An Algorithm for Finding D-efficient Equivalent-Estimation Second-Order Split-Plot Designs

Bradley Jones[1,2]
Peter Goos[2,3]

[1]SAS Institute, SAS Campus Drive, Cary, NC 27513, United States of America.
[2]Faculty of Applied Economics & StatUa Center for Statistics, Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerpen, Belgium.
[3]Erasmus School of Economics, Erasmus Universiteit Rotterdam, Postbus 1738, 3000 DR Rotterdam, Netherlands.

August 12, 2011

Abstract

Many industrial experiments involve restricted rather than complete randomization. This often leads to the use of split-plot designs, which limit the number of independent settings of some of the experimental factors. These factors, named whole-plot factors, are often, in some way, hard to change. The remaining factors, called sub-plot factors, are easier to change. Their levels are therefore independently reset for every run of the experiment. In general, model estimation from data from split-plot experiments requires the use of generalized least squares (GLS). However, for some split-plot designs, the ordinary least squares (OLS) estimator will produce the same factor-effect estimates as the GLS estimator. These designs are called equivalent-estimation split-plot designs and offer the advantage that estimation of the factor effects does not require estimation of the variance components in the split-plot model. While many of the equivalent-estimation second-order response surface designs presented in the literature do not perform well in terms of estimation efficiency (as measured by the D-optimality criterion), Macharia and Goos (2010) showed that, in many instances, it is possible to generate second-order equivalent-estimation split-plot designs that are highly efficient and, hence, provide precise factor-effect estimates. In the present paper, we present an algorithm that allow us to (i) identify equivalent-estimation designs for scenarios where Macharia and Goos (2010) did not find equivalent-estimation designs, and (ii) find equivalent-estimation designs that outperform those of Macharia and Goos (2010) in terms of the D-optimality criterion.

Key Words: coordinate-exchange algorithm, D-optimality, equivalent estimation, generalized least squares, ordinary least squares, split-plot design.

1 Introduction

It is often impossible to conduct industrial experiments using a completely randomized design, which requires that the levels of the experimental factors are set independently for every run. A commonly used alternative to the completely randomized design is a split-plot design, in which only the levels of a subset of experimental factors are reset independently for every run. The levels of the remaining factors are set only a limited number of times to save time or costs. The former factors are usually named the sub-plot factors, while the latter factors are referred to as whole-plot factors. Oftentimes, the whole-plot factors' levels are, in some sense, hard to change. Some of the typical scenarios in which split-plot designs are used are listed in Goos and Jones (2011):

1. When several combinations of some experimental factors' levels are tested while some other factors' levels are not reset, the resulting design is a split-plot design. This happens, for example, when experimenters try out different combinations during one run of an oven. The oven temperature is not reset for all these observations.

2. In two-stage experiments, some of the factors may be applied in the first stage, whereas others are applied in the second stage. In those cases, the first stage of the experiment involves the production of various batches of experimental material using different combinations of levels of the first-stage experimental factors. These batches are then split in sub-batches after the first stage, each of which undergoes a different treatment (dictated by the levels of the second-stage factors) at each run in the second stage.

3. Experiments for simultaneously investigating the impact of ingredients of a mixture and of process factors, are frequently conducted using a split-plot design. This can happen in two ways. First, several batches can be prepared that undergo different process conditions. Second, processing conditions can be held fixed while consecutively trying out different mixtures.

4. When runs of an experimental design are performed one after another, experimenters are often reluctant to change the levels of one or more (hard-to-change) factors because this is impractical, time-consuming or expensive.

5. A special case of an experiment involving hard-to-change factors is a prototype experiment, in which the levels of some of the factors define a prototype, whereas the levels of other factors determine operating conditions under which the prototype is tested. As changing the levels of a prototype factor involves assembling a new prototype, these levels are held constant for several observations under different operating conditions.

6. Robust product experiments often involve crossed arrays, in which case they are frequently run as split-plot designs. This is because the environmental factors (also named noise factors) are usually hard to control and changing their levels is cumbersome.

The design and analysis of split-plot industrial experiments has received considerable attention in the literature in recent years. Letsinger, Myers and Lentner (1996) discussed response surface methods for split-plot designs focusing on the analysis of these designs. They recommended the use of generalized least squares (GLS) and restricted maximum likelihood (REML) for estimating split-plot response surface models. Gilmour and Goos (2009) propose an alternative, Bayesian strategy for split-plot data analysis, which is especially useful when the number of independent settings of the whole-plot factors is small.

Huang, Chen and Voelkel (1998), Bingham and Sitter (1999) and Bingham, Schoen and Sitter (2004) described the construction of two-level fractional factorial split-plot designs using the aberration criterion. Multistratum response surface designs, of which split-plot designs are special cases, are discussed in Trinca and Gilmour (2001). They present a sequential method for constructing these designs, from stratum to stratum and starting in the highest stratum. Kulahci and Bisgaard (2005) illustrated how split-plot designs can be constructed from Plackett-Burman designs. Goos and Vandebroek (2001, 2003, 2004) and Jones and Goos (2007) propose exchange algorithms for constructing D-optimal split-plot designs. Follow-up split-plot designs are discussed by Almimi, Kulahci and Montgomery (2008) and McLeod and Brewster (2008). A review of the recent developments on the design of split-plot experiments can be found in Jones and Nachtsheim (2009).

