Statistica Sinica **5**(1995), 485-497

OPTIMAL REGRESSION DESIGNS UNDER RANDOM BLOCK-EFFECTS MODELS

Ching-Shui Cheng

University of California, Berkeley

Abstract. D-optimal regression designs under random block-effects models are considered. In addition to selecting design points, an experimenter also needs to specify how they are grouped into blocks. We first consider minimum-support designs, which are supported on the minimum number of design points. In this case, it is shown that a D-optimal design can be obtained by combining a D-optimal block design (for treatment comparisons under random block-effects models) with a D-optimal regression design under the usual uncorrelated model. Such a design, however, is not optimal when there are no restrictions on the competing designs. To attack the general problem of constructing optimal designs without restrictions on the competing designs, we sketch an approach based on the approximate theory, and apply it to quadratic regression on $[-1,1]$.

Key words and phrases: Approximate design, balanced incomplete block design, D-optimality, equivalence theorem.

1. Introduction

Optimal regression designs under the usual homoscedastic uncorrelated linear models have been studied extensively. The discovery of the Equivalence Theorem of Kiefer and Wolfowitz (1960) was a major impetus. By now there has been a large and diverse body of work. An excellent account of this rich literature can be found in the recent book by Pukelsheim (1993).

The purpose of this article is to study a variant of the traditional regression design problem. Instead of uncorrelated observations, we shall consider the situation where the observations are taken in *blocks*, wherein it is natural to assume that the observations in the same block are *correlated*. The presence of random block effects complicates the solution and requires special care. Our motivation comes from an optometry experiment, described in Chasalow (1992), for exploring the dependence of a measure of corneal hydration control on the CO_2 level (a continuous-valued treatment) in a gaseous environment applied through a goggle covering a human subject's eyes. A response is measured for each eye, so each human subject's pair of eyes provides a block of two possibly correlated observations. The question is how to allocate the CO_2 levels to estimate the response function efficiently. Khuri (1992) described an experiment for investigating the effects of temperature and curing time on shear strength in the manufacture of an adhesive for bonding galvanized steel bars together. The experiment was run on a random set of days over a 4-month period during which batches of experimental material were drawn randomly from the warehouse supply. In this experiment, the block effect (batches) should be treated as random. While there is a large literature on optimal block designs for treatment comparisons, (see the monograph by Shah and Sinha (1989) for an extensive review), very few results on optimal block designs for regression models, if any, are available. Khuri (1992) discussed the analysis of response surface models with random block effects, but did not consider the design aspect except for orthogonal blocking.

Consider the following general setup. Given a compact design region æ in R^n, let $y_{\boldsymbol{x}}$ be an observation at $\boldsymbol{x} \in$ æ. Assume the usual parametric model

$$E(y_{\boldsymbol{x}}) = \sum_{i=1}^{t} f_i(\boldsymbol{x})\theta_i \quad \text{and} \quad \mathrm{Var}(y_{\boldsymbol{x}}) = \sigma^2, \tag{1.1}$$

where $f_1(\boldsymbol{x}), \ldots, f_t(\boldsymbol{x})$ are real-valued functions defined on æ, and $\theta_1, \ldots, \theta_t$ are unknown parameters. A total of N observations will be taken (hence N points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in$ æ, not necessarily all distinct, are to be selected). Let $\boldsymbol{y} = (y_1, \ldots, y_N)^T$ be the vector of observations at $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$. Then the usual uncorrelated linear model can be expressed as

$$E(\boldsymbol{y}) = \mathbf{A}\boldsymbol{\theta} \tag{1.2}$$

and

$$\mathrm{Cov}(\boldsymbol{y}) = \sigma^2 \mathbf{I}_N, \tag{1.3}$$

where $\mathbf{I}_N$ is the identity matrix of order N, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_t)^T$, and $\mathbf{A}$ is the $N \times t$ matrix with the (i, j)th entry $f_j(\boldsymbol{x}_i)$. For convenience, we denote $(f_1(\boldsymbol{x}), \ldots, f_t(\boldsymbol{x}))^T$ by $\mathbf{f}(\boldsymbol{x})$, and the set of *distinct* points among $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ is called the *support* of the design. A design is said to be D-optimal if it minimizes the determinant of the covariance matrix of the least squares estimators of $\theta_1, \ldots, \theta_t$, or equivalently, maximizes the determinant of the information matrix $\mathbf{A}^T\mathbf{A}$. This problem has been well studied and many results are available. For example, it is clear that in order to estimate all the t parameters, a design must be supported on at least t distinct points, and in many situations, a D-optimal design is supported on exactly t points. One notable example is D-optimal designs for the dth-degree polynomial regression on $[-1, 1]$ with $t = d + 1$, and $f_i(x) = x^{i-1}$, $i = 1, \ldots, d + 1$. In this case, the D-optimal design with $N = t$

consists of the $t = d + 1$ roots of the equation $(1 - x^2)\dot{P}_d(x) = 0$, where $\dot{P}_d(x)$ is the derivative of the dth-degree Legendre polynomial. When $t|N$, a D-optimal design has each of these points replicated the same number of times.

