambda}) & FI(\boldsymbol{\Lambda}, \boldsymbol{\varsigma}) \\ FI(\boldsymbol{\theta}, \boldsymbol{\varsigma}) & FI(\boldsymbol{\Lambda}, \boldsymbol{\varsigma}) & FI(\boldsymbol{\varsigma}, \boldsymbol{\varsigma}) \end{pmatrix}.$$

Let $[\boldsymbol{\Lambda}]_{uv} = \lambda_{u,v}$ denote the $(u,v)^{th}$ element of $\boldsymbol{\Lambda}$. Then, as shown by Lenk et al. (1996),

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

$$FI(\boldsymbol{\theta}, \boldsymbol{\Lambda}) = \mathbf{0}, \quad \text{and} \quad FI(\lambda_{u,v}, \lambda_{r,s}) = \frac{1}{2}\sum_{i=1}^{n} Tr\left(\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\lambda_{u,v}}\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\lambda_{r,s}}\right).$$

Similarly, it can be shown that

$$FI(\boldsymbol{\theta},\boldsymbol{\varsigma}) = \mathbf{0}, \qquad FI(\lambda_{u,v},\sigma_i^2) = \frac{1}{2}Tr(\boldsymbol{\Sigma}_i^{-1}\frac{\partial \boldsymbol{\Sigma}_i}{\partial \lambda_{u,v}}\boldsymbol{\Sigma}_i^{-1}),$$
$$FI(\boldsymbol{\varsigma},\boldsymbol{\varsigma})_{i,i} = FI(\sigma_i^2,\sigma_i^2) = \frac{1}{2}Tr(\boldsymbol{\Sigma}_i^{-2}), \qquad FI(\boldsymbol{\varsigma},\boldsymbol{\varsigma})_{i,j} = FI(\sigma_i^2,\sigma_j^2) = 0.$$

Using the facts that $FI(\boldsymbol{\theta},\boldsymbol{\Lambda}) = \mathbf{0}$ and $FI(\boldsymbol{\theta},\boldsymbol{\varsigma}) = \mathbf{0}$, and following Graybill (1983, Theorem 8.2.1), we obtain

$$\mathbf{AI}(\hat{\boldsymbol{\xi}})^{-1}\mathbf{A}' = FI(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\theta}})^{-1} = \Big[\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{X}_i'(\hat{\sigma}_i^2\mathbf{I} + \mathbf{X}_i\hat{\boldsymbol{\Lambda}}\mathbf{X}_i')^{-1}\mathbf{X}_i\mathbf{Z}_i\Big]^{-1}.$$

Therefore, using (10), it can be shown that the integral (7) can be approximated by

$$\int\left\{-\frac{p}{2}\log(2\pi) + \frac{1}{2}\log\Big|\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{X}_i'(\hat{\sigma}_i^2\mathbf{I} + \mathbf{X}_i\hat{\boldsymbol{\Lambda}}\mathbf{X}_i')^{-1}\mathbf{X}_i\mathbf{Z}_i\Big| - \frac{p}{2}\right\}p(\mathbf{y}|\widetilde{\mathbf{X}})d\mathbf{y}.$$

The integrand depends on $\mathbf{y}$ only through the consistent maximum likelihood estimates of $\boldsymbol{\Lambda}$ and $\boldsymbol{\varsigma} = (\sigma_1^2,\ldots,\sigma_n^2)'$. Following Chaloner and Verdinelli (1995, page 286), a further approximation can be taken where the prior distribution of $\boldsymbol{\Lambda}$ and $\boldsymbol{\varsigma}$ is used to approximate the distributions of $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\varsigma}}$; that is, (7) is further approximated by

(11)
$$-\frac{p}{2}\log(2\pi) - \frac{p}{2} + \int\left\{\frac{1}{2}\log\Big|\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{X}_i'(\sigma_i^2\mathbf{I} + \mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i')^{-1}\mathbf{X}_i\mathbf{Z}_i\Big|\right\}p(\boldsymbol{\Lambda})p(\boldsymbol{\varsigma})d\boldsymbol{\Lambda}d\boldsymbol{\varsigma}.$$

Thus, we seek an optimal design over the class of designs in $\mathcal{D}(m_1,\ldots,m_n)$ that maximizes (11), that is, we seek a design with corresponding model matrix $\widetilde{\mathbf{X}} = (\mathbf{X}_1,\ldots,\mathbf{X}_n)$ that maximizes the integral

$$\int\left\{\log\Big|\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{X}_i'(\sigma_i^2\mathbf{I} + \mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i')^{-1}\mathbf{X}_i\mathbf{Z}_i\Big|\right\}p(\boldsymbol{\Lambda})p(\boldsymbol{\varsigma})d\boldsymbol{\Lambda}d\boldsymbol{\varsigma}. \tag{12}$$

We call this criterion the ψ^I criterion, where the superscript I indicates "integration over the prior distributions of $\boldsymbol{\Lambda}$ and $\boldsymbol{\varsigma}$".

3. Optimal designs for main effects models with fixed $\boldsymbol{\Lambda}$ and $\boldsymbol{\varsigma}$ and $\mathbf{Z_i = I}$. We now fix $\boldsymbol{\Lambda}$ and $\boldsymbol{\varsigma} = [\sigma_1^2, \ldots, \sigma_n^2]'$ equal to the means of their respective prior distributions (given by (4) and (5)). Thus, from (12), we seek an optimal design that maximizes

$$\log \Big| \sum_{i=1}^{n} (\mathbf{Z}_i' \mathbf{X}_i' (\sigma_i^2 \mathbf{I}_{m_i} + \mathbf{X}_i \boldsymbol{\Lambda} \mathbf{X}_i')^{-1} \mathbf{X}_i \mathbf{Z}_i) \Big| . \tag{13}$$

We call this the ψ^F-criterion, where the superscript F indicates fixed $\boldsymbol{\Lambda}$ and $\boldsymbol{\varsigma}$. Some examples are shown in Section 5 where ψ^F-optimal designs are as efficient or almost as efficient as ψ^I-optimal designs obtained under the ψ^I-optimality criterion (12).

3.1. *ψ^F- optimal designs.* First, we take the special case of designs in $\mathcal{D}(m) \in \mathcal{D}(m_1, \ldots, m_n)$, 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) $\mathbf{Z}_i = \mathbf{I}_p$ so that $\boldsymbol{\theta}$ captures the population characteristics (see Section 1).

Then a ψ^F-optimal design in $\mathcal{D}(m)$ that maximizes (13) also maximizes (14) in Lemma 1.

LEMMA 1. *Under model* (6) *with assumptions (i), (ii), (iii), a ψ^F-optimal design in $\mathcal{D}(m)$ maximizes*

$$g(\mathbf{X}) = \log |\mathbf{X}'\mathbf{X}(\sigma^2 \mathbf{I}_p + \boldsymbol{\Lambda}\mathbf{X}'\mathbf{X})^{-1}| = \log \left(\frac{|\mathbf{X}'\mathbf{X}|}{|\sigma^2 \mathbf{I}_p + \boldsymbol{\Lambda}\mathbf{X}'\mathbf{X}|} \right). \tag{14}$$

PROOF. Under (i), (ii), (iii), expression (13) is maximized in $\mathcal{D}(m)$ when

$$g(\mathbf{X}) = \log |\mathbf{X}'(\sigma^2 \mathbf{I}_m + \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}')^{-1}\mathbf{X}| \tag{15}$$

is maximized in $\mathcal{D}(m)$. Let $\mathbf{A} = \sigma^2\mathbf{I}_m$, $\mathbf{B} = \mathbf{X}$ and $\mathbf{C} = \mathbf{\Lambda}$, then from Morrison (1990, page 69),

$$(\sigma^2\mathbf{I}_m + \mathbf{X}\mathbf{\Lambda}\mathbf{X}')^{-1} = \sigma^{-2}\mathbf{I}_m - \sigma^{-4}\mathbf{X}(\mathbf{\Lambda}^{-1} + \sigma^{-2}\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \tag{16}$$

Post-multiplying both sides of Equation (16) by $\mathbf{X}$ leads to the equality

$$\begin{aligned}(\sigma^2\mathbf{I}_m + \mathbf{X}\mathbf{\Lambda}\mathbf{X}')^{-1}\mathbf{X} &= \sigma^{-2}\mathbf{X}[\mathbf{I}_m - \sigma^{-2}(\mathbf{\Lambda}^{-1} + \sigma^{-2}\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}] \\ &= \sigma^{-2}\mathbf{X}[\mathbf{I}_m - (\mathbf{\Lambda}^{-1} + \sigma^{-2}\mathbf{X}'\mathbf{X})^{-1}(\mathbf{\Lambda}^{-1} + \sigma^{-2}\mathbf{X}'\mathbf{X} - \mathbf{\Lambda}^{-1})] \\ &= \sigma^{-2}\mathbf{X}[(\mathbf{\Lambda}^{-1} + \sigma^{-2}\mathbf{X}'\mathbf{X})^{-1}\mathbf{\Lambda}^{-1}] \\ &= \sigma^{-2}\mathbf{X}(\mathbf{I}_p + \sigma^{-2}\mathbf{\Lambda}\mathbf{X}'\mathbf{X})^{-1} \\ &= \mathbf{X}(\sigma^2\mathbf{I}_p + \mathbf{\Lambda}\mathbf{X}'\mathbf{X})^{-1},\end{aligned}$$

and the proof follows. □

Note that, for $\boldsymbol{\theta}$ to be estimable, $\mathbf{X}'\mathbf{X}$ must be non-singular and the maximization of $g(\mathbf{X})$ in (14) over the set of designs such that $\mathbf{X}'\mathbf{X}$ is non-singular is equivalent to the minimization of

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

which is the "mixed-effects model D-criterion" obtained in a non-Bayesian setting by Entholzner et al. (2005) and examined in one or two-factor designs.