A subject that has received substantial attention in the recent split-plot design literature is the equivalence of the ordinary least squares (OLS) and GLS estimators, following the work of Vining, Kowalski and Montgomery (2005). This is because split-plot designs for which OLS and GLS produce the same factor-effect estimates offer the advantage that the estimates of the factor effects do not depend on the estimates of the variance components in the split-plot model. Vining, Kowalski and Montgomery (2005) discussed split-plot arrangements of central composite and Box-Behnken designs for which the OLS and GLS estimators of the model parameters are equivalent and outlined some general conditions for this property to be fulfilled when central composite or Box-Behnken designs are used. These types of designs are nowadays called equivalent-estimation split-plot designs. Parker, Kowalski and Vining (2006, 2007a,b) describe some follow-up work on the construction of equivalent-estimation designs from central composite and Box-Behnken designs.

Goos (2006) compared the efficiency of D-optimal split-plot designs with that of equivalent-estimation designs and reported various instances where the equivalent-estimation designs proposed in the literature were highly inefficient. At the same time, he discovered various D-optimal designs for which OLS and GLS are equivalent. These results inspired Macharia and Goos (2010) to carry out a systematic study of the relationship between D-optimality and the equivalent-estimation property. They discovered many scenarios for which D-optimal designs exist for which OLS and GLS produce the same estimates and for which highly D-efficient equivalent-estimation designs can be found.

In this paper, we present two new algorithms to find D-optimal and D-efficient equivalent-estimation designs. Using these algorithms, we are able to fill the gaps that were left by Macharia and Goos (2010). First, we were able to identify equivalent-estimation designs for cases where

Macharia and Goos (2010) failed to find such designs. Second, in several scenarios, we obtained better equivalent-estimation designs than did Macharia and Goos (2010).

In the next sections, we introduce the split-plot model, specify the condition for the equivalence of the OLS and GLS estimator, and define the D-optimality criterion. Next, we describe our algorithm for generating equivalent-estimation designs and discuss several interesting designs produced by them. Finally, we list all the scenarios for which we found D-efficient or D-optimal equivalent-estimation designs.

2 The Split-Plot Design Model

A key feature of split-plot designs is that their runs appear in groups named whole plots. Within every group, the levels of the whole-plot factors do not change. Every whole plot involves a certain number of runs, often named sub-plots, for each of which the levels of the sub-plot factors are reset independently. The general split-plot design model for an experiment with n runs, b whole plots and $k = n/b$ runs or subplots per whole plot is given by

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of responses, $\mathbf{X}$ is an $n \times p$ model matrix containing the settings of the whole-plot factors, the sub-plot factors and their model expansions, $\boldsymbol{\beta}$ is a p-dimensional vector containing the p fixed effects in the model and $\mathbf{Z}$ is an $n \times b$ matrix of zeros and ones assigning the n runs to the b whole plots (i.e. the $(i,j)^{th}$ element of $\mathbf{Z}$ is one if the i^{th} run belongs to the j^{th} whole plot and zero otherwise). The b-dimensional vector $\boldsymbol{\gamma}$ contains the random effects of the b whole plots. Finally, $\boldsymbol{\epsilon}$ is the n-dimensional vector of the random errors. The random vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\epsilon}$ are assumed to be uncorrelated, with mean zero and covariance matrix $\sigma_\gamma^2\mathbf{I}_b$ and $\sigma_\epsilon^2\mathbf{I}_n$ respectively, where $\mathbf{I}_b$ and $\mathbf{I}_n$ are identity matrices of size b and n. As a result, the assumed covariance matrix for the response vector $\boldsymbol{Y}$ is

$$\mathbf{V} = \sigma_\epsilon^2\mathbf{I}_n + \sigma_\gamma^2\mathbf{Z}\mathbf{Z}'. \tag{2}$$

If the entries of the response vector $\mathbf{y}$ and the model matrix $\mathbf{X}$ are arranged per whole plot, then this covariance matrix can be written as

$$\mathbf{V} = \sigma_\epsilon^2\mathbf{I}_n + \sigma_\gamma^2\mathbf{D}, \tag{3}$$

where

$$\mathbf{D} = \begin{bmatrix} \mathbf{1}_k\mathbf{1}_k' & \mathbf{0}_k & \dots & \mathbf{0}_k \\ \mathbf{0}_k & \mathbf{1}_k\mathbf{1}_k' & \dots & \mathbf{0}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_k & \mathbf{0}_k & \dots & \mathbf{1}_k\mathbf{1}_k' \end{bmatrix}.$$

Here, $\mathbf{1}_k$ is an k-dimensional vector of ones and $\mathbf{0}_k$ is an $k \times k$ zero matrix. The block-diagonal nature of the covariance matrix $\mathbf{V}$ means that responses from the same whole plot are correlated, while those from different whole plots are not.

The GLS estimator of the factor effects is

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{Y}.$$

This estimator has the covariance matrix

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}_{GLS}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}.$$

The OLS estimator is given by

$$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{Y}.$$

For equivalent-estimation split-plot designs, by definition, the OLS and GLS estimators given above are the same, i.e.

$$\hat{\boldsymbol{\beta}}_{OLS} = \hat{\boldsymbol{\beta}}_{GLS}.$$

The equivalence of OLS and GLS is appealing because it implies that the $\mathbf{V}$ matrix and hence the variances σ_γ^2 and σ_ϵ^2 need not be estimated in order to estimate the factor effects. This is especially important for researchers who do not have access to software that allows REML estimation of variance components. However, it is worth noting that knowledge or estimation of the variance components remains essential for statistical inference of the estimated model. The analysis of data from split-plot response surface designs in general is discussed in detail in Goos, Langhans and Vandebroek (2006) and Goos and Jones (2011). Exact inference procedures for data from a specific class of equivalent-estimation designs are discussed in Vining and Kowalski (2008), whereas a check for split-plot model adequacy is proposed by Almimi, Kulahci and Montgomery (2009).

