

Cost Penalized Estimation and Prediction Evaluation for Split-Plot Designs

Li Liang
Virginia Polytechnic Institute and State University, Blacksburg, VA 24060

Christine M. Anderson-Cook
Los Alamos National Laboratory, Los Alamos, NM 87545

Timothy J. Robinson
University of Wyoming, Laramie, WY 82071

1. INTRODUCTION

The use of response surface methods generally begins with a process or system involving a response y that depends on a set of k controllable input variables (factors) x_1, x_2, \dots, x_k . To assess the effects of these factors on the response, an experiment is conducted in which the levels of the factors are varied and changes in the response are noted. The size of the experimental design (number of distinct level combinations of the factors as well as number of runs) depends on the complexity of the model the user wishes to fit. Limited resources due to time and/or cost constraints are inherent to most experiments, and hence, the user typically approaches experimentation with a desire to minimize the number of experimental trials while still being able to adequately estimate the underlying model.

There are many different ways to assess a design's capability to estimate the underlying model. For instance, one can focus on the quality of parameter estimates (often quantified by the D -criterion) or the precision of model predictions such as with the minimization of average prediction variance [(V , Q or IV -criterion) or minimization of worst prediction variance (G -criterion)]. The use of these 'alphabetic optimality' criteria for comparing competing designs is well documented (see Myers and Montgomery 2002 pp. 390-402).

Embedded in the calculation of each of the alphabetic optimality criteria is the information matrix of the estimated model parameters. The information matrix is based

on the error structure and the error structure is a function of how one randomizes the run order of the experimental trials. Frequently, experiments are designed assuming a completely random run order. However, if the levels of one or more factors' are difficult and/or costly to change, the practitioner is not as inclined to run the experiment using a completely randomized run order. Instead, the practitioner may select a run order involving fewer changes of the hard to change factor. The resulting experiment generally involves two separate randomizations: one for the run order of the levels of the hard to change factors and one within each level combination of the hard to change factors, to randomize the run order of all possible combinations of the easy to change variables. When there are separate randomizations for hard to change factors and easy to change factors, the error structure is that of a split plot design. Letsinger, Myers, and Lentner (1996), Ganju and Lucas (1999), and Ju and Lucas (2002) describe split plot designs resulting from factor levels on consecutive runs of an experiment not being reset.

Although the selected experimental design may have nice statistical properties for complete randomization, if the design is run as a split-plot, the design's statistical properties may not be well understood and hence are unlikely to be optimal. Ganju and Lucas (1999) point out that split-plot designs chosen by random run order are not as appealing as those chosen by a good design strategy. Design strategies for split-plot randomization have received considerable attention in the literature of late. Huang, Chen and Voelkel (1998) and Bingham and Sitter (1999) derive minimum aberration two-level fractional factorial designs for screening experiments. Anbari and Lucas (1994) considered the G -optimality criterion for several competing split-plot designs. Goos and Vandebroek (2001 and 2004) proposed exchange algorithms for constructing D -optimal split-plot designs. Liang, Anderson-Cook, and Robinson and Myers (2004, 2005) considered graphical techniques for assessing competing split-plot designs over the design region in terms of G - and V -efficiency. Bisgaard (2000), noting that the benefits of running a split plot design are the savings obtained by reducing the number of whole-plot setups, formulated cost functions indicating the relative costs of performing each of the sub-plot tests to the cost of setting up the individual whole-plot tests.

With the exception of Bisgaard (2000), optimal strategies for split-plot designs, in large part, have been focused on objective criteria that reflect the statistical properties of

the design (the D -criteria involves quality of parameter estimates, while G - and V -optimality involve quality of prediction). The fact that the split-plot error structure is a result of hard-to-change factors often implies that there is greater cost/time involved in changing the levels of the whole plot factors than there is in changing the levels of the subplot factors. As a result, the practitioner may desire to use design selection criteria that not only reflect the statistical properties of the design but also the cost of the experiment. For instance, suppose we have three factors, one hard to change (w) and two easy to change (x_1 and x_2) variables, each at two levels run as a full factorial design. Consider the three competing designs in Table 1 with 2, 4 and 8 whole plots, respectively. The numbers in the columns for each design specify which whole plot will contain that combination of factor combinations. For any design that has more than one observation per whole plot, the level of w remains unchanged within a whole plot, making it possible to collect these observations without changing whole plot levels. Goos and Vandebroek (2004) state that Design 2 is the best possible eight run split plot design for estimating the pure linear model in terms of D -efficiency. The authors go on to state that when there is more variation among whole plots than there is among subplots, Design 1 is always more D -efficient than Design 3, the completely randomized design.

Table 1: Eight runs factorial design in split-plot structure. The Design 1, 2 and 3 indicates the index of the whole plot, and corresponding w level represents the whole plot level.

w	x_1	x_2	Design 1	Design 2	Design 3
-1	-1	-1	1	1	1
-1	-1	+1	1	2	2
-1	+1	-1	1	2	3
-1	+1	+1	1	1	4
+1	-1	-1	2	3	5
+1	-1	+1	2	4	6
+1	+1	-1	2	4	7
+1	+1	+1	2	3	8

Although Design 2 may be the best possible 8 run split-plot design in terms of D -efficiency, a consideration of ‘cost’ in terms of ease of experimentation, time, etc., may suggest other alternatives. For instance, suppose the only appreciable cost in experimentation is due to changes in the hard to change factor. If this is the case, Design 1 is twice as appealing as Design 2 (Design 1 requires two set-ups for the levels of w and

Design 2 requires four) and four times more appealing than Design 3 (Design 3 involves eight set-ups). Clearly there is a trade-off between “good statistical properties” and “cost of experimentation”. Although alphabetic optimality criteria are useful in determining split-plot designs, these criteria do not reflect the different costs which are likely associated with hard to change and easy to change factors.

Dompere (2004) states “there are two important sides to any decision...the two sides are simply the *costs* that may be incurred in order to receive the *benefits* that may be associated with a particular decision.” Relating this to split-plot experimentation, a design that may have nice statistical properties in terms of the estimated model may not be appealing from a cost perspective. In the split-plot setting, we often need to find the right balance of designs with good statistical properties and are within the experimenter’s budget

In this manuscript, we propose cost adjusted D , G , and V optimality criteria for split-plot designs. We adjust the D , G , and V optimality criteria for cost, where the expressions for cost are similar to that of Bisgaard (2000) with adjustments made to allow for unequal whole plot size. With the cost adjusted optimality criteria, the user is presented with single objective functions that simultaneously account for the desired statistical property (efficient parameter estimation or model prediction) and cost of experimentation. Utilizing the new objective functions, we demonstrate strategies for choosing optimal split-plot designs and then illustrate these ideas with two examples. In the next section we discuss the cost formulations and then we develop the appropriate cost adjusted D , G , and V expressions for split-plot designs. Finally, two examples are provided which demonstrate the trade-off between ‘good statistical properties’ and ‘cost reduction’.

2. COST FORMULATIONS

In practice, if a completely randomized experiment is run, it is generally the case that changing the levels of a factor is uniformly difficult across all factors. As a result, the cost or time associated with the experiment is related only to the number of experimental units (EUs). In split-plot experiments, there are generally two types of EUs –whole plots

and sub plots. In industrial experimentation, whole plot factors are those factors whose levels are hard/costly to change and subplot factors are those factors whose levels are relatively easy to change. Thus, in considering the cost of running an industrial split-plot experiment, the total cost is a function of both the cost associated with whole plot units as well as the cost associated with subplot units. Similar to Bisgaard (2000), we write the cost of a split-plot experiment as

$$C = C_{WP}a + C_{SP}N \quad (1)$$

where C denotes the total cost of the experiment, a denotes the number of whole plot units, N is the total number of subplot units, and C_{WP} and C_{SP} are the costs associated with individual whole plot and subplot, respectively. Note that cost for measuring the observation is considered a part of cost of subplot.

In practice, it may be difficult for the practitioner to ascertain the exact costs associated with whole plots or subplots, i.e. precise values for C_{WP} and C_{SP} , but it may be more feasible to specify the relative cost of these quantities, i.e. $r = C_{SP}/C_{WP}$. Hence, the cost of the experiment is proportional to $a + rN$, i.e.

$$C \propto a + rN. \quad (2)$$

Writing the cost in this manner allows flexibility for specifying the relative costs of the two cost components without having to specify their exact values. Generally speaking, C_{WP} is greater than C_{SP} due to the time/effort involved with changing levels of the whole plot factors. As C_{WP} increases relative to C_{SP} , r approaches zero. On the other hand, if obtaining the measurement of the response for each observation is expensive, then C_{SP} may be relatively large compared to C_{WP} and the r will increase. When $C_{WP} = C_{SP}$, $r = 1$.

It is noteworthy that the completely randomized design (CRD) can be thought as a special case of a split-plot design where each observation can be treated as a separate whole plot. Also the number of whole plots and subplots are equal and the total cost of the experiment is given by

$$C = (C_{WP} + C_{SP})N = C_{WP}(1 + r)N. \quad (3)$$

The expression in (3) is proportional then to the standard penalty of N commonly used for the D -, G -, and V -optimality criteria in completely randomized experiments. In the next

section, we review the general model for split-plot designs and present some alphabetic optimality criteria for split-plot designs. We then present cost adjusted D -, G -, and V -criteria which utilize the expressions for cost discussed above.

3. THE SPLIT-PLOT MODEL AND COST ADJUSTED D -, G -, AND V -OPTIMALITY CRITERIA

When the experiment is run as a split-plot design with a whole plots, the following linear mixed model can be written to explain the variation in the $n \times 1$ response vector, \mathbf{y} ,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}. \quad (4)$$

Regarding notation, \mathbf{X} is the $N \times p$ design matrix expanded to model form for p parameters including the intercept; \mathbf{Z} is an $N \times a$ classification matrix of ones and zeroes where the ij^{th} entry is 1 if the i^{th} observation ($i = 1, \dots, N$) belongs to the j^{th} whole plot ($j = 1, \dots, a$); $\boldsymbol{\delta}$ is an $a \times 1$ vector of random effects where the δ_j elements are assumed *i.i.d* $N(0, \sigma_\delta^2)$ with σ_δ^2 denoting the variability among whole plots; $\boldsymbol{\varepsilon}$ is the $n \times 1$ vectors of residual errors where the ε_i elements are assumed *i.i.d* $N(0, \sigma_\varepsilon^2)$ and σ_ε^2 denotes the variation among subplot units. It is also assumed that $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are independent.

