Statistica Sinica **13**(2003), 385-401

OPTIMAL DESIGNS FOR MIXED-EFFECTS MODELS WITH TWO RANDOM NESTED FACTORS

Bruce E. Ankenman, Ana Ivelisse Avilés and José C. Pinheiro

Northwestern University, US Department of Commerce, NIST and Biostatistics, Novartis Pharmaceuticals

Abstract: The main objective of this paper is to provide experimental designs for the estimation of fixed effects and two variance components, in the presence of nested random effects. Random nested factors arise from quantity designations such as lot or batch, and from sampling and measurement procedures. We introduce a general class of designs for mixed-effects models with random nested factors, called assembled designs, where the nested factors are nested under the treatment combinations of the crossed factors. We provide parameters and notation for describing and enumerating assembled designs. Conditions for existence and uniqueness of D-optimal assembled designs for the case of two variance components are presented. Specifically, we show that, in most practical situations, designs that are most balanced (i.e., where the samples are distributed as uniformly as possible among batches) result in D-optimal designs for maximum likelihood estimation.

Key words and phrases: Assembled designs, crossed and nested factors, D-optimality, experimental design, fixed and random effects, hierarchical nested design, maximum likelihood, nested factorials, variance components.

1. Introduction

In many experimental settings, different types of factors affect the measured response. The factors of primary interest can usually be set independently of each other and thus are called crossed factors. For example in a welding operation, clamp pressure and voltage level are crossed factors since each level of voltage can be applied at each level of clamp pressure. These effects are often modeled as fixed effects. Nested factors cannot be set independently because the level of one factor has a different meaning when other factors are changed. Random nested factors arise from quantity designations such as lot or batch, and from sampling procedures that are often inherent in the experimentation. The variances of the random effects of nested factors are called variance components since they are components of the random variation.

The *nested or hierarchical nested design* (HND), used in many sampling and testing situations, is a design where the levels of each factor are nested within the levels of another factor. Balanced HNDs for estimating variance components have an equal number of observations for each branch at a given level of nesting. Not only do these designs tend to require a large number of observations, but they also tend to produce precise estimates of certain variance components and poor estimates of others. Articles that address these issues are Bainbridge (1965), Smith and Beverly (1981), and Naik and Khattree (1998). These articles use unbalanced HNDs, called *staggered nested designs*, to spread the information in the experiment more equally among the variance components. Goldsmith and Gaylor (1970) address the optimality of unbalanced HNDs and Delgado and Iyer (1999) extend this work to obtain nearly optimal HNDs using both a limit argument and numerical optimization.

When the fixed effects of crossed factors and variance components from random nested factors appear in the same model, many analysis techniques are available for estimation and inference (see Searle, Casella and McCulloch (1992) (hereafter SCM) or Pinheiro and Bates (2000)). However, only limited work has been done to determine the experimental designs for such cases. Smith and Beverly (1981) introduced the idea of a nested factorial, which is an experimental design where some factors appear in factorial relationships and others in nested relationships. They proposed placing staggered nested designs at each treatment combination of a crossed factor design and called the resulting designs *staggered nested factorials*. Ankenman, Liu, Karr and Picka (2001) introduced *split factorials*, which split a factorial design into sub-experiments. The nested designs in a split factorial only branch at a single level and thus the effect is to study a different variance component in each sub-experiment.

This paper provides experimental design procedures for the estimation of both fixed effects of crossed factors as well as variance components associated with nested random factors. As an example, consider a concrete mixing experiment where the size of the aggregate and the ratio of water to cement powder are crossed factors and batch-to-batch and sample-to-sample variances are the variance components of interest. The sizes of the variance components indicate the variability of the properties of the concrete throughout concrete structures.

In the next section, we introduce a special class of nested factorials, called *assembled designs*, where the nested factors are random and nested under the treatment combinations of the crossed factors. This class of designs includes both split factorials and staggered nested factorial designs. In Section 3, we describe a linear mixed-effects model for the analysis of assembled designs. The fixed effects and the variance components are estimated using maximum likelihood (ML). In Section 4, we present the information matrix of the simplest assembled design. In Section 5, we provide theorems which show that under most practical situations, the most balanced design is D-optimal for estimating the fixed effects and two variance components. In Section 6, we show with examples how to obtain the D-optimal design and how these designs compare to other alternatives. Conclusions are presented in Section 7.

2. Assembled Designs with Two Variance Components

An assembled design (AD) is a crossed factor design with an HND placed at each design point. The class of assembled designs is large and contains many designs that are too complicated for practical use. We restrict our attention to the case of ADs with two variance components and leave ADs with multiple variance components for future research. The terms *batch* and *sample* will refer to the higher and lower levels of random effects, respectively.

An AD with two variance components is described by various parameters. The number of design points in the crossed factor design is denoted r and s is the number of non-identical HNDs used. If the same number of observations is taken at each design point, it is denoted n. Define B_T as the total number of batches, B_j as the number of batches in the jth HND, $j = 1, \ldots, s$ and r_j as the number of design points that contain the jth HND. Thus, $\sum_{j=1}^{s} r_j B_j = B_T$.

A simplified version of the concrete mixing experiment in Jaiswal, Picka, Igusa, Karr, Shah, Ankenman and Styer (2000) is used to illustrate the assembled design. The objective of the experiment is to determine the effects of certain crossed factors on the permeability of concrete and to estimate the variance of the permeability from batch-to-batch and from sample-to-sample. The design (Figure 1) has three two-level crossed factors, Aggregate Grade, Water to Cement (W/C) Ratio, and Max Size of Aggregate. There are a total of 20 batches ($B_T = 20$). In this application, as in most engineering applications and many other applications, the higher quantity designation (batch) implies higher variance. Thus, we assume that the batch-to-batch variance is at least as large as the sample-to-sample variance.

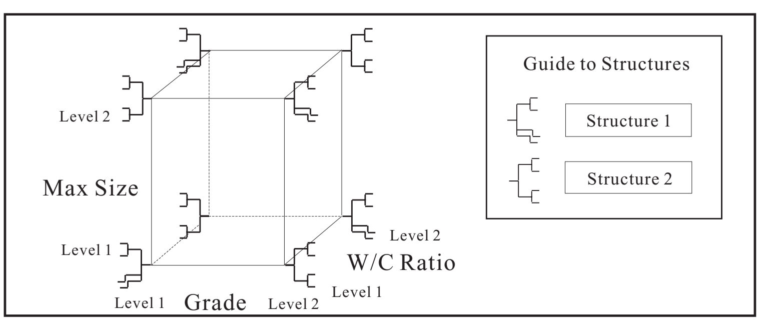

Figure 1. An assembled design ($B_T = 20$, $r = 8$, $n = 4$, $s = 2$).

In Figure 1, each vertex of the cube represents one of the eight possible concrete recipes or design points ($r = 8$) that can be made using the two levels of three factors. Thus, the front lower left-hand vertex represents a concrete recipe with the low level of all three factors.