Now, unlike the usual setup, suppose the N observations must be grouped into $b = N/k$ blocks of size k (hence $k|N$). Furthermore, the observations in the same block are correlated. For simplicity, we shall assume that any two observations in the same block have the same *positive* correlation ρ, and any two observations in different blocks are uncorrelated. This covers the model $\boldsymbol{y} = \mathbf{A}\boldsymbol{\theta} + (\mathbf{I}_b \otimes \mathbf{1}_k)\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with additive random block effects, where $\mathbf{l}_k$ is the $k \times 1$ vector of 1's, $\boldsymbol{\beta}$ is a $b \times 1$ vector of random block effects, and $\boldsymbol{\epsilon}$ is a $bk \times 1$ vector of random errors, with $E(\boldsymbol{\beta}) = \mathbf{0}$, $E(\boldsymbol{\epsilon}) = \mathbf{0}$, $\mathrm{Cov}(\boldsymbol{\beta}) = \sigma_b^2\mathbf{I}_b$, $\mathrm{Cov}(\boldsymbol{\epsilon}) = \sigma_e^2\mathbf{I}_{bk}$, and $\mathrm{Cov}(\boldsymbol{\beta}, \boldsymbol{\epsilon}) = \mathbf{0}$. In this case, $\sigma^2 = \sigma_b^2 + \sigma_e^2$, and $\rho = \sigma_b^2/(\sigma_b^2 + \sigma_e^2)$.

As a first step, assume that ρ is known, and use the generalized least squares estimators $\hat{\theta}_1, \ldots, \hat{\theta}_t$. We consider the problem of determining D-optimal designs for estimating the unknown parameters $\theta_1, \ldots, \theta_t$. Suppose the entries of $\boldsymbol{y}$ are ordered so that the $[(i-1)k + j]$th entry is the jth observation in the ith block, $1 \le i \le b$, $1 \le j \le k$; then instead of (1.3), we have

$$\mathrm{Cov}(\boldsymbol{y}) = \sigma^2(\mathbf{I}_b \otimes \mathbf{V}), \text{ with } \mathbf{V} = [(1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k], \tag{1.4}$$

where $\mathbf{J}_k$ is the $k \times k$ matrix with all the entries equal to 1. Again, a design is called D-*optimal* if it minimizes the determinant of the covariance matrix of $\hat{\theta}_1, \ldots, \hat{\theta}_t$, or equivalently, maximizes the determinant of the information matrix $\mathbf{M} \equiv \mathbf{A}^T(\mathbf{I}_b \otimes \mathbf{V})^{-1}\mathbf{A}$.

Here, in addition to selecting $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, an experimenter also needs to specify how they are grouped into blocks of size k. Suppose there are $s\,(\ge t)$ distinct points among $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, say $\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_s^*$. Then there are two phases in a design: selecting s distinct points $\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_s^*$, and then constructing a block design with b blocks of size k for these s "treatments". Notice that the same $\boldsymbol{x}_i^*$ may be assigned to more than one unit in the same block, i.e., the blocks are not necessarily binary.

Intuitively, it seems natural to start with a D-optimal design for the uncorrelated model (1.2) and (1.3), and then allocate the points in this design into the blocks of an "optimal" block design. For instance, balanced incomplete block designs (BIBD) are well known to be optimal for treatment comparisons with respect to many criteria including the D-criterion (Kiefer (1958, 1975)). Therefore if $\{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_t\}$ is a D-optimal design with t observations for the uncorrelated model, and there exists a BIBD with t treatments and b blocks of size k, then one could try the BIBD formed by the t "treatments" $\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_t$. For example, in the quadratic regression on $[-1, 1]$, a D-optimal design for the uncorrelated model is supported on -1, 0 and 1. If $N = 2b$ observations are to be taken in b correlated pairs, where b is a multiple of three, then the above procedure suggests that one may replicate each of the three blocks $(1, -1)$, $(1, 0)$ and $(0, -1)$ the same number of times. How good is this design? Does it have any optimality property?

It turns out that a design constructed in this way is *not* D-optimal. After all, one would expect the optimal designs to depend on ρ. However, it does have some optimality property. If we consider only designs which are supported on exactly t distinct $\boldsymbol{x}$'s, then the above design is the best under the D-criterion for all positive ρ. Such designs with minimum number of distinct $\boldsymbol{x}$'s will be called *minimum-support designs.* For instance, in the quadratic regression on $[-1, 1]$, the design given in the preceding paragraph is at least as good as any b blocks of size two formed by *any three* values x_1, x_2, $x_3 \in [-1, 1]$ under the D-criterion, for all positive ρ. Without such a restriction on the competing designs, a D-optimal design consists of three different kinds of blocks $(1, -1)$, $(-1, -a)$ and $(1, a)$, where $a\,(< 0)$ and the relative frequencies of these blocks depend on ρ. Note that such a design has *four* distinct x values. Even though a BIBD supported on 1, 0 and -1 is not optimal, it turns out to be highly efficient.

Section 2 is devoted to the result described in the preceding paragraph for BIBD's (and other optimal block designs for treatment comparisons) supported on the design points of D-optimal designs for the uncorrelated model. The problem of determining D-optimal designs without restricting to minimum-support designs is discussed in Section 3. Here we resort to the *approximate theory* of optimal design. Some deviation from the standard theory is needed to deal with the special feature of within-block correlation. Specifically, the observations in the same block can be considered as a *multivariate* response at a point in the design space $æ^k$. A multivariate version of the Equivalence Theorem as described in Fedorov (1972) and Kiefer (1974) can be used. The case of quadratic regression on $[-1, 1]$ is treated in details.