The covariance matrix of the responses in a split-plot design is

$$\text{Var}(\mathbf{y}) = \boldsymbol{\Sigma} = \sigma_\delta^2 \mathbf{Z}\mathbf{Z}' + \sigma_\varepsilon^2 \mathbf{I}_n = \sigma_\varepsilon^2 [d\mathbf{Z}\mathbf{Z}' + \mathbf{I}_n] \quad (5)$$

where \mathbf{I}_n is an $n \times n$ identity matrix and $d = \sigma_\delta^2 / \sigma_\varepsilon^2$ represents the variance component ratio. For simplicity of presentation, we will assume that observations are sorted by the whole plots, implying $\mathbf{Z} = \text{diag}\{\mathbf{1}_{n_1}, \dots, \mathbf{1}_{n_a}\}$, where $\mathbf{1}_{n_j}$ is an $n_j \times 1$ vector of one's and n_j is the size of the j^{th} whole plot. Assuming sorted observations by whole plots allows one to conveniently write the covariance matrix as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Sigma}_a \end{bmatrix} \quad (6)$$

where each $n_j \times n_j$ matrix $\boldsymbol{\Sigma}_j$ is given by

$$\Sigma_j = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_\delta^2 & \cdots & \sigma_\delta^2 \\ \vdots & \ddots & \vdots \\ \sigma_\delta^2 & \cdots & \sigma_\varepsilon^2 + \sigma_\delta^2 \end{bmatrix}. \quad (7)$$

Note that the variance of an individual observation is the sum of the subplot and whole plot error variances, $\sigma_\varepsilon^2 + \sigma_\delta^2$. A popular method for estimating the variance components is restricted maximum likelihood (REML).

The vector of fixed effects parameters, $\boldsymbol{\beta}$, is estimated via generalized least squares, yielding

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{y}. \quad (8)$$

The covariance matrix of the estimated model coefficients is given by

$$\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1}. \quad (9)$$

When the design is completely randomized, $\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}' \mathbf{X})^{-1} \hat{\sigma}^2$. Comparing the expressions for the estimated model coefficients for split-plot designs and CRDs is important as it lends insight into the greater complexity associated with optimal design strategies for split-plot designs vs. the CRD. For example, if one wishes to obtain the optimal design in terms of ability to estimate model parameters, the optimal CRD depends only on the settings of the levels of the terms in \mathbf{X} . The optimal split plot design in terms of parameter estimation will depend on the structure of \mathbf{X} , the variance ratio, d , the number of whole plots, a , and the dimensionality of each of the Σ_j (determined by the number of subplots within each whole plot), and subplot levels arrangements in whole plots.

3.1 Cost Adjusted D -Optimality Criterion

Strategies for choosing an optimal design depend on the goal of the researcher. If the desire is to have quality model parameter estimates, one strategy is to find a design with high D -efficiency. The D -efficiency criterion is defined in terms of the scaled moment matrix. For CRDs, the scaled moment matrix is $\mathbf{M} = \mathbf{X}' \mathbf{X} / N$, scaled by σ^2 / N .

Note scaling by σ^2 (the observation error variance), causes \mathbf{M} to be unitless and the scaling by $1/N$ causes \mathbf{M} to be reflective of design size. Since the cost of a CRD is determined by the design size, the scaling by $1/N$ is essentially a scaling for cost. The D -optimal design is then the design that maximizes $|\mathbf{M}| = \left| (\mathbf{X}'\mathbf{X}) \right| / N^p$ where p denotes the number of model parameters. The moment matrix for split-plot designs is $(\mathbf{X}'\Sigma^{-1}\mathbf{X})$. Scaling the information matrix in a similar fashion to the scaling in the CRD, we can define the scaled moment matrix for split-plot designs as

$$\mathbf{M} = \frac{(\sigma_\delta^2 + \sigma_\varepsilon^2)(\mathbf{X}'\Sigma^{-1}\mathbf{X})}{cost}.$$

Note that $(\sigma_\delta^2 + \sigma_\varepsilon^2)(\mathbf{X}'\Sigma^{-1}\mathbf{X}) = \left(\mathbf{X}' \left(\frac{\Sigma}{\sigma_\delta^2 + \sigma_\varepsilon^2} \right)^{-1} \mathbf{X} \right) = (\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})$, where \mathbf{R} denotes the observational correlation matrix. Rewriting \mathbf{M} we have

$$\mathbf{M} = \frac{(\mathbf{X}'\mathbf{R}\mathbf{X})}{cost}$$

Since \mathbf{R} is the correlation matrix, \mathbf{M} is unitless as desired. Since the cost of a split-plot is not as simple as the design size, N , we must adjust for an expression for cost that allows for potentially different costs associated with whole plots and subplots. A natural divisor is the expression for cost provided in (2), yielding $\mathbf{M} = (\mathbf{X}'\mathbf{R}\mathbf{X}) / (a + rN)$. The cost-adjusted D -efficiency is then defined as

$$D_{eff} = \frac{|\mathbf{M}(D)^{1/p}|}{\text{Max}_{D \in \Omega} |\mathbf{M}(D)^{1/p}|}.$$

Since the upper bound for the determinant is generally unknown, we can consider relative performance for two or more designs by looking at the cost adjusted D -criteria,

$$D = |\mathbf{M}|^{1/p} = \frac{|\mathbf{X}'\mathbf{R}\mathbf{X}|^{1/p}}{a + rN}, \quad (10)$$

where a good design maximizes this criterion.

3.2 Cost Adjusted G - and V -Optimality Criteria

The predicted value of the mean response at any location \mathbf{x}_0 is given by

$$\hat{y}_0 = \mathbf{x}_0' \hat{\boldsymbol{\beta}} = \mathbf{x}_0' \left(\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{y},$$

where \mathbf{x}_0 is the point of interest in the design space expanded to model form. The prediction variance is then given by

$$\text{Var}(\hat{y}_0) = \mathbf{x}_0' \left(\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} \right)^{-1} \mathbf{x}_0. \quad (11)$$

If interest is in finding a design with precise estimates of the mean, G - and V -efficiency of the design are popular choices. As with D -efficiency, the desire is to work with a scale free quantity that provides a penalty for design cost. In the CRD, the prediction variance is given by

$$\text{Var}(\hat{y}) = \sigma^2 \mathbf{x}_0' \left(\mathbf{X}' \mathbf{X} \right)^{-1} \mathbf{x}_0$$

and the proper scaling and cost penalty is σ^2/N (observation error divided by the design size). The scaled prediction variance for the CRD is then given by

$$SPV = \frac{N \text{Var}[\hat{y}(\mathbf{x}_0)]}{\sigma^2} = N \mathbf{x}_0' \left(\mathbf{X}' \mathbf{X} \right)^{-1} \mathbf{x}_0.$$

The scaling of the prediction variance for split-plot designs can be done in a similar fashion by scaling by $(\sigma_\delta^2 + \sigma_\varepsilon^2)/(a+rN)$ (observation error divided by the design cost). The cost penalized scaled prediction variance (CPPV) for the split plot design is then given by

$$\begin{aligned} CPPV &= \frac{(a+rN) \text{Var}[\hat{y}(\mathbf{x}_0)]}{\sigma_\delta^2 + \sigma_\varepsilon^2} = (a+rN) \mathbf{x}_0' \left(\mathbf{X}' \left(\frac{\boldsymbol{\Sigma}}{\sigma_\delta^2 + \sigma_\varepsilon^2} \right)^{-1} \mathbf{X} \right)^{-1} \mathbf{x}_0 \\ &= (a+rN) \mathbf{x}_0' \left(\mathbf{X}' \mathbf{R}^{-1} \mathbf{X} \right)^{-1} \mathbf{x}_0 \end{aligned} \quad (12)$$

By minimizing the average or maximum CPPV over the entire design region, one can obtain the best balance between quality and cost in terms of V - or G -efficiency. Anbari and Lucas (1994) used the lower bound of the maximum scaled prediction variance for CRD (p , the number of model parameters) to evaluate G -efficiency of split-

plot designs and claimed some super-efficient designs. Apparently p is not a reasonable lower bound for split-plot design because the two errors structure and the various values of variance component ratio play role in the G -efficiency. It should be pointed out that the actual bounds for the D -, G -, and V -efficiencies for SPDs needs further investigation, and here we focus more on relative efficiencies for comparisons between competing designs.

In the following sections we examine first and second order model SPDs utilizing the cost adjusted D -, G -, and V -criteria developed above. For simplicity of presentation, we discuss experiments with three factors – one whole plot variable, w , and two subplot variables, x_1 and x_2 . Section 4 involves a study of a first order design in which the candidate set of design points is the 8 runs of a 2^3 factorial design. All possible sequences of run orders are permuted and their corresponding split-plot designs are constructed. By evaluating the estimation and prediction properties of these designs with and without cost penalization, we demonstrate that the selected design is influenced by not only the split-plot error structure but also the relative costs between whole plots and subplots. The best design in terms of a joint consideration of cost and quality is often different from the optimal design when only quality is considered. As might be expected, designs with a smaller number of whole plots are preferred as the whole plots become more expensive. In Section 5, five variations of the central composite design (CCD) are studied for the second order model. The study shows that under different scenarios, the CPPV penalizes designs with larger numbers of whole plots proportional to the relative cost of the whole plots to subplots. We also provide some design strategies for second order split plot designs.

