DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

OUTPERFORMING COMPLETELY
RANDOMIZED DESIGNS
by
P. GOOS
M. VANDEBROEK

Outperforming Completely Randomized Designs

Peter Goos

Martina Vandebroek

Katholieke Universiteit Leuven, Belgium

Abstract

Bi-randomization designs have become increasingly popular in industry be-
cause some of the factors under investigation are often hard-to-change. It is
well-known that the resulting compound symmetric error structure not only
affects estimation and inference procedures but also the efficiency of the ex-
perimental designs used. In this paper, the use of bi-randomization designs is
shown to outperform completely randomized designs in terms of $\mathcal{D}$-efficiency.
This result suggests that bi-randomization designs should be considered as an
alternative to completely randomized designs even if all experimental factors
are easy-to-change.

Keywords: $\mathcal{D}$-optimality, correlated observations, experimental design, hard-
to-change factors, restricted randomization, split-plot design

1   Introduction

Bi-randomization designs (BRDs) are heavily used in industry especially when factor
levels are difficult or costly to change or to control. Typical examples of such factors
are pressure, humidity and process temperature. Rather than conducting a com-
pletely randomized experiment in which pressure has to be moved back and forth
according to the randomization scheme, executing experimental runs with equal
pressure successively is preferred by the experimenter. Alternatively, the exper-
imenter could group all experimental runs with the same temperature level and
execute them simultaneously in one furnace. In doing so, the number of changes in
the levels of the hard-to-change factors is limited to the number of levels. Letsing-
er et al. (1996) point out that the resulting compound symmetric error structure
affects estimation and inference procedures as well as design efficiency. Goos and
Vandebroek (1999) propose an algorithm to construct $\mathcal{D}$-optimal bi-randomization
designs in this context. Ganju and Lucas (1999) describe how bi-randomization
experiments are obtained by not resetting the factor levels for the consecutive runs
of the random run order. They argue that bi-randomization experiments should be
designed rather than being the accidental outcome of a random run order.

By using a bi-randomization design in the presence of hard-to-change factors, the ease of experimentation is significantly increased and precious time and resources can be saved. However, this should not be the only reason to consider using bi-randomization. In this paper, we will show that bi-randomization designs are often statistically much more efficient than completely randomized experiments. Contrary to bi-randomization designs as described by Letsinger et al. (1996) and Goos and Vandebroek (1999), the bi-randomization designs described here allow more factor level changes. The benefits of this approach are fourfold. Firstly, the statistical efficiency of the experiment is increased. Secondly, increasing the number of level changes protects the experimenter against systematic errors which may occur when something goes wrong at a certain hard-to-change factor level. Thirdly, more degrees of freedom are available for the estimation of the whole plot error. Finally, the number of factor level changes is mostly smaller than in a completely randomized design. For all these reasons, the type of bi-randomization design discussed here should not only be considered when some of the experimental factors are hard-to-change, but even more so when no hard-to-change factors are present.

In Section 2, we describe the model assumptions corresponding to a bi-randomization design as it is used in industry when hard-to-change factors are present. The consequences for model analysis and design efficiency are briefly discussed. In Section 3, we allow more flexibility in the design of a bi-randomization experiment and motivate this approach. In Section 4, an efficient algorithm for the construction of $\mathcal{D}$-optimal BRDs is presented. Finally, computational results are discussed in Section 5.

2 Hard-to-change factors

The presence of hard-to-change factors often dictates the necessity to run response surface experiments in a bi-randomization format. In this section, the corresponding response surface model and the underlying assumptions are discussed. It is shown that observations are statistically dependent and have a compound symmetric error structure. As a result, the statistical analysis and the efficiency of the experimental design are affected.

2.1 Model

The set of independent variables in a bi-randomization experiment is divided in two groups. The n_w hard-to-change design variables are denoted by $z_1, \ldots, z_{n_w}$ or simply by $\mathbf{z}$ and are called the whole plot variables. The remaining n_s variables are the sub-plot variables $x_1, \ldots, x_{n_s}$ or $\mathbf{x}$. The bi-randomization experiment has two types of experimental units called whole plots and sub-plots and therefore two randomization procedures. Firstly, the w unique factor level combinations of the whole plot variables are randomly assigned to the w whole plot experimental units, thereby generating the whole plot error variance. It is assumed that there is a one-to-one relation between the whole plots and the w hard-to-change factor level

combinations in the experiment. As a consequence, the level of the hard-to-change factors is changed only w times during the experiment. The second randomization consists of allocating the factor level combinations of the sub-plot variables $\mathbf{x}$ to the sub-plot experimental units, generating the sub-plot error variance. The jth observation within the ith whole plot can be written as

$$Y_{ij} = \mathbf{f}'(\mathbf{z}_i, \mathbf{x}_{ij})\boldsymbol{\beta} + \delta_i + \varepsilon_{ij}, \tag{1}$$

where the $p \times 1$ vector $\mathbf{f}(\mathbf{z}, \mathbf{x})$ represents the polynomial expansion of the experimental variables and the $p \times 1$ vector $\boldsymbol{\beta}$ contains the p model parameters. In matrix notation, we have

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{Y}$ is the $n \times 1$ vector containing the n observations, $\mathbf{A}$ is the $n \times p$ design matrix containing the n vectors $\mathbf{f}(\mathbf{z}_i, \mathbf{x}_{ij})$, $\mathbf{B}$ is the $n \times w$ matrix with (i,j)th entry equal to one if the ith observation belongs to the jth whole plot and equal to zero if not, and $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are the $w \times 1$ and $n \times 1$ vectors with whole plot and sub-plot errors respectively. We assume that $E(\boldsymbol{\delta}) = E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\text{Cov}(\boldsymbol{\delta}) = \sigma_\delta^2 \mathbf{I}_w$, $\text{Cov}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}_n$ and $\text{Cov}(\boldsymbol{\delta}, \boldsymbol{\varepsilon}) = \mathbf{0}$, with $\mathbf{I}_w$ and $\mathbf{I}_n$ the w- and n-dimensional identity matrix. The variance-covariance matrix of the s_i observations within the ith whole plot is the $s_i \times s_i$ compound symmetric matrix

$$\begin{aligned} \mathbf{V}_i &= \sigma_\varepsilon^2 \mathbf{I}_{s_i \times s_i} + \sigma_\delta^2 \mathbf{1}_{s_i} \mathbf{1}'_{s_i}, \\ &= \sigma_\varepsilon^2 (\mathbf{I}_{s_i \times s_i} + d \mathbf{1}_{s_i} \mathbf{1}'_{s_i}), \end{aligned} \tag{3}$$

with d the variance ratio $\sigma_\delta^2/\sigma_\varepsilon^2$, $\mathbf{I}_{s_i}$ the s_i dimensional identity matrix and $\mathbf{1}_{s_i}$ an $s_i \times 1$ vector of ones. Note that $n = \sum_{i=1}^w s_i$. The variance ratio d measures the extent to which observations from the same whole plot are correlated. Observations from different whole plots are uncorrelated. Suppose the entries of $\mathbf{Y}$ are grouped per whole plot, then $\text{Cov}(\mathbf{Y})$ is given by the $n \times n$ block diagonal matrix