3 Split-Plot Design Construction Strategies

The initial work on the design of split-plot response surface experiments by Goos and Vandebroek (2001, 2003, 2004) was based on optimal experimental design ideas. More specifically, the focus was on creating D-optimal split-plot designs. The D-optimality criterion seeks to minimize the generalized variance of the parameter estimates, which is done by minimizing the determinant of the variance-covariance matrix of the factor effects' estimates or, equivalently, by maximizing the determinant of the information matrix about $\boldsymbol{\beta}$. For a split-plot design, the information matrix is given by

$$\mathbf{M} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \tag{4}$$

when the GLS estimator is used.

The D-efficiency of a design is obtained by comparing the determinant of its information matrix with that of the corresponding D-optimal design. Letting $\mathbf{M}_{\text{opt}}$ be the information matrix of the D-optimal design with design matrix $\mathbf{X}_{\text{opt}}$ and $\mathbf{M}$ be the information matrix of a design with design matrix $\mathbf{X}$ for the same design problem, the relative D-efficiency of the design corresponding to $\mathbf{X}$ is defined as

$$D_{\text{eff}} = \left\{ \frac{|\mathbf{M}|}{|\mathbf{M}_{\text{opt}}|} \right\}^{1/p} = \left\{ \frac{|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|}{|\mathbf{X}'_{\text{opt}}\mathbf{V}^{-1}\mathbf{X}_{\text{opt}}|} \right\}^{1/p}, \tag{5}$$

where p is the number of parameters in the model.

Goos and Vandebroek (2001, 2003, 2004) developed several point-exchange algorithms for constructing split-plot designs that require specification of a candidate set containing all the possible allowable combinations of the factor levels (i.e., all possible design points). The point-exchange algorithms start from an initial design, that is partly generated at random and completed by repeatedly adding the point that gives the largest increase in the D-criterion value. The algorithms proceed by exchanging design points from the initial design with points from the candidate set until the design's value for the D-optimality criterion cannot be improved any more. The construction of a candidate set can be problematic when the number of experimental factors is large and/or the experimental space is highly constrained. To avoid this problem, Jones and Goos (2007) described a flexible candidate-set-free coordinate-exchange algorithm for constructing D-optimal split-plot designs. The algorithm starts from an initial design generated randomly and then improves this design coordinate by coordinate until there are no more coordinate exchanges that lead to an increase in the D-optimality criterion. Most often, point-exchange algorithms and coordinate-exchange algorithms do not lead to qualitatively different designs. However, Schoen, Jones and Goos (2011) describe a case study in which a coordinate-exchange algorithm for constructing split-plot designs outperforms a point-exchange algorithm to a considerable extent.

While Goos (2006) and Parker, Kowalski and Vining (2007a) discussed a few instances where D-optimal split-plot design possess the equivalent-estimation property, Macharia and Goos (2010) provided a catalog of equivalent-estimation designs that were either D-optimal or highly D-efficient. They modified the coordinate-exchange algorithm of Jones and Goos (2007) so that it tests every intermediate design for the equivalent-estimation property, and came across many new equivalent-estimation designs.

The necessary and sufficient condition for equivalence of OLS and GLS estimates, as given by McElroy (1967), is the existence of a $p \times p$ nonsingular matrix, $\mathbf{F}$, such that

$$\mathbf{XF} = \mathbf{VX}. \tag{6}$$

Parker, Kowalski and Vining (2007a) give a general form of the equivalence condition tailored to split-plot designs. By substituting Equation (3) in Equation (6) and pre-multiplying by $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ they find that

$$\mathbf{F} = \sigma_\epsilon^2 \mathbf{I}_p + \sigma_\gamma^2 (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{DX}, \tag{7}$$

so that the condition for OLS-GLS equivalence in the case of split-plot designs becomes

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{DX} = \mathbf{DX}. \tag{8}$$

This condition can be rewritten as

$$(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{DX} = \mathbf{0}_{n\times p}, \tag{9}$$

where $\mathbf{0}_{n\times p}$ is an $n \times p$ matrix of zeros. This last expression means that a necessary and sufficient condition for a split-plot design to be an equivalent-estimation design is that the matrix $\mathbf{D}$ has to be in the column space of the design's model matrix $\mathbf{X}$.

To measure how close a design is to being an equivalent-estimation design, it is convenient to introduce a short hand notation for the matrix in Equation (9):

$$\mathbf{C} = (\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{D}\mathbf{X}.$$

To quantify how close this matrix is to a zero matrix, we calculate the sum of its squared elements as $c = \text{trace}(\mathbf{C}'\mathbf{C})$. If we want to stress that this quantity, which is zero for an equivalent-estimation design, is a function of the model matrix $\mathbf{X}$ of a design, we can write $c(\mathbf{X}) = \text{trace}\{\mathbf{C}'(\mathbf{X})\mathbf{C}(\mathbf{X})\}$. A measure of the extent to which the design is an equivalent-estimation design then is

$$E_{\text{eff}} = 1 - \frac{c(\mathbf{X})}{c(\mathbf{X}_{\text{opt}})} = 1 - \frac{\text{trace}\{\mathbf{C}'(\mathbf{X})\mathbf{C}(\mathbf{X})\}}{\text{trace}\{\mathbf{C}'(\mathbf{X}_{\text{opt}})\mathbf{C}(\mathbf{X}_{\text{opt}})\}}. \tag{10}$$

This measure, which we refer to as the equivalent-estimation efficiency, takes the value one for any design with model matrix $\mathbf{X}$ that satisfies the equivalent-estimation condition. Jones and Nachtsheim (2011) use a similarly structured efficiency measure to quantify the extent to which the alias matrix of a design is close to a zero matrix.