2. Optimal Minimum-Support Designs

Suppose $\{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_t\}$ is a D-optimal design with t observations for the uncorrelated model (1.2) and (1.3). In this section, we restrict our attention to minimum-support designs, i.e., those supported on exactly t points (although not necessarily the same as $\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_t$). As mentioned earlier, in the present setup, a design has two phases: selecting support points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t$ and constructing a block design with t treatments and b blocks of size k. The blocks can be described by b $k \times t$ incidence matrices $\mathbf{N}_1, \ldots, \mathbf{N}_b$, where the (i, j)th entry of $\mathbf{N}_s$ is 1 if $\boldsymbol{x}_j$ is assigned to the ith observation in the sth block; otherwise, it is equal to 0. Let $\mathbf{N} = [\mathbf{N}_1^T, \ldots, \mathbf{N}_b^T]^T$. Then the design matrix $\mathbf{A}$ in (1.2) can be expressed as $\mathbf{A} = \mathbf{N}\Phi$, where Φ is the $t \times t$ matrix with the (i, j)th entry equal to $f_j(\boldsymbol{x}_i)$, and the information matrix becomes

$$\mathbf{M} = \Phi^T \mathbf{N}^T (\mathbf{I}_b \otimes \mathbf{V})^{-1} \mathbf{N} \Phi. \tag{2.1}$$

For convenience, for any $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t \in \text{æ}$, let $\mathcal{D}(t, b, k; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)$ be the set of all the designs in b blocks of size k supported on $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t$, and let $\mathcal{D}(t, b, k)$ be the set of all the designs in b blocks of size k supported on t points. We first consider the determination of D-optimal designs in $\mathcal{D}(t, b, k; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)$ with fixed support points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t$. Lemma 2.1 in this section shows that this is equivalent to the classical problem of D-optimal designs for *treatment comparisons*.

In the usual block design set-up, let $\mathcal{B}(t, b, k)$ be the set of all the block designs with t treatments and b blocks of size k. This is, of course, the same as $\mathcal{D}(t, b, k; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)$ if we think of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t$ as the treatments. Much work has been done on optimal block designs for estimating treatment contrasts under the usual additive homoscedastic uncorrelated models with fixed block effects. We cite Kiefer's (1958, 1975) fundamental result on the universal optimality of BIBD's and Cheng's (1978) result on the optimality of some group-divisible designs. Some of these results have been extended to random block-effects models; see, e.g. Mukhopadhyay (1981) and Bagchi (1987). Let $\boldsymbol{\alpha}$ be the $t \times 1$ vector of unknown treatment effects, and let $\mathbf{O} = [\mathbf{1}_t/\sqrt{t} \,\vdots\, \mathbf{P}]^T$ be a $t \times t$ orthogonal matrix, where $\mathbf{1}_t$ is the $t \times 1$ vector of ones. Then the problem is to find an optimal design for estimating $\mathbf{P}^T\boldsymbol{\alpha}$ under the linear model

$$E(\boldsymbol{y}) = \mathbf{N}\boldsymbol{\alpha},\ \mathrm{Cov}(\boldsymbol{y}) = \sigma^2(\mathbf{I}_b \otimes \mathbf{V}),\ \text{where } \mathbf{V} = [(1-\rho)\mathbf{I}_k + \rho \mathbf{J}_k]. \tag{2.2}$$

For example, a D-optimal design minimizes $\det[\mathrm{Cov}(\mathbf{P}^T\hat{\boldsymbol{\alpha}})]$. Many results on optimal block designs for fixed block effects carry over to this case. For instance, Kiefer's (1975) method can be used to show that a balanced incomplete block design remains universally optimal in this setup for all positive ρ; see, e.g., p. 86 of Shah and Sinha (1989).

For convenience, for each $d \in \mathcal{B}(t, b, k)$, we use $d(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)$ to denote the design in $\mathcal{D}(t, b, k; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)$ obtained by allocating $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t$ into the blocks of d. Then we have the following

Lemma 2.1. *Suppose $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t$ are t points in æ such that the vectors $\mathbf{f}(\boldsymbol{x}_1), \ldots$ and $\mathbf{f}(\boldsymbol{x}_t)$ are linearly independent. Let d be D-optimal over $\mathcal{B}(t, b, k)$ for estimating the treatment contrasts $\mathbf{P}^T\boldsymbol{\alpha}$ under* (2.2). *Then $d(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)$ is D-optimal over $\mathcal{D}(t, b, k; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)$ under* (1.2) *and* (1.4).

Proof. By (2.1), $\det(\mathbf{M}) = \det[\Phi^T\mathbf{N}^T(\mathbf{I}_b \otimes \mathbf{V})^{-1}\mathbf{N}\Phi] = [\det(\Phi)]^2 \cdot \det[\mathbf{N}^T(\mathbf{I}_b \otimes \mathbf{V})^{-1}\mathbf{N}]$. Since $\mathbf{f}(\boldsymbol{x}_1), \ldots$ and $\mathbf{f}(\boldsymbol{x}_t)$ are linearly independent, $\det(\Phi) > 0$. Also, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t$ are fixed; so $\det(\Phi)$ is a constant. Therefore we need to

$$\text{find a design in } \mathcal{B}(t, b, k) \text{ which maximizes } \det[\mathbf{N}^T(\mathbf{I}_b \otimes \mathbf{V})^{-1}\mathbf{N}]. \tag{2.3}$$