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{V}_w \end{bmatrix}. \tag{4}$$

2.2 Analysis

The statistical data analysis of bi-randomization experiments differs from that of completely randomized experiments. Under normality, the maximum likelihood estimate of the unknown model parameter $\boldsymbol{\beta}$ in (2) is the generalized least squares estimate instead of the ordinary least squares estimate. In order to investigate the impact of this change in analysis, a distinction between crossed, balanced and non-crossed BRDs has to be made.

Crossed, balanced and non-crossed BRDs

Crossed BRDs —also referred to as split-plot designs— differ from non-crossed BRDs in that every combination of levels of $\mathbf{x}$ appears in each whole plot. Each whole plot then has an equal number of sub-plots, namely $s = n/w$, as well as equal levels of $\mathbf{x}$. In non-crossed BRDs, each whole plot may have a different number of sub-plots and the levels of the sub-plot variables need no longer be identical across whole plots. A BRD is called balanced when the number of sub-plots within each whole plot is equal. Crossed designs are balanced, but not vice versa. In general, the design matrices $\mathbf{A} = (\mathbf{A}_1' | \ldots | \mathbf{A}_w')'$ of crossed, non-crossed and balanced BRDs can be written as

$$
\begin{bmatrix}
\mathbf{f}'(\mathbf{z}_1, \mathbf{x}_1) \\
\vdots \\
\mathbf{f}'(\mathbf{z}_1, \mathbf{x}_s) \\
\hdashline
\mathbf{f}'(\mathbf{z}_2, \mathbf{x}_1) \\
\vdots \\
\mathbf{f}'(\mathbf{z}_2, \mathbf{x}_s) \\
\hdashline
\vdots \\
\hdashline
\mathbf{f}'(\mathbf{z}_w, \mathbf{x}_1) \\
\vdots \\
\mathbf{f}'(\mathbf{z}_w, \mathbf{x}_s)
\end{bmatrix} ,
\begin{bmatrix}
\mathbf{f}'(\mathbf{z}_1, \mathbf{x}_{11}) \\
\vdots \\
\mathbf{f}'(\mathbf{z}_1, \mathbf{x}_{1s_1}) \\
\hdashline
\mathbf{f}'(\mathbf{z}_2, \mathbf{x}_{21}) \\
\vdots \\
\mathbf{f}'(\mathbf{z}_2, \mathbf{x}_{2s_2}) \\
\hdashline
\vdots \\
\hdashline
\mathbf{f}'(\mathbf{z}_w, \mathbf{x}_{w1}) \\
\vdots \\
\mathbf{f}'(\mathbf{z}_w, \mathbf{x}_{ws_w})
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
\mathbf{f}'(\mathbf{z}_1, \mathbf{x}_{11}) \\
\vdots \\
\mathbf{f}'(\mathbf{z}_1, \mathbf{x}_{1s}) \\
\hdashline
\mathbf{f}'(\mathbf{z}_2, \mathbf{x}_{21}) \\
\vdots \\
\mathbf{f}'(\mathbf{z}_2, \mathbf{x}_{2s}) \\
\hdashline
\vdots \\
\hdashline
\mathbf{f}'(\mathbf{z}_w, \mathbf{x}_{w1}) \\
\vdots \\
\mathbf{f}'(\mathbf{z}_w, \mathbf{x}_{ws})
\end{bmatrix}
$$

respectively. Designs that fall within the category of crossed BRDs are the factorial designs. Included within the category of non-crossed BRDs are the central composite design (CCD) and the Box-Behnken designs. BRDs derived from a CCD or a Box-Behnken design are not balanced.

Estimation and inference

Under the assumption of normal errors, the maximum likelihood estimator of the unknown model parameters $\boldsymbol{\beta}$ is given by the generalized least squares (GLS) estimation equation

$$
\hat{\boldsymbol{\beta}} = (\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}\mathbf{y}. \tag{5}
$$

The variance-covariance matrix of the estimators is given by

$$
\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}. \tag{6}
$$

However, (5) and (6) cannot be used directly since the variances σ_δ^2 and σ_ε^2 of the whole plot and sub-plot errors are not known. For saturated designs, crossed BRDs and some specific first order non-crossed BRDs (see Davison (1995)), ordinary and generalized least squares prove to be equivalent, implying that, in these cases, error

4

variance knowledge is no longer necessary for model estimation purposes. In all other cases and for model editing, knowledge of the whole plot and sub-plot error variance remains essential. Error variance estimates are thoroughly described by Letsinger et al. (1996). They recommend restricted maximum likelihood (REML) for error variance estimation because of its robustness across various values of d and because it is a good estimation option when small designs and near full second order models are used.

The risks of improper analysis of BRDs are pointed out by Box and Jones (1992), Davison (1995) and Ganju and Lucas (1997), who extend the results of Kempthorne (1952). By using a BRD, a loss of precision in estimation of whole plot coefficients is incurred, while the opposite is true for the sub-plot coefficients and the whole plot by sub-plot interactions. Analysis of a BRD as a CRD can therefore lead to erroneously considering whole plot effects as significant and sub-plot effects as insignificant. See for instance Nelson (1985).

2.3 Design efficiency

Typically, a standard design is used to perform a bi-randomization experiment. However, these designs were developed to be applied in completely randomized experiments and do not take into account the bi-randomization error structure. Therefore, they are less efficient than designs specifically constructed for bi-randomization experiments. In this paper, we will use the $\mathcal{D}$-optimality criterion to evaluate designs. It is by far the most used optimality criterion and a direct function of $\text{Cov}(\hat{\boldsymbol{\beta}})$.