We first obtain the form of $\mathbf{X}'\mathbf{X}$ for a ψ^F-optimal continuous design where the numbers of observations at the various design points may not be integers. This matrix will be used to obtain an upper bound for the ψ^F-optimality criterion. 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. Following Silvey (1980), we

look for a continuous design η^* such that $\eta^* \in \mathcal{H}$ maximizes the continuous analog of (14), namely

$$\psi(\mathbf{M}(\eta)) = \begin{cases} \log \frac{|\mathbf{M}(\eta)|}{|\frac{\sigma^2}{m}\mathbf{I}_p + \mathbf{\Lambda}\mathbf{M}(\eta)|} & \text{where} \quad \mathbf{M}(\eta) = \frac{1}{m}\mathbf{X}'\mathbf{X}, \\ -\infty & \text{for} \quad \mathbf{X}'\mathbf{X} \ \text{ singular.} \end{cases} \tag{17}$$

Define the set $\mathcal{M}$ to be

$$\mathcal{M} = \{\mathbf{M}(\eta) : \eta \in \mathcal{H}\}. \tag{18}$$

By Silvey (1980, pages 15 and 16), the set $\mathcal{M}$ is a closed convex hull of $\{\mathbf{x}\mathbf{x}' : \mathbf{x} \in \mathcal{X}\}$. Lemma 2, whose proof is given in the Appendix, shows that the function $\psi(\mathbf{M}(\eta))$ in (17) is concave and strictly increasing in $\mathcal{M}$, that is, for $0 \le \lambda \le 1$ and $\mathbf{M}_1, \mathbf{M}_2 \in \mathcal{M}$, $\psi(\lambda\mathbf{M}_1 + (1-\lambda)\mathbf{M}_2) \ge \lambda\psi(\mathbf{M}_1) + (1-\lambda)\psi(\mathbf{M}_2)$ (concave), and, if $\mathbf{M}_1 - \mathbf{M}_2$ is positive definite, then $\psi(\mathbf{M}_1) > \psi(\mathbf{M}_2))$ (strictly increasing).

LEMMA 2. *The ψ^F criterion function $\psi(\mathbf{M}(\eta))$ defined in (17) is concave and strictly increasing in $\mathcal{M}$ where $\mathcal{M}$ is defined in (18).*

Theorem 1 gives a necessary and sufficient condition for a ψ^F-optimal continuous design η^*.

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^F-optimal if and only if*

$$\mathbf{x}'\mathbf{M}(\eta^*)^{-1}[\tfrac{\sigma^2}{m}\mathbf{I}_p + \mathbf{M}(\eta^*)\mathbf{\Lambda}]^{-1}\mathbf{x} \le Tr[(\tfrac{\sigma^2}{m}\mathbf{I}_p + \mathbf{M}(\eta^*)\mathbf{\Lambda})^{-1}], \tag{19}$$

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

The proof of Theorem 1 follows from Lemma 2 and Theorem 3.7 in Silvey (1980), or (5.2.7) in Fedorov and Hackl (1997) using the transformation $\mathbf{D}_0 = \frac{m}{\sigma^2}\mathbf{\Lambda}$ and applying $(\mathbf{D}_0 + \mathbf{M}(\eta)^{-1})^{-1}\mathbf{M}(\eta)^{-1} = (\mathbf{M}(\eta)\mathbf{D}_0 + \mathbf{I}_p)^{-1}$. The theorem will be used in Section 3.2 to obtain the form of $\mathbf{M}(\eta^*)$ for an ψ^F-optimal design η^* for given forms of $\mathbf{\Lambda}$.

3.2. *The main effects model.* We now consider ψ^F-optimal designs under a main effects model. Suppose there are K treatment factors where treatment factor k ($k = 1, \ldots, K$) has h_k levels, then the parameter vector $\boldsymbol{\beta}_i$ for each respondent i in (1) includes the mean parameter and $h_k - 1$ contrast parameters for each factor k. The length of $\boldsymbol{\beta}_i$ is then $1 + \sum_{k=1}^{K}(h_k - 1) = p$. Under assumptions (i), (ii), (iii) in Section 3.1, the hyperparameter vector $\boldsymbol{\theta}$ of length p captures the mean of the individual-level random effects $\boldsymbol{\beta}_i$. For the contrast coefficients in $\mathbf{X}$, we use the coefficients of the "standardized orthogonal main effect contrasts" (c.f. Kuhfeld, 2005) defined so that the jth main effect contrast of factor k is a comparison of the effect of level j with the average of the effects of levels $j+1, \ldots, h_k$, scaled by a normalizing constant.

LEMMA 3. *Consider an experiment with K treatment factors where treatment factor k ($k = 1, \ldots, K$) has h_k levels. Let $\mathbf{H}_k$ denote a $h_k \times (h_k - 1)$ matrix whose jth column contains the coefficients for the jth standardized orthogonal main effect contrast of factor k ($j = 1, \ldots, h_k - 1$), then the sum of squares of the elements in row i of $\mathbf{H}_k$ is $h_k - 1$ ($i = 1, \ldots, h_k,\ k = 1, \ldots, K$).*

PROOF. Consider an arbitrary treatment factor with h levels. The matrix

of coefficients of the standardized main effect contrasts is:

$$\mathbf{H} = \begin{pmatrix} h-1 & 0 & \cdots & 0 & 0 \\ -1 & h-2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & -1 & \cdots & 2 & 0 \\ -1 & -1 & \cdots & -1 & 1 \\ -1 & -1 & \cdots & -1 & -1 \end{pmatrix} \mathbf{D}, \tag{20}$$

where $\mathbf{D}$ is a diagonal matrix of scaling constants so that the column sum of squares equals h for every column; that is, $\mathbf{D} = Diag(d_1, d_2, \ldots, d_{h-1})$ where $d_s = \{h/[(h-s)^2 + h - s]\}^{1/2}$ for $s = 1, \ldots, h-1$.

Ignoring scaling constants, the rth row ($1 \leq r \leq h-1$) of $\mathbf{H}$ contains -1 occurring $(r-1)$ times, $(h-r)$ once, and the remaining entries zero, whereas the last row consists entirely of -1s. In each case the sth element is multiplied by the scaling constant d_s ($s = 1, \ldots, h-1$). The sum of squares of the elements in the rth row of $\mathbf{H}$ ($1 \leq r \leq h-1$) is, therefore,

$$\sum_{s=1}^{r-1} \frac{h}{(h-s)^2 + h - s} + \frac{(h-r)^2 h}{(h-r)^2 + h - r}, \tag{21}$$

and the sum of squares of the elements in the last row is:

$$\sum_{s=1}^{h-1} \frac{h}{(h-s)^2 + h - s}. \tag{22}$$

Using the equality

$$\frac{h}{(h-s)^2 + h - s} = \frac{h}{h-s} - \frac{h}{h-s+1},$$

and by simple algebra both (21) and (22) are equal to $h-1$. □

The model matrix $\mathbf{X}$ contains m rows corresponding to the factor combinations in the design, and these rows are subsets of the rows of the full

model matrix

$$\left[\mathbf{1}_{h_1\times h_2\ldots h_K},\quad \mathbf{H}_1\otimes\mathbf{1}_{h_2\times\ldots h_K},\quad \mathbf{1}_{h_1}\otimes\mathbf{H}_2\otimes\mathbf{1}_{h_3\times\ldots h_K},\quad \ldots,\quad \mathbf{1}_{h_1\times\ldots h_{K-1}}\otimes\mathbf{H}_K\right].$$