4. EFFECT OF COST IN DESIGNS WITH FIRST-ORDER MODEL

In this section, we consider an example with 8 design points for a first-order model with fixed effects modeled as $f(w, x_1, x_2) = \beta_0 + \beta_1 w + \beta_2 x_1 + \beta_3 x_2$. The 8 designs points are the usual factorial runs in a 2^3 factorial experiment with one whole plot factor, w , and two subplot factors x_1 and x_2 . All possible sequences of different run orders are generated and the corresponding SPD can be constructed based on the assumption that

the whole plots level changes are expensive so that the same consecutive whole plot (hard-to-change) factor levels are not reset. Then cost penalized evaluations are performed for the obtained experiments in terms of D -, G - and V -efficiency under different combinations of cost ratio, r , representing the relative cost of the whole plot and subplot, and variance component ratio, d , indicating the relative size of variability from the two components. The comparisons between all the designs with different number of whole plots and sizes demonstrate desired split-plot settings under different conditions of the real experiments.

By permuting the run order (1,2,3,4,5,6,7,8), all possible SPDs with 8 runs can be obtained by matching the values in the sequences to the 8 factorial runs listed in standard order. From the sequences, the split-plot structure with the number and structure of the whole plots can be extracted. Note that this is different from the way the design is set-up (CRD) by Joiner and Campbell (1976), who did not take into account the correct error structure for a design when some factors levels are assumed not reset. Here we assume that consecutive runs with same whole plot level are not reset and consequently are considered part of the same whole plot. Because of the assumed low cost of changing subplot levels, the subplot levels for runs in a same whole plot are always assumed to be reset. This applies regardless of whether the adjacent observations within a whole plot have the same subplot levels or not. Consequently, the observations from different whole plots are assumed to be independent, but observations within the same whole plot are assumed dependent. The generated split-plot designs have a relatively low unbalanced level with the number of whole plots at each of $+1$ and -1 being the same or have at most a difference of 1. For each run order sequence and corresponding split-plot design, the determinant of estimates variances, average and maximum prediction variance with and without cost adjustment are calculated. The quality without cost penalization indicates the precision of the parameters estimates and predicted value. Cost penalized evaluation tells us the desirability of a design when its expense is taken into account. Design performance is strongly related to the variance component ratio, d , through the correlation matrix of observations. Bisgaard and Steinberg (1997) stated that the whole plot variance is usually larger than subplot variance in prototype experiments. Letsinger et al. (1996) studied a split-plot experiment in chemical industry with $\sigma_\delta^2 / \sigma_\epsilon^2 = 1.04$. Vining, Kowalski and

Montgomery (2004) estimated the variance terms using pure error and obtain the variance ratio 5.65. Webb, Lucas and Borkowski (2002) described an experiment with variance ratio 6.92 in a computer component manufacturing company. Kowalski, Cornell and Vining (2002) studied a mixture experiment with process variables where the estimated variance ratio is 0.82. In this paper, specific values of $d=0.5, 1$ and 10 are considered in more details, representing the situations that the whole plot variance is half, same and ten times the subplot variance, respectively. For the cost ratio, $r=0, 0.5, 1$ are considered, and represent the scenarios that subplot costs nothing, the whole plot costs twice as much as a subplot, and they are equally expensive.

An exhaustive search provides 31 distinct SPDs in terms of cost penalized precision of estimation and prediction. We consider all those designs that alternate between the two levels of the whole plot factor on adjacent whole plots. They are listed in the Table 2. Designs are equivalent (in terms of isomorphism) if they have at least one of the following four relations,

- The whole plots are changes from $+/-1$ to $-/+1$.
- The order of any whole plots with same whole plot level can be changed.
- The subplots levels are switched between $+1$ and -1 .
- The subplots within each whole plot are permuted.

Table 2: 31 distinct designs, one combination (w, x_1, x_2) represents a design point and indicates the levels of the three factors at this point, a is the number of whole plots, ID is the identification of the design in the list. The units in the same cell of the table are within the same whole plot.

ID	a	Whole plots			
1	2	$(-1, -1, -1)$	$(1, -1, -1)$		
		$(-1, -1, 1)$	$(1, -1, 1)$		
		$(-1, 1, -1)$	$(1, 1, -1)$		
		$(-1, 1, 1)$	$(1, 1, 1)$		
2	3	$(-1, -1, -1)$	$(1, -1, -1)$	$(-1, -1, 1)$	
		$(-1, -1, 1)$	$(1, -1, 1)$	$(-1, 1, -1)$	
		$(-1, 1, -1)$	$(1, 1, -1)$		
		$(-1, 1, 1)$	$(1, 1, 1)$		
3	3	$(-1, -1, -1)$	$(1, -1, -1)$	$(-1, 1, -1)$	
		$(-1, -1, 1)$	$(1, -1, 1)$	$(-1, 1, 1)$	
		$(-1, 1, -1)$	$(1, 1, -1)$		
		$(-1, 1, 1)$	$(1, 1, 1)$		
4	3	$(-1, -1, -1)$	$(1, -1, -1)$	$(-1, -1, 1)$	
			$(1, -1, 1)$	$(-1, 1, -1)$	
			$(1, 1, -1)$	$(-1, 1, 1)$	
			$(1, 1, 1)$		
5	4	$(-1, -1, -1)$	$(1, -1, -1)$	$(-1, 1, -1)$	$(1, 1, -1)$
		$(-1, -1, 1)$	$(1, -1, 1)$	$(-1, 1, 1)$	$(1, 1, 1)$
6	4	$(-1, -1, -1)$	$(1, -1, -1)$	$(-1, 1, -1)$	$(1, -1, 1)$
		$(-1, -1, 1)$	$(1, 1, -1)$	$(-1, 1, 1)$	$(1, 1, 1)$
7	4	$(-1, -1, -1)$	$(1, -1, -1)$	$(-1, -1, 1)$	$(1, 1, -1)$

			(1, -1, 1)	(-1, 1, -1) (-1, 1, 1)	(1, 1, 1)				
8	4	(-1, -1, -1)	(1, 1, 1)	(-1, -1, 1) (-1, 1, -1) (-1, 1, 1)	(1, -1, -1) (1, -1, 1) (1, 1, -1)				
9	4	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1) (-1, 1, -1) (-1, 1, 1)	(1, -1, 1) (1, 1, -1) (1, 1, 1)				
10	4	(-1, -1, -1)	(1, -1, 1)	(-1, -1, 1) (-1, 1, -1) (-1, 1, 1)	(1, -1, -1) (1, 1, -1) (1, 1, 1)				
11	4	(-1, -1, -1) (-1, 1, -1)	(1, -1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, 1)	(1, -1, 1) (1, 1, -1)				
12	4	(-1, -1, -1)	(1, -1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, -1) (-1, 1, 1)	(1, -1, 1) (1, 1, -1)				
13	4	(-1, -1, -1) (-1, 1, 1)	(1, -1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, -1)	(1, -1, 1) (1, 1, -1)				
14	5	(-1, -1, -1)	(1, -1, -1) (1, -1, 1)	(-1, -1, 1)	(1, 1, -1) (1, 1, 1)	(-1, 1, -1) (-1, 1, 1)			
15	5	(-1, -1, -1)	(1, -1, -1) (1, -1, 1)	(-1, 1, -1)	(1, 1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, 1)			
16	5	(-1, -1, -1)	(1, -1, 1)	(-1, 1, -1)	(1, -1, -1) (1, 1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, 1)			
17	5	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, -1, 1) (1, 1, -1) (1, 1, 1)	(-1, 1, -1) (-1, 1, 1)			
18	5	(-1, -1, -1)	(1, -1, -1) (1, -1, 1)	(-1, 1, 1)	(1, 1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, -1)			
19	5	(-1, -1, -1)	(1, -1, -1)	(-1, 1, 1)	(1, -1, 1) (1, 1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, -1)			
20	5	(-1, -1, -1)	(1, -1, 1)	(-1, 1, 1)	(1, -1, -1) (1, 1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, -1)			
21	5	(-1, -1, -1)	(1, -1, -1) (1, 1, 1)	(-1, -1, 1)	(1, -1, 1) (1, 1, -1)	(-1, 1, -1) (-1, 1, 1)			
22	5	(-1, -1, -1)	(1, -1, -1) (1, 1, 1)	(-1, 1, 1)	(1, -1, 1) (1, 1, -1)	(-1, -1, 1) (-1, 1, -1)			
23	6	(-1, -1, -1)	(1, -1, 1)	(-1, 1, -1)	(1, 1, 1)	(-1, -1, 1) (-1, 1, 1)	(1, -1, -1) (1, 1, -1)		
24	6	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, -1, 1)	(-1, 1, -1) (-1, 1, 1)	(1, 1, -1) (1, 1, 1)		
25	6	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, 1, -1)	(-1, 1, -1) (-1, 1, 1)	(1, -1, 1) (1, 1, 1)		
26	6	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, 1, 1)	(-1, 1, -1) (-1, 1, 1)	(1, -1, 1) (1, 1, -1)		
27	6	(-1, -1, -1)	(1, -1, -1)	(-1, 1, 1)	(1, 1, 1)	(-1, -1, 1) (-1, 1, -1)	(1, -1, 1) (1, 1, -1)		
28	6	(-1, -1, -1)	(1, -1, 1)	(-1, 1, 1)	(1, 1, -1)	(-1, -1, 1) (-1, 1, -1)	(1, -1, -1) (1, 1, 1)		
29	7	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, -1, 1)	(-1, 1, -1)	(1, 1, -1) (1, 1, 1)	(-1, 1, 1)	
30	7	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, 1, 1)	(-1, 1, -1)	(1, -1, 1) (1, 1, -1)	(-1, 1, 1)	
31	8	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, -1, 1)	(-1, 1, -1)	(1, 1, -1)	(-1, 1, 1)	(1, 1, 1)

Among the designs in Table 2, we emphasize two different designs with the same whole plots settings. There are two distinct designs with 3 whole plots with sizes 2, 4 and 2 respectively – design 2 and 3. Both designs are equally appealing from a cost perspective since both involve three whole plots. However, the designs are not equally

appealing in terms of estimation and prediction quality due to the different ways of splitting the four runs with $w=-1$ into two whole plots. In design 2, within the first and third whole plots, both x_1 and x_2 have different levels, hence they obtain the same and as much as amount of information from the contrasts within the whole plots. In design 3, the two subplots have the same level for x_1 within the first and third whole plots, hence less information about x_1 is obtained from the contrasts within the two whole plots in design 3 than in design 2. We define the pattern in design 2 as “pattern A”, design 3 as “pattern B”. We can conclude that pattern A is more efficient for estimation and prediction than pattern B. We can also validate the conclusion from the information matrices. For equal

whole plot and subplot variances ($d=1$), Design 2 has $\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} = \begin{bmatrix} 2.13 & 0.53 & 0 & 0 \\ 0.53 & 2.13 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$ with

determinant equal to 273.07, and design 3 has $\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} = \begin{bmatrix} 2.13 & 0.53 & 0 & 0 \\ 0.53 & 2.13 & 0 & 0 \\ 0 & 0 & 5.33 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$ with

determinant 182.04. We can see the two settings of subplot levels have equivalent effect on the estimate for the whole plot factor w and the subplot factor x_2 , but different effects on the subplot factor x_1 . The estimate of x_1 from pattern A is more precise than from pattern B. Similarly, if a design has four whole plots and each of them has two subplots, the design where the four level combinations of subplot factors x_1 and x_2 with the same whole plot level are split into two whole plots as in pattern A (as for design 13 in Table 2) is the best. This design has symmetric setting for the whole plot levels and subplot levels within all of the whole plots.