For a CRD the $\mathcal{D}$-criterion value is given by $|\mathbf{A}'\mathbf{A}|/\sigma_\varepsilon^2$. Since σ_ε^2 is a constant, it does not affect the efficiency of a design. When an experiment is conducted under a bi-randomization structure, the $\mathcal{D}$-criterion value $|\mathbf{A}'\mathbf{V}^{-1}\mathbf{A}|$ depends on σ_ε^2 and σ_δ^2 through $\mathbf{V}$. This result suggests that $\mathcal{D}$-optimal CRDs are no longer $\mathcal{D}$-optimal as a bi-randomization experiment. In fact, designing a completely randomized experiment only consists of determining the design points. Designing a bi-randomization experiment simultaneously involves choosing the number of whole plots and the number of sub-plots within each whole plot, i.e. determining the structure of the variance-covariance matrix $\mathbf{V}$. Neglecting the bi-randomization structure of a response surface experiment may therefore lead to poor designs and consequent analysis. The dependence of design efficiency on the bi-randomization structure was pointed out by Letsinger et al. (1996), who compare $\mathcal{D}$- and $\mathcal{Q}$-efficiencies of standard first and second order response surface designs under various variance ratios. Goos and Vandebroek (1999) provide an algorithm to compute $\mathcal{D}$-optimal BRDs and point out that the resulting designs are usually not crossed.

3 Increasing the number of level changes

In this section, we extend the class of bi-randomization designs by allowing more changes in the whole plot factor levels. Stated otherwise, we relax the assumption that there is a one-to-one relation between the whole plot factor levels and the whole plot experimental units. Firstly, we illustrate the modification by means of an example and describe how model (2) is changed due to the different assumption. Next, we will motivate the importance of the model and the corresponding design problem.

3.1 Example

In order to clarify the purpose of this paper, consider a factorial experiment from the steel industry described by Andersen and McLean (1974). The experiment involved four furnaces and was meant to investigate the impact of furnace temperature and orientation within the furnace on the strength of three alloys. We assume the model of interest contains all linear factor effects and the quadratic effect of the factor temperature. Temperature had four experimental levels (675, 700, 725 and 750), while the factor orientation had only two levels: random and aligned orientation. Each furnace had a different temperature and within each furnace the investigator randomly placed (orientation 1) samples from each of three alloys and had an aligned arrangement (orientation 2) of other samples from the same three alloys. In other words, all experimental runs for which the temperature was at the same level were conducted in the same furnace. The whole plot factor in this experiment is temperature, whereas alloy type and orientation are the sub-plot factors. As a result, the temperature is set only four times and the experiment involves four whole plots. It is obvious that a one-to-one relation is assumed between the whole plot factor levels (temperatures) and the whole plots (furnaces).

In this paper, we will allow the whole plot factor temperature to be reset more often. In doing so, we increase the number of whole plots in the experiment. For instance, instead of using each furnace only once, the experimenter could use them twice resulting in eight independent heatings and as many whole plots. For a given number of experimental runs, each whole plot would then contain less observations. As before, exactly one temperature corresponds to each whole plot (furnace). However, more than one whole plot corresponds to each whole plot factor level (temperature). Although this experimental setup is more cumbersome than the original, it is still much easier to conduct than a completely randomized experiment. As a matter of fact, the latter would require $3 \times 4 \times 2 = 24$ instead of 8 independent heatings. Moreover, increasing the number of whole plots yields BRDs that are statistically more efficient than the original experiment. As an illustration, a $\mathcal{D}$-optimal BRD with six whole plots is shown in Table 1. For slightly correlated responses ($d = 0.1$), the $\mathcal{D}$-criterion value is improved by 60%. For larger degrees of correlation, the improvement is even more pronounced: when $d = 2$, the $\mathcal{D}$-criterion value of the BRD with six whole plots is three times larger than that of the original experiment with four whole plots

Temp	Or	Alloy	Temp	Or	Alloy	Temp	Or	Alloy
675	1	1	700	1	3	750	1	2
675	2	1	700	2	2	750	2	2
675	1	2	700	1	1	750	1	3
675	2	3	700	2	2	750	2	1
675	1	3	725	1	1	750	1	1
675	2	2	725	2	3	750	2	3
675	1	3	725	1	2	750	1	2
675	2	1	725	2	3	750	2	1

Table 1: $\mathcal{D}$-optimal BRD with 24 runs and 6 whole plots for the steel experiment.

and is even slightly better than that of a completely randomized experiment with the design points of the original experiment. In the sequel of the paper, we will show that situations in which completely randomized experiments are outperformed by bi-randomization experiments in terms of $\mathcal{D}$-efficiency are numerous when the number of whole plots is slightly increased. In addition, the experimenter is better protected against systematic errors. Suppose that something went wrong in one of the furnaces in the original experiment, then all six runs executed in that furnace would be affected. On the contrary, if something went wrong in one of the furnaces in the alternative setup of Table 1, only four observations would be distorted. We will now outline the modified model assumptions and motivate the use of increasing the number of changes in the whole plot factor levels.

3.2 Model

As before, observations within one whole plot all have the same whole plot factor level combination. However, a given whole plot factor level combination can occur in more than one whole plot. The kth observation within the jth whole plot with the ith whole plot factor level combination can be written as

$$Y_{ijk} = \mathbf{f}'(\mathbf{z}_i, \mathbf{x}_{ijk})\boldsymbol{\beta} + \delta_{ij} + \varepsilon_{ijk}. \tag{7}$$

As a direct result of the changed assumption, the bi-randomization experiment can have more different whole plots than it has different whole plot factor level combinations. A special case is the case in which each observation is assigned to a different whole plot. The factor levels are then independently reset for each run and the experiment comes down to a completely randomized experiment. As a matter of fact, each whole plot then contains only one observation such that $\mathbf{V}$ is a diagonal matrix and σ_{ε}^2 and σ_{δ}^2 cannot be distinguished from each other. Similarly, model (2) is obtained from model (7) when there is only one whole plot for each whole plot factor level combination in the design. However, between both approaches lies a largely unexplored spectrum of bi-randomization designs which are often statistically more efficient than both extremes. In general, the designs corresponding to model (7) possess more whole plots than the BRDs corresponding to model (2), but less than a completely randomized experiment. In the sequel of this paper, we will show that

7

they should be considered as an alternative to the completely randomized designs in the absence of hard-to-change factors. Furthermore, we produce a number of good reasons why increasing the number of whole plots may be useful even if some of the experimental factors are hard-to-change.