In the context of an assembled design, HNDs attached to the crossed factor design points will be called *structures*. The structure at each design point represents the batches and samples to be made and tested from that recipe. There are four samples tested per recipe ($n = 4$) and two non-identical structures ($s = 2$). Structure 1 consists of three batches ($B_1 = 3$) where two samples are cast from one batch and one sample is cast from each of the other two. It appears at four of the design points ($r_1 = 4$). Structure 2 appears at the other four design points ($r_2 = 4$) and consists of two batches ($B_2 = 2$) with two samples cast from each batch. A vector-like notation for the two structures is shown in Figure 2, where each element in the vector is the number of samples in a batch. For uniqueness of equivalent structures, the elements are specified in descending order.

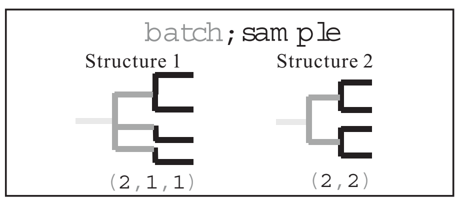

Figure 2. Notation for the structures in the concrete experiment.

The notation is easily extended for any two-level HND. The number of unique HNDs depends on n. For example, if $n = 3$ there are only three unique HNDs: (1,1,1), (2,1) and (3). However, if $n = 7$ there are 15 unique HNDs. The number of unique assembled designs increases even more quickly as n increases since there are r design points each of which can have a different structure.

The notation for an AD is $\sum_{j=1}^{s}$ Structure j@{design points with Structure j}, where the design points must be ordered in some way. We choose to order the design points so that all rows with the same structure are in adjacent rows. This order is called *design order*. Figure 3 shows that structures in the concrete experiment were assigned to the design points using the interaction ABC and the comparison of design order and standard order (see Myers and Montgomery (1995, p.84)) for a two-level factorial. These orderings are for convenience in describing the experiment and in manipulating the expressions of the model and analysis. When conducting the experiment, the order of the observations should be randomly determined whenever possible.

An assembled design has nr observations and thus nr degrees of freedom. There are r degrees of freedom for estimating the fixed effects including the constant. There are $nr-r$ degrees of freedom for estimating variance components. Let d_i be the degrees of freedom for estimating the ith variance component. Hence, $d_1 = \sum_{j=1}^{s} r_j(B_j - 1) = B_T - r$ and $d_2 = nr - d_1 - r = nr - B_T$.

Design Order	Standard Order	A	B	C	ABC	Structure
1	1	-	-	-	-	
2	4	+	+	-	-	
3	6	+	-	+	-	
4	7	-	+	+	-	
5	2	+	-	-	+	
6	3	-	+	-	+	
7	5	-	-	+	+	
8	8	+	+	+	+	

Notation in Design Order: (2,1,1)@{1,2,3,4}+(2,2)@{5,6,7,8}

Notation in Standard Order: (2,1,1)@{1,4,6,7}+(2,2)@{2,3,5,8}

Figure 3. Concrete Exp. (A=Grade, B=W/C Ratio, and C=Max Size); Comparison of Standard and Design Order.

3. Analysis of Assembled Designs

3.1. Model and variance structure

The linear mixed-effects model used to represent the response in an AD with nr observations and two variance components is

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z}_1\boldsymbol{u}_1 + \boldsymbol{Z}_2\boldsymbol{u}_2, \tag{1}$$

where $\boldsymbol{y}$ is a vector of nr observations, $\boldsymbol{X}$ is the fixed-effects model matrix, $\boldsymbol{\beta}$ is a vector of r unknown coefficients including the constant term, $\boldsymbol{Z}_i$ is an indicator matrix associated with the ith variance component, $\boldsymbol{u}_i$ is a vector of normally distributed independent random effects associated with the ith variance component such that $\boldsymbol{u}_i \sim N(\boldsymbol{0}, \sigma_i^2\boldsymbol{I})$. Let $\boldsymbol{V}$ be the $nr \times nr$ variance-covariance matrix of the observations. Assume that the variance components do not depend on the crossed factor levels. Then,

$$\boldsymbol{V} = \mathrm{Var}\,(\boldsymbol{y}) = \sigma_1^2\boldsymbol{Z}_1\boldsymbol{Z}_1' + \sigma_2^2\boldsymbol{Z}_2\boldsymbol{Z}_2'. \tag{2}$$

In keeping with our previous terminology, σ_1^2 is the batch-to-batch variance and σ_2^2 is the sample-to-sample variance.

Let $\boldsymbol{X}_D$ be the full rank $r \times r$ model matrix (including the constant column and all estimable contrasts) for a single replicate of a crossed factor design, where rows are ordered in design order. Also let the observations in $\boldsymbol{X}$ be ordered such that $\boldsymbol{X} = \boldsymbol{X}_D \otimes \mathbf{1}_n$, where $\mathbf{1}_n$ is an n-length vector of ones and $\otimes$ represents the Kronecker product. This ordering in $\boldsymbol{X}$ gives rise to Z-matrices that have the form $\boldsymbol{Z}_i = \overset{r}{\underset{t=1}{\oplus}} \boldsymbol{Z}_{it}$, where $\oplus$ refers to the Kronecker sum and $\boldsymbol{Z}_{it}$ is the indicator matrix related to the observations associated with variance component i for the treatment combination (or design point) t. The $\boldsymbol{Z}_{1t}$ matrix has n rows and as many columns as the number of batches used with treatment combination t. Hence, the batch indicator matrix, $\boldsymbol{Z}_1$, has as many rows as the total number of samples (nr) and as many columns as the total number of batches (B_T). The sample indicator matrix, $\boldsymbol{Z}_2$, is the identity matrix of order nr. For example, recall that in the concrete experiment (introduced in Section 2) there are eight design points ($r = 8$), two structures ($s = 2$), where Structure 1 is (2,1,1) and Structure 2 is (2,2), and four observations at each design point ($n = 4$). Using design order, $\boldsymbol{Z}_2 = \boldsymbol{I}_{32}$,

$$\boldsymbol{Z}_{11} = \boldsymbol{Z}_{12} = \boldsymbol{Z}_{13} = \boldsymbol{Z}_{14} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{Z}_{15} = \boldsymbol{Z}_{16} = \boldsymbol{Z}_{17} = \boldsymbol{Z}_{18} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix},$$

$$\text{and} \quad \boldsymbol{Z}_1 = \begin{bmatrix} \boldsymbol{Z}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z}_{12} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{Z}_{18} \end{bmatrix}.$$

For the fixed effects, $\boldsymbol{X}_t$ is the portion of the $\boldsymbol{X}$ matrix associated with the tth treatment combination. Then, $\boldsymbol{X}_t$ is an $n \times r$ matrix where all n rows are identical. Let $\boldsymbol{x}'_{Dt}$ represent the row of $\boldsymbol{X}_D$ corresponding to the tth treatment combination, then $\boldsymbol{X}_t = \boldsymbol{x}'_{Dt} \otimes \mathbf{1}_n$. Let $\boldsymbol{V}_t$ be the $n \times n$ variance-covariance matrix of the observations associated with treatment combination t, then $\boldsymbol{V}_t = \sigma_1^2 \boldsymbol{Z}_{1t}\boldsymbol{Z}'_{1t} + \sigma_2^2 \boldsymbol{Z}_{2t}\boldsymbol{Z}'_{2t}$. Because of the independence of the observations from different treatment combinations, $\boldsymbol{V}$ in (2), can be written as $\boldsymbol{V} = \overset{r}{\underset{t=1}{\oplus}} \boldsymbol{V}_t$.