Thus, each row vector in the model matrix $\mathbf{X}$ has $1+\sum_{k=1}^{K}(h_k-1)$ elements, corresponding to the mean and a set of standardized orthogonal main effect contrasts for factor k ($k=1,\ldots,K$). Such points form the border points of a compact continuous design space $\mathcal{X}\subseteq\mathcal{R}^p$, in which the first coordinate of all points $\mathbf{x}\in\mathcal{X}$ is constrained to be 1; that is,

$$\mathcal{X}=\Big\{\mathbf{x}=[1,\ldots,x_{k_1},\ldots,x_{k_{(h_k-1)}},\ldots,x_{K_1},\ldots,x_{K_{(h_K-1)}}]' \quad \text{such that} \quad \sum_{s=1}^{h_k-1}x_{k_s}^2\le h_k-1, k=1,\ldots,K.\Big\}. \tag{23}$$

With the design space $\mathcal{X}$ defined as (23), we seek a ψ^F-optimal continuous design η^* over $\mathcal{X}$. We note that any design η^* that is ψ^F-optimal under the standardized orthogonal main effect contrasts coding of model matrix $\mathbf{X}$ is also ψ^F-optimal under any other model matrix $\breve{\mathbf{X}}$ such that

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

and $\mathbf{T}$ is a $p\times p$ non-singular transformation matrix (c.f. Scheffé, 1959, page 31-32).

Theorem 2 shows that, under the main effects model, a continuous design η^* with matrix $\mathbf{M}(\eta^*)=\mathbf{I}$ is ψ^F-optimal when the random effects $\boldsymbol{\beta}_i$ in (1) are independently distributed with equal variances for the contrasts of the same factor. For the proof of Theorem 2, we need the following lemma, which follows from (23) and Lemma 3.

LEMMA 4. *For every design point* $\mathbf{x}$ *in* $\mathcal{X}$,

$$\mathbf{x}'\mathbf{x}=1+\sum_{k=1}^{K}\sum_{s=1}^{h_k-1}x_{k_s}^2\le 1+\sum_{k=1}^{K}(h_k-1)=p. \tag{24}$$

THEOREM 2. *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 (23). Let the random individual-level main effects $\boldsymbol{\beta}_i$ in (1) be independently distributed with equal within-factor variances such that $\boldsymbol{\Lambda}$ is*

$$\boldsymbol{\Lambda} = Diag\left(\lambda_0^2,\ \lambda_1^2\mathbf{I}_{(h_1-1)},\ \ldots,\ \lambda_k^2\mathbf{I}_{(h_k-1)},\ \ldots,\ \lambda_K^2\mathbf{I}_{(h_K-1)}\right), \tag{25}$$

where the k^{th} factor has h_k levels (k=1, ..., K). Given such a form of $\boldsymbol{\Lambda}$, any design η^* that satisfies $\mathbf{M}(\eta^*) = \mathbf{I}$ is ψ^F-optimal.

PROOF. For $\boldsymbol{\Lambda}$ given in (25) and $\mathbf{M}(\eta^*) = \mathbf{I}$, the left hand side of (19) is

$$\begin{aligned}
&(\tfrac{\sigma^2}{m} + \lambda_0^2)^{-1} + \sum_{k=1}^{K}\left[(\tfrac{\sigma^2}{m} + \lambda_k^2)^{-1}\sum_{j=1}^{h_k-1} x_{k_j}^2\right] \\
\leq &(\tfrac{\sigma^2}{m} + \lambda_0^2)^{-1} + \sum_{k=1}^{K}\left[(\tfrac{\sigma^2}{m} + \lambda_k^2)^{-1}(h_k-1)\right] \quad \text{by (23)} \\
= &Tr[(\tfrac{\sigma^2}{m}I + \boldsymbol{\Lambda})^{-1}],
\end{aligned}$$

which is the right hand side of (19). Therefore the statement of the theorem follows from Theorem 1. □

The following corollary follows directly from Theorem 2 by noting that, for a level-balanced orthogonal design, $\mathbf{X}'\mathbf{X} = m\mathbf{I}$, and so $\mathbf{M}(\eta) = \mathbf{I}$.

COROLLARY 1. *Under the conditions of Theorem 2, if a level-balanced orthogonal design exists, it is ψ^F-optimal.*

When the random individual-level main effects $\boldsymbol{\beta}_i$ in (1) are equally correlated and with equal variances, then $\boldsymbol{\Lambda}$ 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. Theorem 3 gives the form of the matrix $\mathbf{M}(\eta^*)$ of the ψ^F-optimal design η^*.

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 (23). Given $\mathbf{\Lambda}$ of the form $\tilde{a}\mathbf{I}+\tilde{d}\mathbf{J}$, a design η^* with $\mathbf{M}(\eta^*) = (1+\epsilon)\mathbf{I}-\epsilon\mathbf{J}$ is ψ^F-optimal, where*

$$\epsilon = \frac{2(a+pd)+1-2d-\sqrt{4(a+1)(a+pd)-4d+1}}{2(a+pd)(p-2)+2d}, \tag{26}$$

$a = m\tilde{a}/\sigma^2$, and $d = m\tilde{d}/\sigma^2$.

PROOF. Let $\mathbf{M}(\eta^*)$ be as defined in the statement of the theorem, then

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

In addition, we have

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

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

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

Using ϵ in (26), the second item in the bracket of the last equation becomes $\mathbf{0}$ and the left hand side of (19) becomes

$$\mathbf{x}'\mathbf{M}(\eta^*)^{-1}[\mathbf{I} + \mathbf{M}(\eta^*)(a\mathbf{I}+d\mathbf{J})]^{-1}\mathbf{x} = \frac{1}{(1+\epsilon)[1+a(1+\epsilon)]}\mathbf{x}'\mathbf{x}. \tag{27}$$

The right hand side of (19) is

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

Using ϵ in (26), the second item in the bracket of the last equation becomes 0 and we obtain

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

Since $\mathbf{x}'\mathbf{x} \leq p$ from (24), the theorem follows from Theorem 1. □

The implication of Theorem 3 is that, when the random effects are equicorrelated (interchangeable), ψ^F-optimal designs are nonorthogonal and unbalanced, that is, the off-diagonal elements of $\mathbf{M}(\eta^*)$ are non-zero. Examples will be given in Section 5. Theorem 3 can easily be extended to more general settings where 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*

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_0 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Lambda}_2 & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{\Lambda}_G \end{pmatrix}, \tag{28}$$

where $\mathbf{\Lambda}_g = \tilde{a}_g \mathbf{I}_{p_g} + \tilde{d}_g \mathbf{J}_{p_g}$ $(g = 1, \dots, G)$, *a design* η^* *is* ψ*-optimal if* $\mathbf{M}(\eta^*)$ *is*

$$\mathbf{M}(\eta^*) = \begin{pmatrix} 1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{M}_2 & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{M}_G \end{pmatrix}, \tag{29}$$

where $\mathbf{M}_g = (1+\epsilon_g)\mathbf{I}_{p_g} - \epsilon_g \mathbf{J}_{p_g}$, *with*

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

$a_g = m\tilde{a}_g/\sigma^2$, *and* $d_g = m\tilde{d}_g/\sigma^2$, $g = 1, \dots, G$.

As for many optimal designs, Theorems 2 and 3, together with Corollary 2, suggest that the matrix $\mathbf{M}(\eta^*)$ of a ψ^F-optimal design η^* will often have a structure similar to the covariance matrix, $\mathbf{\Lambda}$, of the random effects. We explore this further through examples in Section 5.