Since $\mathbf{O}$ is an orthogonal matrix,

$$\det[\mathbf{N}^T(\mathbf{I}_b\otimes\mathbf{V})^{-1}\mathbf{N}]=\det[\mathbf{O}\mathbf{N}^T(\mathbf{I}_b\otimes\mathbf{V})^{-1}\mathbf{N}\mathbf{O}^T]; \tag{2.4}$$

also, in (2.2), we have

$$E(\boldsymbol{y})=\mathbf{N}\boldsymbol{\alpha}=\mathbf{N}\mathbf{O}^T\mathbf{O}\boldsymbol{\alpha}=[\mathbf{N}\mathbf{1}_t/\sqrt{t}\,\vdots\,\mathbf{N}\mathbf{P}]\,[\mathbf{1}_t/\sqrt{t}\,\vdots\,\mathbf{P}]^T\boldsymbol{\alpha}.$$

The information matrix $\mathbf{O}\mathbf{N}^T(\mathbf{I}_b\otimes\mathbf{V})^{-1}\mathbf{N}\mathbf{O}^T$ for $\mathbf{O}\boldsymbol{\alpha}$ can be partitioned as $\begin{bmatrix} A & \mathbf{B}^T \\ \mathbf{B} & \mathbf{D}\end{bmatrix}$, where $A=\mathbf{1}_t^T\mathbf{N}^T(\mathbf{I}_b\otimes\mathbf{V})^{-1}\mathbf{N}\mathbf{1}_t/t$, $\mathbf{B}=\mathbf{P}^T\mathbf{N}^T(\mathbf{I}_b\otimes\mathbf{V})^{-1}\mathbf{N}\mathbf{1}_t/\sqrt{t}$, and $\mathbf{D}=\mathbf{P}^T\mathbf{N}^T(\mathbf{I}_b\otimes\mathbf{V})^{-1}\mathbf{N}\mathbf{P}$. Then the information matrix for estimating $\mathbf{P}^T\boldsymbol{\alpha}$ under (2.2) is equal to $\mathbf{D}-A^{-1}\mathbf{B}\mathbf{B}^T$; in other words, under (2.2), a design $d\in\mathcal{B}(t,\,b,\,k)$ is D-optimal for estimating treatment contrasts $\mathbf{P}^T\boldsymbol{\alpha}$ if it

$$\text{maximizes}\ \det(\mathbf{D}-A^{-1}\mathbf{B}\mathbf{B}^T). \tag{2.5}$$

On the other hand,

$$\det[\mathbf{O}\mathbf{N}^T(\mathbf{I}_b\otimes\mathbf{V})^{-1}\mathbf{N}\mathbf{O}^T]=A\cdot\det(\mathbf{D}-A^{-1}\mathbf{B}\mathbf{B}^T). \tag{2.6}$$

Since $\mathbf{N}$ is a $(0,1)$-matrix in which each row contains exactly one 1 (Only one $\boldsymbol{x}\in\text{æ}$ is assigned to each observation), we have $\mathbf{N}\mathbf{1}_t=\mathbf{1}_{bk}$. Therefore A is a constant independent of the designs. From (2.4) and (2.6), we see that (2.3) and (2.5) are equivalent. This proves Lemma 2.1.

From the proof of Lemma 2.1, we have $\det(\mathbf{M})=\det[(\Phi)]^2\cdot A\cdot\det(\mathbf{D}-A^{-1}\mathbf{B}\mathbf{B}^T)$, in which $A\cdot\det(\mathbf{D}-A^{-1}\mathbf{B}\mathbf{B}^T)$ does not depend on $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_t$, and is maximized by the D-optimal design in $\mathcal{B}(t,\,b,\,k)$. Suppose $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_t$ are no longer fixed. Then $\det(\mathbf{M})$ can be maximized by maximizing $\det(\Phi)$. Recall that Φ is the $t\times t$ matrix with (i,j)th entry $f_j(\boldsymbol{x}_i)$, which is precisely the design matrix under the uncorrelated model (1.2) and (1.3) for t design points. In this case the information matrix is $\Phi^T\Phi$. Since $[\det(\Phi)]^2=\det(\Phi^T\Phi)$, it is maximized by the D-optimal design $\tilde{\boldsymbol{x}}_1,\ldots,\tilde{\boldsymbol{x}}_t$ for the uncorrelated model (1.2) and (1.3). Thus we have proved the following:

Theorem 2.2. *Suppose $\{\tilde{\boldsymbol{x}}_1,\ldots,\tilde{\boldsymbol{x}}_t\}$ is a D-optimal design with t observations for the uncorrelated model* (1.2) *and* (1.3), *and d is a D-optimal design in $\mathcal{B}(t,$ $b,\,k)$ for estimating treatment contrasts $\mathbf{P}^T\boldsymbol{\alpha}$ under* (2.2). *Then the design $d(\tilde{\boldsymbol{x}}_1,\ldots,\tilde{\boldsymbol{x}}_t)$ is D-optimal over $\mathcal{D}(t,\,b,\,k)$ under* (1.2) *and* (1.4).

