Response Surface Design and Analysis in the Presence of Restricted Randomization

Peter A. Parker

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

G. Geoffrey Vining, co-chair
Scott M. Kowalski, co-chair
Christine M. Anderson-Cook
Marion R. Reynolds, Jr.
William H. Woodall

March 29, 2005
Blacksburg, Virginia

Keywords: Response Surface Methodology, Restricted Randomization, Split-Plot Designs, Equivalent Estimation, Central Composite, Box-Behnken

Response Surface Design and Analysis in the
Presence of Restricted Randomization

Peter A. Parker

ABSTRACT

Practical restrictions on randomization are commonplace in industrial experiments due to the presence of hard-to-change or costly-to-change factors. Employing a split-plot design structure minimizes the number of required experimental settings for the hard-to-change factors. In this research, we propose classes of equivalent estimation second-order response surface split-plot designs for which the ordinary least squares estimates of the model are equivalent to the generalized least squares estimates. Designs that possess the equivalence property enjoy the advantages of best linear unbiased estimates and design selection that is robust to model misspecification and independent of the variance components. We present a generalized proof of the equivalence conditions that enables the development of several systematic design construction strategies and provides the ability to verify numerically that a design provides equivalent estimates, resulting in a broad catalog of designs. We explore the construction of balanced and unbalanced split-plot versions of the central composite and Box-Behnken designs. In addition, we illustrate the utility of numerical verification in generating $\mathcal{D}$-optimal and minimal point designs, including split-plot versions of the Notz, Hoke, Box and Draper, and hybrid designs. Finally, we consider the practical implications of analyzing a near-equivalent design when a suitable equivalent design is not available. By simulation, we compare methods of estimation to provide a practitioner with guidance on analysis alternatives when a best linear unbiased estimator is not available. Our goal throughout this research is to develop practical experimentation strategies for restricted randomization that are consistent with the philosophy of traditional response surface methodology.

This work received support from the National Aeronautics and Space Administration's Langley Research Center under the Graduate Studies Program.

Acknowledgments

I wish to express my gratitude to Dr. G. Geoffrey Vining and Dr. Scott M. Kowalski for their guidance, encouragement, and inspiration throughout this research. In addition, I thank Mr. Paul W. Roberts and Mr. Ray D. Rhew, whose support made focused graduate study possible. I would especially like to thank my loving wife, Suzanne, and children, Samantha, Benjamin, Rebekah, Philip, James, and Grace for their prayers, encouragement, patience, and understanding during my studies. Ultimately, I praise God for the opportunities that He has provided me to pursue graduate study. By His grace alone, I have been able to complete this lofty endeavor.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Response Surface Methodology (RSM) is used to characterize and optimize industrial products and processes. Early development of RSM was based on the well-established use of experimental designs for agricultural research. RSM methods exploited the differences between the execution of industrial and agricultural experiments. For example, an agricultural experiment may require an entire growing season to obtain experimental data and subsequently analyze the treatment effects. Conversely, the response from an industrial experiment can be obtained soon after the experiment is conducted; in many cases immediately. Furthermore, many industrial experiments can be conducted in a relatively short period of time, for example days or weeks. In the industrial context, the analysis of small exploratory experiments can be used to guide the design of larger more comprehensive experiments. As each experiment is conducted, knowledge of the experimental system increases. We combine the experimental results with engineering knowledge to determine the next phase. Box (1999) characterizes this process as *sequential learning* in which we start with an initial idea and move between inductive and deductive reasoning to learn what factors are important to our investigation and the settings that are required to improve the response.

There are also differences in the type of factors studied and the analysis methods employed in agricultural and industrial experiments. Agricultural experiments more commonly involve qualitative or categorical variables. In these experiments, an analysis of variance (ANOVA) approach is used to identify significant factors, and contrasts are constructed to study the nature of the factor effects. Alternatively, in industrial experiments, quantitative factors are typically involved, whose levels are carefully chosen from a continuous experimental region. In industrial experiments, regression models are built to provide a parametric equation suitable for response prediction at any level within the experimental region; not restricted to the specific levels included in the experiment. A second-order Taylor series approximation is commonly employed in RSM models, given by

$$E(y) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j \tag{1.1}$$

where y is the response, the β's are the regression coefficients, and the x's are the explanatory variables.

Box and Wilson (1951) recognize the unique opportunities that are available for industrial experiments. They propose adaptations of classical agricultural methods and introduce new experimental strategies. Based on this work, statistical methods have been developed to efficiently plan, execute, analyze, and optimize industrial experiments.

Similarly, this area of research builds on an agricultural design technique known as split-plotting. In agronomic experiments, it is common to have large areas of land subdivided into smaller areas. The large areas are known as whole-plots (WP), and the smaller areas are known as subplots (SP). This sequential subdivision of the land is performed to accommodate the restrictions on the experimental treatments. Some treatments are restricted to large areas, such as an irrigation technique, and are referred to as whole-plot treatments. Alternatively, other treatments allow for a local application to smaller areas, such as seed variety, and are known as subplot treatments. (In agricultural experiments, we use the term treatment because the variables are typically categorical or qualitative in nature. In industrial experiments, quantitative variables are more commonly studied and are referred to as factors rather than treatments.) In general, we can think of a split-plot design as a nesting of two experimental designs with one design based on the whole-plot factors and the other based on the subplot factors. A split-plot design is a generalization of a factorial experiment in which we have two types of experimental units (EU). We define an experimental unit as the smallest unit in which a treatment, or factor combination, can be applied independently of any other treatment. The observational unit (OU) is the unit on which we observe, or measure, the response. However, the observational unit is not necessarily the same as the experimental unit. In a split-plot experiment, the EU for a subset of factors is an observational unit for another subset of factors.

To illustrate the relationship between the EU and OU, consider an agricultural experiment investigating the effect of two irrigation methods and three varieties of seed on crop yield. In this case, our highest level EU's are the large areas of land (WP's) that are randomly assigned an irrigation method. Each whole-plot is subdivided into smaller areas of land (SP's) generating our next level of EU's. Each subplot is randomly assigned a specific variety of seed. We will consider the measure of yield to be the total plant mass for each subplot. Therefore, the EU unit for the irrigation method is the whole-plot and its OU's are the subplots within each whole-plot. The EU for the seed variety is the subplot, which is also its OU. From this example, we see that the OU for the whole-plot factor is the EU for the subplot factor. In addition, note that in a split-plot structure there are more EU's available for the subplot factors than the whole-plot factors.

When planning a split-plot experiment, we see that two randomizations are performed. First, the whole-plot factors are randomly assigned to the whole-plot experimental units generating the whole-plot error variance, σ_δ^2. Then, the subplot factors are randomly assigned to experimental units within each whole-plot and generates the subplot error variance, σ_ϵ^2. The presence of these two error terms creates an experimental situation that is more complex than a typical completely randomized factorial design. Note that the primary reason that we employ a split-plot strategy is to reduce the manipulation of the whole-plot factors. As a result, we have fewer degrees of freedom (df) for the estimation of the whole-plot error variance, which decreases our power to detect significant effects. In comparison, we have relatively high power for the subplot effects since we have more degrees of freedom available from replication of the subplot factors. If we required equal power for all factors, then we would accept the expense of performing the experiment in a completely randomized structure. If we employ a split-plot structure in cases where a completely randomized design is a viable alternative, then we are implicitly considering the subplot factors as more important than the whole-plot factors.

A common situation in industrial experiments involves restrictions on randomization. In many practical situations, there are factors which are difficult, time consuming, or costly to manipulate, referred to as hard-to-change (HTC) factors. Alternatively, some factors may be relatively easy to manipulate, referred to as easy-to-change (ETC) factors. This type of situation arises in batch type processes and in the setting of environmental factors such as temperature, pressure, or humidity. Traditional RSM techniques do not distinguish between these two different types of factors. In fact, all factors are assumed to be equivalent in terms of ease of manipulation and cost of experimentation. In addition, the factors are considered to be of equal importance. Beyond the difficulty in setting the HTC factors, many times there are also sequential restrictions on the levels of these factors. For example, in an experiment with a temperature factor, complete randomization of the experimental run order may require an alternating setting of temperature between its high and low levels that would likely be infeasible.

It is common for a practitioner to order the runs of a particular experiment to achieve the minimum number of level changes of the HTC factors. While this strategy is intuitive and cost effective, it must be properly considered in the analysis of the experimental data. Failure to account for the lack of complete randomization can lead to inappropriate inference about factor significance, as will be discussed in more detail later in this dissertation. Furthermore, the simple reordering of experimental runs may not produce an efficient experimental plan. In this situation, an industrial split-plot approach can be employed. For a more detailed development of split-plot designs see Cochran and Cox (1957), Montgomery (2001), and Myers and Montgomery (2002).

1.2 Applications

The literature presents several examples of industrial or scientific investigations that require a split-plot structure. Box and Jones (1992) consider the design of products that are robust to environmental conditions, known as the robust parameter design (RPD) problem. Their application involved a manufacturer of a cake mix recipe that would be robust to the actual cooking time and temperature. In their example, the environmental factors of time and temperature were considered to be whole-plot factors and are included in the design to determine their interaction with other controllable factors related to the cake mix formulation. Vining, Kowalski, and Montgomery (2005) presents an example of environmental process variables as HTC factors in the sintering process for nickel battery plates used in a nickel-hydrogen battery. In this application, the furnace temperature and belt speed were considered to be HTC factors due to the time required to set their levels. Once the HTC factors are set, two other ETC factors can be manipulated. Box and Jones (1996) provide an example of a batch process involving the application of coatings to steel bars designed to increase corrosion resistance. In this example, the setting of a particular furnace temperature, known as a “heat”, was the HTC factor (whole-plot factor), and the position of the steel bars within the furnace was an ETC factor. In this way, we can consider the furnace “heat” as a batch in which various positions can be randomly assigned.

Kowalski and Potchner (2003) present an application related to the image quality of a printing press. Three factors are of interest; blanket type, cylinder gap, and press speed. The setting of the blanket type requires a partial disassembly of the printing press, which is time consuming and costly, and therefore is considered to be a hard-to-change factor. In contrast, the setting of the cylinder gap and press speed are controlled by dials and are easy to manipulate. This illustrates a HTC factor that is quite different from the usual examples of environmental variables of temperature, pressure, and humidity.

Simpson, Kowalski, and Landman (2004) consider an automotive wind tunnel experiment involving a race car. They study the impact of four factors (front ride height, rear ride height, yaw angle, and grille configuration) on the aerodynamic performance of the car. In this study, the front and rear ride height factors were considered to be hard-to-change in the sense that they required extensive time to manipulate. In contrast, the yaw angle and grille configuration were easy to manipulate. The goal of the study was to find settings of these four factors to optimize the aerodynamic performance of the vehicle. Furthermore, a second-order regression model was developed from the wind tunnel tests that could be used to optimize the performance for a given set of track conditions. In addition, they highlight many practical experimentation and analysis issues related to split-plot experiments.

Parker, Morton, Draper, and Line (2001) and Parker and DeLoach (2001) study the application of RSM applied to a calibration experiment. The application involves the calibration of a high-precision force measurement device, known as a force balance, that is used in wind-tunnel experiments. A force balance is a multiple-axis load cell that provides the simultaneous measurement of three orthogonal components of aerodynamic force (normal, axial, and side force) and three orthogonal components of aerodynamic torque (rolling, pitching, and yawing moments) exerted on a scaled aircraft test article. The force balance consists of a structural element instrumented with a network of strain gauges and is designed to elastically deform under the application of external forces and moments. This deformation results in differential strain across the structure that is sensed by the strain gauges. Six electrical responses are produced by the strain gauges that are proportional to the magnitude and direction of their respective aerodynamic components. A calibration experiment is performed to develop a prediction-oriented second-order model of the relationship between the applied forces and moments (explanatory variables) and the electrical responses.

Due to the extreme temperature environment in which a force balance is subjected during wind-tunnel testing, a model of the response as a function of temperature and applied force has been proposed. The force balance responds to temperature because the strain gages produce an undesirable output caused by the differential thermal expansion of the balance structure. Temperature differentials across the balance induce strain that is indistinguishable from an externally applied force. That is, varying temperatures on the balance will cause the balance to microscopically deform producing an electrical response that indicates the application of an external load, when in fact none has been applied. A temperature-based calibration is performed by conditioning the calibration apparatus to a specified temperature level, factor z. Then combinations of forces and moments are applied to the force balance, factors x_1 through x_6, and the responses are recorded. To achieve a single isothermal condition, one level of z, requires 2-4 hours of conditioning. In contrast, a single combination of forces and moments can be applied within 5 minutes. As a result, only one temperature setting can be achieved in a normal eight-hour workday, however many combinations of the x's can set while at this temperature.

In this application, we see that temperature is the hard-to-change factor, and the applied loads, x's, are relatively easy-to-change. Therefore, the temperature levels form the whole-plots, and the combinations of the applied loads form the subplot runs. This is a 7-factor experiment, and a second-order model is assumed. From a practical perspective, a split-plot experiment in which we change the level of temperature as infrequently as possible is desired. The cost of the experiment is primarily a function of the number of whole-plots performed, rather than the total number of subplot runs. We would like to exploit the ability to perform many subplot runs once a particular temperature level is set. This example illustrates a practical application of a high-dimensional response surface design conducted within a split-plot structure.

1.3 General Form of the Split-Plot Model

We define the general form of the model as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta} + \boldsymbol{\epsilon} \ , \tag{1.2}$$

where $\mathbf{y}$ is the $N \times 1$ vector of responses, $\mathbf{X}$ is the $N \times p$ model matrix, $\boldsymbol{\beta}$ is the $p \times 1$ vector of coefficients, $\boldsymbol{\delta}$ is the $N \times 1$ vector of random whole-plot errors, $\boldsymbol{\epsilon}$ is the $N \times 1$ vector of random subplot errors. We assume that $\boldsymbol{\delta} + \boldsymbol{\epsilon}$ has a mean of $\mathbf{0}$ and variance-covariance matrix

$$\boldsymbol{\Sigma} = \sigma_\epsilon^2 \mathbf{I} + \sigma_\delta^2 \mathbf{J}_b,$$

where σ_ϵ^2 is the subplot error variance, and σ_δ^2 is the whole-plot error variance. In the balanced case (same number of subplot runs for each whole-plot), we can express the $N \times N$ matrix $\mathbf{J}_b$ as

$$\mathbf{J}_b = \mathbf{I}_m \otimes \mathbf{J}_n,$$

where m is the number of whole-plots, n is the number of subplot runs within each whole-plot, $\mathbf{J}_n$ is a $n \times n$ matrix of ones, and N is the total number of subplot runs. For an unbalanced design, it is more convenient to formulate $\mathbf{J}_b$ as a function of the $N \times p$ incidence matrix $\mathbf{Z}$ as

$$\mathbf{J}_b = \mathbf{Z}\mathbf{Z}'.$$

In general, the matrix $\mathbf{J}_b$ has the form of

$$\mathbf{J}_b = \begin{bmatrix} \mathbf{1}_{n_1}\mathbf{1}'_{n_1} & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{n_2}\mathbf{1}'_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_{n_m}\mathbf{1}'_{n_m} \end{bmatrix},$$

where n_i is the number of subplot runs in the i^{th} whole-plot and denotes the length of the vector of ones, $\mathbf{1}$. In the balanced case, $n_i = n$ for $i = 1, \ldots, m$.

Examining the structure of $\mathbf{J}_b$, reveals that observations in different whole-plots are independent, while observations within a whole-plot are correlated. It is convenient to define the ratio of whole-plot to subplot error variance as

$$d = \frac{\sigma_\delta^2}{\sigma_\epsilon^2}.$$

We can express $\mathbf{\Sigma}$ in terms of d as

$$\mathbf{\Sigma} = \sigma_\epsilon^2 \left[\mathbf{I} + d\mathbf{J}_b\right]. \tag{1.3}$$

The ordinary least squares (OLS) estimate of the model coefficients is

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

and the variance-covariance matrix for $\hat{\boldsymbol{\beta}}_{OLS}$ is $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$. The generalized least squares (GLS) estimate is

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y},$$

and the variance-covariance matrix $\hat{\boldsymbol{\beta}}_{GLS}$ is $(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$. When OLS estimation is equivalent to GLS, the variance-covariance matrix for the OLS estimate is $(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$, which is the same as for GLS.

Chapter 2

Literature Review

Myers *et al.* (2004) provide a broad review of current RSM research related to restricted randomization. A subset of this body of literature that is closely related to this dissertation is presented here. Following this overview, four foundational articles are discussed in detail, and the primary contribution of each article is highlighted.

There is a significant body of literature related to split-plot designs used in agricultural experiments. Literature related to their application in industrial experiments is more modest. Wooding (1973) considers the use of a split-plot design to study the effect of three chemical compounds on the sizing properties of paper. He considers the case of inadvertent split-plotting, discussed in more detail later in this review, and the necessity of performing an analysis that properly considers the multiple error terms. Hahn (1978) proposes split-plot designs as a systematic approach to experiments with hard-to-change factors in industry. He described these experimental plans as a *compromise* between complete randomization and failing to randomize, which does not provide protection from extraneous variables that could bias the experimental results. Taguchi (1987) discusses experimental designs with hard-to-change factors and promotes the use of restricted randomization. Box (1996) discusses split-plot experiments as the norm rather than the exception in industrial experimentation. Furthermore, he highlights the benefits of a split-plot design in terms of the increased precision in studying the subplot factors even in the presence of highly variable whole-plot factors and the economy over a completely randomized plan. Moreover, he emphasizes the importance of recognizing and accounting for the split-plot structure in the data analysis. The majority of the industrial literature is limited to a study of first-order or first-order with interaction models that are well-suited for screening designs.