4. Optimal designs when Z_i can be controlled. In this section, the design problem is investigated in which both the design and the covariates can be controlled by the experimenter. For example, in a survey study, the experimenter can control the questions to be administered to respondents (which determines $\mathbf{X}_i$) as well as controlling the sampling of the respondents on the basis of certain demographic information such as age or income (which determines $\mathbf{Z}_i$). Consider the situation in which $\mathbf{X}_i$ and $\mathbf{Z}_i$ can be determined independently of each other, and $\mathbf{Z}_i$ can be expressed as $\mathbf{I}_p \otimes \mathbf{z}_i'$, where $\otimes$ denotes Kronecker product, and $\mathbf{z}_i$ is a vector containing the q/p demographics of respondent i. For example, when 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, so

$$\mathbf{Z}_i = [\mathbf{I}_p,\ a_i\mathbf{I}_p,\ \ h_i\mathbf{I}_p] = \mathbf{I}_p \otimes \mathbf{z}_i', \tag{30}$$

where $\mathbf{z}_i' = [1,\ \ a_i,\ \ h_i]$. The hyperparameter vector $\boldsymbol{\theta}$ in (2) is then of length $3p$. Here, the first set of p hyperparameters corresponds to the first p columns of $\mathbf{Z}_i$ in (30), 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 (30) 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 (30),

and noting that $\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i$ is a $p \times p$ matrix, expression (13) becomes

$$(31)\quad \log\Big|\sum_{i=1}^{n}\left[(\mathbf{I}_p\otimes\mathbf{z}_i')'\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i(\mathbf{I}_p\otimes\mathbf{z}_i')\right]\Big| = \log\Big|\sum_{i=1}^{n}\left[(\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i)\otimes(\mathbf{z}_i\mathbf{z}_i')\right]\Big|.$$

When all subjects receive the same design, and the response errors are homoscedastic, so that $\mathbf{X}_i = \mathbf{X}$, $\sigma_i^2 = \sigma^2$, equation (31) simplifies to

$$\begin{aligned}(32)\quad \log\Big|\sum_{i=1}^{n}\left[(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})\otimes(\mathbf{z}_i\mathbf{z}_i')\right]\Big| &= \log\Big|(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})\otimes\big[\sum_{i=1}^{n}(\mathbf{z}_i\mathbf{z}_i')\big]\Big| \\ &= \log\left\{\Big|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\Big|^{q/p}\Big|\sum_{i=1}^{n}(\mathbf{z}_i\mathbf{z}_i')\Big|^{p}\right\} \\ &= \tfrac{q}{p}\log\Big|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\Big| + p\log\Big|\sum_{i=1}^{n}(\mathbf{z}_i\mathbf{z}_i')\Big|,\end{aligned}$$

where the second equality follows from Graybill (1983, Theorem 8.8.10). With the independence of $\mathbf{z}_i$ and $\mathbf{X}$, and given the number of parameters p and q, the maximization of (32) is achieved through the individual maximization of $\log|\sum_{i=1}^{n}(\mathbf{z}_i\mathbf{z}_i')|$ and $\log|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}|$, the latter of which has been discussed in Section 3. 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, Atkinson and Donev, 1992). Similar comments apply to the ψ^I criterion in (12).

Note that, when $\mathbf{X}_i$ and $\mathbf{Z}_i$ are linked, such as in the modeling of the "level effect" (see, Liu et al., 2006) where $\mathbf{Z}_i$ consists of relative ranges and frequencies of the attribute levels present in $\mathbf{X}_i$, then $\mathbf{Z}_i$ cannot be written in the form of (30) and computer search algorithms need to be used to find the optimal combination of $\mathbf{X}_i$ and $\mathbf{Z}_i$.

5. Simulation studies. In practical applications, exact designs are required in which there are integer numbers of observations at the design points in $\mathcal{X}$. We refer to the exact designs that maximize (14) in $\mathcal{D}(m)$ with $\boldsymbol{\Lambda}$ and σ^2 set to the prior means ($\boldsymbol{\Lambda}_0$ and σ_0^2) as ψ^{EF}-optimal, that is, they

maximize

$$\log\left(\frac{|\mathbf{X}'\mathbf{X}|}{|\sigma_0^2\mathbf{I}_p+\mathbf{\Lambda}_0\mathbf{X}'\mathbf{X}|}\right) \tag{33}$$

Similarly, we call exact designs ψ^{EI}-optimal if they maximize (12) in $\mathcal{D}(m)$ under the assumptions (i), (ii), (iii) in Section 3.1, that is, if they maximize

$$\begin{aligned}\int\left\{\log\left|\mathbf{X}'(\sigma^2\mathbf{I}+\mathbf{X}\mathbf{\Lambda}\mathbf{X}')^{-1}\mathbf{X}\right|\right\}p(\mathbf{\Lambda})p(\sigma^2)d\mathbf{\Lambda}d\sigma^2 \\ =\int\log\left(\frac{|\mathbf{X}'\mathbf{X}|}{|\sigma^2\mathbf{I}_p+\mathbf{\Lambda}\mathbf{X}'\mathbf{X}|}\right)p(\mathbf{\Lambda})p(\sigma^2)d\mathbf{\Lambda}d\sigma^2, \quad \text{by Lemma 1.}\end{aligned} \tag{34}$$

We show through examples that the ψ^{EF}-optimal designs obtained through computer search have efficiencies similar to those of the ψ^{EI}-optimal designs under the ψ^I criterion. Note that, since the optimal designs were obtained by computer search, they may be locally rather than globally optimal. In addition, we examine the robustness of the ψ^{EF}-optimal designs to different prior mean specifications of $\mathbf{\Lambda}$.

5.1. *Design efficiency.* First we define the D-efficiency, the ψ^F-efficiency, and the relative ψ^I-efficiency measures to be used in the remainder of this section. Define the D-efficiency of an exact design with model matrix $\mathbf{X}$ relative to an optimal design η^+ under the D-criterion for a fixed-effects main effects model with standardized orthogonal contrasts coding as

$$D\text{-eff}(\mathbf{X})=\frac{\left|\frac{1}{m}\mathbf{X}'\mathbf{X}\right|^{\frac{1}{p}}}{|\mathbf{M}(\eta^+)|^{\frac{1}{p}}}=\left|\tfrac{1}{m}\mathbf{X}'\mathbf{X}\right|^{\frac{1}{p}}, \tag{35}$$

where the second equality follows since $\mathbf{M}(\eta^+)=\mathbf{I}$ (see, for example, Kuhfeld et al., 1994). Define the ψ^F-efficiency of an exact design with model

matrix $\mathbf{X}$ relative to an optimal continuous design η^* be

(36)
$$\psi^F\text{-eff}(\mathbf{X}) = \frac{\left(\left|\frac{1}{m}\mathbf{X}'\mathbf{X}\right|\left|\frac{\sigma_0^2}{m}\mathbf{I}+\mathbf{\Lambda}_0\mathbf{M}(\eta^*)\right|\right)^{\frac{1}{p}}}{\left(\left|\frac{\sigma_0^2}{m}\mathbf{I}+\mathbf{\Lambda}_0\frac{1}{m}\mathbf{X}'\mathbf{X}\right||\mathbf{M}(\eta^*)|\right)^{\frac{1}{p}}} = \left(\frac{|\mathbf{X}'\mathbf{X}|\left|\frac{\sigma_0^2}{m}\mathbf{I}+\mathbf{\Lambda}_0\mathbf{M}(\eta^*)\right|}{|\mathbf{M}(\eta^*)|\,|\sigma_0^2\mathbf{I}+\mathbf{\Lambda}_0\mathbf{X}'\mathbf{X}|}\right)^{\frac{1}{p}},$$

where $\mathbf{M}(\eta^*)$ is given in Theorems 2 and 3 as well as Corollary 2 under certain specifications of $\mathbf{\Lambda}_0$. Since the ψ^I-optimal continuous design is unknown, in the following examples we use the design η^+ with $\mathbf{M}(\eta^+) = \mathbf{I}$ as the base design and calculate the relative ψ^I-efficiency of an exact design $\mathbf{X}$ as

(37)
$$\text{rel. } \psi^I\text{-eff} = \exp\left\{\frac{1}{p}\int \log\left(\frac{|\mathbf{X}'\mathbf{X}|\left|\frac{\sigma^2}{m}\mathbf{I}+\mathbf{\Lambda}\mathbf{M}(\eta^+)\right|}{|\mathbf{M}(\eta^+)|\,|\sigma^2\mathbf{I}+\mathbf{\Lambda}\mathbf{X}'\mathbf{X}|}\right)p(\mathbf{\Lambda})p(\sigma^2)d\mathbf{\Lambda}d\sigma^2\right\}.$$

We used simple-exchange algorithms (see Mitchell and Miller, 1970; Atkinson and Donev, 1992) to obtain the ψ^{EF}- and ψ^{EI}-optimal designs. Monte Carlo method was used for the integration over the prior distributions of $\mathbf{\Lambda}$ and σ^2 in (34). The codes, written in SAS PROC IML, are available at `http://www.stat.ohio-state.edu/~amd/dissertations.html`.