In particular, we have

Corollary 2.3. *Suppose $\{\tilde{\mathbf{x}}_1,\ldots,\tilde{\mathbf{x}}_t\}$ is a D-optimal design with t observations for the uncorrelated model* (1.2) *and* (1.3), *and there exists a balanced incomplete* *block design d^* with t treatments and b blocks of size k. Then $d^*(\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_t)$ is D-optimal over $\mathcal{D}(t,\ b,\ k)$ under* (1.2) *and* (1.4) *for all positive ρ.*

Chasalow (1992) conducted a computer search to find optimal designs with blocks of size two for the quadratic regression on $[-1, 1]$. Several optimality criteria including the D-criterion were used, but the search was restricted to binary designs supported on 1, 0 and -1. When the number of blocks is a multiple of 3, the D-optimal designs he obtained are BIBD's as expected.

3. Approximate Theory

In the preceding section, for the minimum-support case, it was shown that a D-optimal design can be obtained by combining a D-optimal block design for treatment comparisons with a D-optimal regression design for the uncorrelated model. This design, however, is not optimal if the number of support points is allowed to be more than the number of unknown parameters. The general problem of constructing optimal designs without restricting to those with minimum support is quite difficult. In this section, we sketch an approach based on approximate theory, and apply it to the quadratic regression on $[-1, 1]$.

The standard theory for (1.2) and (1.3) defines an approximate design ξ as a (discrete) probability measure on æ. The information matrix of ξ is

$$\mathbf{M}(\xi) = \int_{\text{æ}} \mathbf{f}(\boldsymbol{x})\mathbf{f}(\boldsymbol{x})^T \xi(d\boldsymbol{x}),$$

and a design ξ^* is said to be D-optimal if it maximizes $\det[\mathbf{M}(\xi)]$. The Kiefer-Wolfowitz Equivalence Theorem implies that ξ^* is D-optimal if and only if it is supported on points where the maximum of $\mathbf{f}(\boldsymbol{x})^T\mathbf{M}(\xi^*)^{-1}\mathbf{f}(\boldsymbol{x})$ over $\boldsymbol{x} \in \text{æ}$ is attained. This is a powerful tool for finding optimal designs.

In the case of random block-effects models considered in this article, it is appropriate to think of each block as a point in æ^k. Therefore a design is to choose a collection of points from æ^k. Instead of æ, now the design region is æ^k. We shall denote each point in æ^k by $\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$. The k observations $y_{\boldsymbol{x}_1}, \ldots, y_{\boldsymbol{x}_k}$ in each block (or at each $\boldsymbol{X} \in \text{æ}^k$) can be thought of as a k-variate response. Denote this k-variate response by $\boldsymbol{Y_X} : \boldsymbol{Y_X} \equiv (y_{\boldsymbol{x}_1}, \ldots, y_{\boldsymbol{x}_k})^T$. Then we have

$$E(\boldsymbol{Y_X}) = [\mathbf{f}(\boldsymbol{x}_1)^T\boldsymbol{\theta}, \ldots, \mathbf{f}(\boldsymbol{x}_k)^T\boldsymbol{\theta}]^T = \mathbf{F}(\boldsymbol{X})^T\boldsymbol{\theta},$$

where $\mathbf{F}(\mathbf{X})$ is the $t \times k$ matrix $[\mathbf{f}(\boldsymbol{x}_1), \ldots, \mathbf{f}(\boldsymbol{x}_k)]$, and

$$\mathrm{Cov}(\boldsymbol{Y_X}) = \sigma^2\mathbf{V}, \text{ where } \mathbf{V} = [(1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k].$$

An approximate design is then a probability measure ξ on æ^k. The information matrix of ξ is

$$\mathbf{M}(\xi) = \int_{\text{æ}^k} \mathbf{F}(\boldsymbol{X})\mathbf{V}^{-1}\mathbf{F}(\boldsymbol{X})^T\xi(d\boldsymbol{X}).$$

Again, a D-optimal design maximizes $\det[\mathbf{M}(\xi)]$. Notice that because the order of the k observations in each block is not important, for each $\boldsymbol{X} \in \text{æ}^k$, all the points obtained by permuting the components $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ of $\boldsymbol{X}$ should also be indistinguishable. The approach described above is appropriate if the weights of all the permutations of the components $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ of $\boldsymbol{X}$ are lumped together as the weight of the corresponding block.

The Equivalence Theorem still applies in this multivariate problem except for some necessary modification. The quadratic form $\mathbf{f}(\boldsymbol{x})^T\mathbf{M}(\xi^*)^{-1}\mathbf{f}(\boldsymbol{x})$ is replaced by $\mathbf{F}(\boldsymbol{X})^T\mathbf{M}(\xi^*)^{-1}\mathbf{F}(\boldsymbol{X})\mathbf{V}^{-1}$. Also, since $\mathbf{F}(\boldsymbol{X})^T\mathbf{M}(\xi^*)^{-1}\mathbf{F}(\boldsymbol{X})\mathbf{V}^{-1}$ is a *matrix*, we need to take its trace: a design ξ^* is D-optimal if and only if it is supported on points where the maximum of $\mathrm{tr}[\mathbf{F}(\boldsymbol{X})^T\mathbf{M}(\xi^*)^{-1}\mathbf{F}(\boldsymbol{X})\mathbf{V}^{-1}]$ over $\boldsymbol{X} \in \text{æ}^k$ is attained. If one can guess a candidate ξ^* for a D-optimal design, then this necessary and sufficient condition can be used to verify the optimality.