The construction of split-plot designs has been primarily considered from two perspectives; that of minimum aberration and $\mathcal{D}$-optimality. Overall, various types of minimum aberration (MA) has been the dominant criterion used in design construction. We can think of the minimum aberration design as achieving the least amount of departure from an orthogonal design within a class of designs defined by certain restrictions, such as the number of whole-plot factors, subplot factors, and the whole-plot size. Minimum aberration designs seek to find the *best* alias structure by minimizing the number of low-order aliased effects for a first-order with interaction model. Huang, Chen, and Voelkel (1998) consider two computational search methods to construct fractional factorial split-plot designs. The first method starts with the construction of a minimum aberration design in the whole-plot factors and then seeks a design in the subplot factors (easy-to-change) that achieves minimum aberration at this level. Their second method attempts to find overall minimum aberration designs using an integer programming technique. In addition, they provide an extensive catalog of 16, 32, and 64 subplot run designs based on their first method of construction. In a similar computational effort, Bingham and Sitter (1999a) employ an algorithm that combines two separate fractional factorial designs; one at the whole-plot level and one at the subplot level. Using this approach, they develop a catalog of 8 and 16 subplot run designs that are ranked according to the MA criterion. Bingham and Sitter (1999b) provide the theoretical support for their search strategy. In a subsequent work, Bingham and Sitter (2001) illustrate how to select a MA split-plot design for an actual industrial experiment. In addition, they discuss the estimation and precision of the parameter estimates from fractional factorial split-plot designs. While minimum aberration may be well-suited for completely randomized designs, it may not be the best criteria in the split-plot context because it inherently assumes that all of the factors are of equal importance. Furthermore, the designs produced by these computational methods tend to have a large number of whole-plots with a small number of subplot runs. This seems to contradict the purpose of employing a split-plot technique in that it fails to consider the HTC and ETC nature of the factors.

Other work on fractional factorials in a split-plot context includes Loeppky and Sitter (2002) who consider the analysis of unreplicated split-plot experiments. Schoen (1999) studies a fractional design with a nested error structure and considers a method to assign the factor effects to different split-plot levels. Schoen and Wolff (1997) present a case study of an asymmetric fractional factorial split-plot experiment.

Box and Jones (1992 and 2001) and Bisgaard and Kulachi (2001) consider the application of split-plot designs in the robust product design context. Bisgaard (2000) proposes the concept of partial resolution that reflects the relative importance of the whole-plot and subplot factors. In addition, Bisgaard draws from the important contribution by Addelman (1964) in the area of split-plot confounding, which exploits the ability to perform different subplot designs across the whole-plots. Kowalski (2002) proposes a class of 24 subplot run designs that are a compromise between the 16 and 32 subplot run designs of Huang, Chen, and Voelkel (1998) and Bingham and Sitter (1999a). A more complete discussion of Bisgaard (2000) and Kowalski (2002) is supplied in a subsequent paragraph.

A common criteria for the construction of completely randomized designs is $\mathcal{D}$-optimality, which minimizes the volume of the joint confidence region of regression coefficients. Goos and Vandebroek (2001a), propose the use of a point exchange algorithm to construct $\mathcal{D}$-optimal blocked designs with random block effects. Many of these designs depend of the level of correlation within the whole-plots, however there are cases when a particular design is optimal regardless of the level of correlation. The practical implication of this finding is that in some cases $\mathcal{D}$-optimal designs could be constructed that do not require precise knowledge of the variance components. They also discuss orthogonal blocking. Goos and Vandebroek (2001b) provide split-plot designs based on the point exchange algorithm for three factors. The construction of the $\mathcal{D}$-optimal designs in this case are dependent on the ratio of whole-plot to subplot variance. Goos and Vandebroek (2003) provide additional designs that are based on this algorithm. Many of these proposed designs have a large number of whole-plots with few subplot runs; a condition similar to the minimum aberration designs. Also, many of the designs lack appealing geometric symmetry and practical levels due to the $\mathcal{D}$-optimality criteria. Goos (2002) provides a thorough discussion and applications of the $\mathcal{D}$-optimal design approach. Goos and Vandebroek (2004) highlight that split-plot experiments can be more efficient than completely randomized designs and suggest them as an alternative to completely randomized designs even if running a completely randomized design is feasible. While $\mathcal{D}$-optimality is a useful criteria, we agree with the robust design philosophy proposed by Box and Draper (1975) and caution against the focus on a single number criterion. They point out that the statistician must be able to make "sensible subjective compromises" between competing criteria in the design selection process.

While the majority of this research has focused on designs that consider HTC factors in the experimental plan, there has been considerable work in the analysis of inadvertent split-plotting due to a failure to properly reset consecutive runs requiring the same level of a factor in a completely randomized design. This situation arises when a completely randomized run order is planned, but the experimenter either fails to realize the importance of resetting all factors before each run, or successive levels of the HTC factor makes it convenient to not reset. Ganju and Lucas (1997 and 1999) call these designs *randomized not reset* (RNR) and study analysis techniques for this situation. They illustrate how analyzing such a design as if it were completely randomized can lead to inappropriate inference about factor significance. Webb, Lucas, and Borkowski (2004) propose factorial designs for the RNR context. Ju and Lucas (2002) develop the expected covariance matrices for the RNR situation and illustrate that classical split-plot blocking can be superior to a completely randomized run order in terms of the precision of the subplot parameter estimates. Ganju and Lucas (1997) present a comparison of the well-known helicopter experiment run as a completely randomized design and in a split-plot mode. The RNR literature exposes a practical implementation issue in industrial experimentation and proposes an analysis-based solution. However, it seems that the RNR problem could be overcome with designs that account for HTC factors and precise experimental protocols requiring the reset of consecutive runs at the same level.

Variations of split-plot designs have been studied in the case of sequential processing. Mee and Bates (1998) discuss split-lot experiments for multi-stage batch processes. Their work was motivated by semi-conductor manufacturing where the integrated circuits undergo a sequence of processing steps. In a similar application where the experiment is run in stages, Miller (1997) discusses the application of strip-plot experiments. A strip-plot structure is represented by a rectangular array of experimental units. We can think of this as a twodimensional split-plot arrangement.

Only recently have second-order designs been considered in a split-plot context. In a review paper, Myers (1999) discusses the need for additional research in the area of restricted randomization. He states that practitioners have adopted a "sweeping under the rug" mentality in using a completely randomized analysis approach to designs that were actually executed in split-plot mode. In the most recent review paper on RSM, Myers *et al.* (2004) state that "More work needs to be done in extending basic split-plot analysis techniques to different types of industrial experiments." These two review papers emphasize the need for new developments in RSM that consider the common occurrence of hard-to-change factors.

In the area of mixture designs, Cornell (1988) discusses the use of split-plot designs in a combined mixture and process variable experiment. In these experiments, the process variables are considered as whole-plots. Kowalski, Cornell, and Vining (2002) propose a new design construction method that produces a class of smaller split-plot designs for this case. They used a simulation study to compare three estimation methods; ordinary least squares (OLS), restricted maximum likelihood (REML), and a pure-error based method. They compared the size of the joint confidence ellipsoids associated with the parameter estimates from the three methods.

Letsinger, Myers, and Lentner (1996) is the first paper to exclusively focus on restricted randomization in RSM. They primarily address analysis issues and do not consider practical design construction strategies. Draper and John (1998) propose modifications of central composite and Box-Behnken designs to be executed in a split-plot mode. They focused on achieving rotatability, or near-rotatability in these designs. Trinca and Gilmour (2001) study multiple layers of split-plotting, which they called *multistratum* designs. The split-plot design is a special case of this more general class. They employ a computer-intensive search algorithm to construct designs in which they employ weighted efficiency criteria while maintaining near-orthogonality between strata. All of these works relied on the assumption that a residual based method was required in the analysis of second-order split-plot designs, except in certain cases where the designs were not practical in size. We now review four of these articles directly related to the current research in more detail.

As previously mentioned, Letsinger, Myers, and Lentner (1996) discusses response surface designs conducted in a split-plot structure, which they call a *bi-randomization* error control format. The term bi-randomization comes from the two levels of randomization that are performed in a split-plot experiment. The authors note that little attention has been given in the RSM literature to techniques that address restricted randomization. Consequently, many practitioners who conduct split-plot designs do not perform a proper analysis that accounts for the error structure, rather they assume it to be a completely randomized design in the analysis. The consequences of this simplifying assumption depends on the ratio of whole-plot to subplot error, and could result in erroneously considering certain whole-plot effects as significant when they are not, resulting in a Type I error. Conversely, we could also declare certain subplot or subplot by whole-plot interaction effects as non-significant when

Table 2.1: Letsinger, Myers, and Lentner Split-Plot CCD with 1 Hard-to-Change and 2 Easy-to-Change Factors.

Whole-Plot	z_1	x_1	x_2
1	-1.682	0	0
2	-1	-1	-1
	-1	1	-1
	-1	-1	1
	-1	1	1
3	0	-1.682	0
	0	1.682	0
	0	0	-1.682
	0	0	1.682
	0	0	0
	0	0	0
4	1	-1	-1
	1	1	-1
	1	-1	1
	1	1	1
5	1.682	0	0

in fact they are significant, resulting in a Type II error. This naïve approach to analysis can result in inappropriate inferences.

Letsinger, Myers, and Lentner define *crossed* and *non-crossed* as two classes of bi-randomization designs. The crossed designs are based on a product array where all combinations of the whole-plot and subplot factors appear together in the design. Construction of this design is based on generating individual designs in the WP and SP factors and then crossing them in a product array. The resulting design has an equal number of subplot runs within each whole plot and has identical subplot designs within each whole-plot. Unfortunately, these designs can be quite large and could result in an inefficient allocation of experimental resources. In contrast, the non-crossed designs may have an unequal number of subplot runs within each whole-plot and the subplot designs would not necessarily be identical. A non-crossed design based on a central composite design (CCD), consistent with this construction philosophy, was presented in Myers (1999) and is shown in Table 2.1. The design contains one HTC factor, z_1, and two ETC factors, x_1, x_2. In the design, the factor levels are coded (scaled), which allows the design to be mapped to the actual factor levels of a particular application.

While this design provides the necessary information to estimate a second-order model, there a several features that do not seem very attractive. The most apparent is that some whole-plots only have one subplot run. This does not reflect that factor z_1 is hard-to-change. It seems logical that once a whole-plot level is set, we should perform at least some minimum number of subplot runs. In addition, the design is unbalanced having a different number of subplot runs per whole-plot. Lastly, the design does not contain replication of the whole-plots to enable a pure-error based estimate of the whole-plot error variance, and therefore must rely on lack-of-fit degrees of freedom for the estimate.

In the area of analysis, the authors consider methods of model estimation for the crossed and non-crossed cases separately. For crossed designs, the authors provide a proof of the equivalence of ordinary least squares (OLS) and generalized least squares (GLS), which simplifies model estimation. In this way, the model estimates do not rely on the variance components, although their variance does. For non-crossed designs, the authors state that the equivalence of OLS and GLS is not attainable for a second-order model, however the equivalence is preserved for first-order models. Additional discussion on the equivalence of OLS and GLS is provided in a later section.

The estimation techniques evaluated for non-crossed designs include; ordinary least squares (OLS), iterated reweighted least squares (IRLS), and restricted maximum likelihood (REML). In the non-crossed case, the use of OLS to estimate a second-order model is considered as the "ignorance" approach because it assumes a completely randomized design was performed. Note, this is different than the use of OLS when it is equivalent to GLS. Based on a simulation study, the authors recommend the use of REML for error variance and model estimation when there is limited prior knowledge of the error variance values, in particular the ratio of whole-plot to subplot error, $\sigma_\delta^2/\sigma_\epsilon^2$. Note that REML is a residual based method that relies on a model assumption. In traditional RSM, pure-error based methods are preferred because they provide model independent estimates of the error variances. They noted that if the ratio of error variances is less than 1, implying that the subplot error is larger than the whole-plot error, then OLS is a reasonable approach, however the basis for this recommendation is not clearly presented.

- The major contribution of Letsinger, Myers, and Lentner (1996) is that it addressed the need for industrial practitioners to recognize a split-plot structure in response surface experiments and to properly consider it in the analysis. By highlighting the need for response surface techniques in the presence of restricted randomization, it provided the impetus for future research in this area.

Also in the area of industrial split-plot design, Bisgaard (2000) presents an expository work on the efficient allocation of resources in split-plot experimentation. Moreover, he provides a practical and intuitive discussion of the use of split-plot designs in industrial applications, especially in the area of robust parameter design (RPD). In a RPD problem, we have control factors (or design factors) and noise factors (or environmental factors) that we wish to investigate simultaneously. Our primary goal in a RPD experiment is the identification and estimation of control factor main effects and control by environmental factor interactions. Bisgaard (2000) restricts his attention to the context of factor screening, and therefore he considers two-level full and fractional factorial designs. He highlights two specific shortcomings of similar robust parameter designs proposed by Taguchi (1987).

1. The size of the experiment is often quite large (many subplot runs), because they are based on a crossed design construction method.

2. The experiments do not maximize the information on a per observation basis. Often, higher resolution designs could be performed and more attractive alias structures are available with the same or fewer subplot runs. In addition, these designs fail to recognize that the subplot effects are estimated with a higher precision, and thus do not properly assign the environmental variables to the whole-plot level.

Bisgaard (2000) discusses two approaches for the construction of a split-plot design. The first approach is to cross the whole-plot and subplot array, and thereby obtain all possible combinations of the factor levels. This is referred to as a Cartesian product by Bisgaard. (This is equivalent to the *crossed* designs discussed by Letsinger, Myers, and Lentner (1996), and *product arrays* discussed by Taguchi (1987).) A second approach is split-plot confounding, where we take advantage of the opportunity to run different subplot fractional designs within the whole-plots. In this way, we are able to break certain alias chains in the defining relation, and thereby increase the partial resolution of the subplot factors. The concept of partial resolution, where the whole-plot factors may have a lower resolution than the subplot factors, emphasizes the relative importance of the two types of factors. In addition, it attains the goal of estimating the control and control by environmental effects. The use of split-plot confounding increases the information provided by a design on a per observation basis.

- The primary contribution of Bisgaard (2000) is the emphasis on the efficient allocation of experimental resources, especially in the selection of subplot designs. In addition, Bisgaard (2000) demonstrates the findings of Letsinger, Myers, and Lentner for the equivalence of OLS and GLS estimation for a first-order model in a design which does not have identical subplots.

Kowalski (2002) also addresses the issue of efficient allocation in a split-plot structure, and discusses designs that are particularly useful for screening and in the RPD context. Recognizing that additional runs at the subplot level are less costly than additional runs at the whole-plot level, Kowalski proposed the augmentation of the subplot design. He considers methods to augment a 16 subplot run design to obtain a 24 subplot run design, while maintaining the same number of whole-plots as the original design. The basic premise of this approach is that the increase from 16 to 24 subplot runs would not significantly increase the cost of the experiment if the number of whole-plots was kept constant. As previously noted, the 24 subplot run designs provide a compromise between the previously proposed 16 and 32 subplot run designs of Huang, Chen, and Voelkel (1998) and Bingham and Sitter (1999a).

Another advantage of these proposed designs is the additional degrees of freedom available in the subplot factors for the estimation of subplot error, rather that relying on the use of a normal probability plot of the factor effects.

Two methods are proposed for the construction of 24 subplot run designs. The first uses a technique called semifolding that augments a 16 subplot run design with 8 additional subplot runs. A typical approach to augment a fractional factorial design is to perform a full foldover in which all of the signs of the factors are reversed, thereby doubling the design size. In contrast to a full foldover, semifolding isolates the subplot runs of a single factor at either the high or low level and reverses the signs of these subplot runs, thereby adding half as many subplot runs as the original design. Semifolding breaks the original alias chains for the subplot factor selected as the basis of the fold. This folding procedure is performed within each whole-plot, and different folding factors can be chosen within sets of whole-plots. For example in a design with four whole-plots, in two of the whole-plots the subplot design is folded on one factor, and in the other two whole-plots a different subplot factor is folded. This approach is particularly useful if prior knowledge is available regarding the likelihood of certain interactions.

A second design construction approach involves the use of balanced incomplete block designs (BIBD). In this approach, we consider the whole-plots as blocks and seek a design with the appropriate allocation of factor combinations. Due to the necessary conditions on the number of blocks and block size for the construction of a BIBD, this approach is not always possible. If a BIBD is not available then a partially balanced incomplete block design can be used. A primary advantage of the BIBD approach is that within each whole-plot there will be an equal number of high and low levels for each subplot factor. This results in a design where the subplot factors provide balanced information across all whole-plots.

- The primary contribution of Kowalski (2002) is the recognition that increasing the size of the subplot design is a cost effective approach in a split-plot experiment. His work also incorporates the ideas of Bisgaard (2000) in using different subplot designs across the whole plots. Moreover, the design construction methods incorporate the HTC factors in an appropriate manner. Similar to Bisgaard (2000), the overriding focus of this article is the efficient allocation of experimental resources for split-plot designs.

Vining, Kowalski, and Montgomery (2005), hereafter referred to as VKM, proposes a new approach to constructing second-order designs in a split-plot structure that results in designs where OLS is equivalent to GLS for model estimation. In addition, these designs provide estimates of pure-error at both the whole-plot and subplot levels, thereby producing a model-independent estimate of the variance-covariance matrix, $\mathbf{\Sigma}$. Traditionally in RSM, pure-error estimates are preferable because we do not know the final form of the model before the experiment is executed. The authors illustrate the construction of designs that are based on the central composite design (CCD) originally proposed by Box and Wilson (1951) and the Box-Behnken design (BBD) proposed by Box and Behnken (1960). These are the two most commonly occurring second-order designs in the practice of RSM. The proposed split-plot designs offer appealing properties such as balance and a minimum number of whole-plots. Their construction method can be extended to various combinations of whole-plot and subplot factors. Moreover, by achieving the equivalence of OLS and GLS estimates, estimation can be performed without using REML. From a practitioner's perspective, this is very appealing because it implies that any software capable of matrix manipulations or OLS estimation can be used to perform the model estimation, rather than more sophisticated statistical software required for REML.