3.2. Estimation method

Maximum likelihood (ML) and restricted maximum likelihood (REML) are commonly used methods for the estimation of parameters in linear mixed-effects models. ML estimation prescribes the maximization of the likelihood function over the parameter space, conditional on the observed data. REML estimation is effectively ML estimation based on the likelihood of the ordinary least-squares residuals of the response vector $\boldsymbol{y}$ regressed on the $\boldsymbol{X}$ matrix of the fixed effects. Because it takes into account the loss of degrees of freedom due to the estimation of the fixed effects, REML estimates of variance components tend to be less biased than ML estimates. However, the expressions for the information matrix for REML estimation are complicated and were found to be analytically intractable for proving D-optimality of assembled designs. In this paper, we use ML estimation for the fixed effects and variance components because of the greater simplicity of the expressions for the associated information matrices. However, since ML and REML estimators are asymptotically equivalent, we expect that the D-optimal designs for ML will also be optimal (or nearly optimal) for REML. This conjecture was confirmed for small designs by exhaustive search. In an empirical study ($r = 8$ and 16, and $n = 3$ to 7), we found that the D-optimal designs for ML and REML estimation were identical as long as $\sigma_1^2 \geq \sigma_2^2$, which, as previously mentioned, is true for most engineering applications.

Conditional on the variance components, the estimation of the fixed effects is a generalized least squares (GLS) problem, with solution $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$ (see Theil (1971, p.238)). Since the variance components need to be estimated from the data, the GLS methodology becomes equivalent to ML. The estimated variance-covariance matrix of the fixed-effects estimators is $(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$, where the unknown variance components are replaced by their ML estimates. In this case, the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is the inverse of the information matrix of $\boldsymbol{\beta}$, denoted $I(\boldsymbol{\beta})$ (see SCM, p.252-254). Since observations are independent at different treatment combinations, $I(\boldsymbol{\beta}) = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X} = \sum_{t=1}^{r} I_t(\boldsymbol{\beta})$, where $I_t(\boldsymbol{\beta}) = \boldsymbol{X}_t'\boldsymbol{V}_t^{-1}\boldsymbol{X}_t$. It follows that $I(\boldsymbol{\beta}) = \sum_{t=1}^{r}(\boldsymbol{x}_{Dt} \otimes \mathbf{1}_n')(1 \otimes \boldsymbol{V}_t^{-1})(\boldsymbol{x}_{Dt}' \otimes \mathbf{1}_n) = \sum_{t=1}^{r}(\boldsymbol{x}_{Dt}\boldsymbol{x}_{Dt}' \otimes \mathbf{1}_n'\boldsymbol{V}_t^{-1}\mathbf{1}_n)$ and, since $\mathbf{1}_n'\boldsymbol{V}_t^{-1}\mathbf{1}_n$ is scalar,

$$I(\boldsymbol{\beta}) = \sum_{t=1}^{r} \mathbf{1}_n'\boldsymbol{V}_t^{-1}\mathbf{1}_n \boldsymbol{x}_{Dt}\boldsymbol{x}_{Dt}'. \tag{3}$$

Denoting the vector of ML variance estimators by $\hat{\boldsymbol{\sigma}}^2$, the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\sigma}}^2$ for an assembled design is the inverse of the information matrix and is given by

$$\mathrm{Var}\,(\hat{\boldsymbol{\sigma}}^2) = \mathrm{Var}\begin{pmatrix}\hat{\sigma}_1^2\\ \hat{\sigma}_2^2\end{pmatrix} \approx 2\begin{bmatrix}\mathrm{tr}(\boldsymbol{V}^{-1}\boldsymbol{Z}_1\boldsymbol{Z}_1'\boldsymbol{V}^{-1}\boldsymbol{Z}_1\boldsymbol{Z}_1') & \mathrm{tr}(\boldsymbol{V}^{-1}\boldsymbol{Z}_1\boldsymbol{Z}_1'\boldsymbol{V}^{-1}\boldsymbol{Z}_2\boldsymbol{Z}_2')\\ \mathrm{tr}(\boldsymbol{V}^{-1}\boldsymbol{Z}_2\boldsymbol{Z}_2'\boldsymbol{V}^{-1}\boldsymbol{Z}_1\boldsymbol{Z}_1') & \mathrm{tr}(\boldsymbol{V}^{-1}\boldsymbol{Z}_2\boldsymbol{Z}_2'\boldsymbol{V}^{-1}\boldsymbol{Z}_2\boldsymbol{Z}_2')\end{bmatrix}^{-1},$$

(see SCM, p.253), where tr() indicates the trace function of a matrix. The information matrix of the two variance components for treatment combination t

is

$$I_t(\boldsymbol{\sigma}^2) = \frac{1}{2}\begin{bmatrix} \mathrm{tr}(\boldsymbol{V}_t^{-1}\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}') & \mathrm{tr}(\boldsymbol{V}_t^{-1}\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_{2t}\boldsymbol{Z}_{2t}') \\ \mathrm{tr}(\boldsymbol{V}_t^{-1}\boldsymbol{Z}_{2t}\boldsymbol{Z}_{2t}'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}') & \mathrm{tr}(\boldsymbol{V}_t^{-1}\boldsymbol{Z}_{2t}\boldsymbol{Z}_{2t}'\boldsymbol{V}_t^{-1}\boldsymbol{Z}_{2t}\boldsymbol{Z}_{2t}') \end{bmatrix}.$$

The information matrix of variance-components estimators is $I(\boldsymbol{\sigma}^2) = \sum_{t=1}^{r} I_t(\boldsymbol{\sigma}^2)$.

The information matrix of $[\boldsymbol{\beta}\ \boldsymbol{\sigma}^2]'$ is block diagonal (see SCM, p.239) and thus the estimators for the fixed effects and the variance components are asymptotically uncorrelated.

4. Simplest Assembled Design

The simplest assembled design with $r = 1$ design point and $s = 1$ structure is an HND. Hence, it is fundamental to study a single HND or the jth structure. For notational simplicity, we use B instead of B_j to represent the number of batches in a structure for the case of $s = 1$. Using the notation from Section 2, an HND with B batches is represented by $\boldsymbol{m} = (m_1, \ldots, m_B)$, where m_i is the number of samples in batch i, $i = 1, \ldots, B$ and $m_i \geq m_{i+1}$. Define M_B as the set of all feasible and non-trivial HNDs:

$$M_B = \{\langle m_1, \ldots, m_B\rangle | m_i \in \mathcal{I}^+;\ m_i \geq m_{i+1}\ i = 1, \ldots, B;\ B < n-1\}, \tag{4}$$

where $\mathcal{I}^+$ denotes the set of positive integers (i.e., at least one sample is taken per batch). Note that, by definition, $\sum_{i=1}^{B} m_i = n$. The HNDs where $B = n$ or $B = n - 1$ are trivial, since there is only one unique HND for each of these cases and thus they must be optimal.