4 Constructing efficient equivalent-estimation designs

The equivalent-estimation designs reported by Macharia and Goos (2010) were all obtained as a mere by-product of the coordinate-exchange algorithm proposed by Jones and Goos (2007). As a result, nothing in their algorithm steered the design construction explicitly in the direction of equivalent-estimation designs. We took a different route and implemented an algorithm that has a dual objective function, i.e. a weighted optimality criterion that does not just focus on finding D-optimal designs but also equivalent-estimation designs.

The objective function we use is

$$\text{obj} = \lambda D_{\text{eff}} + (1 - \lambda)E_{\text{eff}}. \tag{11}$$

The first term in that function, the D-efficiency D_{eff} defined in Equation (5), receives a weight λ between 0 and 1. The second term in the objective function is the equivalent-estimation efficiency E_{eff} we defined in Equation (10), with weight $1 - \lambda$. A similar criterion was used by Jones and Nachtsheim (2011), who constructed efficient designs with minimal aliasing. In their approach, the objective was to obtain highly efficient designs for which the alias matrix was as close as possible to the zero matrix. To find the designs reported in this paper, we set $\lambda = 0.99$ in the objective function to indicate that we desired highly efficient designs. The weight $1 - \lambda = 0.01$ for the equivalent-estimation efficiency E_{eff} in the objective function was large enough to direct the search toward equivalent-estimation designs.

To maximize our dual objective function, we used the split-plot coordinate-exchange algorithm of Jones and Goos (2007). As our primary goal was to find three-level equivalent-estimation designs, we first utilized only -1, 0 and 1 as the possible levels for each of the factors in the algorithm. In some cases, however, we were unable to find three-level equivalent-estimation designs. In these cases, we let the sub-plot factor levels take any value between -1 and 1. To determine the optimal level of each sub-plot factor in each run of the design, we used Brent's minimization algorithm (Brent (1973)). The algorithm, also known as the van Wijngaarden-Deker-Brent method (Press et al. (2007)), is a one-dimensional minimization procedure that combines the golden section search algorithm with the inverse quadratic interpolation method. Convergence of the algorithm is guaranteed provided the given initial interval contains exactly one minimum. Brent's algorithm has been used in the optimal experimental design literature by Rodríguez, Jones, Borror and Montgomery (2010), who constructed G-optimal designs.

5 Illustrations

In this section, we report some of the most noticeable improvements of our results over those in Macharia and Goos (2010). We discuss two scenarios for which no equivalent-estimation design was found before, as well as two scenarios where we obtained an equivalent-estimation design that was substantially better in terms of the D-optimality criterion than the best known equivalent-estimation design.

A 21-run design involving three factors

Our algorithm enabled us to find a three-level equivalent-estimation design involving seven whole plots of three runs for estimating a second-order response surface model in one whole-plot factor w and two sub-plot factors s_1 and s_2. The design is displayed in Table 1. It has a D-efficiency of 89.8%. For this scenario, Macharia and Goos (2010) did not find an equivalent-estimation design.

Two 36-run designs involving three factors

Another case where our algorithm yielded better results than the approach of Macharia and Goos (2010) is where nine whole plots of four runs are used to estimate a second-order response surface model in two whole-plot factors w_1 and w_2 and one sub-plot factor s. The three-level equivalent-estimation design we obtained is shown in the left panel of Table 2. It has a D-efficiency of 98.7%. The right panel of Table 2 shows the equivalent-estimation design obtained by Macharia and Goos (2010). That design is only 93.5% D-efficient. So, in this scenario, the new algorithm yields an equivalent-estimation design that is 5.6% more D-efficient than the best known equivalent-estimation design.

Table 3 shows two other three-level 36-run equivalent-estimation split-plot designs for three experimental factors, each having six whole plots of six runs. Unlike in the previous case, the designs involve one whole-plot factor w and two sub-plot factors s_1 and s_2. The design in the left

Table 1: 21-run equivalent-estimation design involving 7 whole plots of 3 runs, one whole-plot factor w, and two sub-plot factors s_1 and s_2

Whole plot	w	s_1	s_2
1	1	−1	0
1	1	0	1
1	1	1	−1
2	1	1	−1
2	1	−1	0
2	1	0	1
3	−1	1	−1
3	−1	1	1
3	−1	−1	0
4	−1	1	1
4	−1	1	−1
4	−1	−1	0
5	1	1	0
5	1	0	−1
5	1	−1	1
6	0	−1	−1
6	0	0	0
6	0	−1	1
7	0	−1	−1
7	0	−1	1
7	0	0	0

panel of the table was obtained using our algorithm, whereas the design in the right panel was reported by Macharia and Goos (2010). The former design has a D-efficiency of 98.8%, whereas the latter design is only 89.0% D-efficient. Hence, the new design performs 11.0% better in terms of the D-optimality criterion.

A 60-run design involving four factors

In some instances, it is only possible to find equivalent-estimation designs by allowing the factor levels to differ from −1, 0 and 1. This is, for example, the case when the interest is in running a 60-run split-plot design in ten whole plots of six runs for studying the impact of two whole-plot factors w_1 and w_2 and two sub-plot factors s_1 and s_2. This is another instance for which Macharia and Goos (2010), who restricted themselves to three-level designs, could not find an equivalent-estimation design.

The best equivalent-estimation design we were able to find for this scenario involves three levels only for the whole-plot factors, but many more different levels for the sub-plot factors. This can

Table 2: 36-run equivalent-estimation designs involving 9 whole plots of 4 runs, two whole-plot factors w_1 and w_2, and one sub-plot factor s. The design in the left panel was obtained using the algorithm described in this paper, while the design in the right panel was found by Macharia and Goos (2010).