When the random effects $\boldsymbol{\beta}_i$ are believed to be correlated, it is reasonable to assume that the mean of the Inverted Wishart prior distribution of $\mathbf{\Lambda}$ has non-zero off-diagonal elements. First, consider the case of interchangeable random effects so that $\mathbf{\Lambda}_0 = E(\mathbf{\Lambda})$ is of the form $\tilde{a}\mathbf{I} + \tilde{d}\mathbf{J}$. Theorem 3 shows that the matrix $\mathbf{M}(\eta^*)$ of a ψ^F-optimal continuous design η^* is of the form $(1+\epsilon)\mathbf{I} - \epsilon\mathbf{J}$ where ϵ is given in (26). On dividing the numerator and denominator of ϵ by a, this is

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

where $r = d/a = \tilde{d}/\tilde{a}$ and $a = m\tilde{a}/\sigma^2$. As the number of runs m becomes

large while $\tilde{a}$ and σ^2 remain fixed, $1/a$ becomes small and ϵ is almost equal to

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

This suggests that for a fixed number of parameters p, the value of ϵ is mainly determined by the ratio r. The same applies to the ψ^F-optimal designs under the more general forms of $(\mathbf{\Lambda}_0)$ where within-group random effects are interchangeable and inter-group random effects are independent, as in Corollary 2. Thus, without loss of generality, in the following examples we set $E(\sigma^2) = 1$, $\tilde{a} = 1$, and search for the ψ^{EF}-optimal and ψ^{EI}-optimal designs given different values of $r = \tilde{d}/\tilde{a}$.

EXAMPLE 5.1: Consider an experiment with two treatment factors each with two levels under a hierarchical linear model. No covariates are considered ($\mathbf{Z}_i = \mathbf{I}$), response errors are assumed to be homoscedastic ($\sigma_i^2 = \sigma^2$), and each subject i (i=1, ..., n) receives the same treatment allocation ($\mathbf{X}_i = \mathbf{X}$). The individual-level random effects $\boldsymbol{\beta}_i$ in (1) consists of the general mean, the main effects of factors 1 and 2, for subject i. $\boldsymbol{\beta}_i$ is assumed to be randomly distributed according to a multivariate normal distribution with mean $\boldsymbol{\theta}$ and covariance matrix $\mathbf{\Lambda}$ as in (2). The estimation of the hyper-parameter vector $\boldsymbol{\theta}$ is of major interest. In this setting, the design criterion that applies is ψ^I in (34) if $\mathbf{\Lambda}$ and σ^2 are unknown, or ψ^F in (14) if $\mathbf{\Lambda}$ and σ^2 are known or fixed to their respective prior means $\mathbf{\Lambda}_0$ and σ_0^2.

Let the number of observations per subject be $m = 12$. Tables 1 and 2 report the ψ^{EF}-optimal designs obtained through a computer search given different values of $r = \tilde{d}/\tilde{a}$ in $\mathbf{\Lambda}_0$, with $\mathbf{\Lambda}_0$ in the form of

$$\mathbf{\Lambda}_0^{\sharp} = (\tilde{a}\mathbf{I}_3 + \tilde{d}\mathbf{J}_3) \quad \text{and} \quad \mathbf{\Lambda}_0^{\flat} = \begin{pmatrix} \lambda_0 & \mathbf{0} \\ \mathbf{0} & \tilde{a}\mathbf{I}_2 + \tilde{d}\mathbf{J}_2 \end{pmatrix} \tag{38}$$

respectively. As shown in Theorem 3 and Corollary 2, $\mathbf{M}(\eta^*)$ of a ψ^F-optimal continuous design η^* is of the form

$$(1+\epsilon)\mathbf{I}_3 - \epsilon\mathbf{J}_3 \quad \text{and} \quad \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & (1+\epsilon)\mathbf{I}_2 - \epsilon\mathbf{J}_2 \end{pmatrix} \tag{39}$$

respectively, where ϵ is given in (26). The value of ϵ is reported in the second column of each table. In the third column, the ψ^{EF}-optimal design from the computer search is expressed as $(m_{11}, m_{12}, m_{21}, m_{22})$, where m_{ij} is the number of times level i of factor 1 and level j of factor 2 occur together in the design. The resulting matrix $\mathbf{X}'\mathbf{X}$ of the ψ^{EF}-optimal design is reported in the next column. The ψ^F-efficiency for each design is calculated from (36). The D-efficiency in (35) is also reported.

Note that reversing factor levels will change the signs of the corresponding covariance terms in $\mathbf{\Lambda}_0$. In this example, we consider only nonnegative ratios $r = \tilde{d}/\tilde{a}$ in Tables 1 and 2, but in Section 5.2 and 5.3, we examine the robustness of the designs in Tables 1 and 2 to the different specifications of $\mathbf{\Lambda}_0$, which include negative covariances.

The efficiencies listed in Tables 1 and 2 show that the ψ^F efficiencies of the computer-searched ψ^{EF}-optimal designs are very close to their respective theoretical upper bounds. When the random effects β_i are independent, then $\tilde{d}/\tilde{a} = 0$, and the orthogonal design with equal allocation of m_{11}, m_{12}, m_{21} and m_{22} is both D- and ψ^{EF}- optimal. As the random effects β_i become more highly correlated (that is, as $\tilde{d}/\tilde{a}$ increases), the computer-searched ψ^{EF}-optimal design becomes nonorthogonal (Table 2), or nonorthogonal and unbalanced (Table 1). For example, when $\tilde{d}/\tilde{a} = 10$ in Table 1, $m_{22} = 0$ and the ψ^{EF}-optimal design obtained from the computer search does not contain observations on the combination of the second levels of the two factors. 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^F- and the D- efficiency values in the last two columns of Tables 1 and 2 suggest that the optimal designs under the ψ^F-criterion are far from what are considered optimal under the D-criterion when the ratio $r = \tilde{d}/\tilde{a}$ is large (so that the random effects are highly correlated).

Ratio $\tilde{d}/\tilde{a}$	ϵ in $\mathbf{M}(\eta^*)$	Design $(m_{11}, m_{12}, m_{21}, m_{22})$	Matrix $\mathbf{X}'\mathbf{X}$	Efficiency ψ^F	Efficiency D
0	0	(3,3,3,3)	$12\mathbf{I}_3$	1.0000	1.0000
0.5	0.14	(4,3,3,2)	$\begin{pmatrix} 12 & -2 & -2 \\ -2 & 12 & 0 \\ -2 & 0 & 12 \end{pmatrix}$	0.9992	0.9811
2	0.26	(4,4,3,1)	$\begin{pmatrix} 12 & -4 & -2 \\ -4 & 12 & -2 \\ -2 & -2 & 12 \end{pmatrix}$	0.9992	0.9341
10	0.37	(4,4,4,0)	$\begin{pmatrix} 12 & -4 & -4 \\ -4 & 12 & -4 \\ -4 & -4 & 12 \end{pmatrix}$	0.9997	0.8399

TABLE 1

ψ^{EF}-optimal 12-run designs of Example 5.1 with fixed $\mathbf{\Lambda} = \tilde{a}\mathbf{I}_3 + \tilde{d}\mathbf{J}_3$

Ratio $\tilde{d}/\tilde{a}$	ϵ in $\mathbf{M}(\eta^*)$	Design $(m_{11}, m_{12}, m_{21}, m_{22})$	Matrix $\mathbf{X}'\mathbf{X}$	Efficiency ψ^F	Efficiency D
0	0	(3,3,3,3)	$12\mathbf{I}_3$	1.0000	1.0000
0.5	0.17	(3,3,3,3)	$12\mathbf{I}_3$	0.9990	1.0000
2	0.38	(2,4,4,2)	$\begin{pmatrix} 12 & 0 & 0 \\ 0 & 12 & -4 \\ 0 & -4 & 12 \end{pmatrix}$	0.9999	0.9615
10	0.64	(1,5,5,1)	$\begin{pmatrix} 12 & 0 & 0 \\ 0 & 12 & -8 \\ 0 & -8 & 12 \end{pmatrix}$	0.9999	0.8221

TABLE 2

ψ^{EF}-optimal 12-run designs of Example 5.1 with fixed $\mathbf{\Lambda} = \begin{pmatrix} \lambda_0 & \mathbf{0} \\ \mathbf{0} & \tilde{a}\mathbf{I}_2 + \tilde{d}\mathbf{J}_2 \end{pmatrix}$

Tables 3 and 4 provide the ψ^{EI}-optimal designs obtained through computer search that maximize (34), with prior distribution means $E(\sigma^2) = 1$, and $E(\mathbf{\Lambda})$ in the form of $\mathbf{\Lambda}_0^\sharp$ and $\mathbf{\Lambda}_0^\flat$ in (38), respectively, under the following