A proof of the conditions that are required to achieve the equivalence between OLS and GLS was provided by VKM and will be summarized here because it serves as a foundation for the current research. We note that the proof is not model specific, and therefore makes no assumption about the order of the model, only that the design supports the specified model. The proof assumes that the model contains an intercept term and the design is balanced; meaning that each whole plot has same number of subplot runs. McElroy (1967) (see also Graybill (1976, pg. 209)) prove that the OLS and GLS estimates are equivalent if and only if there exists a nonsingular matrix $\mathbf{F}$ such that

$$\mathbf{\Sigma X} = \mathbf{XF}.$$

Recall from the definition of the model,

$$\mathbf{\Sigma} = \sigma\epsilon^2\mathbf{I} + \sigma_\delta^2 \begin{bmatrix} \mathbf{1}_{n_1}\mathbf{1}'_{n_1} & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{n_2}\mathbf{1}'_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_{n_m}\mathbf{1}'_{n_m} \end{bmatrix}.$$

In the balanced case, $n_i = n$ for $i = 1, 2, \ldots, m$, where m is the number of whole-plots and n is the number of subplot runs per whole-plot. Therefore, $\mathbf{1}_{n_i}$ is a column vector of ones of length n and subsequently is denoted simply by $\mathbf{1}_n$. We note that

$$\mathbf{\Sigma X} = \left[\sigma^2\mathbf{I} + \sigma_\delta^2 \begin{bmatrix} \mathbf{1}_n\mathbf{1}'_n & 0 & \cdots & 0 \\ 0 & \mathbf{1}_n\mathbf{1}'_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_n\mathbf{1}'_n \end{bmatrix} \right] \mathbf{X}.$$

VKM partition the $\mathbf{X}$ matrix as

$$\mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{W}_1 & \mathbf{S}_{D_1} & \mathbf{S}_{O_1} \\ \mathbf{1} & \mathbf{W}_2 & \mathbf{S}_{D_2} & \mathbf{S}_{O_2} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{1} & \mathbf{W}_m & \mathbf{S}_{D_m} & \mathbf{S}_{O_m} \end{bmatrix},$$

where $\mathbf{W}_i$ is the i^{th} whole plot model matrix apart from the y-intercept term (represented by the first column of ones), $\mathbf{S}_{D_i}$ is the i^{th} subplot model matrix. $\mathbf{S}_{O_i}$ is the rest of the i^{th} subplot model matrix. In most cases, $\mathbf{S}_{O_i}$ will contain the columns of the model matrix associated with the subplot pure quadratics, although the definition of $\mathbf{S}_{O_i}$ is general in nature.

We now state the first condition of an equivalent design as

$$\mathbf{S}'_{D_i}\mathbf{1} = \mathbf{q} \qquad i = 1, 2, \ldots, m \ ;$$

stating that the columns of the subplot design sum to the vector $\mathbf{q}$. Note that the subplot design of the i^{th} whole-plot does not need to be the same as the design for another whole-plot, say the $i^{*\,th}$ whole-plot. In other words, $\mathbf{S}_{D_i}$ may not equal $\mathbf{S}_{D_i^*}$, but

$$\mathbf{S}'_{D_i}\mathbf{1} = \mathbf{S}'_{D_i^*}\mathbf{1} = \mathbf{q},$$

where $\mathbf{q}$ is a vector of constants. If each subplot design was an orthogonal design, then $\mathbf{q} = \mathbf{0}$.

Now performing the multiplication, we can express $\mathbf{\Sigma X}$ as

$$\begin{aligned}
\mathbf{\Sigma X} &= \sigma^2\mathbf{X} + \sigma^2_\delta \begin{bmatrix} \mathbf{11'1} & \mathbf{11'W}_1 & \mathbf{11'S}_{D_1} & \mathbf{11'S}_{O_1} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{11'1} & \mathbf{11'W}_m & \mathbf{11'S}_{D_m} & \mathbf{11'S}_{O_m} \end{bmatrix} \\
&= \sigma^2\mathbf{X} + \sigma^2_\delta \begin{bmatrix} n\mathbf{1} & \mathbf{11'W}_1 & \mathbf{1q'} & \mathbf{11'S}_{O_1} \\ \vdots & \vdots & \vdots & \vdots \\ n\mathbf{1} & \mathbf{11'W}_m & \mathbf{1q'} & \mathbf{11'S}_{O_m} \end{bmatrix}.
\end{aligned}$$

In the whole-plot designs, the factor levels are keep constant by the nature of the split-plot structure. It then follows that the rows of the model matrix are all the same within the i^{th} whole-plot. Therefore, each whole-plot model matrix can be expressed as $\mathbf{W}_i = \mathbf{1w}'_i$, for $i = 1, 2, \ldots, m$, where $\mathbf{w}'_i$ is a vector containing the whole-plot model terms for the i^{th} whole-plot. Working with this expression we see that

$$\mathbf{11'W}_i = \mathbf{11'1w}'_i = n\mathbf{1w}'_i = n\mathbf{W}_i.$$

Now we can express $\mathbf{\Sigma X}$ as

$$\mathbf{\Sigma X} = \sigma^2\mathbf{X} + \sigma^2_\delta \begin{bmatrix} n\mathbf{1} & n\mathbf{W}_1 & \mathbf{1q'} & \mathbf{11'S}_{O_1} \\ \vdots & \vdots & \vdots & \vdots \\ n\mathbf{1} & n\mathbf{W}_m & \mathbf{1q'} & \mathbf{11'S}_{O_m} \end{bmatrix}.$$

We now need to find a suitable $\mathbf{F}$. Let $\mathbf{F}$ be defined with the following structure

$$\mathbf{F} = \sigma^2\mathbf{I} + \sigma_\delta^2 \begin{bmatrix} n & \mathbf{0}' & \mathbf{q}' & \mathbf{0}' \\ \mathbf{0} & n\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G} \end{bmatrix}.$$

Then we can express $\mathbf{XF}$ as

$$\begin{aligned} \mathbf{XF} &= \sigma^2\mathbf{X} + \sigma_\delta^2 \begin{bmatrix} \mathbf{1} & \mathbf{W}_1 & \mathbf{S}_{D_1} & \mathbf{S}_{O_1} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{1} & \mathbf{W}_m & \mathbf{S}_{D_m} & \mathbf{S}_{O_m} \end{bmatrix} \begin{bmatrix} n & \mathbf{0} & \mathbf{q}' & \mathbf{0} \\ \mathbf{0} & n\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G} \end{bmatrix} \\ &= \sigma^2\mathbf{X} + \sigma_\delta^2 \begin{bmatrix} n\mathbf{1} & n\mathbf{W}_1 & \mathbf{1q}' & \mathbf{S}_{O_1}\mathbf{G} \\ \vdots & \vdots & \vdots & \vdots \\ n\mathbf{1} & n\mathbf{W}_m & \mathbf{1q}' & \mathbf{S}_{O_m}\mathbf{G} \end{bmatrix}. \end{aligned}$$

We note that $\mathbf{\Sigma X} = \mathbf{XF}$ if

$$\mathbf{11}'\mathbf{S}_{O_i} = \mathbf{S}_{O_i}\mathbf{G} \qquad i = 1, 2, \ldots, m.$$

Therefore, we can summarize the conditions of the proof as

1. The design is balanced (each whole-plot contains the same number of subplot runs).
2. The subplot designs are orthogonal or the subplot model matrix, excluding the subplot pure quadratics, sums to a constant vector $\mathbf{q}$, however they are not necessarily the same design.
3. The expression, $\mathbf{11}'\mathbf{S}_{O_i} = \mathbf{S}_{O_i}\mathbf{G} \qquad i = 1, 2, \ldots, m$ holds.

If these conditions are met, then the OLS parameter estimates will be equivalent to the GLS estimates. Subsequently, we will refer to a design where the OLS and GLS estimates are the same as an *equivalent design.*

To relate these results back to earlier literature, we see that the conditions can be achieved if all of the subplot designs are the same. This is the case of *crossed* (Letsinger, Myers, and Lentner (1996)) and *Cartesian product* (Bisgaard (2000)) designs. Moreover, the OLS-GLS equivalence is maintained if the subplot designs are orthogonal within each whole-plot and a first-order plus interaction model is employed. In this case, there is no $\mathbf{S}_{O_i}$ partition required. Therefore, the subplot designs do not need to be identical and can exploit the benefits of *partial confounding* as proposed by Bisgaard (2000). Furthermore, the subplot designs do not need to be full rank. For example a subplot design consisting of all center runs (all zeroes in coded variables) will satisfy the conditions. In practice, the subplot designs could be based on full or fractional factorials or center runs.

More importantly, we note the general nature of these conditions in comparison to that of Letsinger, Myers, and Lentner (1996).

1. The conditions are not dependent on the form of the model (e.g. first- or second-order), only that the design is adequate to support the model selected.

2. The conditions are not a function of the ratio of the whole-plot to subplot error variances, $\sigma_\delta^2/\sigma_\epsilon^2$. This removes the requirement of prior knowledge of this ratio as discussed in Letsinger, Myers, and Lentner (1996).

To consider the implications of these conditions related to second-order models, VKM provide a detailed example that illustrates how to construct a CCD and BBD for one whole-plot factor and two subplot factors. The BBD is shown in Table 2.2 and features 6 whole-plots each with 4 subplot runs. We see that whole-plots 4, 5, and 6 are whole-plot replicates because they contain exactly the same settings for both the whole-plot and subplot factors. This is an equivalent design, meaning that OLS can be used to obtain best linear unbiased estimates (BLUE) of a complete second-order model. It also provides pure-error estimates of both the whole-plot and subplot error variances.

Another strategy proposed by VKM to construct an equivalent design that essentially makes the subplot pure quadratic terms whole-plot effects. In this case, the $\mathbf{S}_{O_i}$ matrices become submatrices of the $\mathbf{W}_i$ matrices. Vining and Kowalski (2004, 2005) discuss this design structure in more detail in the context of testing.

VKM also illustrate how to obtain a pure-error based estimate of the variance-covariance matrix, $\mathbf{\Sigma}$. We obtain a pooled estimate of the subplot error variance as

$$S_{SP}^2 = \frac{\sum_{i=1}^{m_r} S_i^2}{m_r},$$

where S_i^2 is the sample variance for the i^{th} whole-plot replicate, and m_r is the number of whole-plot replicates. For example in Table 2.2, $m_r = 3$. VKM established that $\mathrm{E}(S_{SP}^2) = \sigma_\epsilon^2$. Therefore, $\hat{\sigma}_\epsilon^2 = S_{SP}^2$ is an unbiased estimate of σ_ϵ^2 with $m_r(n-1)$ degrees of freedom. We note that in Table 2.2, there are only replicated subplot centers. The pooled estimate of subplot variance can be easily extended if other subplot factor combinations are replicated within a whole-plot.

Now, we let $\overline{Y}_{i\cdot}$ be the sample mean of the values in the i^{th} whole-plot replicate, and $\overline{Y}_{\cdot\cdot}$ the

Table 2.2: Vining, Kowalski, and Montgomery (2005) Split-Plot BBD Design with 1 Hard-to-Change and 2 Easy-to-Change Factors.

Whole-Plot	z_1	x_1	x_2	Whole-Plot	z_1	x_1	x_2
1	-1	-1	0	4	0	0	0
	-1	1	0		0	0	0
	-1	0	-1		0	0	0
	-1	0	1		0	0	0
2	1	-1	0	5	0	0	0
	1	1	0		0	0	0
	1	0	-1		0	0	0
	1	0	1		0	0	0
3	0	-1	-1	6	0	0	0
	0	1	-1		0	0	0
	0	-1	1		0	0	0
	0	1	1		0	0	0

overall mean and define S^2_{WP} as

$$S^2_{WP} = \frac{\sum_{i=1}^{m_r}\left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2}{m_r - 1}.$$

VKM established that the $\text{E}(S^2_{WP}) = \sigma^2_\delta + \frac{1}{n}\sigma^2_\epsilon$. Therefore, using the method of moments we obtain an unbiased estimate of σ^2_δ as

$$\hat{\sigma}^2_\delta = S^2_{WP} - \frac{1}{n}\hat{\sigma}^2_\epsilon,$$

with $m_r - 1$ degrees of freedom. These pure-error estimates of the variance components are used to form the estimated variance-covariance matrix of the observations, $\widehat{\mathbf{\Sigma}}$. Moreover, these estimates are model independent and are robust to model misspecification.

- The primary contribution of Vining, Kowalski, and Montgomery (2005) is judged to be their approach to design construction that allows for OLS estimation of the model and pure-error estimates of the error variance terms. Prior to this work, the literature assumed that a second-order split-plot design must be analyzed using a residual based technique such as REML, except in certain restrictive cases. In addition, they recognize that the cost of the split-plot experiment is primarily a function of the number of whole-plots, rather than the total number of subplot runs.

Chapter 3

Balanced Designs

3.1 Second-Order Equivalent Estimation Designs

In the literature review we discussed the large body of work related to split-plot designs for first-order models, typically used as screening designs in the initial stages of an experimental investigation to identify active factors. However, when a first-order model is no longer adequate, then we utilize designs that enable us to model curvature. In traditional RSM, these models are based on a second-order Taylor series approximation to the true underlying response surface. Note that as we perform sequential experimentation, moving from a first-order to second-order model, the requirement for a split-plot structure remains. That is, HTC factors identified as significant in the screening design will remain hard to change in subsequent designs. Therefore, second-order split-plot designs are required.

In this research, we explore systematic design construction techniques for second-order equivalent estimation designs. The goal of this work is to develop practical experimentation strategies in the presence of restricted randomization that are consistent with traditional RSM philosophy. We now list the benefits of equivalent estimation designs from the perspective of design selection and estimation.

- Equivalent estimation designs can be constructed such that the equivalence property does not depend on the model. In this way, prior assumptions of a model will not influence the choice of the design; only that the design supports the model. In addition, as we perform model reduction during the analysis phase, we maintain the equivalence.

- Design selection is independent of the variance components. As a result we do not need prior knowledge or an assumption of the ratio of the whole-plot to subplot error variance.

- Model parameter estimates are BLUE (best linear unbiased estimate). Moreover, the

BLUE property is robust to the assumption of normality by the Gauss-Markov Theorem.

- Model parameter estimates are not a function of the variance components. This is a particularly attractive feature considering the typical situation where estimation of the variance components is required.

- Model estimation is simplified using OLS. It can easily be performed with any software capable of matrix manipulations and any statistical software package. Furthermore, OLS estimation is approachable to those with modest statistical training.

- The CCD version is well-suited for sequential experimentation. We have the ability to execute the factorial whole-plots and estimate a first-order plus interaction model, and then decide if augmentation to a second-order model with the current subset of factors is useful.

- Design augmentation providing replication to enable pure-error estimates of the variance components is straightforward. Based on pure-error estimates, we can construct a model-independent non-iterative estimate of the variance-covariance matrix of the parameter estimates.

- Design construction is based on common families of RSM designs which are familiar to practitioners.

In the following sections, we propose classes of equivalent estimation designs. The central composite and Box-Behnken designs are selected as the basis for design construction, because these designs are familiar to industrial practitioners, have been proven effective in many applications, and their desirable properties have been thoroughly studied. In addition, they are consistent with the desirable design attributes listed by Box and Draper (1975) providing guidance in the selection of a response surface design. Particularly applicable to equivalent estimation designs, are the criteria for robustness to model misspecification and the ability to obtain estimates of pure-error. In previous research, reliance on the use of REML violates the criterion of robustness to model assumptions and lacks pure-error estimates.

We propose an augmentation of the original 14-point list provided by Box and Draper (1975) with two additional criteria that are specific to the restricted randomization case. (1) To employ an efficient allocation of resources in the subplot designs. As discussed in VKM (2005), if the hard-to-change factors are truly hard to change, then there is a minimum number of subplots that should be included at these conditions. Also, it is possible that there is a maximum number of subplots that can be run at each setting of the whole-plot factors. Combining these two considerations gives $n_{min} \leq n_i \leq n_{max}$. We consider the value of $n_{min} = 2$ and do not assume a value for n_{max}. However, using our design construction methods, it is often possible to reduce the whole-plot size if dictated by an application specific n_{max}, thereby requiring additional whole-plots.

Additionally, the use of split-plot confounding, as proposed by Bisgaard (2000), and the augmentation of subplot designs, as proposed by Kowalski (2002), are consistent with this criteria. These approaches recognize that the cost of an experiment is not solely determined by the total number of subplot runs, rather by the allocation of subplot runs among the whole-plots.

(2) To utilize a minimum number of whole-plots. Due to the split-plot design structure, the number of whole-plot levels is directly related to the number of whole-plots required. Moreover, each whole-plot requires resetting the HTC factors. Recall, that minimizing the number of times that the HTC factors are required to be reset is the primary reason for employing a split-plot structure. As a result of this criteria, we focus on three-level designs in our examples. However, we develop the methods for the central composite design as a function of α, the distance to the axial points, providing flexibility to the practitioner.

These two proposed criteria summarize the additional design challenges in the restricted randomization context. In the second criteria, we see that using a minimum number of whole-plots is in conflict with the ability to obtain a pure-error estimate of the whole-plot variance. As in most design situations, a compromise will be required between these competing criteria.

In this chapter, we generalize the equivalence proof of VKM (2005), outline systematic design construction techniques to build equivalent estimation designs from the CCD and BBD, and discuss equivalent estimation $\mathcal{D}$-optimal designs. We restrict our attention to balanced designs, where each whole-plot contains the same number of subplot runs, leaving the unbalanced case for Chapter 4.