To develop the information matrix for fixed effects at a single design point, the expressions in terms of the m_i's for a treatment combination t are $\boldsymbol{V}_t = \sigma_1^2\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}' + \sigma_2^2\boldsymbol{Z}_{2t}\boldsymbol{Z}_{2t}' = \bigoplus_{i=1}^{B}[\sigma_1^2\boldsymbol{J}_{m_i} + \sigma_2^2\boldsymbol{I}_{m_i}]$, where $\boldsymbol{J}_n$ is an $n \times n$ square matrix with all elements equal to one. Note that $\boldsymbol{Z}_{2t}\boldsymbol{Z}_{2t}' = \boldsymbol{I}_n$. Also, $\boldsymbol{V}_t^{-1} = \bigoplus_{i=1}^{B}[-\sigma_1^2(\sigma_2^2(\sigma_2^2+\sigma_1^2 m_i))^{-1}\boldsymbol{J}_{m_i} + \sigma_2^{-2}\boldsymbol{I}_{m_i}]$ and, from (3), it can be shown that the information for the fixed-effects estimator of a structure at a design point given n and B is

$$I_t(\beta) = \boldsymbol{X}_t'\boldsymbol{V}_t^{-1}\boldsymbol{X}_t = \boldsymbol{1}_n'\boldsymbol{V}_t^{-1}\boldsymbol{1}_n\boldsymbol{x}_{Dt}\boldsymbol{x}_{Dt}' = \sum_{i=1}^{B}\left(\frac{-\sigma_1^2 m_i^2}{\sigma_2^2(\sigma_2^2+\sigma_1^2 m_i)} + \frac{m_i}{\sigma_2^2}\right) = \sum_{i=1}^{B}\frac{m_i}{\sigma_2^2+\sigma_1^2 m_i}. \tag{5}$$

Note that for a single design point ($r = 1$) $\boldsymbol{x}_{Dt} = 1$ and, thus, β and $I_t(\beta)$ are scalars.

The information matrix for the variance-components estimators from ML is

$$I_t(\boldsymbol{\sigma}^2) = \frac{1}{2}\begin{bmatrix} \mathrm{tr}((\boldsymbol{V}_t^{-1}\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}')^2) & \mathrm{tr}((\boldsymbol{V}_t^{-1})^2\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}') \\ \mathrm{tr}((\boldsymbol{V}_t^{-1})^2\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}') & \mathrm{tr}((\boldsymbol{V}_t^{-1})^2) \end{bmatrix}$$

$$= \frac{1}{2\sigma_2^4}\begin{bmatrix} \sum_{i=1}^{B}\frac{m_i^2}{(1+\tau m_i)^2} & \sum_{i=1}^{B}\frac{m_i}{(1+\tau m_i)^2} \\ \sum_{i=1}^{B}\frac{m_i}{(1+\tau m_i)^2} & n-B+\sum_{i=1}^{B}\frac{1}{(1+\tau m_i)^2} \end{bmatrix}, \qquad (6)$$

where τ is defined as the variance ratio σ_1^2/σ_2^2. The expressions in terms of the m_i's in (6) are derived in Avilés (2001).

5. Optimality

In this section, designs (with equal values of r, n, B_T and s) are compared in terms of their ability to accurately estimate the fixed effects and the two variance components. The goal is to minimize the variance of the estimates of the fixed effects and variance components. The D-optimality criterion (see Myers and Montgomery (1995, p.364) and Pukelsheim (1993, p.136)) has been chosen. A design is D-optimal if it minimizes the determinant of the variance-covariance matrix of the parameter estimates. Because no closed form expressions are available for the variance-covariance matrix of the maximum likelihood estimates in a linear mixed-effects model, we investigate (approximate) D-optimal assembled designs using the asymptotic variance-covariance matrices. Equivalently, we seek the assembled design that maximizes the determinant of the information matrix of the fixed effects and the variance components. Because the information matrix is block diagonal, its determinant is the product of the determinant of the fixed-effects information matrix and the determinant of the variance-components information matrix. Thus, if the same design maximizes the individual determinants, that design is D-optimal overall.

5.1. Fixed-effects optimality

For a single HND, the D-optimality criterion for the fixed effects is the determinant of the matrix defined in (5). Since (5) is a scalar, it is the inverse of the D-optimality criterion. The D-optimal HND for fixed effects can be found for any choice of n and B by solving

Problem I. $\max\limits_{\boldsymbol{m}\in M_B} \sum_{i=1}^{B}\frac{m_i}{\sigma_2^2+\sigma_1^2 m_i}$ subject to $\sum_{i=1}^{B} m_i = n.$

Problem I is non-linear and has implicit integer constraints. By the definition of M_B in (4), $m_i \geq m_{i+1}$, m_1 and m_B are respectively the maximum and minimum number of samples per batch in $\boldsymbol{m}$. Theorem 1 shows that, for a given n and B, the unique solution for Problem 1 is the one and only design such that $m_1 - m_B \leq 1$. This design is shown in (7). This is the most balanced HND confirming that more balanced designs are generally better. Note that the results in this section do not depend on any assumption about the size of varainace components.

Theorem 1. *Let $\boldsymbol{m} \in M_B$ such that $\sum_{i=1}^{B} m_i = n$. The unique solution to Problem* I *is*

$$\boldsymbol{m}^* = \left\langle \underbrace{\lceil n/B \rceil, \ldots, \lceil n/B \rceil}_{\chi = n - \lfloor n/B \rfloor B}, \underbrace{\lfloor n/B \rfloor, \ldots, \lfloor n/B \rfloor}_{\phi = B - \chi} \right\rangle, \tag{7}$$

where $\lfloor x \rfloor$ gives the greatest integer less than or equal to x and $\lceil x \rceil$ gives the smallest integer greater than or equal to x. The χ and ϕ are the number of ceilings and floors in (7)*, respectively.*

Proof. See Appendix 1.

Corollary 1.1. *If $\boldsymbol{V}^*_{n,B}$ is the variance-covariance matrix for the observations from the D-optimal HND represented by $\boldsymbol{m}^*$ in* (7)*, then*

$$I_\beta(\boldsymbol{m}^*) = \mathbf{1}'(\boldsymbol{V}^*_{n,B})^{-1}\mathbf{1} \geq \mathbf{1}'(\boldsymbol{V}_{n,B})^{-1}\mathbf{1}, \tag{8}$$

where $I_\beta(\boldsymbol{m}^*)$ is the fixed-effects information matrix for $\boldsymbol{m}^*$ and $\boldsymbol{V}_{n,B}$ is the variance-covariance matrix for any $\boldsymbol{m} \in M_B$ with n observations.

Proof. The information of the D-optimal design, $\boldsymbol{m}^*$, must be greater than or equal to the information from any other design.

The following theorem uses the results of Theorem 1 and Corollary 1.1 to show that given the number of samples/batch and the number of batches for each structure, the D-optimal AD is the design that places the D-optimal HND at each design point.