Ratio $\tilde{d}/\tilde{a}$	Design $(m_{11}, m_{12}, m_{21}, m_{22})$	Matrix $\mathbf{X}'\mathbf{X}$	Relative Efficiency Rel.ψ^I-eff
0	(3,3,3,3)	$12\mathbf{I}_3$	1.000
0.5	(4,3,3,2)	$\begin{pmatrix} 12 & -2 & -2 \\ -2 & 12 & 0 \\ -2 & 0 & 12 \end{pmatrix}$	1.003(=1/0.997)
2	(4,4,3,1)	$\begin{pmatrix} 12 & -4 & -2 \\ -4 & 12 & -2 \\ -2 & -2 & 12 \end{pmatrix}$	1.012(=1/0.988)
10	(4,4,4,0)	$\begin{pmatrix} 12 & -4 & -4 \\ -4 & 12 & -4 \\ -4 & -4 & 12 \end{pmatrix}$	1.026(=1/0.974)

TABLE 3
ψ^{EI}-optimal 12-run designs of Example 5.1 with $E(\mathbf{\Lambda}) = \tilde{a}\mathbf{I}_3 + \tilde{d}\mathbf{J}_3$

Ratio $\tilde{d}/\tilde{a}$	Design $(m_{11}, m_{12}, m_{21}, m_{22})$	Matrix $\mathbf{X}'\mathbf{X}$	Relative Efficiency Rel. ψ^I-eff
0	(3,3,3,3)	$12\mathbf{I}_3$	1.000
0.5	(3,3,3,3)	$12\mathbf{I}_3$	1.000
2	(2,4,4,2)	$\begin{pmatrix} 12 & 0 & 0 \\ 0 & 12 & -4 \\ 0 & -4 & 12 \end{pmatrix}$	1.007 (=1/0.993)
10	(1,5,5,1)	$\begin{pmatrix} 12 & 0 & 0 \\ 0 & 12 & -8 \\ 0 & -8 & 12 \end{pmatrix}$	1.018 (=1/0.983)

TABLE 4
ψ^{EI}-optimal 12-run designs of Example 5.1 with $E(\mathbf{\Lambda}) = \begin{pmatrix} \lambda_0 & \mathbf{0} \\ \mathbf{0} & \tilde{a}\mathbf{I}_2 + \tilde{d}\mathbf{J}_2 \end{pmatrix}$

prior distributions

$$\sigma_i^2 \sim \text{Inverse Gamma}(3/2, 1/2),$$

$$\mathbf{\Lambda} \sim \text{Inverted Wishart}(5,\ \tilde{a}\mathbf{I}_3 + \tilde{d}\mathbf{J}_3), \quad \text{for Table 3,}$$

$$\mathbf{\Lambda} \sim \text{Inverted Wishart}\left(5,\ \begin{pmatrix} \lambda_0 & \mathbf{0} \\ \mathbf{0} & \tilde{a}\mathbf{I}_2 + \tilde{d}\mathbf{J}_2 \end{pmatrix}\right), \quad \text{for Table 4.}$$

The relative ψ^I-efficiency values (37) reported in Tables 3 and 4 show that as the random effects become more highly correlated, the orthogonal design becomes inferior to nonorthogonal designs. A comparison of the designs in Tables 1 and 2 with those in Tables 3 and 4, respectively, show that the ψ^{EF}-optimal designs are the same as the ψ^{EI}-optimal designs.

We also examined some ψ^{EF}-optimal and ψ^{EI}-optimal designs searched

under other $E(\mathbf{\Lambda})$ values such as those two reported in Table 5. Our results show that ψ^{EF}-optimal designs are not always the same as the corresponding ψ^{EI}-optimal designs but, in each of our examples, their relative ψ^{I}-efficiency has always been close to that of the ψ^{EI}-optimal design.

$E(\mathbf{\Lambda})$	Criterion	$(m_{11}, m_{12}, m_{21}, m_{22})$	$\mathbf{X}'\mathbf{X}$	Relative ψ^{I}-eff.
$\begin{pmatrix} 0.47 & 0.19 & 0.46 \\ 0.19 & 5.50 & 0.24 \\ 0.46 & 0.24 & 0.48 \end{pmatrix}$	ψ^{EF}	(5,1,5,1)	$\begin{pmatrix} 12 & 0 & -8 \\ 0 & 12 & 0 \\ -8 & 0 & 12 \end{pmatrix}$	1.052(=1/0.951)
	ψ^{EI}	(3,2,5,2)	$\begin{pmatrix} 12 & 0 & -6 \\ 0 & 12 & -2 \\ -6 & -2 & 12 \end{pmatrix}$	1.060(=1/0.943)
$\begin{pmatrix} 1.43 & 0.16 & 2.62 \\ 0.16 & 0.10 & 0.31 \\ 2.62 & 0.31 & 5.02 \end{pmatrix}$	ψ^{EF}	(3,2,5,2)	$\begin{pmatrix} 12 & 0 & -6 \\ 0 & 12 & -2 \\ -6 & -2 & 12 \end{pmatrix}$	1.042(=1/0.960)
	ψ^{EI}	(3,2,5,2)	$\begin{pmatrix} 12 & 0 & -6 \\ 0 & 12 & -2 \\ -6 & -2 & 12 \end{pmatrix}$	1.042(=1/0.960)

TABLE 5

ψ^{EF}- and ψ^{EI}-optimal 12-run designs of Example 5.1 with alternative $E(\mathbf{\Lambda})$ values

EXAMPLE 5.2: Consider an experiment with three treatment factors each with two levels. The number of runs allocated to each subject is $m = 8$. The same assumptions are made as those in Example 5.1, and the same design criteria apply. As in Example 5.1, the ψ^{EF}-optimal designs are the same as the ψ^{EI}-optimal designs (see Tables 6 to 7), or almost as efficient as the ψ^{EI}-optimal designs (see Table 8).

The results of Examples 5.1 and 5.2 suggest that the ψ^{F} criterion, with $\mathbf{\Lambda}$ and σ^2 fixed to the means of their respective prior distributions, is a good surrogate for the ψ^{I} criterion. This information is very helpful in practical applications since the search for ψ^{EF}-optimal designs is far less computationally intensive than the search for ψ^{EI}-optimal designs, especially when the numbers of runs, factors and factor levels are large.

5.2. *Design robustness when covariances are all positive.* The results in Section 5.1 show that the ψ^{EF}-optimal designs are highly ψ^{I}-efficient. As

Ratio $\tilde{d}/\tilde{a}$	Design $\binom{m_{111},m_{112},m_{121},m_{122}}{m_{211},m_{212},m_{221},m_{222}}$	Matrix $\mathbf{X}'\mathbf{X}$	Efficiency ψ^F	Efficiency D	Relative Efficiency Rel.ψ^I-eff
0	$\binom{1,1,1,1}{1,1,1,1}$	$8\mathbf{I}_4$	1.000	1.000	1.000
0.5	$\binom{1,1,1,1}{1,1,1,1}$	$8\mathbf{I}_4$	0.997	1.000	1.000
2	$\binom{1,2,1,1}{1,1,1,0}$	$\begin{pmatrix} 8 & -2 & -2 & 0 \\ -2 & 8 & 0 & -2 \\ -2 & 0 & 8 & -2 \\ 0 & -2 & -2 & 8 \end{pmatrix}$	0.997	0.931	1.012(=1/0.988)
10	$\binom{1,2,1,1}{1,1,1,0}$	$\begin{pmatrix} 8 & -2 & -2 & 0 \\ -2 & 8 & 0 & -2 \\ -2 & 0 & 8 & -2 \\ 0 & -2 & -2 & 8 \end{pmatrix}$	0.994	0.931	1.021 (=1/0.980)

TABLE 6

ψ^{EF}- and ψ^{EI}-optimal 8-run designs of Example 5.2 with $E(\mathbf{\Lambda}) = \tilde{a}\mathbf{I}_4 + \tilde{d}\mathbf{J}_4$

Ratio $\tilde{d}/\tilde{a}$	Design $\binom{m_{111},m_{112},m_{121},m_{122}}{m_{211},m_{212},m_{221},m_{222}}$	Matrix $\mathbf{X}'\mathbf{X}$	Efficiency ψ^F	Efficiency D	Relative Efficiency Rel.ψ^I-eff
0	$\binom{1,1,1,1}{1,1,1,1}$	$8\mathbf{I}_4$	1.000	1.000	1.000
0.5	$\binom{1,1,1,1}{1,1,1,1}$	$8\mathbf{I}_4$	0.998	1.000	1.000
2	$\binom{1,1,1,2}{1,1,1,0}$	$\begin{pmatrix} 8 & -2 & 0 & 0 \\ -2 & 8 & -2 & -2 \\ 0 & -2 & 8 & 0 \\ 0 & -2 & 0 & 8 \end{pmatrix}$	0.996	0.950	1.003 (=1/0.997)
10	$\binom{0,2,2,1}{1,1,1,0}$	$\begin{pmatrix} 8 & -2 & 0 & 0 \\ -2 & 8 & -2 & -2 \\ 0 & -2 & 8 & -4 \\ 0 & -2 & -4 & 8 \end{pmatrix}$	0.997	0.847	1.020 (=1/0.981)