3.2 Generalized Equivalence Condition

We assume that there is one level of split-plotting and that the design supports the model selected. We now generalize the conditions given in VKM (2005) to provide a technique to numerically verify the OLS-GLS equivalence of a particular design and to enable the generation of systematic approaches to design construction. We start with the necessary and sufficient condition for equivalence from McElroy (1967) (see also Graybill (1976, p. 209)),

$$\mathbf{XF} = \mathbf{\Sigma X} \tag{3.1}$$

where $\mathbf{X}$ is an $N \times p$ model matrix with rank of p and $\mathbf{F}$ is a $p \times p$ non-singular matrix. Pre-multiplying by $\mathbf{X}'$ we obtain,

$$\mathbf{X}'\mathbf{XF} = \mathbf{X}'\mathbf{\Sigma X}.$$

Expanding the right hand side,

$$\mathbf{X}'\mathbf{X}\mathbf{F} = \mathbf{X}'\left(\sigma_\epsilon^2\mathbf{I} + \sigma_\delta^2\mathbf{J}_b\right)\mathbf{X},$$

and pre-multiplying by $(\mathbf{X}'\mathbf{X})^{-1}$ provides

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{F} = \sigma_\epsilon^2\,(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} + \sigma_\delta^2\,(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{J}_b\mathbf{X}.$$

Simplifying this expression we obtain

$$\mathbf{F} = \sigma_\epsilon^2\mathbf{I} + \sigma_\delta^2\,(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{J}_b\mathbf{X}, \tag{3.2}$$

where $\mathbf{F}$ is non-singular. Therefore,

$$\mathbf{X}\mathbf{F} = \sigma_\epsilon^2\mathbf{X} + \sigma_\delta^2\mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{J}_b\mathbf{X}.$$

Now, we can express the Equation 3.1 as

$$\sigma_\epsilon^2\mathbf{X} + \sigma_\delta^2\mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{J}_b\mathbf{X} = \sigma_\epsilon^2\mathbf{X} + \sigma_\delta^2\mathbf{J}_b\mathbf{X}.$$

By simplification and inspection we see that

$$\mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{J}_b\mathbf{X} = \mathbf{J}_b\mathbf{X}, \tag{3.3}$$

must hold for the equivalence of OLS and GLS estimates. Given a model matrix, $\mathbf{X}$, and the whole plot structure contained in $\mathbf{J}_b$, we can numerically verify equivalence without knowledge of σ_ϵ^2, σ_δ^2, or $\mathbf{F}$, which was required to apply the VKM proof.

To construct equivalent estimation designs, suitable choices for $\mathbf{F}$ that satisfy Equation 3.1 are required. Define $\mathbf{F}$ from Equation 3.2 to be

$$\mathbf{F} = \sigma_\epsilon^2\mathbf{I} + \sigma_\delta^2\mathbf{K}$$

where,

$$\mathbf{K} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{J}_b\mathbf{X}. \tag{3.4}$$

Even though $\mathbf{F}$ is non-singular, the matrix $\mathbf{K}$ is singular. Equation 3.4 provides the ability to compute the form of the $\mathbf{K}$ matrix for a specific design. Expressing Equation 3.3 in terms of $\mathbf{K}$, the condition for equivalence can be expressed as

$$\mathbf{X}\mathbf{K} = \mathbf{J}_b\mathbf{X}. \tag{3.5}$$

This expression holds in general for the balanced or unbalanced case and does not assume a particular structure for the $\mathbf{K}$ matrix.

3.3 A Balanced Form of K for a Second-Order Model

First, we perform a finer partitioning of $\mathbf{X}$ than the one proposed by VKM (2005). We restrict our attention to a second-order Taylor series approximation that is commonly employed in RSM models. The matrix form of the model is given in Equation 1.2. To illustrate the terms in the model matrix, we expand as follows

$$\begin{aligned} E(y) &= \beta_0 + \sum_{i=1}^{w} \beta_i z_i + \sum_{i=1}^{w-1} \sum_{j=i+1}^{w} \beta_{ij} z_i z_j + \sum_{i=1}^{w} + \beta_{ii} z_i^2 \\ &+ \sum_{i=1}^{k} \theta_i x_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \theta_{ij} x_i x_j + \sum_{i=1}^{w} \sum_{j=1}^{k} \gamma_{ij} z_i x_j + \sum_{i=1}^{k} \theta_{ii} x_i^2, \end{aligned} \tag{3.6}$$

where y is the response, z is a whole-plot factor, x is a subplot factor, the β's are the regression coefficients at the whole-plot level, θ's and γ's are the regression coefficients at the subplot level. Note that in some designs subplot level terms will be considered as whole-plot level effects. The model contains w whole-plot factors, k subplot factors, and $f = w + k$ total factors. For a complete second-order model, including the intercept, we can express the total number of model terms, p, as

$$p = 1 + 2f + \frac{f(f-1)}{2} \tag{3.7}$$

Now, let the $N \times p$ matrix $\mathbf{X}$ be partitioned as

$$\mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{W}_{\mathrm{D}_1} & \mathbf{W}_{\mathrm{Q}_1} & \mathbf{S}_{\mathrm{D}_1} & \mathbf{S}_{\mathrm{Q}_1} \\ \mathbf{1} & \mathbf{W}_{\mathrm{D}_2} & \mathbf{W}_{\mathrm{Q}_2} & \mathbf{S}_{\mathrm{D}_2} & \mathbf{S}_{\mathrm{Q}_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{1} & \mathbf{W}_{\mathrm{D}_m} & \mathbf{W}_{\mathrm{Q}_m} & \mathbf{S}_{\mathrm{D}_m} & \mathbf{S}_{\mathrm{Q}_m} \end{bmatrix}, \tag{3.8}$$

where the first column of ones corresponds to the intercept, $\mathbf{W}_{\mathrm{D}_i}$ denotes the whole-plot main effects and two-factor interactions, $\mathbf{W}_{\mathrm{Q}_i}$ contains the whole-plot pure quadratics, $\mathbf{S}_{\mathrm{D}_i}$ includes the subplot main effects, subplot two-factor interactions, and the whole-plot by subplot interactions, and $\mathbf{S}_{\mathrm{Q}_i}$ denotes the subplot pure quadratic terms of the i^{th} whole-plot.

Consider one choice for $\mathbf{K}$ $(p \times p)$ of the form

$$\mathbf{K} = \begin{bmatrix} n_0 & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{m}_0' \\ \mathbf{0} & n_{\mathrm{W}}\mathbf{I}_{\mathrm{D}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & n_{\mathrm{W}}\mathbf{I}_{\mathrm{Q}} & \mathbf{0} & \mathbf{M}_{\mathrm{W}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V}_{\mathrm{S}} \end{bmatrix},$$

where n_0 and n_W are scalars, $\mathbf{m_0}$ is a vector of length k, $\mathbf{I}_\mathrm{D}$ and $\mathbf{I}_\mathrm{Q}$ are $(w + \frac{w(w-1)}{2}) \times (w + \frac{w(w-1)}{2})$ and $w \times w$ identity matrices, respectively, $\mathbf{M}_\mathrm{W}$ is a $w \times k$ matrix, and $\mathbf{V}_\mathrm{S}$ is a $k \times k$ matrix. The elements of $\mathbf{K}$ correspond to the design construction methods presented in the remainder of this chapter.

Consider Equation 3.5 and note that the right-hand side of the equation sums the columns of the model matrix within each whole-plot. These column totals, or unscaled means, must equal the model matrix multiplied by $\mathbf{K}$ to satisfy the condition of equivalence. Expanding Equation 3.5, we obtain

$$\begin{bmatrix} \mathbf{1} & \mathbf{W}_{\mathrm{D}_1} & \mathbf{W}_{\mathrm{Q}_1} & \mathbf{S}_{\mathrm{D}_1} & \mathbf{S}_{\mathrm{Q}_1} \\ \mathbf{1} & \mathbf{W}_{\mathrm{D}_2} & \mathbf{W}_{\mathrm{Q}_2} & \mathbf{S}_{\mathrm{D}_2} & \mathbf{S}_{\mathrm{Q}_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{1} & \mathbf{W}_{\mathrm{D}_m} & \mathbf{W}_{\mathrm{Q}_m} & \mathbf{S}_{\mathrm{D}_m} & \mathbf{S}_{\mathrm{Q}_m} \end{bmatrix} \begin{bmatrix} n_0 & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{m}_0' \\ \mathbf{0} & n_\mathrm{W}\mathbf{I}_\mathrm{D} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & n_\mathrm{W}\mathbf{I}_\mathrm{Q} & \mathbf{0} & \mathbf{M}_\mathrm{W} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V}_\mathrm{S} \end{bmatrix}$$
$$= \begin{bmatrix} n_1\mathbf{1} & \mathbf{11}'\mathbf{W}_{\mathrm{D}_1} & \mathbf{11}'\mathbf{W}_{\mathrm{Q}_1} & \mathbf{11}'\mathbf{S}_{\mathrm{D}_1} & \mathbf{11}'\mathbf{S}_{\mathrm{Q}_1} \\ n_2\mathbf{1} & \mathbf{11}'\mathbf{W}_{\mathrm{D}_2} & \mathbf{11}'\mathbf{W}_{\mathrm{Q}_2} & \mathbf{11}'\mathbf{S}_{\mathrm{D}_2} & \mathbf{11}'\mathbf{S}_{\mathrm{Q}_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ n_m\mathbf{1} & \mathbf{11}'\mathbf{W}_{\mathrm{D}_m} & \mathbf{11}'\mathbf{W}_{\mathrm{Q}_m} & \mathbf{11}'\mathbf{S}_{\mathrm{D}_m} & \mathbf{11}'\mathbf{S}_{\mathrm{Q}_m} \end{bmatrix}. \tag{3.9}$$

This expression provides insight into the subsequent development of design construction strategies. At this point, we offer a three implications of the proposed form of $\mathbf{K}$.

(1) Within each whole-plot, the column sums of the main effects and two-factor interactions in the subplot factors are zero. This is true for an orthogonal subplot design. While orthogonality is not a necessary condition for equivalence, it is a desirable design property, and therefore, it has been assumed in our proposed design construction techniques.

(2) Since for each whole-plot, the whole-plot factors are held constant, then within each whole-plot the column sum of the whole-plot factors equals n_i times the level of the whole-plot model term. For example, if a whole-plot factor is set to one in coded units and $n_i = 4$, then the column sum would be 4. This relationship is expressed in the $n_\mathrm{W}\mathbf{I}_\mathrm{D}$ and $n_\mathrm{W}\mathbf{I}_\mathrm{Q}$ submatrices of $\mathbf{K}$.

(3) We know that for a non-zero level of the whole-plot and subplot factors the pure quadratic terms cannot sum to zero since these terms are always positive. This highlights the complexity in satisfying the equivalent estimation condition in the case of a second-order model. Bisgaard (2000) discusses the construction of split-plot designs for a first-order plus interaction model and demonstrates that the OLS-GLS equivalence can be obtained in both crossed and non-crossed designs. If we reduced our model, thereby eliminating the third and fifth columns of the partitioned $\mathbf{X}$ matrix in Equation 3.8, our $\mathbf{K}$ matrix becomes

$$\mathbf{K} = \begin{bmatrix} n_0 & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & n_\mathrm{W}\mathbf{I}_\mathrm{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{3.10}$$

If the subplot designs are orthogonal, we see that this condition is trivial. In addition, Bisgaard proposes the use of split-plot confounding, where we take advantage of the opportunity to run different subplot designs across the whole-plots. From Equation 3.10, we see that if we have orthogonal subplot designs, they do not need to be identical and therefore the use of split-plot confounding does not disrupt the equivalence condition. Bisgaard's work emphasized the efficient allocation of experimental resources, especially in the selection of subplot designs that we will adopt in our proposed designs.

In general, the conditions to achieve equivalence involve only the relationships between the column sum of the intercept and the column sum of the whole-plot and subplot pure quadratic terms, which defines the selection of the appropriate constants for the $\mathbf{K}$ matrix. It is important to note that the equivalence condition does not require knowledge of the variance components.

3.4 Design Construction Techniques

In this section, we propose two design construction techniques that enable the construction of equivalent estimation SPD's derived from completely randomized response surface designs. These designs support the estimation of a complete second-order model, are efficient in the allocation of experimental resources, and are easily augmented to obtain pure-error estimates of whole-plot and subplot error variances. The first technique is a generalized version of the construction method proposed by Vining, Kowalski, and Montgomery (2005), subsequently referred to as the VKM method. In the second method, we obtain designs that achieve the minimum number of whole-plots, subsequently referred to as the MWP method. For both methods, examples of balanced central composite and Box-Behnken designs are presented.

The notation $\mathrm{D}(w, k)$ is used to denote a design with w whole-plot factors and k subplot factors. Designs with $1 \leq w \leq 3$ in combination with $1 \leq k \leq 4$ are considered to encompass many practical situations. Extension of these methods to more factors is straightforward.

For a central composite design, we begin with a design based on $f = w + k$ factors. For $f < 5$, this design consists of a 2^f full factorial, $2f$ axial points, and n_c center points. Note that for $f \geq 5$ we could choose to use a fractional factorial of resolution V or greater. We use α and β to denote the distance to the axial points for the subplot and whole-plot factors respectively. We assume that the value of α is the same for all of the subplot factors and similarly β is the same for all of the whole-plot factors.

Similarly for a Box-Behnken design, we begin with a CRD based on f factors. We restrict our attention to $f \leq 5$ since the BBD for $f > 5$ is based on a partially balanced incomplete block design and does not produce a consistent form of $\mathbf{K}$. However, with a suitable assignment of the whole-plot factors to the columns of the CRD version of the BBD, the construction methods discussed can be applied directly to produce equivalent estimation designs.

Table 3.1: VKM Method - Central Composite Design ($\alpha = \beta = 1$) for D$(2, 2)$.

Whole-Plot	z_1	z_2	x_1	x_2	Whole-Plot	z_1	z_2	x_1	x_2
1	-1	-1	-1	-1	6	1	0	0	0
	-1	-1	1	-1		1	0	0	0
	-1	-1	-1	1		1	0	0	0
	-1	-1	1	1		1	0	0	0
2	1	-1	-1	-1	7	0	-1	0	0
	1	-1	1	-1		0	-1	0	0
	1	-1	-1	1		0	-1	0	0
	1	-1	1	1		0	-1	0	0
3	-1	1	-1	-1	8	0	1	0	0
	-1	1	1	-1		0	1	0	0
	-1	1	-1	1		0	1	0	0
	-1	1	1	1		0	1	0	0
4	1	1	-1	-1	9	0	0	-1	0
	1	1	1	-1		0	0	1	0
	1	1	-1	1		0	0	0	-1
	1	1	1	1		0	0	0	1
5	-1	0	0	0	10	0	0	0	0
	-1	0	0	0		0	0	0	0
	-1	0	0	0		0	0	0	0
	-1	0	0	0		0	0	0	0

3.4.1 VKM Method

Central Composite Designs

To introduce the VKM method, consider a split-plot design with two hard-to-change factors, denoted by z_1 and z_2, and two easy-to-change factors, denoted by x_1 and x_2. A cuboidal ($\alpha = \beta = 1$) central composite design (CCD) in coded units is provided in Table 3.1 that enables equivalent estimation of a complete second-order model.

The design in Table 3.1 was built from a CCD with 4 factors comprised of 16 factorial points, 8 axial points, and 1 center run; a total of 25 experimental runs. First, the design was sorted such that WP factors remain constant within each WP and the subplot designs are orthogonal for a first-order plus interaction model. We see that the axial points of the WP factors are placed in a WP by themselves. To maintain balance, subplot runs are added as required to each whole-plot. The resulting split-plot design features 10 whole-plots of size 4. The first 4 whole-plots contain the factorial runs, whole-plots 5, 6, 7, and 8 contain the axial points for the whole-plot factors, whole-plot 9 provides the axial points for the subplot factors, and whole-plot 10 includes the center points.

While the overall size of the split-plot design has increased over the CRD by 15 subplot runs, the primary cost of split-plot experiment is dominated by the manipulation of the HTC factors. A better approach for judging the size of a split-plot design is to compare the number of whole-plots to the design size for a CRD based solely on the whole-plot factors. For example, a CCD with 2 factors requires 9 experimental runs, $N_{CRD_w} = 9$, or $m = 9$ whole-plots. From this perspective, we see that the design in Table 3.1 has only increased this by one whole-plot. In fact, this is a general result of the VKM method that the number of whole-plots $m = N_{CRD_w} + 1$, when $w \geq 2$ and the number of subplot factors, k, is a power of 2. If k is not a power of 2, then $m = N_{CRD_w} + k$.

The design in Table 3.1 provides 15 df for the estimation of a pure-error based estimate of subplot error variance and 10 degrees of freedom (df) for testing lack-of-fit. Unfortunately, there are no df available for obtaining an pure-error estimate of whole-plot error variance. As proposed by VKM (2005), the design can be easily augmented by replicating whole-plot number 10 to provide this estimate.

The form of the $\mathbf{K}$ matrix for a balanced VKM design based on the cuboidal central composite design when the number of subplot factors is a power of 2 is

$$\mathbf{K} = \begin{bmatrix} n_0 & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & n_{\mathrm{W}}\mathbf{I}_{\mathrm{D}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & n_{\mathrm{W}}\mathbf{I}_{\mathrm{Q}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V}_{\mathrm{S}} \end{bmatrix}, \tag{3.11}$$

where $n_0 = n_{\mathrm{W}} = n$ is the number of subplot runs per whole-plot. Based on this form of $\mathbf{K}$ and Equation 3.9, the condition that must be satisfied is

$$\mathbf{1}\mathbf{1}'\mathbf{S}_{\mathrm{Q}_i} = \mathbf{S}_{\mathrm{Q}_i}\mathbf{V}_{\mathrm{S}} \qquad \forall\ i. \tag{3.12}$$

For a central composite design, the $k \times k$ matrix $\mathbf{V}_{\mathrm{S}}$ was defined by VKM to be

$$\mathbf{V}_{\mathrm{S}} = \frac{n}{k}\mathbf{1}_k\mathbf{1}_k'.$$

For the design in Table 3.1, $n = 4$ and $k = 2$.