3.3   Motivation

Traditionally, observations from an experiment were assumed to be statistically independent. However, this assumption is often not realistic and the impact of correlated observations on the efficiency of experimental designs has recently received considerable attention in the literature. It turns out that the presence of correlation between observations can be beneficial to the design efficiency. Anbari and Lucas (1994) show that proper blocking of two level factorial designs in the presence of hard-to-change factors is a better approach in terms of $\mathcal{G}$-efficiency than running them in a random order. Goos and Vandebroek (1999) found that $\mathcal{D}$-optimal BRDs are often more efficient than $\mathcal{D}$-optimal completely randomized experiments. Moreover, they prove that design efficiency always benefits from correlated observations when the designs are saturated. Box and Jones (1992) and Davison (1995) show that using a BRD increases the precision in estimation of the sub-plot coefficients and the whole plot by sub-plot interaction coefficients, but decreases the precision of the whole plot coefficients. This is especially true when the BRD possesses few whole plots, making it hard to detect significant whole plot factor effects. However, all other things being equal, slightly increasing the number of whole plots largely solves this problem.

Using the algorithm described by Goos and Vandebroek (1999) to construct $\mathcal{D}$-optimal BRDs by choosing design points from a 21×21 grid on $\chi = [-1, 1]^2$, we obtained strong indications that substantial efficiency gains could be realized by relaxing the one-to-one relation between the combinations of whole plot factor levels and the whole plots. Consider the $\mathcal{D}$-optimal BRDs for a full quadratic model in one whole plot variable z and one sub-plot variable x in Figure 1 obtained by using the default 3×3 grid on χ. Choosing the finer 21×21 grid yields the BRDs in Figure 2.

For a given variance ratio d, the BRDs in Figure 2 have a much larger $\mathcal{D}$-efficiency than the BRDs in Figure 1. For instance, the BRD in Figure 2a is 29% more efficient than the BRD in Figure 1b when $d = 0.1$. They also possess larger numbers of whole plot levels and thus of whole plots. Besides -1, 0 and +1, the design construction algorithm also chooses $\pm 0.1, \pm 0.2, \pm 0.8$ and ± 0.9 as whole plot levels. The optimal sub-plot levels remain 0 and ± 1. However, grouping the observations which are close to each other yields the designs displayed in Figure 1. This result suggests that the optimal whole plot levels are indeed at the 0 and ± 1 levels, but that not all observations at a certain whole plot level should be put in the same whole plot. Instead, it strongly recommends the use of more whole plots in order to decrease the number of correlated observations.
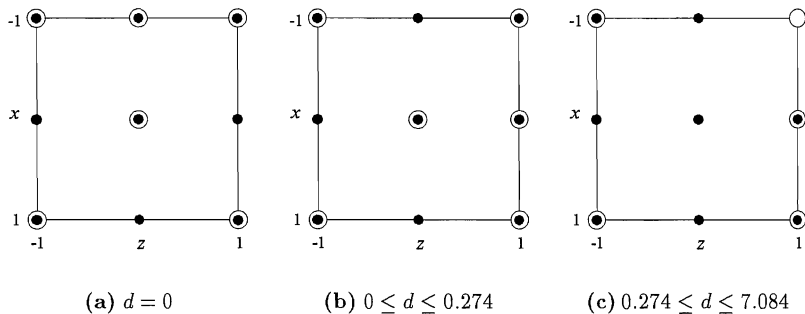
(a) $d = 0$ (b) $0 \le d \le 0.274$ (c) $0.274 \le d \le 7.084$

Figure 1: $\mathcal{D}$-optimal 15-point BRDs for the full quadratic model in 2 variables. Candidate design points are points on a 3×3 grid on $\chi = [-1, 1]^2$. $\bullet$ is a design point, $\odot$ is a design point replicated twice and $\bigcirc$ is a design point replicated three times.
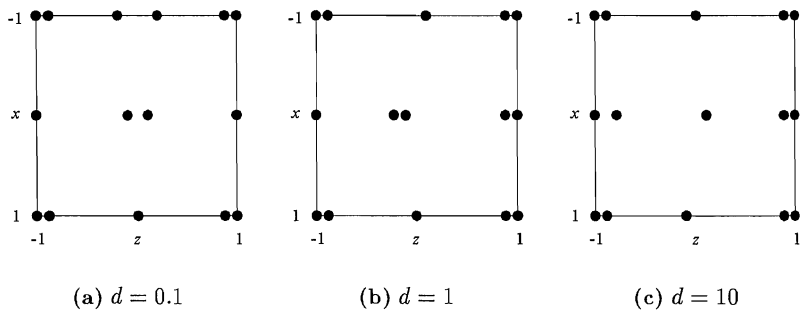

(a) $d = 0.1$ (b) $d = 1$ (c) $d = 10$

Figure 2: $\mathcal{D}$-optimal 15-point BRDs for the full quadratic model in 2 variables. Candidate design points are points on a 21×21 grid on $\chi = [-1, 1]^2$. $\bullet$ is a design point.

9

Another advantage of increasing the number of whole plots is that more degrees of freedom are available for the estimation of the whole plot error. The whole plot degrees of freedom are given by the difference between the rank of $[\mathbf{A}|\mathbf{B}]$ and the rank of $\mathbf{A}$. All other things being equal, increasing the number of whole plots increases the number of columns in $\mathbf{B}$ and thus the rank of $[\mathbf{A}|\mathbf{B}]$. However, one should not forget that the sub-plot error has to be estimated as well. The sub-plot error degrees of freedom are given by the number of observations n minus the rank of $[\mathbf{A}|\mathbf{B}]$. As a result, increasing the number of whole plots leads to a decrease in the degrees of freedom available for sub-plot error estimation. Ganju and Lucas (1999) plead for properly designed bi-randomization experiments that allow decent estimation of both error components. Of course, better estimates for the variance components will lead to better estimates of the model parameters too.