First we consider D -efficiency. If the cost of the experiments is not a consideration, the best 5 designs are listed in Table 3 based on the determinant of estimates variances. These designs focus on the estimation property and minimize the volume of the confidence region for the regression coefficients. The highly efficient designs for different d values have only minor differences, which fortunately implies that the relative estimation performance of these top designs is quite robust to the changes in the variance components ratio.

Table 3: The best 5 designs with the best performances, i.e., the 5 highest values of $|\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X}|^{1/4}$. Higher value indicates more information for the parameters and thus more precise estimates. The sequence of whole plot sizes are listed in (n_1, n_2, \dots, n_a) , where a is the number of whole plots of the design.

	a	ID	Whole plot sizes		a	ID	Whole plot sizes		a	ID	Whole plot sizes
$d=0.5$	4	13	2,2,2,2	$d=1$	4	13	2,2,2,2	$d=10$	4	13	2,2,2,2
	5	22	2,2,1,2,1		5	22	2,2,1,2,1		5	22	2,2,1,2,1
	6	28	1,1,2,2,1,1		6	28	1,1,2,2,1,1		4	12	1,2,3,2
	6	27	1,1,2,2,1,1		6	27	1,1,1,1,2,2		6	28	1,1,2,2,1,1
	7	30	1,1,1,1,1,2,1		5	21	2,2,1,2,1		5	21	2,2,1,2,1

Interestingly, the best quality design is not the completely randomized design (CRD) (design 31 in Table 2), which implies that often split-plot designs are more efficient than completely randomized designs when hard-to-change factors exist. Note that this is true even when the higher expense and inconvenience of running the CRD is not accounted for. Goos and Vandebroek (2004) gave similar conclusions regarding the superiority of the split-plot scheme over the CRDs.

The balanced design with 4 whole plots and each with whole plot size 2 (design 13 in Table 2) provides the most precise parameter estimates. Notice this design follows “pattern A” as described above, where the 4 runs with same w level should be split according to pattern A for the two subplot factors, insuring that both x_1 and x_2 can obtain maximal information from the contrasts within each of the 4 whole plots, and thus resulting in the highest efficiency in terms of the parameter estimates.

Taking into account the cost associated with changing the levels of the whole plot variable, the best 5 designs under different cost scenarios using equation (10) are provided in Table 4. The different cost scenarios select different best designs, but these designs are relatively robust to the changes of variance component ratio. This means that even if the guess of variance component ratio, d , provided by the practitioner before any data is collected is not precise, the best design obtained can still be close to or equal to the best one for the actual d value. For instance, if the guess is $d=1$ and the actual value is found to be $d=3$ after the data are collected from the experiment, the selected design based on $d=1$ is still optimal or at least highly efficient. This is good news for

practitioner, who frequently must select a design with little idea of the true value of d in real life experiments.

Table 4: The best 5 designs with the best estimation precision with cost adjust (calculated by equation (10)). The larger value indicates higher cost penalized D-efficiency and thus more desirable.).

	a	ID	Whole plot sizes		a	ID	Whole plot sizes		a	ID	Whole plot sizes
$r=0$ $d=0.5$	2	1	4,4	$r=0$ $d=1$	2	1	4,4	$r=0$ $d=10$	2	1	4,4
	3	2	2,4,2		3	2	2,4,2		3	2	2,4,2
	3	4	1,4,3		3	4	1,4,3		3	4	1,4,3
	3	3	2,4,2		3	3	2,4,2		4	13	2,2,2,2
	4	13	2,2,2,2		4	13	2,2,2,2		3	3	2,4,2
$r=0.5$ $d=0.5$	2	1	4,4	$r=0.5$ $d=1$	2	1	4,4	$r=0.5$ $d=10$	4	13	2,2,2,2
	3	2	2,4,2		3	2	2,4,2		3	2	2,4,2
	4	13	2,2,2,2		4	13	2,2,2,2		2	1	4,4
	3	4	1,4,3		3	4	1,4,3		4	12	1,2,3,2
	3	3	2,4,2		4	12	1,2,3,2		3	4	1,4,3
$r=1$ $d=0.5$	4	13	2,2,2,2	$r=1$ $d=1$	4	13	2,2,2,2	$r=1$ $d=10$	4	13	2,2,2,2
	3	2	2,4,2		3	2	2,4,2		3	2	2,4,2
	2	1	4,4		4	12	1,2,3,2		4	12	1,2,3,2
	4	12	1,2,3,2		2	1	4,4		5	22	1,2,2,2,1
	3	4	1,4,3		5	22	1,2,2,2,1		2	1	4,4

When the practitioner only cares about the cost of whole plots ($r=0$), the design with 2 whole plots is the best and the designs with small number of whole plots are preferred. However, this design has the unappealing property of non-estimability of whole plot error, which may present problems in the analysis and consequently is not a desirable design in practice. Hence, the best design selected turns out to be design 2 from Table 2. If the whole plot is slightly more or equally expensive to the subplot ($r=1$), designs with moderate number of whole plots are preferred. The balanced design with 4 whole plots is best for estimation, and design 2 with 3 whole plots turn out to be second best design. Compared to the best 5 designs without cost penalizations, the desirable designs have smaller number of whole plots.

We now consider V - and G -efficiency for the 31 possible designs. The 5 designs with the smallest average and maximum prediction variance (PV) are listed in Table 5. Note that all the designs assume a cuboidal region, thus the average prediction variance is

calculated by $\frac{1}{8} \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 (PV \text{ or } CPPV) dw dx_1 dx_2$, and the maximum prediction variance is found in the design region $\{(w, x_1, x_2): -1 \leq w \leq 1, -1 \leq x_1 \leq 1 \text{ and } -1 \leq x_2 \leq 1\}$.

The SPDs with large number of whole plots (more than 6 in this example) are highly V -efficient, while the high G -efficiency designs prefer a smaller or moderate number of whole plots. When the variability of whole plots accounts for small or medium proportion of the observational variance ($d=0.5$ or 1), the CRD is the most V -efficient design and the second highest G -efficient design. When the whole plot variance dominates ($d=10$), the CRD is no longer the best, but is still highly V - and G -efficient. This contradicts the results based on D -efficiency that show the SPDs' superiority over CRD for estimation as shown above and by Goos and Vandebroek (2004). This example demonstrates that the CRD is still superior to the SPDs when one is interested in overall prediction performance in the design region and the variance component ratio is small. If the relative size of the whole plot variance is large, SPDs become superior to the CRD, but does not dominate as much as for D -efficiency.

Table 5: The best 5 designs in terms of average and maximum prediction variance. Smaller value indicates better performance in terms of V - and G -efficiency.

	Average prediction variance			Maximum prediction variance		
	a	ID	Whole plot sizes	a	ID	Whole plot sizes
$d=0.5$	8	31	1,1,1,1,1,1,1,1	6	28	1,1,1,1,2,2
	7	30	1,1,1,1,1,2,1	8	31	1,1,1,1,1,1,1,1
	6	28	1,1,1,1,2,2	4	13	2,2,2,2
	6	27	1,1,1,1,2,2	4	11	2,2,2,2
	7	29	1,1,1,1,1,1,2	5	21	1,2,1,2,2
$d=1$	8	31	1,1,1,1,1,1,1,1	6	28	1,1,1,1,2,2
	7	30	1,1,1,1,1,2,1	8	31	1,1,1,1,1,1,1,1
	6	28	1,1,1,1,2,2	4	13	2,2,2,2
	6	27	1,1,1,1,2,2	4	11	2,2,2,2
	7	29	1,1,1,1,1,1,2	5	21	1,2,1,2,2
$d=10$	6	28	1,1,1,1,2,2	6	28	1,1,1,1,2,2
	7	30	1,1,1,1,1,2,1	6	25	1,1,1,1,2,2
	6	25	1,1,1,1,2,2	8	31	1,1,1,1,1,1,1,1
	6	26	1,1,1,1,2,2	4	13	2,2,2,2
	8	31	1,1,1,1,1,1,1,1	4	11	2,2,2,2

When all the designs are evaluated using cost penalized prediction variance, the best 5 designs are listed in Table 6 and 7 for average and maximum CPPV using equation

(12) over the entire cuboidal region. The design with two whole plots is best for the cost penalized V - and G -efficiency for $r=0$, but this design has the critical limitation noted above for data analysis and should likely be avoided. Thus if only the whole plot cost is important ($r=0$) and whole plots variance account for small or medium portion of the observation variance, design 2 with whole plots sizes (2,4,2) is the highest V -efficient. The balanced design with 4 whole plots is the best or the second best if the whole plot cost is moderately more than or comparable to the subplot cost ($r=0.5$ or 1) and under small variance component ratio ($d=0.5$ or 1). If the whole plot variance proportion increases (larger d values), the designs with more whole plots are preferred for overall prediction property. Compared to the estimation performance, the good prediction requires more whole plots.