As an example, we shall specialize to the quadratic regression on $[-1,1]$ with blocks of size two. Then $\text{æ} = [-1,1]$, $k = 2$, and $f_1(x) = 1$, $f_2(x) = x$, $f_3(x) = x^2$. Consider a BIBD supported on 1, 0, and -1, which was shown in Section 2 to be D-optimal over $\mathcal{D}(3, b, 2)$. An approximate version of this design is the uniform measure on the three points $(1,0)^T$, $(-1,0)^T$ and $(1,-1)^T$. Denote this design by ξ_B. Then its information matrix is

$$\mathbf{M}(\xi_B) = \frac{1}{3}\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1}\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$
$$+ \frac{1}{3}\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}.$$

Let

$$g(x,y) \equiv \mathrm{tr}\begin{bmatrix} 1 & x & x^2 \\ 1 & y & y^2 \end{bmatrix}[\mathbf{M}(\xi_B)]^{-1}\begin{bmatrix} 1 & 1 \\ x & y \\ x^2 & y^2 \end{bmatrix}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1}.$$

One can verify that $g(1,0) = g(-1,0) = g(1,-1)$. Therefore, if the design region is æ^2, where $\text{æ} = \{0, 1, -1\}$, instead of $[-1,1]^2$, then ξ_B is D-optimal. This confirms what we have shown earlier that a BIBD supported on 0, 1, -1 is D-optimal over all the block designs supported on 0, 1, -1. However, it can easily be seen that the maximum of $g(x,y)$ over $-1 \leq x, y \leq 1$ is *not* achieved at $(x,y) = (1,0)$, $(-1,0)$ or $(1,-1)$. For example, the maximum of $g(1,y)$ over $-1 \leq y \leq 1$ is attained at a certain y with $-1 < y < 0$. Therefore ξ_B is not D-optimal without restrictions on the competing designs. Of course, not being optimal over the approximate designs does not automatically rule out its optimality over exact designs, but later on we shall use the optimal approximate design to obtain an exact design which performs better than the BIBD supported on 1, 0 and -1. Since the maximum of $g(1,y)$ over $-1 \le y \le 1$ is attained at a certain y with $-1 < y < 0$, this seems to suggest that a D-optimal approximate design may have $(1,a)$ in its support, where a is a certain number such that $-1 < a < 0$. Then by symmetry, $(-1,-a)$ should also be in the support with the same weight as $(1,a)$. This leads to the guess that a D-optimal design may be supported on $(-1,-a)$, $(1,a)$, each with weight $\epsilon/2$, and $(-1,1)$ with weight $1-\epsilon$, where $0 < \epsilon < 1$. Denote such a design by ξ^*, where a and ϵ are to be determined. If we can calculate a and ϵ, then the Equivalence Theorem can be used to verify whether it is, indeed, D-optimal. The power and beauty of the Equivalence Theorem is that it can also be used to determine a and ϵ to complete the guess. We shall demonstrate that this guess does lead to a solution.

We first compute

$$\begin{aligned}\mathbf{M}(\xi^*) = (1-\epsilon)&\begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1}\begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\\ &+\frac{\epsilon}{2}\begin{bmatrix} 1 & 1 \\ -1 & -a \\ 1 & a^2 \end{bmatrix}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1}\begin{bmatrix} 1 & -1 & 1 \\ 1 & -a & a^2 \end{bmatrix}\\ &+\frac{\epsilon}{2}\begin{bmatrix} 1 & 1 \\ 1 & a \\ 1 & a^2 \end{bmatrix}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1}\begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \end{bmatrix}.\end{aligned}$$

Then, except for a multiplicative constant,

$$[\mathbf{M}(\xi^*)]^{-1} \propto \begin{bmatrix} \alpha & 0 & \delta \\ 0 & \beta & 0 \\ \delta & 0 & \gamma \end{bmatrix}, \tag{3.1}$$

where

$$\alpha = [\epsilon(1-2\rho a+a^2)+2(1-\epsilon)(1+\rho)\,[\epsilon(1-2\rho a^2+a^4)+2(1-\epsilon)(1-\rho)], \tag{3.2}$$
$$\beta = (1-\rho)[2\epsilon(1-2\rho a^2+a^4)+4(1-\epsilon)(1-\rho)-(1-\rho)(\epsilon a^2+2-\epsilon)^2], \tag{3.3}$$
$$\gamma = 2(1-\rho)[\epsilon(1-2\rho a+a^2)+2(1-\epsilon)(1+\rho)], \tag{3.4}$$
$$\delta = -\,(1-\rho)(\epsilon a^2+2-\epsilon)[\epsilon(1-2\rho a+a^2)+2(1-\epsilon)(1+\rho)]. \tag{3.5}$$

Let $g(x,\,y;\,a,\,\epsilon,\,\rho) = \operatorname{tr}\begin{bmatrix} 1 & x & x^2 \\ 1 & y & y^2 \end{bmatrix}\begin{bmatrix} \alpha & 0 & \delta \\ 0 & \beta & 0 \\ \delta & 0 & \gamma \end{bmatrix}\begin{bmatrix} 1 & 1 \\ x & y \\ x^2 & y^2 \end{bmatrix}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1}$. If ξ^* is D-optimal, then by the Equivalence Theorem, the maximum of $g(x,y;a,\epsilon,\rho)$ over

$-1 \le x, y \le 1$ must be attained at $(-1,1)$, $(-1,-a)$ and $(1,a)$. In particular, we must have

$$g(-1,\, 1;\, a,\, \epsilon,\, \rho) = g(-1,\, -a;\, a,\, \epsilon,\, \rho) \tag{3.6}$$

and

$$\frac{\partial}{\partial y}\, g(-1,\, y;\, a,\, \epsilon,\, \rho)|_{y=-a} = 0. \tag{3.7}$$