TABLE 7

ψ^{EF}- and ψ^{EI}-optimal 8-run designs of Example 5.2 with $E(\mathbf{\Lambda}) = \begin{pmatrix} \lambda_0 & \mathbf{0} \\ \mathbf{0} & \tilde{a}\mathbf{I}_3+\tilde{d}\mathbf{J}_3 \end{pmatrix}$

$E(\mathbf{\Lambda})$	Criterion	Design	$\mathbf{X}'\mathbf{X}$	Relative ψ^I-eff.
$\begin{pmatrix} 5.81 & 1.54 & 3.50 & 1.82 \\ 1.54 & 2.04 & 2.57 & 0.17 \\ 3.50 & 2.57 & 5.62 & 1.63 \\ 1.82 & 0.17 & 1.63 & 1.32 \end{pmatrix}$	ψ^{EF}	$\binom{1,1,2,0}{1,2,1,0}$	$\begin{pmatrix} 8 & 0 & -2 & -2 \\ 0 & 8 & -2 & 2 \\ -2 & -2 & 8 & -4 \\ -2 & 2 & -4 & 8 \end{pmatrix}$	1.017(=1/0.983)
	ψ^{EI}	$\binom{1,0,2,1}{1,2,1,0}$	$\begin{pmatrix} 8 & 0 & 0 & -2 \\ 0 & 8 & -4 & 2 \\ 0 & -4 & 8 & -2 \\ -2 & 2 & -2 & 8 \end{pmatrix}$	1.029(=1/0.972)
$\begin{pmatrix} 4.43 & 1.53 & 3.44 & 0.53 \\ 1.53 & 2.41 & 2.57 & 1.47 \\ 3.44 & 2.57 & 4.21 & 0.73 \\ 0.53 & 1.47 & 0.73 & 4.22 \end{pmatrix}$	ψ^{EF}	$\binom{1,1,0,1}{2,2,1,0}$	$\begin{pmatrix} 8 & 0 & -4 & 0 \\ 0 & 8 & -4 & 0 \\ -4 & -4 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix}$	1.036(=1/0.965)
	ψ^{EI}	$\binom{1,1,0,1}{2,2,1,0}$	$\begin{pmatrix} 8 & 0 & -4 & 0 \\ 0 & 8 & -4 & 0 \\ -4 & -4 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix}$	1.036(=1/0.965)

TABLE 8

ψ^{EF}- and ψ^{EI}-optimal 8-run designs of Example 5.2 with alternative $E(\mathbf{\Lambda})$ values

it is seldom the case that the experimenter has complete knowledge of the mean of the covariance matrix $\mathbf{\Lambda}$, we examine in this section the robustness of the ψ^{EF}-optimal designs under different specifications of $E(\mathbf{\Lambda})$.

Let D_{05}, D_2 and D_{10} denote, respectively, the nonorthogonal and unbalanced ψ^{EF}-optimal design in Table 1 for $\tilde{d}/\tilde{a} = 0.5$, $\tilde{d}/\tilde{a} = 2$ and $\tilde{d}/\tilde{a} = 10$ with respective treatment allocation (4,3,3,2), (4,4,3,1) and (4,4,4,0). Similarly, let T_2 and T_{10} denote the nonorthogonal ψ^{EF}-optimal design in Table 2 for $\tilde{d}/\tilde{a} = 2$ and $\tilde{d}/\tilde{a} = 10$ with treatment allocation (2,4,4,2) and (1,5,5,1), respectively. Using the orthogonal design with treatment allocation (3,3,3,3) as the base design, we examine the relative ψ^I-efficiency in (37) of each of these designs under different specifications of $E(\mathbf{\Lambda})$ and $E(\sigma^2) = 1$.

We start with the scenario where the random effects are all positively correlated, and we generate the variances (diagonal elements) in $E(\mathbf{\Lambda})$ independently from a uniform (0,10) distribution. For the covariances (off-diagonal elements) in $E(\mathbf{\Lambda})$, we generate correlation values from a uniform (0,1) distribution, and multiply these with the square root of the corresponding variances to obtain the covariances. The generation of $E(\mathbf{\Lambda})$ is done 10,000 times, and for each $E(\mathbf{\Lambda})$ the relative ψ^I-efficiency in (37) is calculated for each of the designs D_{05}, D_2, D_{10}, T_2 and T_{10}, and boxplots of the respective distributions are shown in Figure 1. The boxplots show that, over the 10,000 simulated values of $E(\mathbf{\Lambda})$, the nonorthogonal and unbalanced designs D_{05} and D_2, obtained under positively correlated $\mathbf{\Lambda}_0$ of form $\mathbf{\Lambda}_0^\sharp$ in (38) with moderate and equal correlations, are more likely to be ψ^I-efficient than the orthogonal design, whereas D_{10}, T_2 and T_{10} are less likely to be as ψ^I-efficient as the orthogonal design. Specifically, D_{05} is superior to the orthogonal design on 77% of occasions and is never below 97.87% efficiency. D_2 is superior to the orthogonal design 63% of the time. On the other hand,

design D_{10} is inferior to the orthogonal 68% of the time, T_2 is inferior 60% of the time, and T_{10} is inferior 89% of the time. The findings in Example 5.2, as shown in Figure 2, are less clear cut but they still slightly favor D_2. These findings imply that, when all pairs of random effects are believed to be positively correlated, but the sizes of the variances and covariances are unknown, a ψ^{EF}-optimal design with high ψ^I-efficiency is more likely to be achieved if it is searched under positively correlated $\mathbf{\Lambda}_0$ with moderate and equal correlations.

5.3. *Design robustness when covariances may be negative.* Similar simulation studies were conducted when covariances in $E(\mathbf{\Lambda})$ may be negative (see Liu et al., 2005). The findings are that, in general, when certain random effects may be negatively correlated, ψ^{EF}-optimal designs searched under positively-correlated $\mathbf{\Lambda}_0$s are less likely to be ψ^I-efficient than the orthogonal design. This implies that the knowledge on the signs of the correlations of the random effects should be used in the search of ψ^{EF}-optimal designs to achieve higher ψ^I-efficiency.

6. Summary and conclusion. In this paper, we have investigated optimal designs for the estimation of hyperparameters in hierarchical linear models. A design criterion (ψ^I-criterion) was derived which requires the integration over the prior distribution of the random effects covariance matrix $\mathbf{\Lambda}$ and the response error σ^2. We focused on the maximization of the integrand (ψ^F-criterion) by fixing $\mathbf{\Lambda}$ and σ^2 equal to their respective prior means. The form of $\mathbf{M}(\eta^*)$ of a ψ^F-optimal continuous design was derived for some special scenarios. We showed that level-balanced orthogonal designs, if they exist, are optimal when the random effects are independently distributed. However, when the random effects are correlated, nonorthogonal designs are

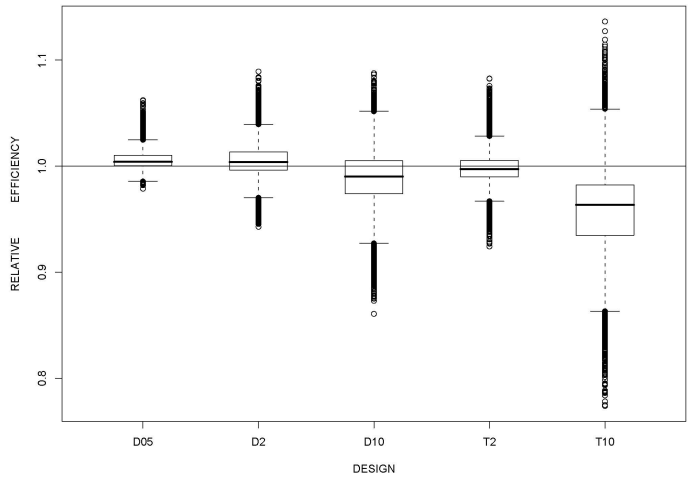


FIG 1. *Relative ψ^I-efficiency under 10,000 different specifications of $E(\mathbf{\Lambda})$ in Example 5.1 where all covariance terms in $E(\mathbf{\Lambda})$ are positive*

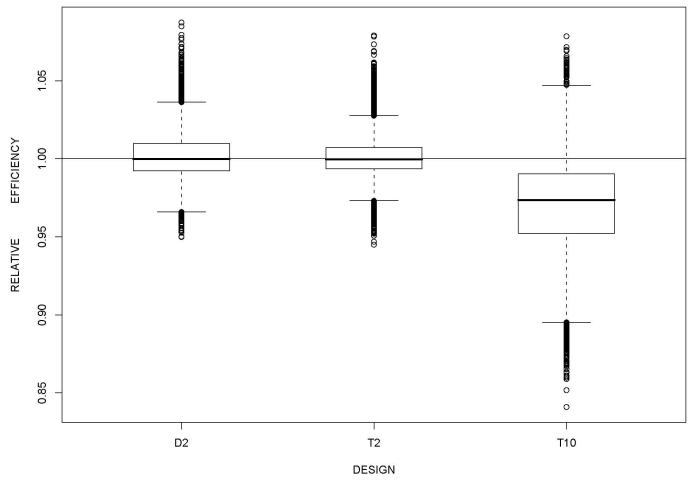


FIG 2. *Relative ψ^I-efficiency under 10,000 different specifications of $E(\mathbf{\Lambda})$ in Example 5.2 where all covariance terms in $E(\mathbf{\Lambda})$ are positive*

optimal. We also discussed briefly optimal designs for the situation when both the treatment allocation and the subject covariate selection are under the control of the experimenter.