Theorem 2. *Given n and B_j, $j = 1, \ldots, s$, such that $B_T = \sum_{j=1}^{s} r_j B_j$, the D-optimal AD for fixed effects is* (*in design order*)*:*

$$\sum_{j=1}^{s} \left\langle \underbrace{\lceil n/B_j \rceil, \ldots, \lceil n/B_j \rceil}_{\chi_j = n - \lfloor n/B_j \rfloor B_j}, \underbrace{\lfloor n/B_j \rfloor, \ldots, \lfloor n/B_j \rfloor}_{\phi_j = B_j - \chi_j} \right\rangle @\{R_{j-1}+1, \ldots, R_j\}, \tag{9}$$

where $R_j = \sum_{h=1}^{j} r_h$ *and* $R_0 = 0$. *The number of ceilings and the number of floors in the optimal structure, given* n *and* B_j, *respectively are* χ_j *and* ϕ_j.

Proof. See Appendix 2.

5.2. Variance-components optimality

For a single HND, the D-optimality criterion for estimation of the two variance components is the determinant of the information matrix defined in (6). The D-optimal HND can be found for any choice of n and B by solving Problem II for two variance components.

Problem II. $\max_{\boldsymbol{m}\in M_B} \det\left(\frac{1}{2}\begin{bmatrix} \mathrm{tr}((\boldsymbol{V}_t^{-1}\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}')^2) & \mathrm{tr}((\boldsymbol{V}_t^{-1})^2\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}') \\ \mathrm{tr}((\boldsymbol{V}_t^{-1})^2\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}') & \mathrm{tr}((\boldsymbol{V}_t^{-1})^2) \end{bmatrix}\right)$ subject to $\sum_{i=1}^{B} m_i = n$.

As with Problem I, Problem II is non-linear and has integer constraints. However, unlike Problem I, the information for the variance components is a matrix, rather than a scalar. Also unlike Problem I, the solution of Problem II does depend on the relative size of the variances. Recall that in most engineering applications, the batch-to-batch variance is at least as large as the sample-to-sample variance (i.e., $\tau \geq 1$). Theorem 3 indicates that, when the variance ratio is at least one, the most balanced HND is D-optimal for two variance components except in cases where a single sample is taken from a large number of batches in that HND. In other words, given B batches and n samples, the HND in (7) is D-optimal unless $\lfloor n/B \rfloor = 1$ and ϕ is large.

Theorem 3. *Let* $\boldsymbol{m} \in M_B$ *such that* $\sum_{i=1}^{B} m_i = n$. *If* $\tau \geq 1$, *then the unique solution to Problem* II *is shown in* (7) *as long as* $M_1 < 1 + \sum_{i:m_i>1} m_i(5m_i - 7)(m_i+1)^{-1}$, *where* M_1 *is the number of batches in* (7) *with a single sample.*

Proof. See Appendix 4.

Note that when at least two samples are taken from each batch then $M_1 = 0$ and the condition $M_1 < 1 + \sum_{i:m_i>1} m_i(5m_i - 7)(m_i+1)^{-1}$ is automatically satisfied. If there are batches with only a single sample, then the condition is still satisfied for most practical situations. When there are batches with single samples in (7), then all other batches have two samples. If we denote the number of batches in (7) with two samples as M_2, then $M_1+M_2 = B$. It can be shown that as long as $M_1 < 1+2M_2$, the condition $M_1 < 1+\sum_{i:m_i>1} m_i(5m_i-7)(m_i+1)^{-1}$ is satisfied. We have also verified empirically that for practical cases (up to 100 batches) when $M_1 < 1 + 6M_2$, then (7) is the solution to Problem II. Since designs with large numbers of batches having only a single sample are not likely to be used in practice, the most balanced HND, as shown in (7), is the D-optimal design in almost all applications.

We now look at ADs with $r > 1$ and again find that in most practical cases, placing the most balanced HND at each design point is the D-optimal design. Consider an AD with r design points and s structures. Assume that constraints are imposed on the number of batches at each design point. That is, r_j design points will have B_j batches, $j = 1, \ldots, s$. Let m_{ij} be the number of samples in batch i, $i = 1, \ldots, B_j$ of structure j. For an assembled design, a structure j with B_j batches can be represented by $(m_{1j}, \ldots, m_{Hj}, \ldots, m_{Lj}, \ldots, m_{B_j j})$, where $m_{ij} \geq m_{(i+1)j}$.

Theorem 4 indicates that, when the variance ratio is at least one, the most balanced HND at each design point, as shown in (9), is D-optimal for variance components except in cases where a single sample is taken from a large number of batches in the AD. Thus, as long as there are no structures that have more than twice as many batches with a single sample as batches with two samples, then (9) can be shown to be the D-optimal design.

Theorem 4. *Suppose n and B_j, $j = 1, \ldots, s$, satisfy $B_T = \sum_{j=1}^{s} r_j B_j$. If $\tau \geq 1$, then the D-optimal AD for variance components is shown in* (9) *as long as $M_{1j} < \sum_{i:m_{ij}>1} m_{ij}(5m_{ij} - 7)(m_{ij} + 1)^{-1}$ for all j, where M_{1j} is the number of batches with a single sample for the jth structure in an AD.*

Proof. See Appendix 5.

Taken together, the results of Theorems 2 and 4 show that, in most practical situations, the most balanced AD is the D-optimal design for ML estimation of fixed effects and two variance components. Since the most balanced design is always the D-optimal design for fixed effects, it could be argued that even in those rare situations when the variance ratio is less than one or there are a large number of batches with a single sample, the balanced design will still perform reasonably well even if it is not optimal for variance components.

6. Examples

The following examples show how to obtain the D-optimal ADs for the requirements of an experiment.

Example 1. In pharmaceutical production, an experiment could be designed to estimate the effects of certain machine settings (fixed effects) on the hardness of a tablet and the variability of hardness from batch-to-batch and from sample-to-sample (variance components). It is likely that the batch-to-batch variance is at least as large as the sample-to-sample variance ($\tau \geq 1$) since there are likely to be larger differences between batches than within batches of tablets. The production schedule only allows for producing a total of 24 experimental batches ($B_T = 24$) and the engineers want to take six samples ($n = 6$) from each of eight treatment combinations ($r = 8$). For this experiment, only one structure is needed ($s = 1$), since $B_1 = B_T/r_1 = 24/8 = 3$ batches can be produced ($B_1 = B = 3$) in each of the eight treatment combinations ($r_1 = r = 8$). For the given parameters, the assembled design in (9) is $(2, 2, 2)@\{1, 2, 3, 4, 5, 6, 7, 8\}$, since $\chi_1 = n - \lfloor n/B_1 \rfloor \times B_1 = 6 - \lfloor 6/3 \rfloor \times 3 = 6 - 2 \times 3 = 0$ and $\phi_1 = B_1 - \chi_1 = 3 - 0 = 3$. Since $\tau \geq 1$ and there are no batches with single samples, then the D-optimal assembled design is $(2, 2, 2)@\{1, 2, 3, 4, 5, 6, 7, 8\}$.