Table 6: The best 5 designs in terms of average CPPV. Smaller value indicates better design and higher cost adjusted V -efficiency.

	a	ID	Whole plot sizes		a	ID	Whole plot sizes		a	ID	Whole plot sizes
$r=0$ $d=0.5$	2	1	4,4	$r=0$ $d=1$	2	1	4,4	$r=0$ $d=10$	2	1	4,4
	3	2	2,4,2		3	2	2,4,2		4	13	2,2,2,2
	3	3	2,4,2		3	3	2,4,2		4	11	2,2,2,2
	3	4	1,4,3		4	13	2,2,2,2		4	12	1,2,3,2
	4	13	2,2,2,2		3	4	1,4,3		4	6	2,2,2,2
$r=0.5$ $d=0.5$	4	13	2,2,2,2	$r=0.5$ $d=1$	4	13	2,2,2,2	$r=0.5$ $d=10$	6	28	1,1,1,2,2,1
	4	11	2,2,2,2		4	11	2,2,2,2		6	25	1,1,1,1,2,2
	2	1	4,4		4	12	1,2,3,2		6	26	1,1,1,1,2,2
	3	2	2,4,2		5	22	1,2,1,2,2		5	22	1,2,1,2,2
	4	12	1,2,3,2		4	6	2,2,2,2		5	21	1,2,1,2,2
$r=1$ $d=0.5$	4	13	2,2,2,2	$r=1$ $d=1$	4	13	2,2,2,2	$r=1$ $d=10$	6	28	1,1,1,2,2,1
	4	11	2,2,2,2		5	22	1,2,1,2,2		6	25	1,2,1,2,2
	5	22	1,2,1,2,2		6	28	1,1,1,1,2,2		6	26	1,2,1,2,2
	4	12	1,2,3,2		4	11	2,2,2,2		7	30	1,1,1,1,1,1,2
	5	21	1,2,1,2,2		5	21	1,2,1,2,2		5	22	1,2,1,2,2

To minimize the worst prediction variance, the best five designs in terms of maximum cost adjusted prediction variance (CPPV) are listed in Table 7. The balanced four whole plot design (Design 13) is shown to be best in most situations and second best for $d=10$ when subplot and whole plot costs are comparable.

Table 7: The best 5 designs in terms of maximum CPPV. Smaller value indicates better design and higher cost adjusted G -efficiency).

	a	ID	Whole plot sizes		a	ID	Whole plot sizes		a	ID	Whole plot sizes
$r=0$ $d=0.5$	2	1	4,4	$r=0$ $d=1$	2	1	4,4	$r=0$ $d=10$	2	1	4,4
	4	13	2,2,2,2		4	13	2,2,2,2		4	13	2,2,2,2
	3	2	2,4,2		4	11	2,2,2,2		4	11	2,2,2,2
	3	3	2,4,2		4	6	2,2,2,2		4	6	2,2,2,2
	3	4	1,4,3		3	2	2,4,2		4	10	1,1,3,3
$r=0.5$ $d=0.5$	2	1	4,4	$r=0.5$ $d=1$	4	13	2,2,2,2	$r=0.5$ $d=10$	6	28	1,1,1,1,2,2
	4	13	2,2,2,2		4	11	2,2,2,2		4	13	2,2,2,2
	4	11	2,2,2,2		4	6	2,2,2,2		4	11	2,2,2,2
	4	6	2,2,2,2		2	1	4,4		4	6	2,2,2,2
	3	2	2,4,2		6	28	1,1,1,1,2,2		6	25	1,1,1,1,2,2
$r=1$ $d=0.5$	4	13	2,2,2,2	$r=1$ $d=1$	4	13	2,2,2,2	$r=1$ $d=10$	6	28	1,1,1,1,2,2
	4	11	2,2,2,2		4	11	2,2,2,2		4	13	2,2,2,2
	4	6	2,2,2,2		6	28	1,1,1,1,2,2		6	25	1,1,1,1,2,2
	2	1	4,4		4	6	2,2,2,2		4	11	2,2,2,2
	6	28	1,1,1,1,2,2		5	21	1,2,1,2,2		4	6	2,2,2,2

The comparison of the designs in this example shows that taking the relative costs of whole plots and subplots into consideration makes an important difference for the evaluation of the split-plot experimental design. Frequently in industrial settings, the cost of the whole plots dominates the experimental cost. In these cases, split-plot designs with a minimum number of whole plots can have the best overall cost penalized quality. However, other necessary properties, like the ability to estimate variance components, should also be considered. When cost and quality are both important, split-plot designs with a moderate number of whole plots are preferred. These designs perform well for a wide range of d values, which gives robustness to the initial guess of the variance component ratio. This exploration provides insights into how the split-plot structure and different scenarios of cost functions that are realistic in practice for split-plot designs influence choice of design, and thus provides a good foundation for more understanding of cost evaluation for SPDs.

Discussion

For this simple example with only a manageable number of run combinations, it is possible to consider all possible run orders. However, when the experimental design size becomes large, it is hard to search the design space exhaustively. Sampling is a feasible alternative in this situation. A large number of designs can be sampled from the design

space by randomly generating sequences of run orders. Joiner and Campbell (1976) outline this approach, but did not provide information about how this helps with design selection. Modified sampling is implemented following the rules: the level of each factor is generated at random and independently, and the only restriction is that the sampling of the eight runs is performed without replacement. Between runs, the decision of whether to change the factor level or not is determined by draw from a Bernoulli distribution with a fixed probability of changing level, P_w for whole plot factors and P_x for subplot factors. The different factors could have different probabilities of changing levels. We may set the probability as 0.5 for each factor to accommodate as many designs as possible, or set the probability for each factor according to the extent that the factor is hard to change. The harder to change the factors will then have a smaller probability of changing levels. This rule is based on the assumption that the hard-to-change whole plot factors have levels that are more costly to change, thus the experimenter intends to reset the levels for this type of factor as small times as possible. The design selected is the best from among those sampled.

Although this sampling is frequently more realistic than examining all permutations to find the best design, further study was needed to find the probability of finding the optimal or near-optimal designs by sampling. The chance to get the best or highly efficient designs can be explored by examining the frequency of the distinct designs (in Table 2) in the design space. Simulation shows approximately equal chances of getting efficient designs by permutation to by sampling with $P_w=0.5$ for all variables and the results are displayed in Figure 1, where the frequency of the designs in Table 2 is plotted versus the design efficiency. In the plots, the x -axis gives the relative efficiency (RE), which shows how close the design is to the most efficient design among the 31 designs in terms of D -, V - and G -efficiency with or without cost penalty, and y -axis represents the accumulated frequency of obtaining the designs that are at least as efficient as the design with $RE=x$.

Relative D -efficiency is $D_{RE} = \frac{\mathbf{M}(D)}{\text{Max}_{D \in \Omega} |\mathbf{M}(D)|}$, where Ω is the design space including the 31 distinct designs, and the relative V - and G -efficiency are $V_{RE} = \frac{\text{Min}_{D \in \Omega} \{\text{average prediction variance}(D)\}}{\text{average prediction variance}(D)}$, and $G_{RE} = \frac{\text{Min}_{D \in \Omega} \{\text{maximum prediction variance}(D)\}}{\text{maximum prediction variance}(D)}$, respectively. For instance, if a point ($x=0.9, y=0.5$) is on the curve, then the chance of

obtaining a design that is at least 90% efficient is 50%. If the sampling technique is applied, suppose there are m generated designs in the sample, then 50% of the m designs have at least relative efficiency 0.9, which infers the sampling works well for searching good designs. In Figure 1, the top plots represent the evaluation not taking the cost of experiment into consideration, the middle plots correspond to the situation that the cost of whole plot dominates the experiment cost ($r=0$), and the bottom plots for the situations that whole plot is even expensive as subplot ($r=1$). The two variance ratios, $d=1$ and 10 are on the ledge and right sides, respectively.