Tedious calculations lead to the following two equations:

$$\begin{aligned}&(2-2a^2+4\rho+4\rho a)[\epsilon(1-2\rho a^2+a^4)+2(1-\epsilon)(1-\rho)]\\ &+(\epsilon a^2+2-\epsilon)^2(a^2-1-2a\rho+2a\rho^2-a^2\rho+2\rho^2-\rho)-2(1-a^2)[\epsilon(1-2\rho a+a^2)\\ &+2(1-\epsilon)(1+\rho)]\,[(1-\rho)(\epsilon a^2+2-\epsilon)-a^2-1+2\rho]=0\end{aligned} \tag{3.8}$$

and

$$\begin{aligned}&2(a-\rho)[\epsilon(1-2\rho a^2+a^4)+2(1-\epsilon)(1-\rho)]-(1-a^2)(2a-a\epsilon+a\epsilon\rho)\\ &\cdot[2\epsilon(1-2\rho a+a^2)+4(1-\epsilon)(1+\rho)]+(1-\rho)(\rho-a)(\epsilon a^2+2-\epsilon)^2=0.\end{aligned} \tag{3.9}$$

For any given ρ, we have two equations in two unknowns, which can be used to calculate ϵ and a. Therefore both a and ϵ depend on ρ. After (3.8) and (3.9) are solved, one can verify whether the maximum of the resulting g function over (x,y), $-1 \le x, y \le 1$, is indeed attained at the three points $(-1,1)$, $(-1,-a)$ and $(1,a)$ in the support of ξ^*. This task is considerably simplified by the following result:

Lemma 3.1. *Suppose for a given ρ, (3.8) and (3.9) have a solution (a_ρ, ϵ_ρ) with $0 < \epsilon_\rho < 1$ and $-1 < a_\rho < 0$. Suppress a_ρ, ϵ_ρ, ρ, and write $g(x,\, y;\, a_\rho,\, \epsilon_\rho,\, \rho)$ as $g(x,y)$. Also, let α_ρ, β_ρ, γ_ρ, δ_ρ be the values of α, β, γ and δ in (3.2)-(3.5) when a_ρ and ϵ_ρ are substituted for a and ρ, respectively. If $\beta_\rho + \gamma_\rho + 2\delta_\rho > 0$, then the maximum of $g(x,y)$ over $-1 \le x, y \le 1$, is attained at $(-1,1)$, $(1,a_\rho)$ and $(-1,-a_\rho)$.*

Proof. It is clear that $g(-1,-a_\rho) = g(1,a_\rho)$. Also, by (3.6), $g(-1,-a_\rho) = g(-1,1)$. So it is enough to verify that the maximum is attained at $(x,y) = (-1,-a_\rho)$.

We first show that the maximum is attained at a point (x,y) with $x = -1$ and $0 \le y \le 1$. By (3.1), we can write $g(x,y)$ as $p(x,y) + q(x,y)$, where

$$p(x,y) = (1-\rho)[2\alpha_\rho + 2\delta_\rho x^2 + 2\delta_\rho y^2 + \beta_\rho x^2 + \beta_\rho y^2 + \gamma_\rho x^4 + \gamma_\rho y^4]$$

and

$$q(x,y) = \rho[\beta_\rho(x-y)^2 + \gamma_\rho(x^2-y^2)^2].$$

Since $\mathbf{M}[(\xi^*)]^{-1}$ is positive definite,

$$\alpha_\rho, \beta_\rho, \gamma_\rho > 0. \tag{3.10}$$

Clearly,

$$p(x,y) = p(-x,y) = p(x,-y) = p(-x,-y) = p(y,x), \tag{3.11}$$

and

$$q(x,y) = q(-x,-y) = q(y,x), \tag{3.12}$$

while (3.10) implies that $q(-x,y) \geq q(x,y)$ if $xy \geq 0$. Therefore, without loss of generality, we may assume that $x \leq 0$ and $y \geq 0$. By using the assumption $\beta_\rho + \gamma_\rho + 2\delta_\rho > 0$, it is easy to see that

$$\text{for any } -1 \leq x \leq 0 \text{ and } 0 \leq y \leq 1,\ p(x,y) \leq p(-1,y). \tag{3.13}$$

Now suppose at least one of x^2 and y^2 is $\leq 1/2$. We shall show that $g(x,y) \leq g(-1,y_0)$ for some $0 \leq y_0 \leq 1$. By (3.11) and (3.12), we may assume that $0 \leq y \leq 1/\sqrt{2}$. Then clearly $q(-1,y) \geq q(x,y)$. This together with (3.13) imply that $g(-1,y) \geq g(x,y)$. Therefore we only have to consider the case where both x^2 and y^2 are $> 1/2$. Repeating the same argument over and over again, we can successively eliminate the cases that both of x^2 and y^2 are < 1, but at least one is $\leq 3/4$, $7/8$, $15/16$, etc, and conclude that the maximum must be attained at a point (x,y) with $x = -1$ and $0 \leq y \leq 1$.

Therefore one can fix x at -1, and consider the maximization of $g(-1,y)$ over $0 \leq y \leq 1$. Since $g(-1,y)$ is a fourth-degree polynomial in y with a positive leading coefficient, it has at most one local maximum and two local minima. Now observe that

(i) $\frac{\partial}{\partial y} g(-1,y)|_{y=1} > 0$, and
(ii) $g(-1,1) > g(-1,-1)$.