Whole plot	w_1	w_2	s	w_1	w_2	s
1	1	1	1	−1	0	−1
1	1	1	−1	−1	0	−1
1	1	1	−1	−1	0	1
1	1	1	1	−1	0	0
2	1	−1	−1	1	1	1
2	1	−1	1	1	1	−1
2	1	−1	−1	1	1	1
2	1	−1	1	1	1	−1
3	−1	1	1	1	−1	0
3	−1	1	−1	1	−1	1
3	−1	1	−1	1	−1	−1
3	−1	1	1	1	−1	−1
4	−1	−1	−1	−1	−1	1
4	−1	−1	−1	−1	−1	1
4	−1	−1	1	−1	−1	−1
4	−1	−1	1	−1	−1	−1
5	0	1	−1	1	−1	1
5	0	1	1	1	−1	−1
5	0	1	0	1	−1	0
5	0	1	0	1	−1	1
6	1	0	1	−1	1	1
6	1	0	0	−1	1	1
6	1	0	0	−1	1	−1
6	1	0	−1	−1	1	−1
7	−1	−1	1	1	0	−1
7	−1	−1	−1	1	0	−1
7	−1	−1	−1	1	0	1
7	−1	−1	1	1	0	0
8	0	−1	1	0	1	−1
8	0	−1	0	0	1	1
8	0	−1	−1	0	1	1
8	0	−1	0	0	1	0
9	−1	0	−1	0	−1	1
9	−1	0	0	0	−1	1
9	−1	0	1	0	−1	0
9	−1	0	0	0	−1	−1

Table 3: 36-run equivalent-estimation designs involving 6 whole plots of 6 runs, one whole-plot factor w, and two sub-plot factors s_1 and s_2. The design in the left panel was obtained using the algorithm described in this paper, while the design in the right panel was found by Macharia and Goos (2010).

Whole plot	w	s_1	s_2	w	s_1	s_2
1	1	0	−1	−1	−1	−1
1	1	1	1	−1	1	−1
1	1	1	−1	−1	1	1
1	1	−1	1	−1	−1	0
1	1	−1	0	−1	−1	1
1	1	1	1	−1	0	−1
2	−1	1	−1	1	1	−1
2	−1	0	−1	1	1	1
2	−1	1	1	1	−1	−1
2	−1	−1	1	1	1	0
2	−1	1	1	1	−1	1
2	−1	−1	0	1	0	1
3	−1	1	−1	−1	1	1
3	−1	−1	1	−1	−1	−1
3	−1	−1	−1	−1	1	−1
3	−1	1	1	−1	−1	1
3	−1	1	0	−1	0	1
3	−1	0	1	−1	−1	0
4	1	1	1	0	−1	1
4	1	−1	−1	0	−1	−1
4	1	1	−1	0	1	1
4	1	−1	1	0	0	0
4	1	0	1	0	1	−1
4	1	1	0	0	0	0
5	0	1	−1	1	1	1
5	0	−1	1	1	−1	−1
5	0	−1	−1	1	1	−1
5	0	1	0	1	0	−1
5	0	−1	−1	1	−1	1
5	0	0	0	1	−1	0
6	1	−1	−1	−1	−1	1
6	1	1	1	−1	−1	−1
6	1	1	−1	−1	1	1
6	1	−1	1	−1	1	−1
6	1	0	1	−1	0	−1
6	1	1	0	−1	1	0

be seen from Table 4, where the design is shown. Obviously, running the design requires that the sub-plot factors are highly controllable. The design has a D-efficiency of 84.2% only, which is due to the fact that the levels of the sub-plot factors are, in some cases, substantially different from -1, 0 and 1.

Table 4: 60-run equivalent-estimation design involving 10 whole plots of 6 runs, two whole-plot factors w_1 and w_2, and two sub-plot factors s_1 and s_2.

Whole plot	w_1	w_2	s_1	s_2	Whole plot	w_1	w_2	s_1	s_2
1	-1	0	-0.25288	-0.71040	6	0	0	0.97861	0.98608
1	-1	0	-0.72584	-0.79763	6	0	0	-0.11000	-0.15295
1	-1	0	0.83486	0.54092	6	0	0	-0.97978	0.98263
1	-1	0	0.36873	0.86236	6	0	0	-0.01820	0.02622
1	-1	0	-0.98195	0.90678	6	0	0	0.99953	-0.99723
1	-1	0	0.90698	-0.99326	6	0	0	-1.00000	-1.00000
2	-1	1	-0.31684	-0.38804	7	1	-1	0.99962	-0.96872
2	-1	1	-1.00000	-0.79542	7	1	-1	1.00000	-1.00000
2	-1	1	1.00000	0.29917	7	1	-1	-0.92972	-1.00000
2	-1	1	-1.00000	0.95800	7	1	-1	1.00000	1.00000
2	-1	1	1.00000	-0.98796	7	1	-1	-0.97702	1.00000
2	-1	1	-0.04849	1.00000	7	1	-1	-1.00000	0.01574
3	0	1	-0.80187	0.29070	8	1	1	-1.00000	-0.71426
3	0	1	0.93467	0.75794	8	1	1	-0.54625	0.90832
3	0	1	-0.93572	-0.99059	8	1	1	0.08232	0.99987
3	0	1	0.99053	-0.20564	8	1	1	0.48161	-0.90413
3	0	1	0.15517	-1.00000	8	1	1	0.98423	-0.94444
3	0	1	-0.99421	0.99095	8	1	1	-0.95217	-0.30111
4	1	-1	-0.92930	0.99724	9	-1	-1	-1.00000	-0.71224
4	1	-1	1.00000	-0.99715	9	-1	-1	-0.20007	0.94995
4	1	-1	0.99966	0.97250	9	-1	-1	0.85249	-0.95141
4	1	-1	-0.98391	0.07376	9	-1	-1	1.00000	-0.42612
4	1	-1	-0.99357	-0.99933	9	-1	-1	-1.00000	-0.88513
4	1	-1	1.00000	-1.00000	9	-1	-1	1.00000	1.00000
5	0	1	-0.73483	0.90590	10	0	0	0.96057	-0.11086
5	0	1	0.10336	-0.09903	10	0	0	-1.00000	0.97592
5	0	1	0.99975	-0.99837	10	0	0	0.99870	0.99732
5	0	1	-0.97198	0.12662	10	0	0	-1.00000	-0.01926
5	0	1	0.94820	0.90729	10	0	0	-0.09485	-0.99837
5	0	1	-0.99593	-0.99905	10	0	0	0.00575	-1.00000