Finally, increasing the number of factor level changes reduces the risk that systematic errors distort the results of the experiment. For example, if something goes wrong at a given level of a hard-to-change factor, this will influence all the observations at this level since it is not independently reset for each experimental run. As experiments in the presence of hard-to-change factors typically use a minimum of factor levels, this may have dramatic consequences for the analysis of the experiment. In other words, even if factor levels are hard to change, it is worthwhile to increase the number of whole plots.

Several authors have shown that correlation can be beneficial for the efficiency of a response surface design. In addition, computational results suggest that increasing the number of whole plots may further increase the efficiency of bi-randomization designs. In the next section, we present an algorithm to construct the bi-randomization designs we have discussed in this section.

4   Design construction

The construction of $\mathcal{D}$-optimal BRDs is more complicated than the computation of $\mathcal{D}$-optimal CRDs because design efficiency depends on the compound symmetric error structure. In this section, we will explain why this is so. We also describe a generic point exchange algorithm for optimal BRDs.

Designing a completely randomized experiment only consists of determining the design points of the experiments. Numerous algorithms have been developed for this design problem. The most famous construction algorithms for response surface designs may be classified as point exchange algorithms and include the algorithm of Fedorov (1972), the DETMAX algorithm of Mitchell (1974) and the BLKL algorithm of Atkinson and Donev (1992). Designing a bi-randomization experiment is complicated by the fact that exchanging design points affects the design efficiency in two ways. Like in the CRD case, exchanging design points modifies the design matrix $\mathbf{A}$. In addition, exchanging design points of a BRD changes $\mathbf{V} = \mathrm{Cov}(\mathbf{Y})$. As in

Goos and Vandebroek (1999), the algorithm presented here autonomously computes the optimal whole plot sizes and gives the optimal allocation of sub-plot levels to each whole plot for a given number of observations n and degree of correlation d. In addition, we now allow more than one whole plot per whole plot factor level combination. For this reason, the number of whole plots per whole plot factor level combination needs to be determined as well. In our algorithm, we allow the possibility to restrict the number of whole plots. This is particularly important when the whole plot variables are hard-to-change. Another reason to put a constraint on the number of whole plots is to leave enough degrees of freedom for estimating the sub-plot error. It should be pointed out that algorithms for blocking response surface, such as the BLKL algorithm and the algorithm of Cook and Nachtsheim (1989), are inappropriate for generating BRDs because they were developed for a totally different design problem. Not only do they assume uncorrelated observations, but they also require specification of the number of blocks and of the block size. Moreover, the composition of the blocks does not depend on the factor levels, as is the case here.

The input to the BRD construction algorithm consists of the desired number of observations n, the number of tries, the order of the model, the number of model parameters p, the number of explanatory variables k and the structure of their polynomial expansion $\mathbf{f}$. The whole plot and sub-plot factors need to be identified and an estimate of the variance ratio d must be provided as well. Since the computed designs are optimal for a range of d-values, a reasonable guess is satisfactory for the purpose of design construction. In most industrial environments, information on d is available from prior experiments. Depending on how they were analyzed, prior bi-randomization experiments may directly or indirectly contain valuable information on the variance ratio in a specific experimental setting. If they were properly analyzed as a BRD, prior guesses for d are obtained by $\hat{\sigma}_\delta^2/\hat{\sigma}_\varepsilon^2$, where $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_\delta^2$ are the estimates of the sub-plot and whole plot error variance respectively. If they were improperly analyzed as a CRD, the data from the experiments can be recovered in order to analyze them properly and thereby obtain estimates of σ_ε^2 and σ_δ^2. Bisgaard and Steinberg (1997) point out that the whole plot error variance is usually larger than the sub-plot error variance in prototype experiments. Letsinger et al. (1996) describe an experiment from the chemical industry and obtain $d = 1.04$.

A more detailed description of the design construction algorithm is given in the Appendix. A Fortran 77 implementation of the algorithm can be obtained from the authors. In the next section, computational results demonstrate the benefits of using BRDs instead of CRDs.

5   Computational results

In this section, the computational results are discussed. A small example illustrates the features of the optimal BRDs and the substantial gains that can be achieved by conducting bi-randomization experiments instead of completely randomized experi-

ments. The effect of limiting the number of whole plots for economical or practical reasons is investigated as well. A factorial experiment shows that large efficiency gains can be realized for pure linear models, linear models with two-factor interactions and quadratic models.

Consider a full quadratic model in one whole plot variable z and one sub-plot variable x. For $0 \leq d \leq 0.7011$, $0.7011 \leq d \leq 0.9113$ and $d \geq 0.9113$, the 10-point $\mathcal{D}$-optimal BRDs obtained by our algorithm are given by

$$
\begin{bmatrix}
-1 & -1 \\
-1 & +1 \\
\cdots\cdots \\
-1 & 0 \\
\cdots\cdots \\
-1 & +1 \\
\cdots\cdots \\
0 & -1 \\
\cdots\cdots \\
0 & 0 \\
\cdots\cdots \\
0 & +1 \\
\cdots\cdots \\
+1 & 0 \\
\cdots\cdots \\
+1 & -1 \\
+1 & +1
\end{bmatrix},
\quad
\begin{bmatrix}
-1 & -1 \\
-1 & +1 \\
\cdots\cdots \\
-1 & 0 \\
-1 & +1 \\
\cdots\cdots \\
0 & -1 \\
\cdots\cdots \\
0 & 0 \\
\cdots\cdots \\
0 & +1 \\
\cdots\cdots \\
+1 & 0 \\
\cdots\cdots \\
+1 & -1 \\
+1 & +1
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
-1 & -1 \\
-1 & +1 \\
\cdots\cdots \\
-1 & 0 \\
\cdots\cdots \\
0 & 0 \\
\cdots\cdots \\
0 & +1 \\
\cdots\cdots \\
+1 & -1 \\
+1 & +1 \\
\cdots\cdots \\
+1 & -1 \\
+1 & 0 \\
+1 & +1
\end{bmatrix}
\tag{8}
$$

respectively. We will refer to these three designs as BRD1, BRD2 and BRD3 respectively. These designs have a couple of striking features that hold generally. Firstly, the optimal number of whole plots decreases as the correlation increases. The 10-point designs shown in (8) have eight, seven and six whole plots. Apparently, the higher the correlation between observations within the same whole plot, the better it is to group more experimental runs and thereby induce more correlated observations. Otherwise, the lower the correlation, the more the optimal design will resemble a completely randomized experiment. Secondly, observations at the zero levels of the whole plot variables are assigned to separate whole plots, such that whole plots containing more than one observation only occur at $\mathbf{z} = \pm\mathbf{1}$. Thirdly, the designs are neither crossed nor balanced.