Example 2. For the pharmaceutical example, let there now be 28 batches ($B_T = 28$) and 10 samples ($n = 10$) at each of 8 treatment combinations ($r = 8$). Since $B_1 = B_T/r_1 = 28/8 = 3.5$ batches is not feasible, more than one structure is needed ($s \neq 1$). The engineers decide to produce four batches at four of the design points ($B_1 = 4$ and $r_1 = 4$) and three batches at the other design points ($B_2 = 3$ and $r_2 = 4$), and thus $d_1 = \sum_{j=1}^{s} r_j(B_j - 1) = B_T - r = 20$ and $d_2 = n \times r - d_1 - r = n \times r - B_T = 52$. Hence, $\chi_1 = 2$, $\phi_1 = 2$ and, since $n/B_1 = 10/4 = 2.5$, the first structure is $(3, 3, 2, 2)$. Since $\chi_2 = 1$, $\phi_2 = 2$, and $n/B_2 = 10/3 = 3.33$, the second structure is $(4, 3, 3)$. The assembled design in (9) is $(3, 3, 2, 2)@\{1, 2, 3, 4\} + (4, 3, 3)@\{5, 6, 7, 8\}$, where the design points are in design order. By Theorem 2 and Theorem 4, since $\tau \geq 1$, $M_{11} = 0 < \sum_{i:m_{i1}>1} m_{i1}(5m_{i1} - 7)(m_{i1} + 1)^{-1} = 16$, and $M_{12} = 0 < \sum_{i:m_{i2}>1} m_{i2}(5m_{i2} - 7)(m_{i2} + 1)^{-1} = 22.4$, the design is D-optimal for ML estimation of both fixed effects and variance components. In order to demonstrate the benefit of using the D-optimal design, a comparison is made with the most unbalanced design with the same n, B_1, and B_2, namely $(7, 1, 1, 1)@\{1, 2, 3, 4\} + (8, 1, 1)@\{5, 6, 7, 8\}$. This unbalanced design is slightly better at estimating the sample-to-sample variance, but significantly worse in estimating the fixed effects and the batch-to-batch variance. Table 1 shows the standard error of the parameter estimates at various values of the batch-to-batch variance assuming that the sample-to-sample variance is 1.

Table 1. Standard error of estimators for most balanced and least balanced assembled designs in Example 2. Given $r = 8$, $n = 10$, $B_1 = 4$, and $B_2 = 3$, the most balanced design is $(3, 3, 2, 2)@\{1, 2, 3, 4\} + (4, 3, 3)@\{5, 6, 7, 8\}$ and the least balanced design is $(7, 1, 1, 1)@\{1, 2, 3, 4\} + (8, 1, 1)@\{5, 6, 7, 8\}$.

	Standard Error of Fixed Effects			Standard Error of Batch Variance			Standard Error of Sample Variance		
Batch Variance	Balanced Design	Unbalanced Design	% Difference	Balanced Design	Unbalanced Design	% Difference	Balanced Design	Unbalance Design	% Difference
1	0.2219	0.2437	9.8%	0.3705	0.4338	17.1%	0.1959	0.1923	-1.8%
3	0.3496	0.3671	5.0%	0.9024	0.9926	10.0%	0.1961	0.1954	-0.3%
5	0.4417	0.4564	3.3%	1.4362	1.5318	6.7%	0.1961	0.1958	-0.1%

7. Conclusions and Discussion

We have introduced assembled designs, a general class of experimental designs for mixed-effects models with random nested factors. The assembled designs are crossed factor designs with a hierarchical nested design (HND) placed at each treatment combination of the crossed factor design. We have shown that when there are two levels of nesting, distributing the samples as uniformly as possible among the batches of the design results in D-optimal designs for maximum likelihood estimation, except in a few relatively uncommon situations. Although we believe that this principle extends to REML estimation and to more than two levels of nesting, proofs of these extensions are left to future research.

Acknowledgements

This material is based upon work supported under a National Science Foundation Graduate Fellowship. The authors would like to thank Jeff Wu, Randy Tobias, Derek Bingham, Steve Gilmour, and Dennis Lin for helpful discussions during this research, and Hyunsoo Suh for helping to prepare the examples.

Appendix 1: Proof of Theorem 1

Define $I_\beta(\boldsymbol{m})$ as the fixed-effects information matrix for a feasible HND, $\boldsymbol{m} \in M_B$. By Equation (5), $I_\beta(\boldsymbol{m}) = \sum_{i=1}^{B} m_i(\sigma_2^2 + \sigma_1^2 m_i)^{-1}$. Let $\boldsymbol{m}_{\mathrm{I}} = \langle m_1, m_2, \ldots, m_{B-1}, m_B\rangle$ and $\boldsymbol{m}_{\mathrm{II}} = order[\langle m_1 - 1, m_2, \ldots, m_{B-1}, m_B + 1\rangle]$, where $order[\,]$ rearranges $\boldsymbol{m}_{\mathrm{II}}$ into decreasing order and $\boldsymbol{m}_{\mathrm{I}}, \boldsymbol{m}_{\mathrm{II}} \in M_B$. It can be shown that $sign(I_\beta(\boldsymbol{m}_{\mathrm{II}}) - I_\beta(\boldsymbol{m}_{\mathrm{I}})) = sign(m_1 - m_B - 1)$, where $sign[x]$ is -1, 0, or 1 depending on whether x is negative, zero, or positive. Thus, $m_1 - m_B = 0$ implies $I_\beta(\boldsymbol{m}_{\mathrm{II}}) < I_\beta(\boldsymbol{m}_{\mathrm{I}})$, $m_1 - m_B = 1$ implies $I_\beta(\boldsymbol{m}_{\mathrm{II}}) = I_\beta(\boldsymbol{m}_{\mathrm{I}})$, and $m_1 - m_B \geq 2$ implies $I_\beta(\boldsymbol{m}_{\mathrm{II}}) > I_\beta(\boldsymbol{m}_{\mathrm{I}})$. Since any $\boldsymbol{m} \in M_B$ such that $m_1 - m_B \geq 2$ can be improved upon, it cannot be a solution to Problem I. Given n and B, there is only one $\boldsymbol{m}^* \in M_B$ such that $m_1 - m_B \leq 1$ and $\sum_{i=1}^{B} m_i = n$. Therefore, $\boldsymbol{m}^* \in M_B$, shown in (7), is the unique solution to Problem I.