When the number of subplot factors is not a power of 2, we consider two options for constructing an equivalent design. The first approach assigns the axial points for each subplot factor to their own whole-plot and are replicated as required to achieve balance. For example, consider a design for the $\mathrm{D}(1,3)$ case shown in Table 3.2. In this designs, the subplot pure quadratics are whole-plot level effects and $\mathbf{V}_{\mathrm{S}}$ is a diagonal matrix. Similarly, when $k = 1$, the subplot pure quadratics are whole-plot level effects. The implications of this type of design in regard to inference is discussed in Vining and Kowalski (2004, 2005).

Table 3.2: VKM Method - Central Composite Design ($\alpha = \beta = 1$) for D$(1,3)$ with Subplot Pure Quadratics as Whole-Plot Terms.

Whole-Plot	z_1	x_1	x_2	x_3	Whole-Plot	z_1	x_1	x_2	x_3
1	-1	1	-1	-1	5	0	-1	0	0
	-1	-1	1	-1		0	1	0	0
	-1	-1	-1	1		0	-1	0	0
	-1	1	1	1		0	1	0	0
	-1	-1	-1	-1		0	-1	0	0
	-1	1	1	-1		0	1	0	0
	-1	1	-1	1		0	-1	0	0
	-1	-1	1	1		0	1	0	0
2	1	1	-1	-1	6	0	0	-1	0
	1	-1	1	-1		0	0	1	0
	1	-1	-1	1		0	0	-1	0
	1	1	1	1		0	0	1	0
	1	-1	-1	-1		0	0	-1	0
	1	1	1	-1		0	0	1	0
	1	1	-1	1		0	0	-1	0
	1	-1	1	1		0	0	1	0
3	-1	0	0	0	7	0	0	0	-1
	-1	0	0	0		0	0	0	1
	-1	0	0	0		0	0	0	-1
	-1	0	0	0		0	0	0	1
	-1	0	0	0		0	0	0	-1
	-1	0	0	0		0	0	0	1
	-1	0	0	0		0	0	0	-1
	-1	0	0	0		0	0	0	1
4	1	0	0	0	8	0	0	0	0
	1	0	0	0		0	0	0	0
	1	0	0	0		0	0	0	0
	1	0	0	0		0	0	0	0
	1	0	0	0		0	0	0	0
	1	0	0	0		0	0	0	0
	1	0	0	0		0	0	0	0
	1	0	0	0		0	0	0	0

An alternate approach, that maintains the subplot pure quadratics as subplot effects, is to construct a design with the whole-plot size equal to a multiple of the whole-plot containing all of the subplot axial points, $n_i = 2k$, and full or partial replicate of the subplot factorial points. For example in the $D(1,3)$ case we have $2^4/2 = 8$ factorial combinations for each whole-plot level and $2k = 6$ subplot axial runs. We can construct a design, with $n = 12$ by employing a one-half replicate of the subplot factorial design. The number of whole-plots in this case would be 6; a reduction of 2 over the previous method. While the resulting design possesses the equivalence property, it is not orthogonal in all of the main-effects and two-factor interactions. A correlation exists between the subplot by subplot and the subplot by whole-plot two-factor interactions. To rectify this situation, two additional whole-plots containing the alternate one-half fraction could be added to the design to balance the fractional factorial in the subplot factors with the whole-plot levels, resulting in $m = 8$. This design is illustrated in Table 3.3, where the vector notation in whole-plot number 6, 7, and 8 represents 12 replicated subplot runs. In applications where a larger whole-plot size is admissible and cost effective, this design construction approach is a viable alternative.

Box-Behnken Designs

Now consider a Box-Behnken design (BBD) for the D(2,2) case. As with the CCD, we sort a completely randomized BBD in four factors such that the whole-plot factors are constant within each whole-plot and we ensure orthogonal subplot designs. Design points with whole-plot factors simultaneously at their high or low level in combination with subplot centers are placed in a whole-plot by themselves. To maintain balance, subplot runs are added as required to each whole-plot. The resulting design in Table 3.4 features 10 whole-plots of size 4.

The **K** matrix has the same form as Equation 3.11, where $n_0 = n_{\mathrm{W}} = n$ is the number of subplot runs per whole-plot. To minimize the number of whole-plots, we have a single whole-plot with the HTC factors simultaneously set at it their center level. The size of this whole-plot, for $f \leq 5$, is

$$n_{max} = 4 \begin{pmatrix} k \\ 2 \end{pmatrix} = 2k(k-1).$$

In the other whole-plots, with the HTC factors at ± 1, replicates of the base subplot design may be required to balance the design. To define $\mathbf{V}_{\mathrm{S}}$, we consider the column sums of the pure quadratic terms in each whole-plot and the condition specified in Equation 3.12 to find

$$\begin{aligned} \mathbf{V}_{\mathrm{S}} &= 2 \left(\frac{4 \begin{pmatrix} k \\ 2 \end{pmatrix}}{2k} \right) \mathbf{1}_k \mathbf{1}_k' \\ \mathbf{V}_{\mathrm{S}} &= 2(k-1)\mathbf{1}_k \mathbf{1}_k'. \end{aligned}$$

Table 3.3: VKM Method - Central Composite Design ($\alpha = \beta = 1$) for D$(1, 3)$ Maintaining Subplot Pure Quadratics as Subplot Terms.

Whole-Plot	z_1	x_1	x_2	x_3	Whole-Plot	z_1	x_1	x_2	x_3
1	-1	1	-1	-1	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		-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	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
	-1	-1	1	1		-1	1	1	1
2	1	1	-1	-1	5	0	-1	0	0
	1	-1	1	-1		0	1	0	0
	1	-1	-1	1		0	0	-1	0
	1	1	1	1		0	0	1	0
	1	-1	-1	-1		0	0	0	-1
	1	1	1	-1		0	0	0	1
	1	1	-1	1		0	-1	0	0
	1	-1	1	1		0	1	0	0
	1	1	-1	-1		0	0	-1	0
	1	-1	1	-1		0	0	1	0
	1	-1	-1	1		0	0	0	-1
	1	1	1	1		0	0	0	1
3	1	1	-1	-1	6	$\mathbf{-1}$	**0**	**0**	**0**
	1	-1	1	-1	7	$\mathbf{+1}$	**0**	**0**	**0**
	1	-1	-1	1	8	**0**	**0**	**0**	**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					
	1	1	1	-1					
	1	1	-1	1					
	1	-1	1	1					

Table 3.4: VKM Method - Box-Behnken Design for D(2, 2)

Whole-Plot	z_1	z_2	x_1	x_2	Whole-Plot	z_1	z_2	x_1	x_2
1	-1	0	0	-1	6	1	-1	0	0
	-1	0	0	1		1	-1	0	0
	-1	0	-1	0		1	-1	0	0
	-1	0	1	0		1	-1	0	0
2	-1	1	0	0	7	1	0	0	-1
	-1	1	0	0		1	0	0	1
	-1	1	0	0		1	0	-1	0
	-1	1	0	0		1	0	1	0
3	0	-1	-1	0	8	-1	-1	0	0
	0	-1	1	0		-1	-1	0	0
	0	-1	0	-1		-1	-1	0	0
	0	-1	0	1		-1	-1	0	0
4	0	0	-1	-1	9	1	1	0	0
	0	0	1	-1		1	1	0	0
	0	0	-1	1		1	1	0	0
	0	0	1	1		1	1	0	0
5	0	1	-1	0	10	0	0	0	0
	0	1	1	0		0	0	0	0
	0	1	0	-1		0	0	0	0
	0	1	0	1		0	0	0	0

Table 3.5: VKM Method Box-Behnken Design for D(1, 3).

Whole-Plot	z_1	x_1	x_2	x_3	Whole-Plot	z_1	x_1	x_2	x_3
1	-1	-1	-1	0	3	0	-1	-1	0
	-1	1	-1	0		0	1	-1	0
	-1	-1	1	0		0	-1	1	0
	-1	1	1	0		0	1	1	0
	-1	-1	0	-1		0	-1	0	-1
	-1	1	0	-1		0	1	0	-1
	-1	-1	0	1		0	-1	0	1
	-1	1	0	1		0	1	0	1
	-1	0	-1	-1		0	0	-1	-1
	-1	0	1	-1		0	0	1	-1
	-1	0	-1	1		0	0	-1	1
	-1	0	1	1		0	0	1	1
2	1	-1	-1	0	4	0	0	0	0
	1	1	-1	0		0	0	0	0
	1	-1	1	0		0	0	0	0
	1	1	1	0		0	0	0	0
	1	-1	0	-1		0	0	0	0
	1	1	0	-1		0	0	0	0
	1	-1	0	1		0	0	0	0
	1	1	0	1		0	0	0	0
	1	0	-1	-1		0	0	0	0
	1	0	1	-1		0	0	0	0
	1	0	-1	1		0	0	0	0
	1	0	1	1		0	0	0	0

Recall, that a Box-Behnken design is derived from a balanced incomplete block design. For $f > 6$, the block size is 3, and the BBD is based on a partially balanced incomplete block design. The VKM construction can be applied when $f > 6$, but the $\mathbf{K}$ matrix takes on a slightly different form due to the partially balanced nature of the parent design. In the case of D(3, 3) and D(3, 4), the subplot pure quadratics are whole-plot level terms.

There are two cases within the range explored where a VKM construction BBD has not been found. These are D(1, 3) and D(2, 3). For the D(1, 3) case, a pseudo-crossed approach is proposed in Table 3.5 where a BBD in the subplot factors is crossed with the three levels of the whole-plot factor. The subplot designs are identical for the first three whole-plots and the fourth whole-plot contains all centers. This strategy can also be applied for the D(2, 3) case.

Table 3.6: Summary of VKM method Central Composite and Box-Behnken Designs.

		CCD		BBD	
w	k	m	n	m	n
1	1	6	2		
1	2	6	4	4	4
1	3	8	8	4	12
1	4	6	8	4	24
2	1	10	2	9	2
2	2	10	4	10	4
2	3	12	4	10	12
2	4	10	8	10	8
3	1	16	2	19	2
3	2	16	4	20	4
3	3	18	4	19	4
3	4	16	8	21	4

We now summarize the features and restrictions of the VKM construction method.

1. The conditions for equivalence for the CCD are not a function of α and β, therefore there is no restriction on their values. In addition, an independent selection of the α and β distances can be made for the whole-plot and subplot factors. For example, to minimize the number of whole-plot levels we could choose $\beta = 1$, however at the subplot level we could choose $\alpha > 1$ generating addition lack-of-fit degrees of freedom.

2. A subplot design that contains center points, zero coded level, must contain only center points.

3. In a CCD with the number of subplot factors that is not a power of 2, additional whole-plots are required.

4. For a CCD with $k = 1$, the subplot pure quadratics are whole-plot level effects, typically resulting in less power to detect them as significant.

Table 3.6 provides a summary of the design size for the VKM method construction split-plot CCD's and BBD's for all combinations of $1 \leq w \leq 3$ and $1 \leq k \leq 4$. A BBD for $\text{D}(1, 1)$ does not exist, even for a completely randomized design.

3.4.2 MWP Method

In this section we propose a class of equivalent estimation designs that require a minimum number of whole-plots. As previously discussed, in a split-plot experiment, we assume that the cost is primarily a function of the number of whole-plots, rather than the overall design size. Therefore, our goal is to find a class of split-plot designs that requires the minimum number of whole-plots.

Central Composite Design, $w = 1$ **and** $\beta = 1$

For the one whole-plot factor case, we choose $\beta = 1$, defining a cuboidal region in the whole-plot factors. Such a choice allows us to reduce the number of whole-plots, and the resulting design will require only three levels of the whole-plot factor. The value for α for the subplot factors is not restricted.

First, we sort the rows of the design by the whole-plot factor column. This results in 3 groups of design points - one for each level of the whole-plot factor. To obtain a balanced design, we add subplot center points within each whole-plot. Furthermore, subplot centers may be added one-by-one to the whole-plots to increase the degrees of freedom for a subplot pure-error estimate. To complete the design, we may choose to include whole-plot replicates.

A MWP design for D(1, 3) is provided in Table 3.7, and it features 3 whole-plots of size 9. Comparing this design structure with the VKM method, we see two distinct differences. First, we have subplot designs that do not consist entirely of subplot centers. Second, there is no whole-plot that contains all whole-plot and subplot centers. These differences are allowed due to a change in the structure of the $\mathbf{K}$ matrix.

For the MWP class of designs, the general form of $\mathbf{K}$ is given by

$$\mathbf{K} = \begin{bmatrix} n_0 & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{m_0}' \\ \mathbf{0} & n_{\mathrm{W}}\mathbf{I}_{\mathrm{D}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & n_{\mathrm{W}}\mathbf{I}_{\mathrm{Q}} & \mathbf{0} & \mathbf{M}_{\mathrm{W}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{3.13}$$

where $n_0 = n_{\mathrm{W}} = n$, $\mathbf{m_0}$ is a vector of length k, and $\mathbf{M}_{\mathrm{W}}$ is a $w \times k$ matrix. Based on this form of $\mathbf{K}$ and Equation 3.9, the condition is

$$\mathbf{1}_n\mathbf{1}_n'\mathbf{S}_{\mathrm{Q}_i} = \mathbf{1}_n\mathbf{m_0}' + \mathbf{W}_{\mathrm{Q}_i}\mathbf{M}_{\mathrm{W}} \qquad \forall\, i. \tag{3.14}$$

Equation 3.14 illustrates the relationship between the whole-plot and subplot pure quadratic terms that must be satisfied for every whole-plot design. Let s_{spa} be the the column sum of $\mathbf{S}_{Q_i}$ for the whole-plot containing the whole-plot factor at its center level (zero level). If the

Table 3.7: MWP Method - Central Composite Design ($\alpha = \beta = 1$) for $D(1, 3)$.

Whole-Plot	z_1	x_1	x_2	x_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	1	1	1
	-1	0	0	0
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
	1	1	1	-1
	1	1	1	1
	1	0	0	0
3	0	-1	0	0
	0	1	0	0
	0	0	-1	0
	0	0	1	0
	0	0	0	-1
	0	0	0	1
	0	0	0	0
	0	0	0	0
	0	0	0	0

subplot axial runs are not replicated within this whole-plot, then $s_{spa} = 2\alpha^2$. Therefore, for the whole-plot with the HTC factor at its center level,

$$\mathbf{1}_n\mathbf{1}_n'\mathbf{S}_{\mathrm{Q}_i} = s_{spa}\mathbf{1}_n\mathbf{1}_k' = 2\alpha^2\mathbf{1}_n\mathbf{1}_k'.$$

In addition for these whole-plots, $\mathbf{W}_{\mathrm{Q}_i}\mathbf{M}_{\mathrm{W}}$ is a vector of zeroes. Therefore, we find

$$\mathbf{m_0} = s_{spa}\mathbf{1}_k = 2\alpha^2\mathbf{1}_k \tag{3.15}$$

Similarly, let s_{spf} equal the column sum of $\mathbf{S}_{Q_i}$ for the whole-plot containing the whole-plot factor at its factorial level (± 1), which is also its axial level due to our choice of $\beta = 1$. Therefore, s_{spf} equals to the number of subplot factorial points contained in these whole-plots, n_f. For these whole-plots,

$$\mathbf{1}_n\mathbf{1}_n'\mathbf{S}_{\mathrm{Q}_i} = s_{spf}\mathbf{1}_n\mathbf{1}_k' = n_f\mathbf{1}_n\mathbf{1}_k'.$$

Since the HTC factor remains constant at plus or minus one during a whole-plot, $\mathbf{W}_{\mathrm{Q}_i} = \mathbf{1}_n\mathbf{1}_w'$, however since $w = 1$ this becomes $\mathbf{1}_n$. Now combining Equation 3.14 and Equation 3.15, we find

$$\mathbf{M}_{\mathrm{W}} = (s_{spf} - s_{spa})\,\mathbf{1}_k' = \left(n_f - 2\alpha^2\right)\mathbf{1}_k'.$$

For the example design presented in Table 3.7, $2\alpha^2 = 2$ and $n_f = 8$. Therefore, $\mathbf{m_0} = 2\mathbf{1}_3$ and $\mathbf{M}_{\mathrm{W}} = 6\mathbf{1}_3'$. We see that there is no restriction on our choice of α in the $w = 1$ case.

From the derivation, we see that the relationship to achieve equivalence is not dependent on the whole-plot size, n; rather it is dependent on the column sum of the pure quadratic terms within each whole-plot. Therefore, we can add subplot center runs to each whole-plot and not change the column sums, thereby maintaining the equivalence. In this way, we can add subplot centers to all of the whole-plots one-by-one. In addition, because the equivalence is independent whole-plot size, we are able to overcome the restriction in the VKM method for cases where k is not a power of 2.

A potentially unattractive feature of these designs is that they cannot contain a whole-plot consisting entirely of whole-plot and subplot centers (all zeroes). This would result in a different column sum of the subplot pure quadratic model terms with the center setting of the whole-plot factor, thereby violating the conditions stated in the derivation.

Central Composite Design, $w > 1$ or $\beta \neq 1$

In this section, we extend the previous results and do not restrict the choice of the values of α or β. However, it is shown that the equivalence condition is a function of α and β. In the case of $w > 1$ or $\beta \neq 1$, we cannot combine the whole-plot axial runs with the whole-plot factorial runs in a single whole-plot. Therefore, the minimum number of whole-plots required is given by

$$m_{min} = 2^w + 2w + 1.$$

This is based on 2^w whole-plot factorial combinations, $2w$ whole-plot axial points, and one with the whole-plot center point. We assume that a full factorial is present in the whole-plot factors. As with the case of one whole-plot factor, it is desirable to include whole-plot replication to obtain a pure-error estimate of whole-plot variance. In this case, we would need to increase m_{min} by the number of degrees of freedom associated with the whole-plot pure-error estimate. As an example, consider a design $\mathrm{D}(2,2)$. A minimum of $2^2 + 2(2) + 1 = 9$ whole-plots are required without whole-plot replication, and 11 whole-plots are required if we choose to have two degrees of freedom available for estimating whole-plot error variance.