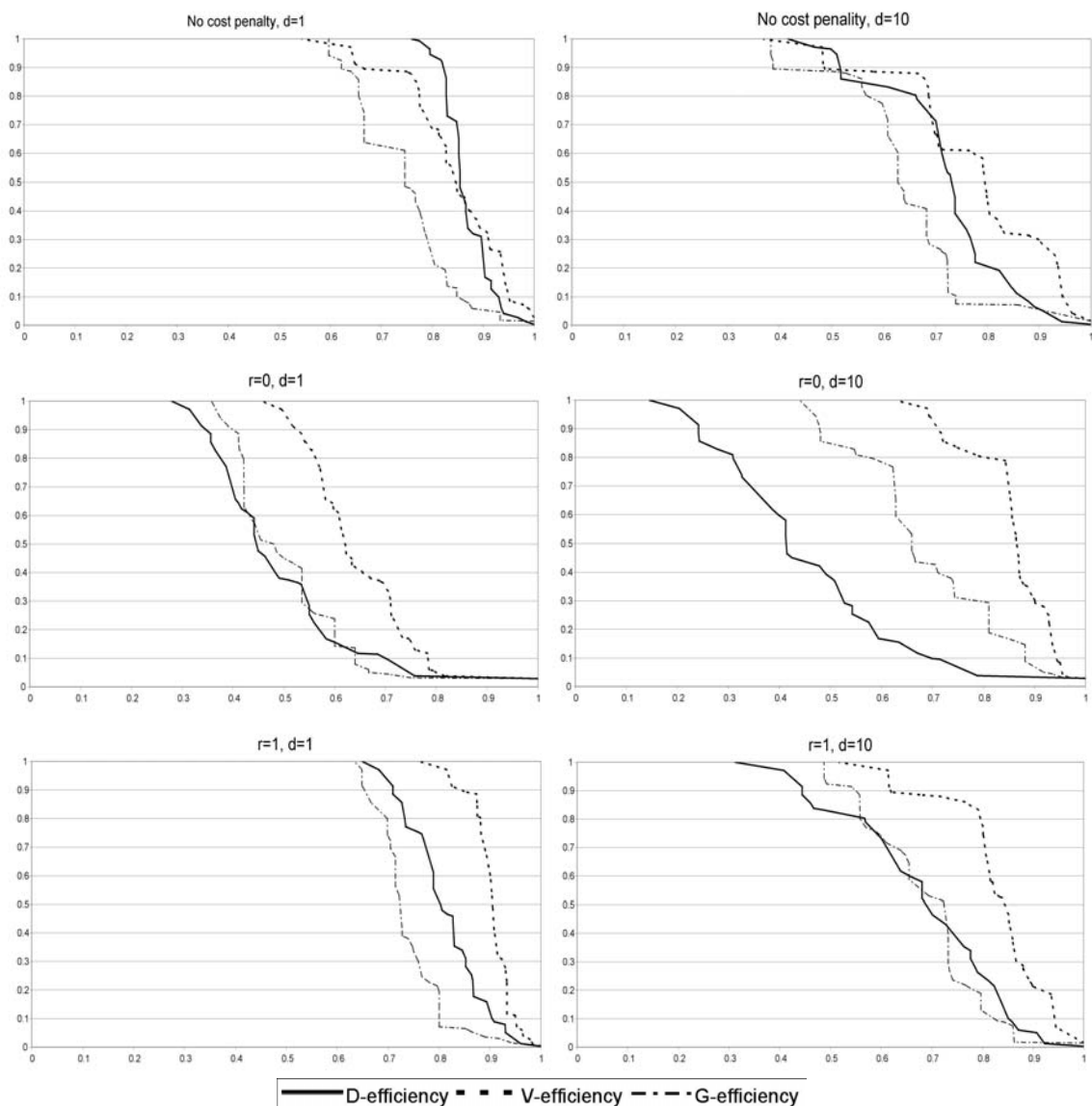


Figure 1: Accumulated frequency of getting highly efficient design for D -, G - and V -optimality under different combinations of cost ratio and variance component ratio.

In the plots, the performances of sampling for V -efficiency are good in most situations, for instance, we have 20%-80% chance to get designs at least 90% highly V -efficient, except for $d=1$ and $r=0$. For D - and G -efficiency, the performance of sampling is less efficient. When sample size increases, we might still obtain highly efficient design, but this is a less desirable choice. The chance of finding the highly efficient designs by sampling is substantially lower when whole plots are extremely expensive ($r=0$), which indicate that the inadvertent split-plot designs are not desirable under some conditions. However, if the desirable properties of superior split-plot designs are known, we may improve the chance to get highly efficient design by adjusting the probability of level changes for each factor. For instance, from the quality summary of the designs in Table 7 we know that the design with small number of whole plots is desirable when $r=0$. Therefore, by reducing the probability of changing whole plot levels appropriately, i.e., $P_w=0.1$ or 0.2 , the chance of obtaining more efficient designs would be enhanced. However, this technique is based on the knowledge of the desirable characteristics of optimal design. The more knowledge we have about the characteristics of good designs, more efficient the sampling can be made, when the exhaustive search of the best design in the design space is not feasible.

5. COST ADJUSTED EVALUATION FOR CENTER COMPOSITE DESIGNS FOR SECOND ORDER MODEL

For designs exploring the relationship of response with the factors or finding the optimal operation conditions for the factors, a second-order model should be considered. Some response surface designs, such as central composite design (CCD), Box-Behnken design (BBD), are popularly used, and are known to be highly efficient under a complete randomization structure. However, if the experiment is implemented with a split-plot structure, the designs can become less efficient for estimation of parameters and prediction of responses. The variance component ratio and structure of the variance matrix play important roles in performance. Moreover, the cost of the experiment is frequently an issue and is a function of the number of whole plots and subplots. The experimenter may want to balance between the quality of estimation and prediction of a

design and its economy. Letsinger, Myers, and Lentner (1996) compared popularly used response surface designs, such as CCD, BBD, small composite design and full factorial design, under split-plot scheme and concluded that central composite design is the most desirable design under various values of d for estimation and overall prediction over the design region. In this section, we focus on the study of different variations of central composite designs to show strategies for selecting split-plot designs when taking the cost into consideration.

In this example we consider an experiment with one whole plot variable and two subplot variables for a second order model with the fixed effect in the form $f(w, x_1, x_2) = \beta_0 + \beta_1 w + \beta_2 x_1 + \beta_3 x_2 + \beta_{12} w x_1 + \beta_{13} w x_2 + \beta_{23} x_1 x_2 + \beta_{11} w^2 + \beta_{22} x_1^2 + \beta_{33} x_2^2$. When central composite design is run as a SPD, the simplest way to implement is to put the observations with same levels of whole plot factors in the same whole plot, so each whole plot level has only one whole plot, which is called restricted split-plot design (RSPD) (Letsinger et al. 1996, Goos and Vandebroek 2004) and named D1 or “standard CCD” in this paper (Table 8). Because previously the properties of CCD in split-plot structure were not well understood, this has been a common choice of SPD in practice. However, because there is no replications for the whole plot levels, the whole plot pure error is non-estimable for this design. In addition, the design is quite unbalanced, where “balanced” for SPD means we have the same whole plot size for all whole plots. For the standard CCD, the size of the axial whole plots is one, while the size of whole plot with $w=0$ is four plus the number of subplot center runs. This might be wasteful if whole plots are limited, and possibly not appealing to the practitioner.

Table 8: D1 - standard CCD with one whole plot variable, w , and two subplot variables, x_1 and x_2 , axial levels for all variables are $\pm\alpha = \pm\sqrt{3} \approx \pm 1.732$, total number of runs is 16 ($N=16$), it has 5 whole plots ($a=5$) and the whole plot sizes are (4,4,1,1,6).

Whole plot	w	x_1	x_2	No. of Runs per whole plot
1	-1	± 1	± 1	4
2	1	± 1	± 1	4
3	1.732	0	0	1
4	-1.732	0	0	1
5	0	± 1.732	0	2
	0	0	± 1.732	2
	0	0	0	2

Vining, Kowalski and Montgomery (2004) recommended imposing minimum and maximum whole plot size restrictions. They proposed a modified CCD (given in Table 9), which satisfies the analysis condition that the generalized least square (GLS) estimates are equal to the ordinary least square (OLS) estimates. When whole plots are very expensive relative to the subplot cost, the total number of runs and whole plot sizes are not critical and only the number of whole plots may matter to the practitioner. Under this situation, they state the modified CCD can be a desirable design. In addition, because of the replications for the whole plot and subplot levels, the pure error can be estimated independently of the model. The quality of estimation and prediction of this design was not studied in their paper. Liang et al. (2004, 2005) used graphical tools to study and compare this design with the standard CCD. In terms of scaled prediction variance (SPV), the modified CCD performs poorly compared to the standard CCD.

Table 9: D2 - Modified CCD, total number of runs is 24, 6 whole plots and the whole plot sizes are (4,4,4,4,4,4) – balanced design.

Whole plot	w	x_1	x_2	No. of runs per whole plot
1	-1	± 1	± 1	4
2	1	± 1	± 1	4
3	1.732	0	0	4
4	-1.732	0	0	4
5	0	± 1.732	0	2
	0	0	± 1.732	2
6	0	0	0	4

Other variations of the CCD may also be of interest. Design D3 in Table 10 has the same number of runs as D1 but with one more whole plot. In D3, the six whole plot center runs is assigned into two whole plots in such a specific way to improve estimation and prediction for the whole plot and subplot terms. This is expected to bring benefit when the subplot cost dominates the cost of the experiment and the total number of runs is important.

Because the center runs help the estimation of the quadratics, design D4 in Table 11 augments one subplot center run within the whole plot factorial and center run. The whole plot axial runs are replicated three times. When the whole plots are extremely expensive, this design is expected to obtain better performance by increasing the number

of subplots accommodated in the whole plot while keeping the same number of whole plots.

Similarly, D5 in Table 12 is obtained by adding one subplot center run to each whole plot of D2, and some of the redundant replications are removed to control the total size of the design. It has the same size of observations and same number of whole plots as D2, but we will see that this design performs much better than the modified CCD.

Table 10: D3 – same number of observations as D1, but there is one more whole plot; total number of runs is 16, 6 whole plots and the whole plot sizes are (4,4,1,1,3,3).

Whole plot	w	x_1	x_2	No. Runs per whole plot
1	-1	± 1	± 1	4
2	1	± 1	± 1	4
3	1.732	0	0	1
4	-1.732	0	0	1
5	0	± 1.732	0	2
	0	0	0	1
6	0	0	± 1.732	2
	0	0	0	1

Table 11: D4 - $N=22$, 5 whole plots and the whole plot sizes are (5,5,3,3,6).

Whole plot	w	x_1	x_2	No. Runs per whole plot
1	-1	± 1	± 1	4
		0	0	1
2	1	± 1	± 1	4
		0	0	1
3	1.732	0	0	3
4	-1.732	0	0	3
5	0	± 1.732	0	2
		0	± 1.732	2
		0	0	2

Table 12: D5 – $N=24$, 6 whole plots and the whole plot sizes are (5,5,3,3,5,3).

Whole plot	w	x_1	x_2	No. Runs per whole plot
1	-1	± 1	± 1	4
		0	0	1
2	1	± 1	± 1	4
		0	0	1
3	1.732	0	0	3
4	-1.732	0	0	3
5	0	± 1.732	0	2
		0	± 1.732	2
		0	0	1
6	0	0	0	3

Different combinations of the variance component ratio, d , and the cost ratio, r , are presented below, which represent possible situations in practice where a split-plot experiment may be appropriate. The relative performance of the five candidates designs are of interest under these different scenarios. The cost penalty and quality are weighted in the evaluation for the following combinations: $d=0.5, 1$ and 10 and $r=0, 0.5$ and 1 . The cost penalization by N represents the standard scaled prediction variance evaluation, which is popularly used for CRD, corresponding to the situations that only subplot/measurement cost is important and whole plot cost is negligible.

Table 13 compares the five candidates in terms of cost adjusted D , G , and V -efficiency. The best and second best designs are in “Bold” fonts and identified by “*” and “+”, respectively. When $r=0$, the cost measure focuses only on the number of whole plots, D4 is the best or second best for estimation and prediction, and the superiority is true for all variance ratio values. This design has the least number of whole plots with competitive performance and thus is superior to other designs. Therefore, augmenting the replications of subplot levels within the whole plot is helpful if the resource of whole plots is extremely limited. D5 has the best overall prediction performance for moderate or large variance ratio values. This design is much better compared to modified CCD “D2”, which implies the strategy of putting the subplot axial points with center runs in the same whole plot helps.