6 A catalog of equivalent-estimation designs

One of our goals was to find three-level equivalent-estimation designs for the 111 scenarios investigated by Macharia and Goos (2010) with up to three whole-plot factors and three sub-plot factors. The maximum number of runs we considered was 72, while the largest number of whole plots was 12. The scenarios are all listed in Table 5, along with the D-efficiencies of the newly found equivalent-estimation designs (in the column labeled "New") and those of the designs reported by Macharia and Goos (2010) (in the column labeled "M & G"). We should note that two of the D-efficiencies reported in Macharia and Goos (2010) were off by 0.2%. These cases are indicated using the subscripts a and b in the last column of Table 5.

A remarkable result is that, in each of the 111 scenarios, we obtained an equivalent-estimation design. This is unlike Macharia and Goos (2010), who failed to find an equivalent-estimation design in 25 of the scenarios. As pointed out earlier, this is because these authors did not use an algorithm geared towards finding equivalent-estimation designs. The designs they obtained were a by-product of the coordinate-exchange algorithm for finding D-optimal split-plot designs. The 25 scenarios for which no equivalent-estimation design was found previously are indicated using a "I" in the last column of Table 5.

Besides these 25 scenarios, there are 37 scenarios where we are able to present new equivalent-estimation designs with a higher D-efficiency than the design of Macharia and Goos (2010). These scenarios are indicated using a "II" in the last column of Table 5. The largest improvement in D-efficiency was as large as 11% and was achieved for scenario 33, involving one whole-plot factor, two sub-plot factors, and six whole plots of six runs. The old and new equivalent-estimation design for this scenario are shown in Table 3. The second largest improvement was for scenario 90, where we obtained an 8.6% increase in D-efficiency. In most of the 37 scenarios, however, the improvements in D-efficiency are smaller than 1%.

In the remaining 49 scenarios, we could not improve upon the designs of Macharia and Goos (2010). In 10 cases, indicated using a "III" in the last column of Table 5, this is because the D-optimal design happens to be an equivalent-estimation design. The D-efficiency of the best equivalent-estimation design in these cases is 100%. In the remaining 39 cases, indicated using a "IV" in the last column of Table 5, the best equivalent-estimation design is not D-optimal.

Table 5: List of 111 scenarios for which an equivalent-estimation split-plot design exists.

Scenario	Number of WP factors	Number of SP factors	b	k	D-efficiency New	D-efficiency M & G	Case
1	1	1	4	2	93.5	93.3	II
2	1	1	4	3	100.0	100.0	III
3	1	1	4	4	99.4	99.4	IV
4	1	1	4	5	100.0	100.0	III
5	1	1	4	6	100.0	100.0	III
6	1	1	5	2	93.4	93.4	IV
7	1	1	5	3	100.0	100.0	III
8	1	1	5	4	99.1	98.9	II
9	1	1	5	5	100.0	100.0	III

Continued on next page

Table 5: List of 111 scenarios for which an equivalent-estimation split-plot design exists (continued).

Scenario	Number of WP factors	Number of SP factors	b	k	D-efficiency New	D-efficiency M & G	Case
10	1	1	5	6	99.8	99.8	IV
11	1	1	6	2	97.1	97.1	IV
12	1	1	6	3	100.0	100.0	III
13	1	1	6	4	99.4	99.1	II
14	1	1	6	5	99.9	99.9	IV
15	1	1	6	6	99.8	99.8	IV
16	1	1	7	2	94.5	94.0	II
17	1	1	7	3	100.0	100.0	III
18	1	1	7	4	99.2	99.2	IV
19	1	1	7	5	99.9	99.9	IV
20	1	1	7	6	99.9	99.9	IV
21	1	2	4	3	96.5	96.4	II
22	1	2	4	4	98.3	98.3	IV
23	1	2	4	5	100.0	100.0	III
24	1	2	4	6	99.6	99.6	IV
25	1	2	5	3	92.4	92.1	II
26	1	2	5	4	97.8	95.8	II
27	1	2	5	5	99.7	99.7	IV
28	1	2	5	6	99.0	99.0	IV
29	1	2	6	2	96.4	96.4	IV
30	1	2	6	3	70.9		I
31	1	2	6	4	97.3	93.4	II
32	1	2	6	5	99.6	99.6	IV
33	1	2	6	6	98.8	89.0	II
34	1	2	7	2	89.6		I
35	1	2	7	3	89.8		I
36	1	2	7	4	82.6		I
37	1	2	7	5	99.5	99.5	IV
38	1	2	7	6	98.3	97.9	II
39	1	3	4	4	92.9	92.9	IV
40	1	3	4	5	99.4	98.8	II
41	1	3	4	6	98.5	98.5	IV
42	1	3	5	4	96.6	96.6	IV
43	1	3	5	5	98.9	98.9	IV
44	1	3	5	6	66.5		I
45	1	3	6	3	89.4	88.9	II
46	1	3	6	4	66.4		I
47	1	3	6	5	59.3		I

Continued on next page

Table 5: List of 111 scenarios for which an equivalent-estimation split-plot design exists (continued).