For this class of designs, the general form of $\mathbf{K}$ and the equivalence conditions are the same as $w = 1$ given in Equations 3.13 and 3.14, respectively. In these designs, there are three cases of whole-plot types defined by;

1. All WP factors at their center level in combination with SP factors at their axial level and center level.
2. All WP factors at their factorial level with SP factors at their factorial and center level.
3. A WP factor at its axial level in combination with the other WP and SP factors at their center level.

As with $w = 1$, $s_{spa} = 2\alpha^2$ if there is one replicate of the subplot axial points and applies to case (1). There are situations when the subplot size allows for replicates of the subplot axial points. If the subplot axial points are replicated r_α times within a whole-plot, then $s_{spa} = 2r_\alpha\alpha^2$. Also as before, $s_{spf} = n_f$ applies to case (2). Lastly, the subplot quadratic column sum for case (3) is zero. For case (1), $\mathbf{W}_\mathrm{Q}$ is a matrix of zeroes. Therefore,

$$\mathbf{m_0} = s_{spa}\mathbf{1}_k = 2r_\alpha\alpha^2\mathbf{1}_k.$$

For case (2), $\mathbf{1}_n\mathbf{1}_n'\mathbf{S}_{\mathrm{Q}_i} = n_f\mathbf{1}_n\mathbf{1}_k'$. Since the HTC factor remains constant at plus or minus one in a whole-plot, then $\mathbf{W}_{\mathrm{Q}_i} = \mathbf{1}_n\mathbf{1}_w'$. For this case,

$$\mathbf{1}_n\mathbf{1}_w'\mathbf{M}_\mathrm{W} = \left(n_f - 2\alpha^2\right)\mathbf{1}_n\mathbf{1}_k'$$

where $\mathbf{1}_w'\mathbf{M}_\mathrm{W}$ represents a row vector of the column sums of the $\mathbf{M}_\mathrm{W}$ matrix. Due to symmetry of the design, we can assume that $\mathbf{M}_\mathrm{W}$ is of the form $\phi\mathbf{1}_w\mathbf{1}_k'$, where ϕ is a constant. Therefore, the row vector of column sum is equal to $w\phi\mathbf{1}_k'$ and we can write

$$w\phi\mathbf{1}_n\mathbf{1}_k' \quad = \quad \left(n_f - 2r_\alpha\alpha^2\right)\mathbf{1}_n\mathbf{1}_k',$$

$$\begin{aligned}\phi \mathbf{1}_n \mathbf{1}_k' &= \frac{(n_f - 2r_\alpha \alpha^2)}{w} \mathbf{1}_n \mathbf{1}_k' \\ \phi &= \frac{(n_f - 2r_\alpha \alpha^2)}{w}. \end{aligned} \tag{3.16}$$

For case (3), $\mathbf{1}_n \mathbf{1}_n' \mathbf{S}_{\mathrm{Q}_i} = \mathbf{0}_{n\times k}$. Since one of the HTC factor remains constant at plus or minus β during a whole-plot,

$$\mathbf{W}_{\mathrm{Q}_i} = \beta^2 \left[\begin{array}{cccc} \mathbf{1}_n & \mathbf{0} & \cdots & \mathbf{0} \end{array} \right]_{n\times w}$$

where the position of the vector of ones will change depending on which WP factor is at its axial level. Based on this structure, we see that

$$\mathbf{W}_{\mathrm{Q}_i} \mathbf{M}_{\mathrm{W}} = \beta^2 \phi \mathbf{1}_n \mathbf{1}_k'$$

Therefore, we obtain

$$\begin{aligned} 2r_\alpha \alpha^2 \mathbf{1}_n \mathbf{1}_k' &= -\beta^2 \phi \mathbf{1}_n \mathbf{1}_k' \\ \phi &= -\frac{2r_\alpha \alpha^2}{\beta^2}. \end{aligned} \tag{3.17}$$

Both Equation 3.16 and 3.17 must be satisfied simultaneously. Therefore, we have

$$\phi = \frac{(n_f - 2r_\alpha \alpha^2)}{w} = -\frac{2r_\alpha \alpha^2}{\beta^2}. \tag{3.18}$$

From this derivation, we see that there is a relationship between the values of α and β. Rearranging Equation 3.18 , we see that

$$n_f = 2r_\alpha \alpha^2 - \frac{w(2r_\alpha \alpha^2)}{\beta^2}. \tag{3.19}$$

If we assume that $n_f \geq 2$ then we find that $2r_\alpha \alpha^2 - \frac{w(2r_\alpha \alpha^2)}{\beta^2} \geq 2$. Simplifying this expression does not result in any practical restriction on α and β. If we desire a design with $\alpha = \beta$ then using Equation 3.19 we find that

$$\alpha^2 = \beta^2 = \frac{n_f}{2r_\alpha} + w. \tag{3.20}$$

Alternatively, solving Equation 3.19 for α^2 we obtain

$$\alpha^2 = \frac{n_f}{2r_\alpha \left(1 - \frac{w}{\beta^2}\right)}.$$

Table 3.8: MWP Method - Axial Point Distances for $\alpha = \beta$.

w	k	q	$\alpha = \beta$
1	1	0	1.4142
1	2	0	1.7321
1	3	0	2.2361
1	4	1	2.2361
2	1	0	1.7321
2	2	0	2.0000
2	3	1	2.0000
2	4	1	2.4495
3	1	0	2.0000
3	2	0	2.2361
3	3	1	2.2361
3	4	1	2.6458

Based on the constraint that $\alpha^2 > 0$, we find that

$$\left(1 - \frac{w}{\beta^2}\right) > 0 \Rightarrow \beta > \sqrt{w}.$$

Furthermore, solving Equation 3.19 for β^2 we obtain

$$\beta^2 = \frac{w(2r_\alpha \alpha^2)}{2r_\alpha \alpha^2 - n_f}.$$

Based on the constraint that $\beta^2 > 0$, we find that

$$2r_\alpha \alpha^2 - n_f > 0 \Rightarrow \alpha > \sqrt{\frac{n_f}{2r_\alpha}}.$$

From Equation 3.20, we provide a summary of the $\alpha = \beta$ values in Table 3.8. These values depend on the fraction used in the factorial portion of the design indicated by q in a 2^{f-q} fractional factorial. In all cases, the value is equivalent to the rotatable or spherical choices for a completely randomized design, except for D$(1, 3)$ and D$(3, 3)$. In these two cases, the value remains reasonable. This table demonstrates that the restriction on the value of α and β does not create any practical issues if they are chosen to be equal.

As an example, consider an MWP design for D$(2, 2)$ supplied in Table 3.9. Based on our previous calculation of $m_{min} = 9$, we see that this design achieves the minimum number of whole-plots. The first four whole-plots contain the factorial points, whole-plots 5 - 8 contain the axial points of the HTC factors, and whole-plot 9 contains the axial points of the ETC factors and a center point. In this design, $\alpha = \beta = 2$ provides a spherical and rotatable design in the context of a completely randomized design. Note that this design contains a single center point in all factors and that the general recommendation for the number of center points in a CRD central composite design is 3 to 5. The design can be augmented by adding 2 subplot center points in each whole-plot to achieve a total of 3 subplot center points in whole-plot number 9.

To specify the **K** matrix for this case, we apply Equation 3.18 and find that $\phi = -2$ and $r_\alpha = 1$; therefore,

$$\begin{aligned} \mathbf{m_0} &= 8\mathbf{1}_2 \\ \mathbf{M}_{\mathrm{W}} &= -2\mathbf{1}_2\mathbf{1}_2'. \end{aligned}$$

Box-Behnken Designs

The MWP design construction can also be applied to Box-Behnken designs (BBD). To compute the minimum number of whole-plots required for a BBD we consider all combinations of the WP factors at their three-levels (-1, 0, $+1$). For $w = 1$, $m_{min} = 3$, and for $w > 1$ and $f \leq 5$,

$$m_{min} = 4\begin{pmatrix} w \\ 2 \end{pmatrix} + 2w + 1.$$

For example, the minimum number of whole-plots for the D(2, 2) case requires, $m_{min} = 4(1) + 4 + 1 = 9$. Note that for $f > 5$, the BBD is based on a partially balanced incomplete block design and the number of whole-plots depends on the assignment of the whole-plot factors to the columns of the completely randomized design. The development that follows is based on $f \leq 5$. The construction of designs for $f > 5$ is not described; however, a summary table of these designs is provided at the end of this section.

To generate an equivalent SPD from a CRD BBD, we proceed as discussed by sorting and ensuring orthogonal subplot designs. Consider an equivalent estimation BBD for D(2, 2) provided in Table 3.10.

The MWP form of **K** from Equation 3.13 and the condition specified in Equation 3.14 still hold.

As with the CCD, we consider the types of whole-plots based on the level and combinations of the whole-plot (WP) and subplot (SP) factors defined by the following three cases;

1. All WP factors simultaneously at their center level in combination with SP factors at ±1 and center level.

Table 3.9: MWP Method - Central Composite Design ($\alpha = \beta = 2$) for D$(2, 2)$.

Whole-Plot	z_1	z_2	x_1	x_2	Whole-Plot	z_1	z_2	x_1	x_2
1	-1	-1	-1	-1	6	2	0	0	0
	-1	-1	-1	1		2	0	0	0
	-1	-1	1	-1		2	0	0	0
	-1	-1	1	1		2	0	0	0
	-1	-1	0	0		2	0	0	0
2	-1	1	-1	-1	7	0	-2	0	0
	-1	1	-1	1		0	-2	0	0
	-1	1	1	-1		0	-2	0	0
	-1	1	1	1		0	-2	0	0
	-1	1	0	0		0	-2	0	0
3	1	-1	-1	-1	8	0	2	0	0
	1	-1	-1	1		0	2	0	0
	1	-1	1	-1		0	2	0	0
	1	-1	1	1		0	2	0	0
	1	-1	0	0		0	2	0	0
4	1	1	-1	-1	9	0	0	-2	0
	1	1	-1	1		0	0	2	0
	1	1	1	-1		0	0	0	-2
	1	1	1	1		0	0	0	2
	1	1	0	0		0	0	0	0
5	-2	0	0	0					
	-2	0	0	0					
	-2	0	0	0					
	-2	0	0	0					
	-2	0	0	0					

Table 3.10: MWP Method - Box-Behnken Design for D(2, 2).

Whole-Plot	z_1	z_2	x_1	x_2	Whole-Plot	z_1	z_2	x_1	x_2
1	-1	-1	0	0	6	1	0	0	-1
	-1	-1	0	0		1	0	0	1
	-1	-1	0	0		1	0	-1	0
	-1	-1	0	0		1	0	1	0
	-1	-1	0	0		1	0	0	0
2	1	-1	0	0	7	0	-1	-1	0
	1	-1	0	0		0	-1	1	0
	1	-1	0	0		0	-1	0	-1
	1	-1	0	0		0	-1	0	1
	1	-1	0	0		0	-1	0	0
3	-1	1	0	0	8	0	1	-1	0
	-1	1	0	0		0	1	1	0
	-1	1	0	0		0	1	0	-1
	-1	1	0	0		0	1	0	1
	-1	1	0	0		0	1	0	0
4	1	1	0	0	9	0	0	-1	-1
	1	1	0	0		0	0	-1	1
	1	1	0	0		0	0	1	-1
	1	1	0	0		0	0	1	1
	1	1	0	0		0	0	0	0
5	-1	0	0	-1					
	-1	0	0	1					
	-1	0	-1	0					
	-1	0	1	0					
	-1	0	0	0					

2. Two WP factors at their ± 1 level simultaneously with SP factors at their center level.
3. A WP factor at ± 1 in combination with the other WP factors at their center level and one SP factor at its ± 1 level.

Note Case (2) is not present for $w = 1$, and Case (3) is a result of the block size of two for $f < 5$ in the construction of the BBD based on a balanced incomplete block design.

We follow a similar analysis as that described in detail for the central composite design to define the elements of $\mathbf{K}$. It can be shown, that the column sum of the subplot pure quadratic terms for the Case (1) is $s_1 = 4(k-1)$ and $\mathbf{W}_{\mathrm{Q}}$ is a matrix of zeroes. Therefore, $\mathbf{m_0} = 4(k-1)\mathbf{1}_k$. For Case (2), the column sum $s_2 = 0$ and $\mathbf{W}_{\mathrm{Q}_i} = \begin{bmatrix} \mathbf{1}_n & \mathbf{1}_n & \cdots & \mathbf{0} \end{bmatrix}_{n\times w}$, where the position of the vector of two vectors of ones will change depending on which two WP factor are at their ± 1 level. For Case (3), $s_3 = 2$ and $\mathbf{W}_{\mathrm{Q}_i} = \begin{bmatrix} \mathbf{1}_n & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}_{n\times w}$, where the position of the vector of ones will change depending on which WP factor is at its ± 1 level. As before, we define $\mathbf{M}_{\mathrm{W}}$ to be of the form $\phi \mathbf{1}_w \mathbf{1}'_k$. For Case (2), from Equation 3.14 we have

$$
\begin{aligned}
\mathbf{1}_n \mathbf{1}'_n \mathbf{S}_{\mathrm{Q}_i} &= \mathbf{1}_n \mathbf{m}'_{\mathbf{0}} + \mathbf{W}_{\mathrm{Q}} \mathbf{M}_{\mathrm{W}} \\
\mathbf{1}_n \mathbf{m}'_{\mathbf{0}} &= -\mathbf{W}_{\mathrm{Q}} \mathbf{M}_{\mathrm{W}} \\
4(k-1)\mathbf{1}_n \mathbf{1}'_k &= -\phi \begin{bmatrix} \mathbf{1}_n & \mathbf{1}_n & \cdots & \mathbf{0} \end{bmatrix}_{n\times w} \mathbf{1}_w \mathbf{1}'_k \\
4(k-1)\mathbf{1}_n \mathbf{1}'_k &= -\phi 2 \mathbf{1}_n \mathbf{1}'_k \\
\phi &= -\frac{4(k-1)}{2}
\end{aligned}
$$

For Case (3), from Equation 3.14 we have

$$
\begin{aligned}
2\mathbf{1}_n \mathbf{1}'_k &= 4(k-1)\mathbf{1}_n \mathbf{1}'_k + \phi \begin{bmatrix} \mathbf{1}_n & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}_{n\times w} \mathbf{1}_w \mathbf{1}'_k \\
2\mathbf{1}_n \mathbf{1}'_k &= 4(k-1)\mathbf{1}_n \mathbf{1}'_k + \phi \mathbf{1}_n \mathbf{1}'_k \\
\phi &= 2 - 4(k-1) = 6 - 4k.
\end{aligned}
$$

Therefore, we summarize these expressions for ϕ as

$$
\phi = \begin{cases} 6 - 4k & \text{if } w = 1 \\ -\frac{4(k-1)}{2} = 6 - 4k & \text{if } w > 1,\ k > 1 \end{cases} .
$$

As a result of the restriction on ϕ, MWP method BBD's are not permissible for D$(2,3)$, D$(2,4)$, D$(3,3)$, and D$(3,4)$. In addition, other special cases include D$(2,1)$ and D$(3,1)$, where the MWP BBD is equivalent to the VKM construction BBD.

For the design in Table 3.10, we have $w = k = 2$, and $\phi = -2$. To define the $\mathbf{K}$ matrix we have, $\mathbf{m_0} = 4\mathbf{1}_2$ and $\mathbf{M}_{\mathrm{W}} = -2\mathbf{1}_2\mathbf{1}_2'$.

Once again, as with the central composite designs, we note that the relationship to achieve equivalence is not dependent on the whole-plot size, n. Therefore, we can add subplot center runs to each whole-plot and not change the column sums, thereby maintaining the equivalence. We can summarize the features and restrictions of the minimum whole-plot construction method as follows.

1. It establishes the lower boundary for the number of whole-plots required to conduct a central composite and Box-Behnken design in split-plot mode.
2. The MWP design structure allows for subplot centers to be included within each whole-plot, enabling separate testing of curvature at the subplot level and whole-plot level (assuming there are whole-plot replicates). In addition, the subplot centers can be added one-by-one to each whole-plot. Note that this feature is not available in the cases of a BBD for D$(2,1)$ and D$(3,1)$, which are identical to the VKM method designs and therefore do not allow subplot centers to be added one-by-one.
3. For a CCD with $w > 1$ or $\beta \neq 1$, the equivalence of the design is dependent of the choice of α and β.
4. A whole-plot containing all center points for both the whole-plot and subplot factors violates the equivalence condition.

Table 3.11 provides a summary of the design size for the MWP method construction split-plot CCD's and BBD's for all combinations of $1 \leq w \leq 3$ and $1 \leq k \leq 4$, where a design has been identified.

3.5 $\mathcal{D}$-optimal Split-Plot Designs

In the previous sections, the approach to design construction was based on an analytical derivation of the form of the $\mathbf{K}$ matrix to satisfy the equivalence condition. An additional application of Equation 3.3 is the ability to verify numerically the equivalence of a particular design without knowledge of $\mathbf{K}$. In fact, if the design is found to be equivalent, then the form of $\mathbf{K}$ can be computed directly. This approach to design construction involves generating a split-plot design and then testing it for equivalence. The parent design can be generated using a numerical search algorithm, as is the case with a $\mathcal{D}$-optimal design, or based on

Table 3.11: Summary of MWP method Central Composite and Box-Behnken Designs.

		CCD		BBD	
w	k	m	n	m	n
1	1				
1	2	4	4	3	5
1	3	4	12	3	13
1	4	4	24	3	25
2	1	9	2	9	2
2	2	10	4	9	5
2	3	10	12		
2	4	10	8		
3	1	19	2	19	2
3	2	20	4	19	5
3	3	19	4		
3	4	21	4		

other families of completely randomized designs. In this section we illustrate this approach by generating minimal size balanced equivalent estimation $\mathcal{D}$-optimal designs.