In order to obtain the optimal BRDs in (8), no constraint was imposed on the number of whole plots. However, limited resources often impose such constraint. Suppose for instance that no more than three whole plots are allowed. In this case, the $\mathcal{D}$-optimal BRDs are given in Figure 3. We will refer to these alternative design options as BRD4 and BRD5 respectively.
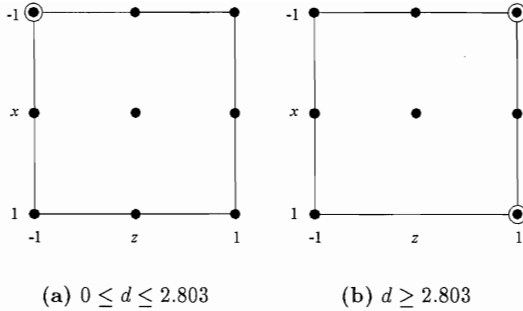
12

(a) $0 \le d \le 2.803$ **(b)** $d \ge 2.803$

Figure 3: $\mathcal{D}$-optimal 10-point BRDs with 3 whole plots for the full quadratic model in one whole plot variable z and one sub-plot variable x. • is a design point, ⊙ is a design point replicated twice.

In order to visualize the superiority of bi-randomization experiments to completely randomized experiments and to investigate the effect of limiting the number of whole plots, we have computed the relative efficiencies

$$\frac{|\mathbf{A}'\mathbf{V}^{-1}\mathbf{A}|}{|\mathbf{C}'\mathbf{C}|} \tag{9}$$

for d-values between 0 and 3, i.e. from zero correlation to largely correlated observations, holding $\sigma_\varepsilon^2 + \sigma_\delta^2 = 1$. The numerator of (9) gives the $\mathcal{D}$-criterion value of the BRD under investigation. The denominator gives the $\mathcal{D}$-criterion value of the $\mathcal{D}$-optimal CRD when $\sigma_\varepsilon^2 = 1$. The design points of the $\mathcal{D}$-optimal CRD are the same as those of BRD1, BRD2 and BRD4. The relative efficiencies of BRD1, BRD2, BRD3, BRD4 and BRD5 are displayed in Figure 4. BRD1 and BRD2 are superior to the CRD for variance ratios between 0 and 3. BRD3 is better than the CRD when $d > 0.323$. It also turns out that restricting the number of whole plots has a substantial negative impact on design efficiency. Nevertheless, both BRD4 and BRD5 outperform the CRD as soon as $d \ge 1.3$.

We have performed a factorial experiment to investigate the role of several model characteristics on the properties of the $\mathcal{D}$-optimal designs. We have computed designs for a pure linear model, a linear model with two-factor interactions and for a full quadratic model in three variables for six different degrees of correlation ($d = 0.1, 0.25, 0.5, 0.75, 1, 2$). For each combination, we have computed a nearly saturated design, a design with twice as much observations and one in between. Also, we have investigated the effect of the number of whole plot variables. The results are shown in Tables 2 and 3. The first column of the tables shows the degree of correlation as measured by d. For each model under investigation, the number of design points n, the number of whole plots w and the relative efficiency of the BRDs with respect to the $\mathcal{D}$-optimal CRD are displayed. The factorial experiment confirms the main results of the 10-point example in (8) and provides additional insights.
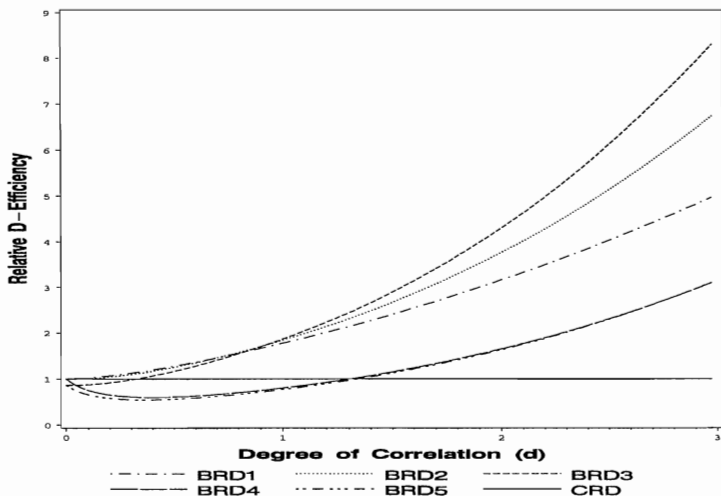
13

Figure 4: Comparison of the $\mathcal{D}$-efficiency of the 10-point BRDs in (8) and Figure 3 to the $\mathcal{D}$-optimal CRD for the full quadratic model in one whole plot and one sub-plot variable for various degrees of correlation.

Table 2 shows that efficiency gains for a pure linear model lie between 2% and 236% and between 14% and 12839% for a full quadratic model when there is one whole plot variable. It turns out that the higher the correlation, the larger the efficiency gain of using a BRD instead of a CRD. Also, the more complex the model, the larger the improvement in efficiency. For all models, the number of whole plots in the optimal designs decreases as the correlation increases. In other words, the higher the degree of correlation, the more observations will tend to be grouped in the same whole plot. Table 3 shows that efficiency gains in the presence of two whole plot factors are smaller. This is because, for a given number of observations, there are more whole plot factor levels when there are two whole plot variables instead of one. The larger number of whole plot factor levels reduces the possibility to group observations and to benefit from the correlation. It turns out that the CRD is the optimal design option in many cases when there are two whole plot variables. In Table 3, these are the cases in which the number of whole plots w equals the number of observations n. However, the number of whole plots in the optimal design becomes smaller than the number of observations as the model moves from pure linear to quadratic and as the degree of correlation increases. The results also indicate that observations at the zero levels of the whole plot variables are seldomly grouped in a whole plot. In holding these observations statistically independent, the variance of the parameters corresponding to the quadratic whole plot factor effects is kept low. In general, the optimal designs are not crossed. For the pure linear model and linear model with two-way interactions the optimal BRDs may be balanced, especially as n is a power of 2. A striking result is that the 2^3 factorial design is not necessarily