Also, by (3.6) and (3.7), we have $g(-1,1) = g(-1,-a_\rho)$ and $\frac{\partial}{\partial y} g(-1,y)|_{y=-a_\rho} = 0$. Using these facts, we can easily show that $g(-1,-a_\rho)$ is a local maximum, and $g(-1,1) = g(-1,-a_\rho)$ is the maximum of $g(-1,y)$ over $0 \leq y \leq 1$.

Therefore to show that the maximum of $g(x,y)$ over $-1 \leq x, y \leq 1$ is attained at $(-1,1)$, $(-1,-a_\rho)$ and $(1,a_\rho)$, it is enough just to verify $\beta_\rho + \gamma_\rho + 2\delta_\rho > 0$, which is a simple task.

At the end of this article is a table containing solutions (a_ρ, ϵ_ρ) of (3.6) and (3.7) for $\rho = 0.1, 0.2, \ldots, 0.9$. In each case, there is a solution with $0 < \epsilon_\rho < 1$ and $-1 < a_\rho < 0$; furthermore, the condition $\beta_\rho + \gamma_\rho + 2\delta_\rho > 0$ is satisfied. Notice that the positive values $-a_\rho$, instead of a_ρ, are tabulated. For instance, for $\rho = 0.5$, the D-optimal approximate design puts a weight of 0.28818 on the block $(-1,1)$, and 0.35591 each on $(1,-0.131269)$ and $(-1,0.131269)$. It is clear from the table that as ρ decreases, both $-a_\rho$ and ϵ_ρ decrease. When ρ gets closer to zero, a_ρ tends to 0, ϵ_ρ tends to 2/3, and the optimal design converges to that for the uncorrelated case.

We define the D-efficiency of any given design ξ as $\{\det[\mathbf{M}(\xi)]/\det[\mathbf{M}(\xi^*)]\}^{1/t}$ ($t=3$ for the quadratic regression), where ξ^* is the D-optimal design. The table also contains efficiencies of the design ξ_U with the same weight 1/3 on each of the three blocks in the support of the D-optimal ξ^*. It turns out that even when optimal weights are not used, the efficiencies are extremely high. Finally, we also calculate the efficiencies of the design ξ_B with uniform weights on $(1,0)$, $(-1,0)$ and $(-1,1)$, the approximate version of a BIBD supported on 1, 0, -1. We see that ξ_B is slightly less efficient than ξ_U; therefore, a BIBD supported on 1, 0 and -1 is *not* an optimal *exact* design. This design does not have optimal support points nor optimal weights, but the table shows that its efficiencies are also very high. In practice, one might prefer such a design since its support points and weights do not depend on the correlation ρ. So our original idea of combining optimal block designs with optimal regression designs for uncorrelated models does seem to work quite well. Such designs are not optimal over all possible competing designs, but are expected to be highly efficient.

Even for the simplest quadratic regression model on $[-1,1]$, the solution presented here is already quite complex. We hope to extend the results in this article to other models, design regions and block sizes in further studies.

Table. Optimal designs for the quadratic regression on $[-1,1]$ with blocks of size two; efficiencies of ξ_U and ξ_B

ρ	$-a_\rho$	ϵ_ρ	efficiency of ξ_U	efficiency of ξ_B
0	0	2/3	1	1
0.1	0.031300	0.669064	0.999997	0.999503
0.2	0.059255	0.675536	0.999956	0.998204
0.3	0.084799	0.685203	0.999807	0.996314
0.4	0.108635	0.697439	0.999472	0.993969
0.5	0.131269	0.711820	0.998871	0.991262
0.6	0.153065	0.728065	0.997932	0.988257
0.7	0.174281	0.745990	0.996590	0.985001
0.8	0.195104	0.765487	0.994785	0.981532
0.9	0.215667	0.786501	0.992463	0.977876

Acknowledgement

This research was supported by National Science Foundation Grant No.

DMS-9100938.

References

Bagchi, S. (1987). On the optimalities of the MBGDDs under mixed effects model. Comm. Statist. Theory Methods **16**, 3565-3576.

Chasalow, S. D. (1992). Exact optimal response surface designs with random block effects. Ph.D. Dissertation, University of California, Berkeley.

Cheng, C. S. (1978). Optimality of certain asymmetrical experimental designs. Ann. Statist. **6**, 1239-1261.

Fedorov, V. V. (1972). Theory of Optimal Experiments. Academic Press, New York.

Khuri, A. I. (1992). Response surface models with random block effects. Technometrics **34**, 26-37.

Kiefer, J. (1958). On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. Ann. Math. Statist. **29**, 675-699.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). Ann. Statist. **2**, 849-879.

Kiefer, J. (1975). Construction and optimality of generalized Youden designs. In A Survey of Statistical Design and Linear Models (Edited by J. N. Srivastava), 333-353, North-Holland, Amsterdam.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. Canada J. Math. **14**, 363-366.

Mukhopadhyay, S. (1981). On the optimality of block designs under mixed effects model. Cal. Statist. Assoc. Bull. **30**, 171-185.

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

Shah, K. R. and Sinha, B. K. (1989). Theory of Optimal Designs. Springer-Verlag.

Department of Statistics, University of California, Berkeley, CA 94720, U.S.A.

(Received November 1993; accepted November 1994)