Consider D$(1, k)$ where $k \in \{1, 2, 3, 4\}$. As previously discussed, $m_{min} = 3$ in the $w = 1$ case. To find the subplot size, we first compute p, the number of terms in a complete second-order model, including the intercept from Equation 3.7. To support the model and provide at least one degree of freedom above saturation, we require a minimum of $p+1$ unique design points. Therefore, to achieve a balanced design the minimum subplot size is the next largest integer value of

$$n_{min} = \frac{p+1}{m_{min}}.$$

As an example, consider D$(1, 3)$ with $p = 15$ that results in a design with $m_{min} = 3$ whole-plots and $n = 6$ subplot runs per whole-plot.

To generate a design, we apply the point-exchange algorithm proposed by Goos (2002) available at the author's website. Our candidate set of points is a full three-level factorial in four factors. Following the advice offered in Goos (2002), we set the ratio of whole-plot to subplot variance equal to one. He states that in most cases, the design selection is relatively robust to the choice of this ratio. We have restricted the search algorithm to designs that have 3 whole-plots of size 6. A proposed design for this case is provided in Table 3.12.

Applying Equation 3.3 it can be shown that the design in Table 3.12 satisfies the condition, and therefore the OLS and GLS estimates are equivalent. Note that the search algorithm is not constrained to equivalent estimation designs, and therefore the design is $\mathcal{D}$-optimal

Table 3.12: $\mathcal{D}$-optimal Design for D$(1,3)$.

Whole-Plot	z_1	x_1	x_2	x_3
1	-1	-1	-1	-1
	-1	1	-1	1
	-1	-1	1	1
	-1	1	1	-1
	-1	1	1	1
	-1	0	0	-1
2	1	1	-1	1
	1	-1	1	1
	1	1	1	-1
	1	-1	-1	0
	1	-1	0	-1
	1	0	-1	-1
3	0	1	1	1
	0	-1	-1	1
	0	-1	1	-1
	0	1	-1	-1
	0	1	0	0
	0	0	1	0

over all designs in the class of D$(1,3)$ with $m = 3$ and $n = 6$. While it is not necessary to compute $\mathbf{K}$ to verify equivalence, it was computed from Equation 3.4. Inspecting the structure of $\mathbf{K}$ revealed that this design features the ability to add subplot centers one-by-one to each whole-plot without disturbing the equivalence.

This class of designs possesses the following desirable attributes;

- provides equivalent estimation enabling variance-independent parameter estimates
- employs a minimum number of whole-plots and whole-plot levels
- achieves minimum overall size for a non-saturated balanced design
- provides efficiency of estimation as measured by the $\mathcal{D}$-criterion.

However, the design does not feature any replicates at either the whole-plot or subplot level excluding the possibility of pure-error, model-independent estimates of the variance components. Whole-plot replicates can be performed as desired. To obtain subplot replicates, we exploit the ability to add subplot centers one-by-one and augment each whole-plot with two subplot center runs, resulting in a design that contains 3 whole-plots of size 8. Consequently,

Table 3.13: Summary of $\mathcal{D}$-optimal Split-Plot Designs for $w = 1$.

w	k	m	n
1	1	3	3
1	2	3	4
1	3	3	6
1	4	3	8

the relative $\mathcal{D}$-efficiency of the resulting design has been deteriorated, which highlights the design trade-offs that must be considered.

Considering the design in Table 3.12 from a cost perspective reveals that it establishes a lower-bound for a balanced design because it achieves minimum size. If we compare the design in Table 3.12 to those in Table 3.6 and 3.11, we see that it contains the same or fewer number of whole-plots, and the subplot size is typically smaller. In this case, even a comparison of overall size, as measured by the total number of subplot runs, reveals that this split-plot design has fewer subplot runs than a completely randomized CCD or BBD in 4 factors.

An interesting feature of this design is that the subplot designs are not orthogonal. As previously stated, orthogonality is not a necessary condition, although it is desirable. This example demonstrates the flexibility in generating equivalent estimation designs.

A complete class of $\mathcal{D}$-optimal designs for $\mathrm{D}(1, k)$, $k \in \{1, 2, 3, 4\}$ has been generated using this method and is summarized in Table 3.13. It is interesting to note that the $\mathcal{D}$-optimal design for $\mathrm{D}(1, 1)$ is identical to the MWP method and crossed construction CCD with $\alpha = \beta = 1$. The ability to numerically verify that a design provides equivalent estimates using Equation 3.3 has proven useful in exploring other families of RSM designs that do not provide a consistent form of $\mathbf{K}$, for example equiradial and small-composite designs proposed by Hartley (1959).

3.6 Cost-based Criterion

In many cases, there are competing equivalent split-plot designs available for the same number of whole-plot and subplot factors. If there are no practical restrictions, a cost-based approach to design selection is proposed. As in the case of a completely randomized design, there are numerous criteria that can be used to compare competing designs. However, the proposed cost-based approach is unique to the split-plot situation where we have different types of experimental units, with different associated costs. Note that we are not proposing a single criterion for design selection; rather we are considering an additional criterion when hard-to-change factors are present. Other criteria to compare these designs include the available degrees-of-freedom for variance component estimation, the efficiency of parameter estimates, and the distribution of the prediction variance. In the following approach, we will assume that all other criteria are not necessarily equal, however they are acceptable for the designs to be considered in competition.

To form a cost-based criterion we modify an approach originally proposed by Bisgaard (2000). We assume that there is one level of split-plotting, thus only two types of experimental units. In addition, we assume that the design is balanced, and we make the common assumption that the cost of changing a whole-plot factor exceeds the cost of changing a subplot factor. In the case of more than one whole-plot or subplot factor, we consider the cost of setting the combination of these factors, not the setting of each factor individually. In other words, we refer to the whole-plot cost as the cost associated with setting all of the whole-plot factors for a particular design point. Similarly, this is true with the subplot factors. Let the ratio of the cost be defined as

$$\xi = \frac{W_c}{S_c}$$

where W_c and S_c are the cost associated with changing the whole-plot and subplot settings, respectively. In situations where obtaining a response from a subplot run is costly, S_c represents the composite cost of setting the subplot factors and obtaining a response. We assume that a simple expression can be used to represent the cost of the entire experiment as

$$\text{Cost}_d = W_c(m_d) + S_c(m_d \times n_d)$$

where m_d is the number of whole-plots and n_d is the number of subplots runs within each whole-plot. These values represent the number of times that the whole-plot and subplot factors are set during the experiment. The subscript d denotes the cost for design d. Without loss of generality, we will assume that $S_c = 1$, and therefore $\xi = W_c$. In this way, ξ represents the whole-plot cost as a multiple of the subplot cost. Now, we can express the Cost_d as

$$\text{Cost}_d = m_d(\xi + n_d).$$

Consider the cost for two competing designs, denoted as **1** and **2**. We want to find the value of ξ that would cause us to prefer Design **1** over Design **2** purely from a cost perspective. In other words, we desire the value of ξ that satisfies the following inequality,

$$\text{Cost}_1 < \text{Cost}_2.$$

Expressing this as a ratio and simplifying, we would prefer Design **1** over Design **2** from a cost perspective if

$$\xi < \frac{m_2 n_2 - m_1 n_1}{m_1 - m_2} = \frac{N_2 - N_1}{m_1 - m_2}.$$

This expression applies when $N_1 \neq N_2$ and $m_1 \neq m_2$. If $N_1 = N_2$, then we would prefer the design with fewer whole-plots. Alternatively, if $m_1 = m_2$, then we would prefer the design with the smaller number of subplot runs.

As an example, consider two competing central composite designs for 2 whole-plot factor and 2 subplot factors. Design **1** is the VKM design in Table 3.1 with $m_1 = 10$ and $n_1 = 4$. Design **2** is the MWP in Table 3.9 with $\alpha = \beta = 2$, $m_2 = 9$, and $n_2 = 5$. Design **1** is smaller overall, however it has an additional whole-plot. Recognizing the higher cost of whole-plots, we would tend to prefer the design with the smaller number of whole-plots, even though the total design size is larger. In order for us to prefer Design **1** based on cost, we see that

$$\xi < \frac{9(5) - 10(4)}{10 - 9}$$

$$\xi < 5.$$

If the cost of resetting the whole-plots is less than 5 times larger than the cost of resetting the subplots, we would prefer Design **1** even though it requires more whole-plots.

Chapter 4

Unbalanced and Minimal Point Designs

In Chapter 3, we proposed systematic design construction techniques for balanced equivalent estimation designs. To achieve balance in these designs, subplot runs were replicated within each whole-plot. Unbalanced versions of these designs can reduce this replication of the subplot runs allowing more efficient allocation of experimental resources at the subplot level. In addition, when the whole-plot size is large, replicating subplot runs to achieve balance generates an excessive number of subplot error variance degrees of freedom. The proposed unbalanced designs provide attractive alternatives to the practitioner.

In this chapter, we outline systematic design construction techniques to build unbalanced equivalent estimation designs from the CCD and BBD, and illustrate minimal point split-plot designs. In most cases, the design points and the number of whole-plots are unchanged from the balanced versions. However, replication of subplot runs within each whole-plot has been reduced, thereby reducing the size of some whole-plots. This development does not attempt to generalize all of the admissible forms of unbalance. Instead, it provides a strategy to reduce the size of the whole-plots containing subplot replicates, especially those containing replicated subplot centers.

4.1 An Unbalanced Form of K for a Second-Order Model

We employ the same second-order model from Equation 3.6, and partitioning of the model matrix from Equation 3.8. A refined form of $\mathbf{K}$ for the unbalanced case is,

$$\mathbf{K}_{(p\times p)} = \begin{bmatrix} n_0 & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{m_0}' \\ \mathbf{0} & n_\mathrm{W}\mathbf{I}_\mathrm{D} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{u}_\mathrm{W} & \mathbf{0} & n_\mathrm{W}\mathbf{I}_\mathrm{Q} & \mathbf{0} & \mathbf{M}_\mathrm{W} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{u}_\mathrm{S} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V}_\mathrm{S} \end{bmatrix}. \tag{4.1}$$

where n_0 and n_W are scalars, $\mathbf{m_0}$ is a vector of length k, $\mathbf{I}_\mathrm{D}$ and $\mathbf{I}_\mathrm{Q}$ are $(w+\frac{w(w-1)}{2})\times(w+\frac{w(w-1)}{2})$ and $w\times w$ identity matrices, respectively, $\mathbf{u}_\mathrm{W}$ is a vector of length w, $\mathbf{u}_\mathrm{S}$ is a vector of length k, is a $w\times k$ matrix, and $\mathbf{V}_\mathrm{S}$ is a $k\times k$ matrix. The definition of $\mathbf{m_0}$, $\mathbf{M}_\mathrm{W}$, and $\mathbf{V}_\mathrm{S}$ were defined in Chapter 3 for balanced designs. The proposed elements, $\mathbf{u}_\mathrm{W}$, and $\mathbf{u}_\mathrm{S}$, are specific to unbalanced designs and are defined in this chapter.

Inserting Equations 3.8 and 4.1 into Equation 3.5, we obtain

$$\begin{bmatrix} n_1\mathbf{1} & \mathbf{11}'\mathbf{W}_{\mathrm{D}_1} & \mathbf{11}'\mathbf{W}_{\mathrm{Q}_1} & \mathbf{11}'\mathbf{S}_{\mathrm{D}_1} & \mathbf{11}'\mathbf{S}_{\mathrm{Q}_1} \\ n_2\mathbf{1} & \mathbf{11}'\mathbf{W}_{\mathrm{D}_2} & \mathbf{11}'\mathbf{W}_{\mathrm{Q}_2} & \mathbf{11}'\mathbf{S}_{\mathrm{D}_2} & \mathbf{11}'\mathbf{S}_{\mathrm{Q}_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ n_m\mathbf{1} & \mathbf{11}'\mathbf{W}_{\mathrm{D}_m} & \mathbf{11}'\mathbf{W}_{\mathrm{Q}_m} & \mathbf{11}'\mathbf{S}_{\mathrm{D}_m} & \mathbf{11}'\mathbf{S}_{\mathrm{Q}_m} \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{1} & \mathbf{W}_{\mathrm{D}_1} & \mathbf{W}_{\mathrm{Q}_1} & \mathbf{S}_{\mathrm{D}_1} & \mathbf{S}_{\mathrm{Q}_1} \\ \mathbf{1} & \mathbf{W}_{\mathrm{D}_2} & \mathbf{W}_{\mathrm{Q}_2} & \mathbf{S}_{\mathrm{D}_2} & \mathbf{S}_{\mathrm{Q}_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{1} & \mathbf{W}_{\mathrm{D}_m} & \mathbf{W}_{\mathrm{Q}_m} & \mathbf{S}_{\mathrm{D}_m} & \mathbf{S}_{\mathrm{Q}_m} \end{bmatrix} \begin{bmatrix} n_0 & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{m_0}' \\ \mathbf{0} & n_\mathrm{W}\mathbf{I}_\mathrm{D} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{u}_\mathrm{W} & \mathbf{0} & n_\mathrm{W}\mathbf{I}_\mathrm{Q} & \mathbf{0} & \mathbf{M}_\mathrm{W} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{u}_\mathrm{S} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V}_\mathrm{S} \end{bmatrix}, \tag{4.2}$$

where the vectors of ones have length n_i in the i^{th} whole-plot (explicit subscripts on these vectors are omitted for clarity).

4.2 Balanced Design Summary

The elements of $\mathbf{K}$, referring to Equation 4.1, derived in Chapter 3 are summarized in Table 4.1. For the balanced case, $n_0 = n_\mathrm{W} = n$ and is the number of subplot runs per whole-plot. We define $\mathbf{M}_\mathrm{W}$ to be of the form $\phi\mathbf{1}_w\mathbf{1}_k'$. Note that the values for the VKM construction method apply to cases where the number of subplot factors is a power of 2, and the values for the MWP method apply to $w = 1$, since these are the unbalanced designs cases we consider.

Table 4.1: Elements of $\mathbf{K}$ for Balanced Case.

	VKM-CCD	VKM-BBD	MWP-BBD
$\mathbf{V}_{\mathrm{S}}$	$\frac{n}{k}\mathbf{1}_k\mathbf{1}'_k$	$2(k-1)\mathbf{1}_k\mathbf{1}'_k$	$\mathbf{0}$
$\mathbf{m_0}$	$\mathbf{0}$	$\mathbf{0}$	$4(k-1)\mathbf{1}_k$
ϕ	0	0	$6-4k$

4.3 Unbalanced VKM Method CCD

To construct an unbalanced design, we first construct the balanced version according to the prescribed method described in Chapter 3. Then, we reduce the size of the whole-plot containing all factors at their center level. We denote the size of this whole-plot as n_c in the derivation, and choose $n_c = 2$ in the examples presented.

Consider a split-plot design with two hard-to-change factors, denoted by z_1 and z_2, and two easy-to-change factors, denoted by x_1 and x_2. An unbalanced central composite design (CCD) with $\alpha = \beta = 2$ in coded units is provided in Table 4.2 that enables equivalent estimation of a complete second-order model. A balanced version of this design would be identical except that whole-plot number 10 would contain 4 subplot runs.

The small reduction in design size may not appear to be significant. However, we suggest augmenting the base design by replication whole-plot number 10 to enable a pure-error based estimate of whole-plot error variance and to increase the degrees of freedom for a pure-error estimate of subplot variance. If we desire df_w degrees of freedom for this whole-plot error estimate, it would require $df_w + 1$ replicates of whole-plot 10. Therefore, we see that the reduction in design size is not merely 2 subplot runs, rather it is $2(df_w + 1)$. Moreover, the smaller size of these replicated whole-plots should be more appealing to a practitioner who is contemplating the cost-benefit trade-off of performing replicated whole-plots.

For an unbalanced VKM method CCD, when k is a power of 2, we define $\mathbf{K}$ as

$$\mathbf{K} = \begin{bmatrix} n_0 & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & n_{\mathrm{W}}\mathbf{I}_{\mathrm{D}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{u}_{\mathrm{W}} & \mathbf{0} & n_{\mathrm{W}}\mathbf{I}_{\mathrm{Q}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{u}_{\mathrm{S}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V}_{\mathrm{S}} \end{bmatrix}. \tag{4.3}$$

Based on this form of $\mathbf{K}$ and Equation 4.2, the conditions that must be satisfied are

$$\begin{aligned} \mathbf{1}_{n_i}\mathbf{1}'_{n_i}\mathbf{S}_{\mathrm{Q}_i} &= \mathbf{S}_{\mathrm{Q}_i}\mathbf{V}_{\mathrm{S}} \\ n_i\mathbf{1}_{n_i} &= n_0\mathbf{1}_{n_i} + \mathbf{W}_{\mathrm{Q}_i}\mathbf{u}_{\mathrm{W}} + \mathbf{S}_{\mathrm{Q}_i}\mathbf{u}_{\mathrm{S}} \qquad \forall\ i. \end{aligned} \tag{4.4}$$

Table 4.2: Unbalanced VKM Method - Central Composite Design for D$(2, 2)$.