Another extreme situation is that in the split-plot experiment the cost of whole plots are negligible and the cost of subplots/measurements are emphasized. Under this condition, D1 and D3 are the best and second best designs for estimation and prediction, respectively. This is the scenario where total number of runs of the experiment is most important.

In practice, more realistic situations are frequently between the above two extreme scenarios where both the cost of whole plots and cost of subplot have to be considered, such as, $r=0.1, 0.5$ or 1 . For the case $r=1$ or 0.5 , D1 is the best for D - and G -efficiency, and it has the second best average prediction (V -efficiency). D3 is the best for overall prediction and the second best for D - and G -efficiency. The two designs have same least total number of runs, which makes the design more competitive when the total size of design is important. D3 has one more whole plot than D1. We can see that high V -

efficiency designs have more whole plots than the high D - and G -efficiency designs. This is consistent with the conclusions of example 1 for the first-order model.

Table 13: The cost penalized D -, G - and V -efficiency, which are CPD, maximum and average cost penalized prediction variance (CPPV), for different combinations of d and r value, including the extreme case of penalization by the total number of runs, N . The best design and second best designs are in “Bold” fonts.

	Design	Cost penalization				
		$r = 0$	$r = 0.1$	$r = 0.5$	$r = 1$	N
$d=0.5$						
CPD	D1 (4,4,1,1,6)	0.51+	0.384+	0.195*	0.121*	0.158*
	D2 (4,4,4,4,4,4)	0.47	0.334	0.156	0.093	0.117
	D3 (4,4,1,1,3,3)	0.42	0.327	0.178	0.113+	0.156+
	D4 (5,5,3,3,6)	0.582*	0.404*	0.182+	0.108	0.132
	D5 (5,5,3,3,5,3)	0.502	0.358	0.167	0.100	0.125
Average CPPV	D1 (4,4,1,1,6)	2.3	3.04	5.98+	9.661+	7.36+
	D2 (4,4,4,4,4,4)	2.218	3.11	6.65	11.089	8.872
	D3 (4,4,1,1,3,3)	2.531	3.21	5.91*	9.28*	6.749*
	D4 (5,5,3,3,6)	1.966*	2.83+	6.29	10.617	8.65
	D5 (5,5,3,3,5,3)	2.004+	2.805*	6.012	10.020	8.016
Maximum CPPV	D1 (4,4,1,1,6)	3.737	4.933	9.716*	15.695*	11.958+
	D2 (4,4,4,4,4,4)	3.649	5.108	10.946	18.244	14.595
	D3 (4,4,1,1,3,3)	4.433	5.615	10.343	16.253+	11.820*
	D4 (5,5,3,3,6)	3.047*	4.388*	9.751+	16.454	13.407
	D5 (5,5,3,3,5,3)	3.458+	4.841+	10.373	17.289	13.831
$d=1$						
CPD	D1 (4,4,1,1,6)	0.598+	0.453+	0.23*	0.142*	0.187*
	D2 (4,4,4,4,4,4)	0.507	0.362	0.169	0.102	0.127
	D3 (4,4,1,1,3,3)	0.482	0.381	0.207	0.132+	0.181+
	D4 (5,5,3,3,6)	0.666*	0.463*	0.208+	0.123	0.151
	D5 (5,5,3,3,5,3)	0.571	0.408	0.190	0.114	0.143
Average CPPV	D1 (4,4,1,1,6)	2.331	3.077	6.062+	9.792+	7.459+
	D2 (4,4,4,4,4,4)	2.351	3.291	7.053	11.755	9.404
	D3 (4,4,1,1,3,3)	2.455	3.11	5.728*	9.002*	6.547*
	D4 (5,5,3,3,6)	2.065+	2.973+	6.607	11.148	9.086
	D5 (5,5,3,3,5,3)	2.028*	2.840*	6.085	10.141	8.113
Maximum CPPV	D1 (4,4,1,1,6)	3.915	5.168	10.180+	16.445+	12.529+
	D2 (4,4,4,4,4,4)	3.75	5.25	11.25	18.75	15
	D3 (4,4,1,1,3,3)	4.645	5.883	10.838	17.031	12.386*
	D4 (5,5,3,3,6)	2.980*	4.291*	9.535*	16.091*	13.111
	D5 (5,5,3,3,5,3)	3.310+	4.634+	9.930	16.550	13.240
$d=10$						
CPD	D1 (4,4,1,1,6)	1.854+	1.405*	0.713*	0.442*	0.579*
	D2 (4,4,4,4,4,4)	1.203	0.859	0.401	0.241	0.301
	D3 (4,4,1,1,3,3)	1.455	1.149	0.623+	0.397+	0.546+
	D4 (5,5,3,3,6)	1.956*	1.358+	0.611	0.362	0.445
	D5 (5,5,3,3,5,3)	1.656	1.183	0.552	0.331	0.414

Average CPPV	D1 (4,4,1,1,6)	2.356	3.109	6.124+	9.893+	7.539+
	D2 (4,4,4,4,4,4)	2.677	3.748	8.072	13.387	10.708
	D3 (4,4,1,1,3,3)	2.149+	2.722*	5.013*	7.878*	5.731*
	D4 (5,5,3,3,6)	2.303	3.316	7.368	12.434	10.133
	D5 (5,5,3,3,5,3)	2.058*	2.881+	6.173	10.289	8.231
Maximum CPPV	D1 (4,4,1,1,6)	4.193+	5.535*	10.901*	17.610*	13.417+
	D2 (4,4,4,4,4,4)	5.591	7.827	16.773	27.955	22.364
	D3 (4,4,1,1,3,3)	4.982	6.311	11.626+	18.269+	13.286*
	D4 (5,5,3,3,6)	3.972*	5.720+	12.710	21.448	17.476
	D5 (5,5,3,3,5,3)	4.723	6.612	14.168	23.614	18.891

*: best designs; +: second best designs.

The additional center runs to the axial levels in the same whole plot help with the estimation of quadratic terms. Splitting the $w=0$ whole plot of D1 into two whole plots for D3 does not substantially compromise the estimation of the subplot quadratic terms, however, it does improve the estimation of the whole plot quadratics. On the other hand, although the modified CCD (D2) has good properties from analysis standpoint, the quality of estimation and prediction is poor, because the subplot axial points are assigned separately from subplot center runs and hence the estimation of subplot quadratics is with whole plot error. However, by adjusting the locations of the designs points, augmenting subplot center run to the whole plot with subplot axial points or reducing the redundant replications of design points as D5 does, the performance can be improved.

Liang et al. (2004, 2005) studied the prediction performance of split-plot designs using three-dimensional variance dispersion graphs (3D VDGs) and fraction of design space (FDS) plots. These graphical tools show the best and weakest prediction over the entire design region and for any particular sub region. In Figure 2, the global FDS plots for the five designs are displayed. See Zahran, Anderson-Cook and Myers (2003) for more details on FDS plots. The horizontal “FDS” axis represents the fraction of the design space with cost adjusted prediction variance at or below the given values by the vertical “CPPV” axis. Therefore, the maximum prediction variance is displayed by the y value at $FDS=1$ and the average of the curve shows the average prediction variance over the entire region, and thus the values of G and V -efficiency summarized in Table 13 can be read from the plots. For the case $d=1$ and $r=0$, D5 is the best design with the FDS curve having the smallest values and a flatter slope. As whole plot variability increases relatively to the subplot variance, D5 is still the most desirable design, which shows the

choice of the best design is robust to the changes of variance component ratio. D3 with the 6 whole plots and 16 observations performs much better for $d=10$ and $r=0$, because the larger size of whole plots helps the performance of SPDs when the proportion of whole plot variance in the system increases. This design is superior to the other designs under the situations that subplots are relatively expensive. Meanwhile, the superiority of the design over others is robust to the changes of d value.

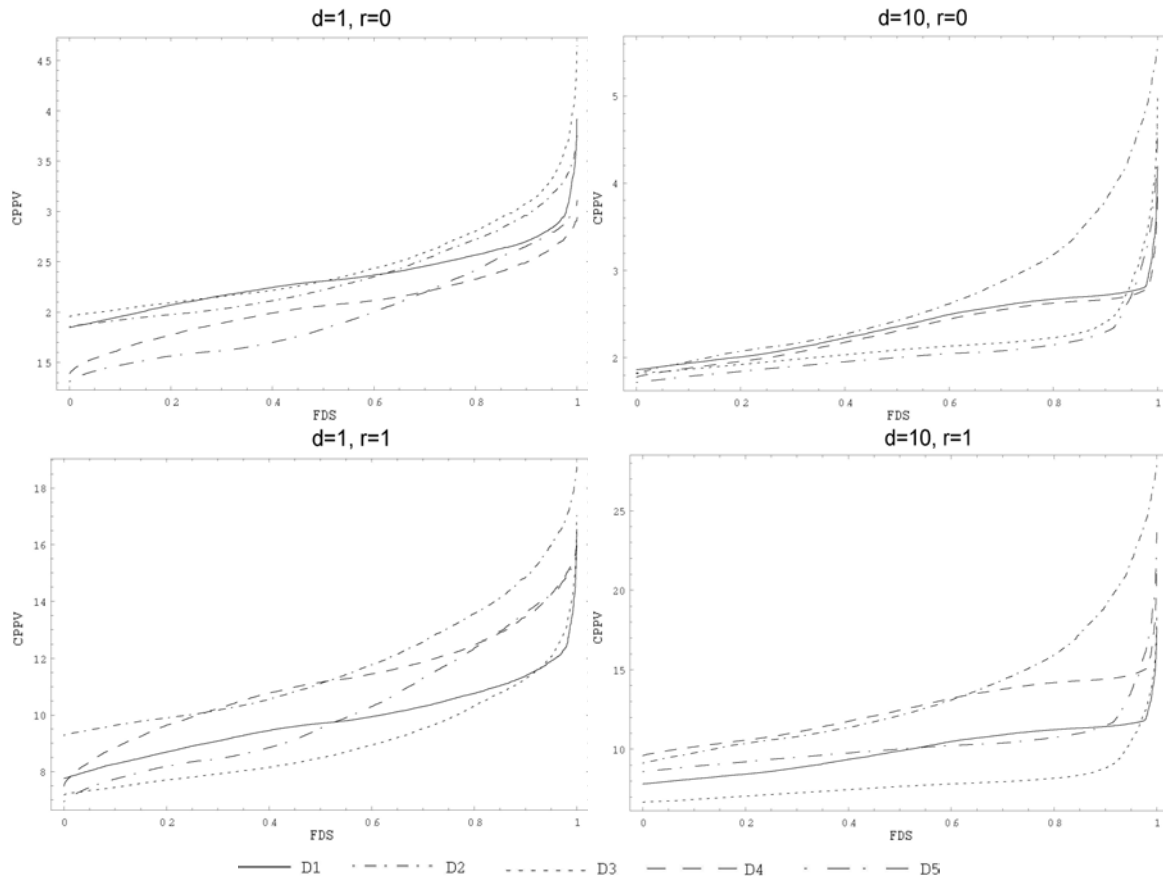


Figure 2: FDS plots for the five candidate designs under different situations of cost ratio and variance ratio