Appendix 2: Proof of Theorem 2

$\boldsymbol{X}_D$ can be partitioned into $\boldsymbol{X}_D = [\boldsymbol{x}_{D1}\ \boldsymbol{x}_{D2}\ \cdots\ \boldsymbol{x}_{Dr}]'$. Define $\boldsymbol{A}$ as the information matrix for $\boldsymbol{\beta}$. From (3), $\boldsymbol{A} = \sum_{i=1}^{r} \mathbf{1}_n' \boldsymbol{V}_t^{-1} \mathbf{1}_n \boldsymbol{x}_{Dt}\boldsymbol{x}_{Dt}'$. If $\boldsymbol{w}$ is an r-dimensional multivariate normal variate with mean zero and variance-covariance matrix $\boldsymbol{A}^{-1}$, we have $(\det(\boldsymbol{A}))^{1/2} \int e^{-\boldsymbol{w}'\boldsymbol{A}\boldsymbol{w}/2} d\boldsymbol{w} = (2\pi)^{r/2}$. Consequently $\det(\boldsymbol{A}) \geq \det(\boldsymbol{B})$ if $\boldsymbol{w}'\boldsymbol{A}\boldsymbol{w} \geq \boldsymbol{w}'\boldsymbol{B}\boldsymbol{w}$ for all $\boldsymbol{w}$ and some alternative information matrix $\boldsymbol{B}$. Let $\lambda_{n,B}^* = \mathbf{1}'(\boldsymbol{V}_{n,B}^*)^{-1}\mathbf{1}$, as defined in (8). Since $\boldsymbol{w}'\boldsymbol{x}_{Dt}\boldsymbol{x}_{Dt}'\boldsymbol{w} = \|\boldsymbol{x}_{Dt}\boldsymbol{w}\|^2 \geq 0$ and $\boldsymbol{A}_{OPT} = \sum_{t=1}^{r} \lambda_{n,B_t}^* \boldsymbol{x}_{Dt}\boldsymbol{x}_{Dt}'$, from Corollary 1.1, $\boldsymbol{w}'\boldsymbol{A}\boldsymbol{w} = \sum_{t=1}^{r} \mathbf{1}_n' \boldsymbol{V}_t^{-1} \mathbf{1}_n \boldsymbol{w}'\boldsymbol{x}_{Dt}\boldsymbol{x}_{Dt}'\boldsymbol{w} \leq \sum_{t=1}^{r} \lambda_{n,B_t}^* \boldsymbol{w}'\boldsymbol{x}_{Dt}\boldsymbol{x}_{Dt}'\boldsymbol{w} = \boldsymbol{w}'\boldsymbol{A}_{OPT}\boldsymbol{w}$.

So $\det(\boldsymbol{A}_{OPT}) \geq \det(\boldsymbol{A})$ for all $\boldsymbol{A}$. Therefore, given n samples and B_j ($j = 1, \ldots, s$) batches, placing the optimal structure (HND) at each design point is the optimal AD for fixed effects.

Appendix 3: Lemma 1

Given B and n, consider two HNDs ($\boldsymbol{m} \in M_B$ such that $\sum_{i=1}^{B} m_i = n$ and $\boldsymbol{m}_U \in M_B$) that are identical except for two batches as in the following definition: $\boldsymbol{m} = \langle m_1, m_2, \ldots, m_H, \ldots, m_L, \ldots, m_{B-1}, m_B\rangle$; $\boldsymbol{m}_U = order[\langle m_1, m_2, \ldots, m_H + 1, \ldots, m_L - 1, \ldots, m_{B-1}, m_B\rangle]$, where $m_H \geq m_L$. Let $I_\sigma(\boldsymbol{m})$ and $I_\sigma(\boldsymbol{m}_U)$ represent the information matrix of $\boldsymbol{m}$ and $\boldsymbol{m}_U$, respectively. Then $\det(I_\sigma(\boldsymbol{m}_U)) < \det(I_\sigma(\boldsymbol{m}))$, provided $M_1 < 1 + \sum_{i:m_i>1} m_i(5m_i - 7)/(m_i + 1)$.

Proof. A more detailed proof is available in Avilés (2001). Here a sketch is provided. The variance-components information matrix, denoted by $I_\sigma(\boldsymbol{m})$, can be written as

$$I_\sigma(\boldsymbol{m}) = \frac{1}{2}\begin{bmatrix} \mathrm{tr}((\boldsymbol{V}_t^{-1}\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}')^2) & \mathrm{tr}((\boldsymbol{V}_t^{-1})^2\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}') \\ \mathrm{tr}((\boldsymbol{V}_t^{-1})^2\boldsymbol{Z}_{1t}\boldsymbol{Z}_{1t}') & \mathrm{tr}((\boldsymbol{V}_t^{-1})^2) \end{bmatrix} = \begin{bmatrix} D & F \\ F & E \end{bmatrix},$$

where $D = (2\sigma_2^4)^{-1}\sum_{i=1}^{B} d(m_i,\tau)$, $E = (2\sigma_2^4)^{-1}\sum_{i=1}^{B} e(m_i,\tau)$, $F = (2\sigma_2^4)^{-1}\sum_{i=1}^{B} f(m_i,\tau)$, $d(m_i,\tau) = m_i^2(1+\tau m_i)^{-2}$, $e(m_i,\tau) = m_i((1+(m_i-1)\tau)^2+(m_i-1)\tau^2)(1+\tau m_i)^{-2}$, and $f(m_i,\tau) = m_i(1+\tau m_i)^{-2}$.

To show that $\det(I_\sigma(\boldsymbol{m}_U)) < \det(I_\sigma(\boldsymbol{m}))$, it suffices to show that for $0 < \delta \leq 1$,

$$\frac{\partial}{\partial\delta}[\det(I_\sigma(\boldsymbol{h}))] = \frac{\partial}{\partial\delta}\left[\det\begin{pmatrix} D_\delta & F_\delta \\ F_\delta & E_\delta \end{pmatrix}\right] < 0, \tag{A3.1}$$

where

$$\boldsymbol{h} = \langle h_1, h_2, \ldots, h_H, \ldots, h_L, \ldots, h_{B-1}, h_B\rangle \quad \text{and} \quad \begin{cases} h_i = m_i + \delta & i = H \\ h_i = m_i - \delta & i = L \\ h_i = m_i & \text{else} \end{cases} \tag{A3.2}$$

and D_δ, E_δ and F_δ are the elements of the information matrix of $\boldsymbol{h}$. If we let $\alpha_1 = E_\delta + \tau F_\delta$ and $\alpha_2 = F_\delta + \tau D_\delta$, then it can be shown that

$$\frac{\partial}{\partial\delta}[\det(I_\sigma(\boldsymbol{h}))] = 2\left[\frac{\alpha_1 h_H - \alpha_2}{(1+h_H\tau)^3} - \frac{\alpha_1 h_L - \alpha_2}{(1+h_L\tau)^3}\right] = 2[g(h_H) - g(h_L)], \tag{A3.3}$$

where $g(x) = (\alpha_1 x - \alpha_2)(1+x\tau)^{-3}$. Hence if $g(x)$ is a decreasing function in x, then (A3.1) is satisfied. Since $\frac{\partial}{\partial x}[g(x)] = (\alpha_1 + 3\tau\alpha_2 - 2\tau\alpha_1 x)/(1+\tau x)^4$, then as long as $x > (2\tau)^{-1} + 3\alpha_2(2\alpha_1)^{-1}$, it can be shown that $\frac{\partial}{\partial x}[g(x)] < 0$ and, thus, $g(x)$ is decreasing in x. Notice that m_L must be at least 2, since

$m_L = 1$ implies that $\boldsymbol{m}_U$ would have only $B-1$ batches, which is infeasible because we are comparing HNDs with B batches. For $\boldsymbol{h}$, $\alpha_1 > \alpha_2$ and it follows that for $\tau \geq 1$, $(2\tau)^{-1} + 3\alpha_2(2\alpha_1)^{-1} < 2$, implying that $g(x)$ is decreasing for $x \geq 2$. When $m_L \geq 3$, then $h_H > h_L \geq 2$ and it follows that (A3.3) is negative and (A3.1) is satisfied, provided $\tau \geq 1$. Only the case of $m_L = 2$ remains. In Avilés (2001), it is shown that if $\tau \geq 1$ and $m_L = 2$, (A3.1) is satisfied if $M_1 < 1 + \sum_{i:m_i>1} m_i(5m_i - 7)/(m_i + 1)$. This completes the sketch of proof.