Scenario	Number of WP factors	Number of SP factors	b	k	D-efficiency New	D-efficiency M & G	Case
48	1	3	6	6	43.5		I
49	2	1	7	2	94.0	93.9	II
50	2	1	7	3	99.6	99.5	II
51	2	1	7	4	98.5	98.5	IV
52	2	1	7	5	99.9	99.8	II
53	2	1	7	6	99.8	99.7	II
54	2	1	8	2	92.4	92.4	IV
55	2	1	8	3	99.9	99.9	IV
56	2	1	8	4	99.6	99.6	IV
57	2	1	8	5	100.0	100.0	III
58	2	1	8	6	99.7	99.5	II
59	2	1	9	2	92.7	92.0	II
60	2	1	9	3	99.7	99.7	IV
61	2	1	9	4	98.7	93.5	II
62	2	1	9	5	100.0	100.0	III
63	2	1	9	6	99.7	99.7	IV
64	2	1	10	2	93.2	92.9	II
65	2	1	10	3	99.6	99.6	IV
66	2	1	10	4	90.8		I[c]
67	2	1	10	5	99.8	99.8	IV
68	2	1	10	6	83.1		I[c]
69	2	1	11	2	90.9		I
70	2	1	11	3	99.6	99.6	IV
71	2	1	11	4	91.2		I
72	2	1	11	5	99.9	99.8	II
73	2	1	11	6	67.6		I[c]
74	2	1	12	2	80.3		I
75	2	1	12	3	99.8	99.8	IV
76	2	1	12	4	94.2		I
77	2	1	12	5	100.0	99.9	II
78	2	1	12	6	47.5		I[c]
79	2	2	7	3	97.2	97.2	IV
80	2	2	7	4	98.7	98.7	IV
81	2	2	7	5	99.9	99.9	IV
82	2	2	7	6	99.6	99.5	II
83	2	2	8	2	96.3	96.3	IV[a]
84	2	2	8	3	95.1	91.3	II
85	2	2	8	4	97.7	97.0	II

Continued on next page

Table 5: List of 111 scenarios for which an equivalent-estimation split-plot design exists (continued).

Scenario	Number of WP factors	Number of SP factors	b	k	D-efficiency New	D-efficiency M & G	Case
86	2	2	8	5	99.9	99.9	IV
87	2	2	8	6	99.0	96.2	II
88	2	2	9	2	93.3	93.1	II
89	2	2	9	3	75.2		I
90	2	2	9	4	97.1	89.4	II
91	2	2	9	5	99.8	99.8	IV
92	2	2	9	6	96.8	96.5	II
93	2	2	10	2	95.4	93.1	II
94	2	2	10	3	55.1		I
95	2	2	10	4	50.2		I[c]
96	2	2	10	5	99.6	99.4	II
97	2	2	10	6	84.2		I[c]
98	2	2	11	2	94.2	94.2	IV[b]
99	2	2	11	3	38.5		I[c]
100	2	2	11	4	60.6		I[c]
101	2	2	11	5	66.3		I[c]
102	2	2	11	6	55.7		I[c]
103	2	3	7	4	99.7	99.7	IV
104	2	3	7	5	100.0	99.7	II
105	2	3	7	6	99.5	99.4	II
106	2	3	8	3	93.0	87.9	II
107	2	3	8	4	99.4	98.5	II
108	2	3	8	5	99.5	99.2	II
109	2	3	8	6	44.9		I[c]
110	3	2	11	2	99.5	99.5	IV
111	3	3	12	4	93.7	93.3	II

[a] Macharia and Goos (2010) reported an incorrect D-efficiency of 96.5%.
[b] Macharia and Goos (2010) reported an incorrect D-efficiency of 99.4%.
[c] Equivalent-estimation design with more than three levels, obtained using Brent's method for sub-plot factor levels.

In 100 of the 111 scenarios, we managed to find a three-level equivalent-estimation design. The 11 scenarios in which more than three levels were required are indicated using the subscript c in the last column of Table 5.

A file containing all the equivalent-estimation designs listed in Table 5 is available from the authors.

7 Discussion

The structure of a split-plot design generally requires the use of generalized least squares (GLS) to estimate the model. This estimation approach is not always implemented in the statistical software available to practitioners. This has led to the development of various methods for constructing split-plot designs for which the OLS and GLS estimators produce the same point estimates. Often, these designs provided statistically inefficient estimates of the factor effects. Macharia and Goos (2010) showed that, in many cases, it is possible to obtain highly D-efficient three-level equivalent-estimation split-plot designs such that the loss of precision in parameter estimates is negligible if OLS is the estimation technique utilized.

In this paper, we present an algorithm that enabled us to find efficient three-level equivalent-estimation design in 100 out of 111 scenarios. Allowing the sub-plot factors to take more than three levels yielded equivalent-estimation designs for the remaining 11 scenarios. As a result, for second-order response surface designs, it seems possible to generate an equivalent-estimation design for every split-plot scenario.

As stated in Macharia and Goos (2010), we also note that whether or not a design is an equivalent-estimation design depends on the model actually fitted, just like the D-optimality of a design depends on the specified model. For example, dropping any of the terms associated with the whole-plot factors destroys the OLS-GLS equivalence. Thus, reducing the model complexity may lead to the loss of the OLS-GLS equivalence property. This is counterintuitive, as desirable theoretical properties are usually easier to achieve for simple models.

Another point worth noting is that some of the equivalent-estimation designs we obtained involve a large number of factor levels. An example of such a design is given in Table 4. Running such a design requires perfect control of the levels of the sub-plot factors. In many applications, however, it will be required that the factor levels have to be rounded. The resulting design will then no longer possess the OLS-GLS equivalence property. However, the OLS and GLS estimates will not be very different in such cases. We recommend using the equivalent-estimation efficiency measure defined in Equation (10) to quantify the extent to which rounding affects the equivalent-estimation property.