From the surface plots of 3D VDG in Figure 3, we can see the distribution of the cost adjusted prediction variance for the two best designs under different cost and variance structures for split-plot design. See Liang et al. (2004, 2005) for more details on these plots. In the plots, “ w ” indicates the distance of the location from the center in the whole plot space, and “ x ” represents the location in the subplot space. The vertical axis tells the maximum cost adjusted prediction variance at that location in the combined

space. The plots show a common characteristic of the prediction variance distribution for desirable designs, namely that the prediction variance is stable and relatively small in the broad center area and prediction becomes worse for a small portion at the edge of the whole plot space. These plots can be helpful to understand the advantage and weakness of the designs for predicting response.

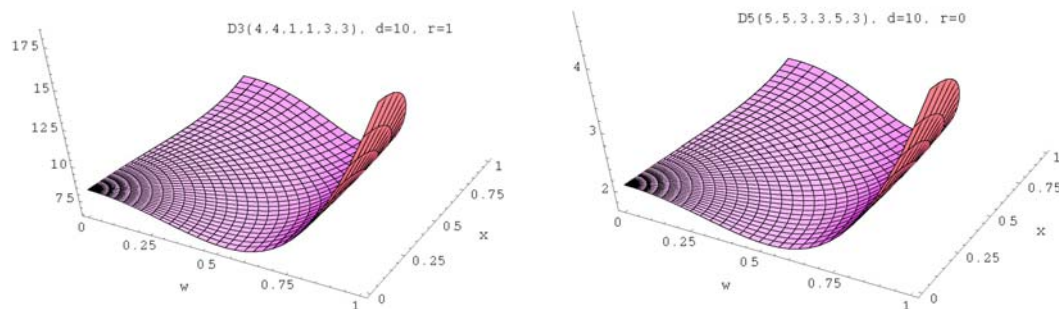


Figure 3: Surface plots of maximum CPPV for D3 and D5 at different scenarios

From comparisons of the five designs one can learn that under different weighting conditions of cost and quality, the best design differs significantly. When the whole plots are extremely expensive, one may set the number of whole plots as the largest number available and achieve the best performance by assigning as much subplots as possible in each whole plot. When the subplot/measurement are comparably expensive to the cost of whole plots, the design with fewer runs is desirable. When the subplot/measurement cost dominates and cost of whole plot is negligible, the scaled prediction variance can be used to evaluate the designs, and smaller sized designs are desirable.

In addition, the example provides helpful information for practitioner to choose a split-plot design for second order model. The standard CCD has good performance when cost is incorporated, which provides theoretical support for extending this type of response surface design from completely randomized design to more complicated experiments in the way of restricted split-plot structure. Moreover, when the whole plot size is limited, for instance, the maximal number of subplot accommodated in the whole plot can't exceed six in practice, running all combinations of subplot factors within each whole plot is not feasible, and thus we might have incomplete subplot levels within whole plots. The intuition would lead to the most balanced setting in the whole plots, such as the modified CCD does. However, this example shows that exact balanced or

symmetric setting of the subplot levels is less efficient. For instance, it is better to assign the subplot center runs with axial or factorial levels in the same whole plot rather than separating the axial points and center runs.

6. CONCLUSIONS AND DISCUSSION

For different problems or under different conditions of the experiment in real life, the split-plot experiment designs may focus on different aspects of the performance or cost. Incorporating the split-plot structure and cost structure into the evaluation of split-plot design is helpful to better understand their effect on the desirability of the design. The proposed cost penalized D -efficiency, average and maximum cost penalized prediction variance (cost adjusted V and G -efficiency), including a special case of the scaled prediction variance, provide strategies for the practitioners rather than choosing the designs arbitrarily based on the available resources. The different weighting system between the practitioner's interest on the cost and quality of estimation and prediction requires the experimenter to evaluate relative costs, as measured in time, effort or money for the whole plot and subplot units, for changing the levels of two types of experimental variables and for measuring the observations, based on the understanding of the practical conditions for running the experiments. This allows for more realistic design selection. From the study, we also learn that the standard scaled prediction variance evaluation for completely randomized design only makes sense under some special cases for split-plot design, and thus the generalization of SPV from CRD to SPD should be done carefully.

Although the constraints on time/cost require the practitioner to use as small size of design as possible, other desirable properties, like the ability to estimate the whole plot and subplot error terms, and affordable precision for the whole plot factors should also be taken into account.

In industrial experiments, an important problem for the practitioner is to select a response surface design with a desirable structure when there are restrictions on randomization. This study shows that by adapting central composite designs in a variety of ways, can help improve performance for different cost and variance ratio situations. Some desirable strategies for assigning subplot levels within whole plots are also

provided, which argue against the intuition that balanced design is always preferable. However, there may be some benefits during analysis.

The estimation of cost ratio can be obtained from understanding of the conditions for a given split-plot experiment from experienced scientists or engineers. The variance component ratio is probably not as easy to estimate, if a pilot study or previous data are not available. However, the study in the two examples implies that the choice of a highly efficient design is frequently quite robust to the change of variance component ratio value. If the guessed or estimated d value is slightly different from the actual value in the experiment, we can still search for the optimal design based on this value and guarantee the obtained design highly efficient. This robustness means that good performance is likely even when the split-plot design is selected based on limited information about the variance component ratio.

References:

1. Anbari, F. T. and Lucas, J. M. (1994). "Super-Efficient Designs: How to Run Your Experiment for Higher Efficiency and Lower Cost". *ASQC Technical Conference Transactions*, pp. 852-863.
2. Bingham, D. and Sitter, R. S. (1999). "Minimum-Aberration Two-Level Fractional Factorial Split-Plot designs". *Technometrics*, 41, pp. 62-70.
3. Bisgaard, S. (2000). "The design and analysis of $2^{k-p} \times 2^{q-r}$ split-plot experiments". *Journal of Quality Technology*, 32, pp. 39-56.
4. Bisgaard, S. and Steinberg, D. M. (1997). "The design and Analysis of $2^{k-p} \times s$ Prototype Experiments". *Technometrics* 39, pp. 52-62.
5. Dompere (2004) ???
6. Ganju, J. and Lucas, J. M. (1999). "Detecting randomization restrictions caused by factors". *Journal of statistical planning and inference*, 81, pp. 129-140.
7. Goos, P. and Vandebroek, M. (2001). "Optimal Split-Plot Designs". *Journal of Quality Technology*, 33, No. 4, pp. 436-450.
8. Goos, P. and Vandebroek, M. (2004). "Outperforming Completely Randomized Designs". *Journal of Quality Technology*, 36, No. 1, pp. 12-26.

9. Huang, P., Chen, D., and Voelkel, J. O. (1998). "Minimum-Aberration Two-Level Split-Plot Designs". *Technometrics*, 40, pp. 314-326.
10. Joiner, B. L. and Campbell, C. (1976). "Designing Experiments When Run Order is Important". *Technometrics*, Vol 18, No. 3, pp. 249-259.
11. Ju, H. L. and Lucas, J. M. (2002). "L^k Factorial Experiments With Hard-To-Change and Easy-To-Change Factors". *Journal of Quality Technology*, 34, No. 4, pp. 411-421.
12. Kowalski, S. M., Cornell, J. A. and Vining, G. G. (2002). "Split-Plot Designs and Estimation Methods for Mixture Experiments with Process Variables". *Technometrics* 44, pp. 72-79.
13. Letsinger, J. D., Myers, R. H., and Lentner, M. (1996). "Response Surface Methods for Bi-Randomization Structure". *Journal of Quality Technology*, 28, pp. 381-397.
14. Liang, L., Anderson-Cook, C. M., Robinson, T., and Myers, R. H. (2004), "Three-Dimensional Variance Dispersion Graphs For Split-Plot Designs," Technical report 04-4, Dept. of Statistics, Virginia Tech, Blacksburg, VA.
15. Liang, L., Anderson-Cook, C. M., Robinson, T., and Myers, R. H. (2005), "Fraction of Design Space Plots For Split-Plot Designs," Technical report 05-2, Dept. of Statistics, Virginia Tech, Blacksburg, VA.
16. Myers, R. H. and Montgomery, D. C. (2002). *Response Surface Methodology*. John Wiley & Sons, Inc.: New York.
17. Vining, G. G., Kowalski, S. and Montgomery, D. C. (2004). "Response Surface Designs within a Split-plot Structure". *Journal of Quality Technology* (in press).
18. Webb, D., Lucas, J. M. and Borkowski, J. J. (2004). "The Prediction Variance and Design Strategies for Factorial Experiments When Factors Are Not Reset". *Journal of Quality Technology*,
19. Zahran, A. R., Anderson-Cook, C. M., and Myers, R. H. (2003). "Fraction of Design Space to Assess Prediction Capability of Response Surface Designs". *Journal of Quality Technology*, 35, pp. 377-386.