Appendix 4: Proof of Theorem 3

Observe that any design in M_B such that $m_1 - m_B \geq 2$ can be represented (with a possible difference in order) by $\boldsymbol{h}$ as defined in (A3.2) with $H < L$ and $\delta = 1$. Lemma 1 (see Appendix 3) shows that such a design can be improved upon in terms of D-optimality for ML estimation of two variance components by choosing $\delta = 0$, provided $M_1 < 1 + \sum_{i:m_i>1} m_i(5m_i - 7)/(m_i + 1)$ and $\tau \geq 1$. Under these conditions, the D-optimal design then must have $m_1 - m_B \leq 1$. For a given value of n and B, the design in (7) is the unique design such that $m_1 - m_B \leq 1$. Thus, it must be the D-optimal design under the conditions of the theorem.

Appendix 5: Proof of Theorem 4

As in Appendix 3, we sketch a proof and refer to Avilés (2001) for the complete proof. We show that $\det[I_\sigma(\boldsymbol{m}_1, \ldots, \boldsymbol{m}_u, \ldots, \boldsymbol{m}_r)]$ is a decreasing function of δ, where $\boldsymbol{m}_u = \boldsymbol{h}$ as defined in (A3.2) with $B = B_u$ and $m_i = m_{iu}$. Denote the information of variance components associated with the tth design point as $I_\sigma(\boldsymbol{m}_t)$. Because the design points are independent,

$$I_\sigma(\boldsymbol{m}_1, \ldots, \boldsymbol{h}, \ldots, \boldsymbol{m}_r) = \sum_{t=1}^{r} I_\sigma(\boldsymbol{m}_t) = \sum_{t=1}^{r} \begin{bmatrix} D_t & F_t \\ F_t & E_t \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{r} D_t & \sum_{t=1}^{r} F_t \\ \sum_{t=1}^{r} F_t & \sum_{t=1}^{r} E_t \end{bmatrix} \equiv \begin{bmatrix} \overline{D} & \overline{F} \\ \overline{F} & \overline{E} \end{bmatrix},$$

where D_t, E_t and F_t are the elements of the information matrix of $\boldsymbol{m}_t$. Note that when $t = u$, $I_\sigma(\boldsymbol{m}_t) = I_\sigma(\boldsymbol{h})$. Now, following similar steps as in the proof of Lemma 1,

$$\frac{\partial}{\partial \delta}[\det(I_\sigma(\boldsymbol{m}_1, \ldots, \boldsymbol{h}, \ldots, \boldsymbol{m}_r))] = 2[\overline{g}(h_H) - \overline{g}(h_L)], \qquad \text{(A5.1)}$$

where $\overline{\alpha}_1 = \overline{E} + \tau\overline{F}$, $\overline{\alpha}_2 = \overline{F} + \tau\overline{D}$, and $\overline{g}(x) = (\overline{\alpha}_1 x - \overline{\alpha}_2)(1 + x\tau)^{-3}$. Hence, if $\overline{g}(x)$ is a decreasing function in x, then $\frac{\partial}{\partial\delta}[\det(\inf_\sigma(\boldsymbol{m}_1, \ldots, \boldsymbol{h}, \ldots, \boldsymbol{m}_r))] < 0$. This last is true if and only if $x > (2\tau)^{-1} + 3\overline{\alpha}_2(2\overline{\alpha}_1)^{-1}$. Since $\overline{\alpha}_1 = \sum_{t=1}^{r} \alpha_{1t}$ and $\overline{\alpha}_2 = \sum_{t=1}^{r} \alpha_{2t}$, where $\alpha_{1t} = E_t + \tau F_t$ and $\alpha_{2t} = F_t + \tau D_t$, it follows from the proof of Lemma 1 that $\overline{\alpha}_1 - \overline{\alpha}_2 = \sum_{t=1}^{r}(\alpha_{1t} - \alpha_{2t}) \geq \alpha_{1u} - \alpha_{2u} > 0$, provided

$h_L \geq 2$. When $m_{Lu} \geq 3$, then $h_H > h_L \geq 2$ and it follows that (A5.1) is negative, provided $\tau \geq 1$. For the case of $\tau \geq 1$ and $m_{Lu} = 2$, Avilés (2001) shows that (A5.1) is satisfied if $M_{1j} < \sum_{i:m_{ij}>1} m_{ij}(5m_{ij} - 7)(m_{ij} + 1)^{-1}$ for all j, $j = 1, \ldots, s$. This concludes the sketch.

References

Ankenman, B. E., Liu, H., Karr, A. F. and Picka, J. D. (2001). A class of experimental designs for estimating a response surface and variance components. *Technometrics* **44**, 45-54.

Avilés, A. I. (2001). Optimal designs for mixed-effects models with random nested factors. Ph.D. Thesis, Northwestern University, Evanston, Illinois.

Bainbridge, T. R. (1965). Staggered nested designs for estimating variance components. *Indust. Quality Control* **22**, 12-20.

Delgado, J. and Iyer, H. (1999). Search for optimal designs in a three stage nested random model. *Statist. Comput.* **9**, 187-193.

Goldsmith, C. H. and Gaylor, D. W. (1970). Three stage nested designs for estimating variance components. *Technometrics* **12**, 487-498.

Jaiswal, S. S., Picka, J. D., Igusa, T., Karr, A. F. Shah, S. P., Ankenman, B. E. and Styer, P. (2000). Statistical studies of the conductivity of concrete using ASTM C1202-94. *Concrete Sci. Engrg.* **2**, 97-105.

Myers, R. H. and Montgomery, D. C. (1995). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments.* Wiley, New York.

Naik, D. N. and Khattree, R. (1998). A computer program to estimate variance components in staggered nested designs. *J. Quality Technology* **30**, 292-297.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-Plus.* Springer-Verlag, New York.

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

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components.* Wiley, New York.

Smith, J. R. and Beverly, J. M. (1981). The use and analysis of staggered nested factorial designs. *J. Quality Technology* **13**, 166-173.

Theil, H. (1971). *Principles of Econometrics.* Wiley, New York.

Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Rd. Evanston, IL 60208, U.S.A.

E-mail: ankenman@northwestern.edu

US Department of Commerce, NIST, 820 W. Diamond Avenue, 820/353, Gaithersburg, MD 20899-8980, U.S.A.

E-mail: aviles@nist.gov

Biostatistics, Novartis Pharmaceuticals, One Health Plaza, Room 419/2115, East Hanover, NJ 07936, U.S.A.

E-mail: jose.pinheiro@pharma.novartis.com

(Received October 2000; accepted September 2002)