We showed examples which suggest that the ψ^{EF}-optimal designs are highly ψ^{I}-efficient. In addition, the robustness study under different specification of $E(\mathbf{\Lambda})$ showed that, when all pairs of random effects are positively correlated, ψ^{EF}-optimal designs obtained under positively correlated $\mathbf{\Lambda}_0$ with moderate and equal correlations are more likely to be ψ^{I}-efficient than the orthogonal design. However, these designs do not perform as well as the orthogonal design if some of the random effects are negatively correlated.

Practical implications of our findings are that (i) highly ψ^{I}-efficient designs can be obtained by fixing $\mathbf{\Lambda}$ and σ^2 to their prior means $\mathbf{\Lambda}_0$ and σ_0^2 and maximizing ψ^{F} in (14), a much faster approach than the direct search of ψ^{I}-optimal designs which requires the integration over the prior distributions of $\mathbf{\Lambda}$ and σ^2; (ii) when the signs of the correlations of the random effects are believed to be known but the approximate sizes of the variances and covariances are unknown, $\mathbf{\Lambda}_0$ should be specified with moderate sized correlations with the anticipated signs; (iii) when random effects are correlated, nonorthogonal designs tend to be more efficient than orthogonal designs under the ψ^{I} criterion.

APPENDIX

Proof of Lemma 2.

PROOF. Let $\mathbf{M}(\eta)$ be nonsingular, then the criterion function $\psi(\mathbf{M}(\eta))$ in (17) can be expressed as

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

Let $\mathbf{M}_1$ denote $\mathbf{M}(\eta_1)$ and $\mathbf{M}_2$ denote $\mathbf{M}(\eta_2)$, then $\mathbf{M}_1$ and $\mathbf{M}_2$ are both positive definite. From Minkowski's determinant theorem (Theorem 28, Magnus and Neudecker, 1999, page 227), if $\mathbf{M}_1 - \mathbf{M}_2$ is positive definite, then $|\mathbf{M}_1| > |\mathbf{M}_2|$, that is, the determinant function is strictly increasing. Also, from Theorem 12.2.14 of Graybill (1983), $\mathbf{M}_2^{-1} - \mathbf{M}_1^{-1}$ is positive definite, which implies that $(\frac{\sigma^2}{m}\mathbf{M}_2^{-1} + \mathbf{\Lambda}) - (\frac{\sigma^2}{m}\mathbf{M}_1^{-1} + \mathbf{\Lambda})$ is also positive definite for positive σ^2 and m. Since the determinant function is strictly increasing, we obtain

$$\left|\frac{\sigma^2}{m}\mathbf{M}_2^{-1} + \mathbf{\Lambda}\right| > \left|\frac{\sigma^2}{m}\mathbf{M}_1^{-1} + \mathbf{\Lambda}\right|. \tag{40}$$

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

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

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

Therefore, the log function of the above two positive determinants are defined. Since the log function is strictly increasing, we obtain from (40)

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

that is, $\psi(\mathbf{M}_1) > \psi(\mathbf{M}_2)$ for positive definite $(\mathbf{M}_1 - \mathbf{M}_2)$. Therefore $\psi(\mathbf{M}(\eta))$ is a strictly increasing function of $\mathbf{M}(\eta) \in \mathcal{M}$.

We now show that $\psi(\mathbf{M}(\eta))$ is a concave function in $\mathcal{M}$. Silvey (1980) Appendix 1 shows that $\mathbf{M}(\eta)^{-1}$ is convex, i.e.,

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

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

sides of the inequality, we obtain,

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

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

Acknowledgement. The authors would like to thank Kathryn Chaloner for her helpful comments on the paper.

References.

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

[2] Allenby, G. M. and Lenk, P. J. (1994). Modeling household purchase behavior with logistic normal regression. *Journal of American Statistical Association 89*, 1218–1229.

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

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

[5] Bradlow, E. T. and Rao, V. R. (2000). A hierarchical bayes model for assortment choice. *Journal of Marketing Research 37 (2)*, 259–268.

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

[7] Entholzner, M., Benda, N., Schmelter, T., and Schwabe, R. (2005). A note on designs for estimating population parameters. *Biometrical Letters – Listy Biometryczne*, 25–41.

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

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

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

[11] Han, C. and Chaloner, K. (2005). Design of population studies of hiv dynamics, chapter 21. In *Deterministic and Stochastic models of AIDS Epidemics and HIV infections with intervention*. Eds: Tan, W-Y. and WU, H., World Scientific Publishing Company, 525–548.

[12] Horney, J., Osgood, W. D., and Marshall, I. H. (1995). Criminal careers in the short-term: Intra-individual variability in crime and its relation to local life circumstances. *American Sociological Review 60 (5)*, 655–673.

[13] 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.

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

[15] Kuhfeld, W. F., Tobias, R. D., and Garratt, M. (1994). Efficient experimental design with marketing research applications. *Journal of Marketing Research 31 (4)*, 545–557.

[16] 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.

[17] Lenk, P. J. and Rao, A. G. (1990). New models from old: Forecasting product adoption by hierarchical bayes procedures. *Marketing Science 9 (1)*, 42–53.

[18] Lindley, D. and Smith, A. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society Series B 34*, 1–41.

[19] Liu, Q., Dean, A. M., and Allenby, G. M. (2005). Optimal experimental designs for hyperparameter estimation in hierarchical linear models. Tech. rep., 762, The Ohio State University, `http://www.stat.ohio-state.edu/~qliu`.

[20] Liu, Q., Dean, A. M., Bakken, D., and Allenby, G. M. (2006). Optimal ex-

perimental design for hyperparameter estimation: Learning when effect-sizes are large. Working Paper, `http://www.stat.ohio-state.edu/~qliu`.

[21] Magnus, J. R. and Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. New York: John Wiley.

[22] Manchanda, P., Ansari, A., and Gupta, S. (1999). The "shopping basket": A model for multicategory purchase incidence decisions. *Marketing Science 18 (2)*, 95–114.

[23] Mentré, F., Mallet, A., and Baccar, D. (1997). Optimal design in random-effects regression models. *Biometrika 84 (2)*, 429–442.

[24] Montgomery, A. L., Li, S., Srinivasan, K., and Liechty, J. C. (2004). Modeling online browsing and path analysis using clickstream data. *Marketing Science 23 (4)*, 579–595.

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

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

[27] Raudenbush, S. W. and Sampson, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology 29*, 1–41.

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

[29] Rumberger, R. W. (1995). Dropping out of middle schools: A multilevel analysis of students and schools. *American Educational Research Journal 32 (3)*, 583–625.

[30] Sándor, Z. and Wedel, M. (2002). Profile construction in experimental choice designs for mixed logit models. *Marketing Science 21 (4)*, 455–475.

[31] Scheffé, H. (1959). *The Analysis of Variance*. Wiley, New York.

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

[33] Tod, M., Mentré, F., Merlé, Y., and Mallet, A. (1998). Robust optimal design for the estimation of hyperparameters in population pharmacokinetics. *Journal of Pharmacokinetics and Biopharmaceuics 26*, 689–716.

[34] 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 Statistics
The Ohio State University
1958 Neil Avenue
Columbus, OH 43210
E-mail: qliu@stat.ohio-state.edu
amd@stat.ohio-state.edu

Department of Marketing
The Ohio State University
2100 Neil Avenue
Columbus, OH 43210
E-mail: allenby_1@cob.osu.edu