14

d	Pure Linear			Linear + interactions			Quadratic		
	n	w	rel eff	n	w	rel eff	n	w	rel eff
0.1	5	4	1.03	8	4	1.13	12	6	1.16
0.25		4	1.09		4	1.41		6	1.59
0.5		3	1.27		4	2.14		6	2.92
0.75		3	1.50		3	3.52		5	5.89
1		3	1.78		3	5.69		5	11.54
2		3	3.24		3	29.16		5	113.52
0.1	8	4	1.02	12	6	1.13	18	11	1.18
0.25		4	1.09		5	1.44		10	1.63
0.5		4	1.27		5	2.37		8	3.14
0.75		4	1.50		4	4.02		8	6.39
1		4	1.78		4	6.64		8	12.58
2		4	3.24		4	35.64		7	129.39
0.1	10	7	1.04	16	7	1.13	24	14	1.14
0.25		7	1.13		7	1.41		12	1.54
0.5		7	1.31		7	2.22		11	3.04
0.75		6	1.56		6	3.72		10	6.17
1		6	1.85		6	6.07		10	12.4
2		6	3.36		6	31.75		10	127.41

Table 2: Properties of $\mathcal{D}$-optimal designs for a pure linear model, a linear model with interactions and a full quadratic model in one whole plot variable and two sub-plot variables.

d	Pure Linear			Linear + interactions			Quadratic		
	n	w	rel eff	n	w	rel eff	n	w	rel eff
0.1	5	5	1.00	8	8	1.00	12	11	1.01
0.25		5	1.00		8	1.00		11	1.04
0.5		5	1.00		4	1.07		8	1.17
0.75		5	1.00		4	1.29		8	1.44
1		5	1.00		4	1.58		8	1.80
2		4	1.20		4	3.50		7	4.37
0.1	8	8	1.00	12	12	1.00	18	18	1.00
0.25		8	1.00		10	1.02		14	1.03
0.5		8	1.00		8	1.19		14	1.24
0.75		8	1.00		8	1.47		14	1.56
1		8	1.00		8	1.84		14	1.96
2		7	1.05		8	4.28		13	4.64
0.1	10	10	1.00	16	16	1.00	24	24	1.00
0.25		10	1.00		13	1.01		20	1.03
0.5		10	1.00		12	1.14		20	1.20
0.75		10	1.00		12	1.36		18	1.44
1		9	1.00		11	1.63		18	1.78
2		8	1.15		9	3.50		15	4.15

Table 3: Properties of $\mathcal{D}$-optimal designs for a pure linear model, a linear model with interactions and a full quadratic model in two whole plot variables and one sub-plot variable.

the best design option in these cases. For example, the 8-point $\mathcal{D}$-optimal BRD for the pure linear model in one whole plot variable and two sub-plot variables is given by two replicates of a 2^{3-1} fractional factorial design and not by the 2^3 factorial design. On the contrary, the optimal 8-point BRD for a linear model with interactions has the points of the 2^3 full factorial experiment as its design points. Whereas the eight observations for the pure linear model are allocated to four whole plots of two observations each no matter what d is specified, the allocation of observations to the whole plots depends on the degree of correlation and makes a substantial difference in design efficiency when interactions are included in the model. For small degrees of correlation, the eight observations are divided in four whole plots of two observations. However, for $d > 0.5$ only three whole plots are used: one with four observations and two with two observations. Finally, all other things being equal, efficiency gains do not depend on the number of observations n.

In this section, we have shown that the CRD is often outperformed by properly designed BRDs. The benefits of using BRDs increase as the degree of correlation and as the model complexity increases. The largest efficiency improvements can be realized if there is only one whole plot variable.

6   Conclusion

Bi-randomization designs are heavily used in industry, especially when hard-to-change factors are present. Key property of bi-randomization designs is that the levels of the hard-to-change factors —also referred to as whole plot factors— are changed as little as possible. In doing so, the ease of experimenting is increased and time and costs are saved. In this paper, it is shown that bi-randomization designs may be a good alternative to a completely randomized design if no hard-to-change factors are present. In addition, it is argued that (slightly) increasing the number of whole plots is useful when some of the factors are hard-to-change. Firstly, statistical efficiency is significantly improved. It turns out that in most design problems, completely randomized experiments are inferior to properly designed bi-randomization experiments. Secondly, the experimenter is better protected against systematic errors when the factor levels are reset more often. Thirdly, more degrees of freedom are available for whole plot error estimation. Fourthly, even when the number of level changes is increased, bi-randomization designs are still easier to conduct than completely randomized designs. These benefits were illustrated by means of a simple example. A factorial experiment was carried out to investigate the impact of the degree of correlation, the number of observations and the model on the features of the $\mathcal{D}$-optimal bi-randomization designs and on their efficiency relative to a completely randomized experiment. It turns out that the number of correlated observations in the optimal bi-randomization design increases as the degree of correlation increases and that the higher the correlation, the larger the efficiency gain of using a BRD instead of a CRD. Also, the more complex the model, the larger the improvement in efficiency.

Acknowledgement

Appendix

Our exchange algorithm starts with the generation of a non-singular n-point starting design and then adds and deletes design points in order to achieve increases in $|\mathbf{A}'\mathbf{V}^{-1}\mathbf{A}|$. Design points are chosen from a predefined set of candidate points which covers the entire region of interest χ. In order to avoid being stuck in a local optimum, the search is repeated for a number of different starting designs. Each repetition is called a try. The generation of starting designs includes a random component after which the starting design is completed by sequentially adding the candidate points with the highest prediction variance. Taking into account a possible restriction on the number of whole plots, the design points are randomly divided in a number of whole plots. The singularity when $n < p$ is overcome by using $\mathbf{A}'\mathbf{V}^{-1}\mathbf{A} + r\mathbf{I}$ instead of $\mathbf{A}'\mathbf{V}^{-1}\mathbf{A}$ with r a small positive number.