Acknowledgment

The first author of this article acknowledges the financial support of the Research Foundation Flanders (FWO). This paper was finalized at the Isaac Newton Institute for Mathematical Sciences at Cambridge, UK, during the Design and Analysis of Experiments research programme hosted by the institute.

Short biographies

Dr. Bradley Jones is a Principal Research Fellow at the SAS Institute and a Guest Professor at the Faculty of Applied Economics of the Universiteit Antwerpen. He is a senior member of the American Society for Quality. His email address is bradley.jones@jmp.com.

Dr. Peter Goos is Full Professor at the Faculty of Applied Economics and StatUa Center for Statistics of the Universiteit Antwerpen and at the Erasmus School of Economics of the Erasmus Universiteit Rotterdam. He is a senior member of the American Society for Quality. His email address is peter.goos@ua.ac.be.

References

Almimi, A. A., Kulahci, M. and Montgomery, D. C. (2008). Follow-up designs to resolve confounding in split-plot experiments, *Journal of Quality Technology* **40**: 154–166.

Almimi, A. A., Kulahci, M. and Montgomery, D. C. (2009). Checking the adequacy of fit of models from split-plot designs, *Journal of Quality Technology* **41**: 272–284.

Bingham, D. R., Schoen, E. D. and Sitter, R. R. (2004). Designing fractional factorial split-plot experiments with few whole-plot factors, *Journal of the Royal Statistical Society: Series C* **53**: 325–339. Corrigendum, **54**, 955-958.

Bingham, D. R. and Sitter, R. R. (1999). Minimum-aberration two-level fractional factorial split-plot designs, *Technometrics* **41**: 62–70.

Brent, R. P. (1973). *Algorithms for Minimization without Derivatives*, Englewood Cliffs: Prentice Hall.

Gilmour, S. G. and Goos, P. (2009). Analysis of data from nonorthogonal multi-stratum designs in industrial experiments, *Journal of the Royal Statistical Society: Series C* **58**: 467–484.

Goos, P. (2006). Optimal versus orthogonal and equivalent-estimation design of blocked and split-plot experiments, *Statistica Neerlandica* **60**: 361–378.

Goos, P. and Jones, B. (2011). *Design of Experiments: A Case-Study Approach*, New York: Wiley.

Goos, P., Langhans, I. and Vandebroek, M. (2006). Practical inference from industrial split-plot designs, *Journal of Quality Technology* **38**: 162–179.

Goos, P. and Vandebroek, M. (2001). Optimal split-plot designs, *Journal of Quality Technology* **33**: 436–450.

Goos, P. and Vandebroek, M. (2003). D-optimal split-plot designs with given numbers and sizes of whole plots, *Technometrics* **45**: 235–245.

Goos, P. and Vandebroek, M. (2004). Outperforming completely randomized designs, *Journal of Quality Technology* **36**: 12–26.

Huang, P., Chen, D. and Voelkel, J. (1998). Minimum-aberration two-level split-plot designs, *Technometrics* **40**: 314–326.

Jones, B. and Goos, P. (2007). A candidate-set-free algorithm for generating D-optimal split-plot designs, *Journal of the Royal Statistical Society: Series C* **56**: 347–364.

Jones, B. and Nachtsheim, C. J. (2009). Split-plot designs: What, why, and how, *Journal of Quality Technology* **41**: 340–361.

Jones, B. and Nachtsheim, C. J. (2011). Efficient designs with minimal aliasing, *Technometrics* **53**: 62–71.

Kulahci, M. and Bisgaard, S. (2005). The use of Plackett-Burman designs to construct split-plot designs, *Technometrics* **47**: 495–501.

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

Macharia, H. and Goos, P. (2010). D-optimal and D-efficient equivalent-estimation second-order split-plot designs, *Journal of Quality Technology* **42**: 358–372.

McElroy, F. W. (1967). A necessary and sufficient condition that ordinary least squares estimators be best linear unbiased, *Journal of the American Statistical Association* **62**: 1302–1304.

McLeod, R. G. and Brewster, J. F. (2008). Optimal foldover plans for two-level fractional factorial split-plot designs, *Journal of Quality Technology* **40**: 227–240.

Parker, P. A., Kowalski, S. M. and Vining, G. G. (2006). Classes of split-plot response surface designs for equivalent estimation, *Quality and Reliability Engineering International* **22**: 291–305.

Parker, P. A., Kowalski, S. M. and Vining, G. G. (2007a). Construction of balanced equivalent estimation second-order split-plot designs, *Technometrics* **49**: 56–65.

Parker, P. A., Kowalski, S. M. and Vining, G. G. (2007b). Unbalanced and minimal point equivalent estimation second-order split-plot designs, *Journal of Quality Technology* **39**: 376–388.

Rodríguez, M., Jones, B., Borror, C. M. and Montgomery, D. C. (2010). Generating and assessing exact G-optimal designs, *Journal of Quality Technology* **42**: 1–18.

Schoen, E., Jones, B. and Goos, P. (2011). A split-plot experiment with factor-dependent whole-plot sizes, *Journal of Quality Technology* **43**: 66–79.

Trinca, L. A. and Gilmour, S. G. (2001). Multi-stratum response surface designs, *Technometrics* **43**: 25–33.

Vining, G. G. and Kowalski, S. M. (2008). Exact inference for response surface designs within a split-plot structure, *Journal of Quality Technology* **40**: 394–406.

Vining, G. G., Kowalski, S. M. and Montgomery, D. C. (2005). Response surface designs within a split-plot structure, *Journal of Quality Technology* **37**: 115–129.