In order to speed up the algorithm, powerful update procedures were used. From (3) we have that

$$\mathbf{V}_i^{-1} = \frac{1}{\sigma_\varepsilon^2}(\mathbf{I}_{s_i \times s_i} - \frac{d}{1 + s_i d}\mathbf{1}_{s_i}\mathbf{1}'_{s_i}). \tag{A1}$$

Since

$$\mathbf{A}'\mathbf{V}^{-1}\mathbf{A} = \sum_{i=1}^{w} \mathbf{A}'_i\mathbf{V}_i^{-1}\mathbf{A}_i, \tag{A2}$$

we can write the information matrix of a BRD as

$$\mathbf{A}'\mathbf{V}^{-1}\mathbf{A} = \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{w} \mathbf{A}'_i(\mathbf{I}_{s_i \times s_i} - \frac{d}{1 + s_i d}\mathbf{1}_{s_i}\mathbf{1}'_{s_i})\mathbf{A}_i,$$

$$= \frac{1}{\sigma_\varepsilon^2}\left\{\sum_{i=1}^{w}\sum_{j=1}^{s_i} \mathbf{f}(\mathbf{z}_i, \mathbf{x}_{ij})\mathbf{f}'(\mathbf{z}_i, \mathbf{x}_{ij}) - \sum_{i=1}^{w} \frac{d}{1 + s_i d}(\mathbf{A}'_i\mathbf{1}_{s_i})(\mathbf{A}'_i\mathbf{1}_{s_i})'\right\}. \tag{A3}$$

Expression (A3) was used to update the information matrix after design changes and to some extent reduces the computational burden. As an illustration, consider addition of an observation to the ith whole plot. Let $\mathbf{A}^*$, $\mathbf{V}^*$ and $\mathbf{A}_i^{*\prime}$ denote the new design matrix, the corresponding variance-covariance matrix and the part of the new design matrix corresponding to the ith whole plot respectively. The relationship

between the information matrix before and after adding the observation is given by

$$
\begin{aligned}
\mathbf{A}^{*\prime}\mathbf{V}^{*-1}\mathbf{A}^{*\prime} = {}& \mathbf{A}'\mathbf{V}^{-1}\mathbf{A} \\
& + \mathbf{f}(\mathbf{z}_i, \mathbf{x}_{(i,s_i+1)})\mathbf{f}'(\mathbf{z}_i, \mathbf{x}_{(i,s_i+1)}) \\
& + \frac{d}{1 + s_i d}(\mathbf{A}'_i \mathbf{1}_{s_i})(\mathbf{A}'_i \mathbf{1}_{s_i})' \\
& - \frac{d}{1 + (s_i + 1)d}(\mathbf{A}^{*\prime}_i \mathbf{1}_{(s_i+1)})(\mathbf{A}^{*\prime}_i \mathbf{1}_{(s_i+1)})'.
\end{aligned}
\tag{A4}
$$

Adding an observation in a new whole plot is a special case of (A4) where $s_i = 0$, $\mathbf{A}_i = \mathbf{0}$ and $\mathbf{A}^*_i = \mathbf{f}'(\mathbf{z}_i, \mathbf{x}_{i1})$. The adapted information matrix can then be calculated using

$$
\mathbf{A}^{*\prime}\mathbf{V}^{*-1}\mathbf{A}^{*\prime} = \mathbf{A}'\mathbf{V}^{-1}\mathbf{A} + \frac{1}{1 + d}\mathbf{f}(\mathbf{z}_{w+1}, \mathbf{x}_{(w+1,1)})\mathbf{f}'(\mathbf{z}_{w+1}, \mathbf{x}_{(w+1,1)}).
\tag{A5}
$$

These expressions show that the information matrix can be updated by adding and subtracting outer products. Using the well-known matrix results that for any positive definite matrix $\mathbf{M}$ and vector $\mathbf{u}$,

$$
|\mathbf{M} + \mathbf{u}\mathbf{u}'| = |\mathbf{M}|(1 + \mathbf{u}'\mathbf{M}^{-1}\mathbf{u}),
\tag{A6}
$$

and

$$
\mathbf{M}^{-1} = \mathbf{M}^{-1} - \frac{(\mathbf{M}^{-1}\mathbf{u})(\mathbf{u}'\mathbf{M}^{-1})}{1 + \mathbf{u}'\mathbf{M}^{-1}\mathbf{u}},
\tag{A7}
$$

the inverse and the determinant of the information matrix can be updated at a low computational cost after adding, deleting or exchanging design points.

By default, the design region is taken as $\chi = [-1, +1]^k$ and the candidate points are chosen from the $2^k, 3^k, \ldots$ factorial design depending on whether the model contains linear, quadratic or higher order terms. Alternatively, the grid of candidate points can be specified by the user. Note that construction of a non-singular design requires $n \geq p$. Finally, the user may impose a restriction on the number of whole plots in the experiment. The algorithm was implemented in Fortran 77 and can be obtained from the authors.

References

Anbari, F. and Lucas, J. (1994). Super-efficient designs: How to run your experiment for higher efficiency and lower cost, *ASQC Technical Conference Transactions*, pp. 852–863.

18

Anderson, V. and McLean, R. (1974). *Design of Experiments: A Realistic Approach*, New York: Marcel Dekker.

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

Bisgaard, S. and Steinberg, D. (1997). The design and analysis of $2^{k-p} \times s$ prototype experiments, *Technometrics* **39**: 52–62.

Box, G. and Jones, S. (1992). Split-plot designs for robust product experimentation, *Applied Statistics* **19**: 3–26.

Cook, R. and Nachtsheim, C. (1989). Computer-aided blocking of factorial and response-surface designs, *Technometrics* **31**: 339–346.

Davison, J. (1995). *Response Surface Designs and Analysis for Bi-Randomization Error Structures*, Ph.d. thesis, Virginia Polytechnic Institute and State University.

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

Ganju, J. and Lucas, J. (1997). Bias in test statistics when restrictions in randomization are caused by factors, *Communications in Statistics: Theory and Methods* **26**: 47–63.

Ganju, J. and Lucas, J. (1999). Detecting randomization restrictions caused by factors, *Journal of Statistical Planning and Inference* **81**: 129–140.

Goos, P. and Vandebroek, M. (1999). Optimal bi-randomization designs. Submitted for publication.

Kempthorne, O. (1952). *The Design and Analysis of Experiments*, New York: Wiley.

Letsinger, J., Myers, R. and Lentner, M. (1996). Response surface methods for bi-randomization structures, *Journal of Quality Technology* **28**: 381–397.

Mitchell, T. (1974). An algorithm for the construction of D-optimal experimental designs, *Technometrics* **16**: 203–210.

Nelson, L. (1985). What do low F ratios tell you?, *Journal of Quality Technology* **17**: 237–238.