Whole-Plot	z_1	z_2	x_1	x_2	Whole-Plot	z_1	z_2	x_1	x_2
1	-1	-1	-1	-1	6	2	0	0	0
	-1	-1	1	-1		2	0	0	0
	-1	-1	-1	1		2	0	0	0
	-1	-1	1	1		2	0	0	0
2	1	-1	-1	-1	7	0	-2	0	0
	1	-1	1	-1		0	-2	0	0
	1	-1	-1	1		0	-2	0	0
	1	-1	1	1		0	-2	0	0
3	-1	1	-1	-1	8	0	2	0	0
	-1	1	1	-1		0	2	0	0
	-1	1	-1	1		0	2	0	0
	-1	1	1	1		0	2	0	0
4	1	1	-1	-1	9	0	0	-2	0
	1	1	1	-1		0	0	2	0
	1	1	-1	1		0	0	0	-2
	1	1	1	1		0	0	0	2
5	-2	0	0	0	10	0	0	0	0
	-2	0	0	0		0	0	0	0
	-2	0	0	0					
	-2	0	0	0					

The first condition involves the column sums of the subplot pure quadratic terms. The second condition expresses a relationship between each row of the whole-plot model matrix and the number of subplot runs in that whole-plot, n_i. We denote the j^{th} row of the model matrix for the whole-plot and subplot pure quadratic terms within the i^{th} whole-plot as $\mathbf{W}_{\mathrm{Q}_{ij}}$ and $\mathbf{S}_{\mathrm{Q}_{ij}}$, respectively. We define the general form of $\mathbf{u}_{\mathrm{W}}$ to be $\lambda\mathbf{1}_w$ and $\mathbf{u}_{\mathrm{S}}$ to be $\rho\mathbf{1}_k$, where λ and ρ are suitable constants. Therefore, we express the unbalanced condition as

$$n_i = n_0 + \lambda\mathbf{W}_{\mathrm{Q}_{ij}}\mathbf{1}_w + \rho\mathbf{S}_{\mathrm{Q}_{ij}}\mathbf{1}_k \quad \forall\, i, j. \tag{4.5}$$

Rearranging Equation 4.5 provides

$$n_i - n_0 = \lambda\mathbf{W}_{\mathrm{Q}_{ij}}\mathbf{1}_w + \rho\mathbf{S}_{\mathrm{Q}_{ij}}\mathbf{1}_k \quad \forall\, i, j\ . \tag{4.6}$$

To find the expression for n_0, λ, and ρ we consider the forms of $\mathbf{W}_{\mathrm{Q}_{ij}}$ and $\mathbf{S}_{\mathrm{Q}_{ij}}$ for each type of whole-plot found in a CCD constructed according to the VKM method, summarized in Table 4.3. Note that for the whole-plot and subplot axial runs the position of α^2 and β^2 will change depending on the factor involved, however there is always a single element in each row vector.

Table 4.3: VKM Method CCD Cases of Whole-Plot Type.

Case	WP Type	$\mathbf{W}_{\mathrm{Q}_{ij}}$	$\mathbf{S}_{\mathrm{Q}_{ij}}$	n_i
c	all centers	$\begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}_w$	$\begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}_k$	n_c
1	WP centers, SP axials	$\begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}_w$	$\begin{bmatrix} \alpha^2 & 0 & \cdots & 0 \end{bmatrix}_k$	$n_1 = 2k$
2	WP axials, SP centers	$\begin{bmatrix} \beta^2 & 0 & \cdots & 0 \end{bmatrix}_w$	$\begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}_k$	$n_2 = n_f$
3	factorial	$\begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}_w$	$\begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}_k$	$n_3 = n_f$

From case (c) and Equation 4.6, we see that $n_0 = n_c$. Based on the structure of the row vectors for the other cases, we can express Equation 4.6 as

$$\begin{aligned} n_1 - n_c &= \rho\alpha^2 \\ n_2 - n_c &= \lambda\beta^2 \\ n_3 - n_c &= \lambda w + \rho k. \end{aligned}$$

Since $n_1 = 2k$ and $n_2 = n_3 = n_f$ in a CCD, we find

$$\rho = \frac{2k - n_c}{\alpha^2}$$

$$\lambda = \frac{n_f - n_c}{\beta^2}.$$

In addition from case (2) and (3), we see that $\lambda\beta^2 = \lambda w + \rho k$. Solving this expression for α and β provides,

$$\alpha = \sqrt{\frac{k(2k - n_c)}{(n_f - n_c)\left(1 - \frac{w}{\beta^2}\right)}} \tag{4.7}$$

$$\beta = \sqrt{\frac{w\alpha^2(n_f - n_c)}{\alpha^2(n_f - n_c) - k(2k - n_c)}} \; . \tag{4.8}$$

If we desire a design with $\alpha = \beta$, then we obtain

$$\alpha = \beta = \sqrt{w + \frac{k(2k - n_c)}{(n_f - n_c)}} \; .$$

Performing an analysis of Equations 4.7 and 4.8, we find that our choice of α and β are constrained by

$$\alpha > \sqrt{\frac{k(2k - n_c)}{n_f - n_c}}$$
$$\beta > \sqrt{w} \; .$$

In addition from Equations 4.7 and 4.8, we see that the relationship between α and β is inversely proportional. Using these expressions, we can define $\mathbf{K}$ for an unbalanced VKM CCD when k is a power of 2.

As an example, consider the $\mathbf{K}$ matrix for the design in Table 4.2. We have $w = k = 2$, $n_c = 2$, $n_f = 4$, and $\alpha = \beta = 2$, resulting in $n_0 = 2$, $\mathbf{u}_{\mathrm{W}} = \frac{1}{2}\mathbf{1}_w$, and $\mathbf{u}_{\mathrm{S}} = \frac{1}{2}\mathbf{1}_k$. Note that $\mathbf{V}_{\mathrm{S}} = \frac{n}{k}\mathbf{1}_k\mathbf{1}_k' = 2\mathbf{1}_k\mathbf{1}_k'$, as it was for the balanced case, where n refers to the number of subplot runs in the whole-plots without all centers.

When k is not a power of 2, an unbalanced design can be generated if we eliminate one set of whole-plots containing the whole-plot factor at its axial level. The resulting design retains the capability to estimate a complete second-order model, however it does not contain all the design points found in a classical central composite design. An attractive feature of this design is that a single whole-plot can contain all of the subplot axial runs, thus making the subplot pure quadratics subplot level terms.

For example, consider a design for $\mathrm{D}(1,3)$ in Table 4.4. The form of $\mathbf{K}$ for this type of design requires a non-zero $\mathbf{M}_{\mathrm{W}}$, which is not found in a classical VKM construction. Therefore, we consider this design structure as a mixture of the MWP and VKM methods. A general development of this form of $\mathbf{K}$ is not provided, however for the design in Table 4.4, $\mathbf{K}$ is defined by $n_0 = 2$, $\mathbf{u}_{\mathrm{W}} = -6$, $\mathbf{u}_{\mathrm{S}} = 4\mathbf{1}_k$, $\mathbf{V}_{\mathrm{S}} = 2\mathbf{1}_k\mathbf{1}_k'$, and $\mathbf{M}_{\mathrm{W}} = 2\mathbf{1}_k'$.

Table 4.4: Unbalanced VKM Method - Composite Design for $\mathrm{D}(1,3)$.

Whole-Plot	z_1	x_1	x_2	x_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	-1	1	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
	1	1	-1	1
	1	-1	1	1
3	0	-1	0	0
	0	1	0	0
	0	0	-1	0
	0	0	1	0
	0	0	0	-1
	0	0	0	1
4	0	0	0	0
	0	0	0	0

Table 4.5: Summary of Unbalanced VKM Method Central Composite Designs.

Factors w	k	m	n_i	N
1	1			
1	2	6	4,2	22
1	3	4	8,6,2	24
1	4	6	8,2	42
2	1			
2	2	10	4,2	38
2	3	8	6,4,2	32
2	4	10	8,2	74
3	1			
3	2	16	4,2	62
3	3	14	6,4,2	56
3	4	16	8,2	122

A summary table of the unbalanced VKM CCD's is provided in Table 4.5. The entries in this table for $\mathrm{D}(w, 3)$ are based on removing one set of whole-plot axial points, as previously discussed, and therefore require two fewer whole-plots than other designs for the same w.

4.4 Unbalanced VKM Method BBD

Considering the balanced VKM method BBD's for $w = 1$ summarized in Table 3.6, we see that the whole-plot size becomes quite large with $k > 2$. This is a result of our construction philosophy to utilize the minimum number of whole-plots. As a result, these large whole-plots may not be practical in certain applications, thereby requiring a design with additional whole-plots of a smaller size. Nevertheless, in this section we assume that these minimum whole-plot designs are permissible and consider methods to reduce the size of the whole-plots containing replicated subplot runs.

Consider a design for the $\mathrm{D}(1, 4)$ case given in Table 4.6. To construct this design, we begin with a balanced BBD consisting of 4 whole-plots of size 24. Recall that to minimize the number of whole-plots, we have a single whole-plot with the HTC factor set at its center level. The size of this whole-plot, for $f \leq 5$, is

$$n_{max} = 4 \begin{pmatrix} k \\ 2 \end{pmatrix} = 2k(k-1).$$

To achieve balance, the subplot designs are replicated as required in the other whole-plots. In whole-plots 1 and 2, there are 3 replicates of the base subplot design and there were 24

Table 4.6: Unbalanced VKM Method - Box-Behnken Design for D(1, 4).

Whole-Plot	z_1	x_1	x_2	x_3	x_4	Whole-Plot	z_1	x_1	x_2	x_3	x_4
1	$\mathbf{-1}$	$\pm\mathbf{1}$	0	0	0	3	0	0	-1	-1	0
	$\mathbf{-1}$	0	$\pm\mathbf{1}$	0	0		0	0	-1	1	0
	$\mathbf{-1}$	0	0	$\pm\mathbf{1}$	0		0	-1	0	0	-1
	$\mathbf{-1}$	0	0	0	$\pm\mathbf{1}$		0	-1	0	0	1
	$\mathbf{-1}$	$\pm\mathbf{1}$	0	0	0		0	0	0	-1	-1
	$\mathbf{-1}$	0	$\pm\mathbf{1}$	0	0		0	0	0	1	-1
	$\mathbf{-1}$	0	0	$\pm\mathbf{1}$	0		0	0	0	-1	1
	$\mathbf{-1}$	0	0	0	$\pm\mathbf{1}$		0	0	0	1	1
	$\mathbf{-1}$	$\pm\mathbf{1}$	0	0	0		0	-1	-1	0	0
	$\mathbf{-1}$	0	$\pm\mathbf{1}$	0	0		0	1	-1	0	0
	$\mathbf{-1}$	0	0	$\pm\mathbf{1}$	0		0	-1	1	0	0
	$\mathbf{-1}$	0	0	0	$\pm\mathbf{1}$		0	1	1	0	0
2	$\mathbf{+1}$	$\pm\mathbf{1}$	0	0	0		0	0	-1	0	-1
	$\mathbf{+1}$	0	$\pm\mathbf{1}$	0	0		0	0	1	0	-1
	$\mathbf{+1}$	0	0	$\pm\mathbf{1}$	0		0	0	-1	0	1
	$\mathbf{+1}$	0	0	0	$\pm\mathbf{1}$		0	0	1	0	1
	$\mathbf{+1}$	$\pm\mathbf{1}$	0	0	0		0	-1	0	-1	0
	$\mathbf{+1}$	0	$\pm\mathbf{1}$	0	0		0	1	0	-1	0
	$\mathbf{+1}$	0	0	$\pm\mathbf{1}$	0		0	-1	0	1	0
	$\mathbf{+1}$	0	0	0	$\pm\mathbf{1}$		0	1	0	1	0
	$\mathbf{+1}$	$\pm\mathbf{1}$	0	0	0	4	0	0	0	0	0
	$\mathbf{+1}$	0	$\pm\mathbf{1}$	0	0		0	0	0	0	0
	$\mathbf{+1}$	0	0	$\pm\mathbf{1}$	0						
	$\mathbf{+1}$	0	0	0	$\pm\mathbf{1}$						

replicates of the center run in whole-plot 4. The vector notation denotes two subplot runs for each row in whole-plots 1 and 2. As with the CCD, we have reduced the size of the whole-plot containing all factors at their center level. The overall number of subplot runs is reduced by $22(df_w + 1)$, which is very significant due to the large whole-plot size. Moreover, it is apparent that the likelihood of running whole-plot replicates is improved considerably.

The general form of $\mathbf{K}$ in Equation 4.3 and the conditions expressed in Equation 4.4 hold for the BBD case. We summarize the forms of $\mathbf{W}_{\mathrm{Q}_{ij}}$ and $\mathbf{S}_{\mathrm{Q}_{ij}}$ found in each type of whole-plot in Table 4.7. For cases (1)-(3), the position of the ones will change depending on the factors involved, however the form is consistent. Note that case (3) is not present when $w = 1$.

From case (c) and Equation 4.6, we see that $n_0 = n_c$. From case (1), we find

$$n_1 - n_c = 2\rho$$

Table 4.7: VKM Method BBD Whole-Plot Types.

Case	$\mathbf{W}_{Q_{ij}}$	$\mathbf{S}_{Q_{ij}}$	n_i
c	$\begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}_w$	$\begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}_k$	n_c
1	$\begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}_w$	$\begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \end{bmatrix}_k$	$n_1 = 4\begin{pmatrix} k \\ 2 \end{pmatrix} = 2k(k-1)$
2	$\begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}_w$	$\begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}_k$	n_2
3	$\begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \end{bmatrix}_w$	$\begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}_k$	n_3

$$\rho = \frac{n_1 - n_c}{2} = \frac{2k(k-1) - n_c}{2} \ .$$

For case (2), we find

$$n_2 - n_0 = \lambda + \rho$$
$$\lambda = n_2 - n_0 - \rho \ .$$

Similarly for case (3), we find

$$\lambda = \frac{2k(k-1) - n_c}{2} \ .$$

If $n_1 = n_2 = n_3 = 2k(k-1)$, then $\lambda = \rho$.

Using these expressions, we can define $\mathbf{K}$ for an unbalanced VKM BBD when k is a power of 2. For the design in Table 4.6, we have $w = 1$, $k = 4$, $n_c = 2$, resulting in $n_0 = 2$, $\mathbf{u}_W = 11$, and $\mathbf{u}_S = (11)\mathbf{1}_k$. We note that $\mathbf{V}_S = 6\mathbf{1}_k\mathbf{1}'_k$, as it was for the balanced case.

A potentially unattractive feature of the design in Table 4.6 are the replicated subplot designs in whole-plots 1 and 2. We can reduce the size of these whole-plots by adding a non-zero $\mathbf{M}_W$ to $\mathbf{K}$. A general development of this form of $\mathbf{K}$ is not provided, however to illustrate this approach consider the design in Table 4.8. The first two whole-plots now contain 8 subplot runs each with no replication. For this design, $\mathbf{K}$ is defined by $n_0 = 2$, $\mathbf{u}_W = -5$, $\mathbf{u}_S = (11)\mathbf{1}_k$, $\mathbf{V}_S = 6\mathbf{1}_k\mathbf{1}'_k$, and $\mathbf{M}_W = -4\mathbf{1}'_k$.

The unbalanced design in Table 4.8 has achieved a significant reduction with $m = 4$ and $N = 42$ compared to Table 4.6 with $m = 4$ and $N = 70$, and the balanced version with $m = 4$ and $N = 96$. Even though the number of whole-plots has not been reduced, the reduction in design size and particularly the size of the proposed replicated whole-plots makes this an attractive alternative to the balanced version.

A summary table of the unbalanced VKM BBD's is provided in Table 4.9. For $\mathrm{D}(1,k)$, these designs require the same or fewer whole-plots than their CCD competitors in Table 4.5, however the larger whole-plot size may not be permissible in some applications. In contrast, the $\mathrm{D}(3,k)$ designs require additional whole-plots over the unbalanced CCD's.

Table 4.8: Unbalanced VKM Method, Minimum Size - Box-Behnken Design for D(1, 4).

Whole-Plot	z_1	x_1	x_2	x_3	x_4
1	$\mathbf{-1}$	$\mathbf{\pm 1}$	0	0	0
	$\mathbf{-1}$	0	$\mathbf{\pm 1}$	0	0
	$\mathbf{-1}$	0	0	$\mathbf{\pm 1}$	0
	$\mathbf{-1}$	0	0	0	$\mathbf{\pm 1}$
2	$\mathbf{+1}$	$\mathbf{\pm 1}$	0	0	0
	$\mathbf{+1}$	0	$\mathbf{\pm 1}$	0	0
	$\mathbf{+1}$	0	0	$\mathbf{\pm 1}$	0
	$\mathbf{+1}$	0	0	0	$\mathbf{\pm 1}$
3	0	0	-1	-1	0
	0	0	-1	1	0
	0	-1	0	0	-1
	0	-1	0	0	1
	0	0	0	-1	-1
	0	0	0	1	-1
	0	0	0	-1	1
	0	0	0	1	1
	0	-1	-1	0	0
	0	1	-1	0	0
	0	-1	1	0	0
	0	1	1	0	0
	0	0	-1	0	-1
	0	0	1	0	-1
	0	0	-1	0	1
	0	0	1	0	1
	0	-1	0	-1	0
	0	1	0	-1	0
	0	-1	0	1	0
	0	1	0	1	0
4	0	0	0	0	0
	0	0	0	0	0

Table 4.9: Summary of Unbalanced VKM Method Box-Behnken Designs.

Factors				
w	k	m	n_i	N
1	1			
1	2	4	4,2	14
1	3	4	12,2	38
1	4	4	24,2	74
2	1			
2	2	10	4,2	38
2	3	10	12,2	110
2	4	10	8,2	74
3	1			
3	2	20	4,2	78
3	3	19	4,2	74
3	4	21	4,2	82

4.5 Unbalanced MWP Method BBD

In this section, we propose a class of unbalanced equivalent estimation designs that require a minimum number of whole-plots for the $w = 1$ case. We begin with a balanced design constructed using the MWP method. This method does not require the subplot centers to be contained in a whole-plot by themselves, rather they can be included in other whole-plots and added one-by-one without disturbing the equivalence property of the design.

Consider a design for $\text{D}(1, 3)$ in Table 4.11. This design was derived from a balanced design with $m = 3$, $n = 13$, and $N = 39$. The unbalanced version still has 3 whole-plots, however the overall size has been reduced to $N = 26$, by removing the replicated design points in the first two whole-plots. This is similar to the design in Table 4.8 except we have now been able to include the subplot centers in whole-plot number 3, instead of requiring an additional whole-plot.

Recall that a completely randomized version of a BBD with $f = 4$ contains $N = 24+n_c$ design points. Therefore, the equivalent estimation design in Table 4.11 achieves the minimum number of whole-plots to support a second-order model and requires no additional subplot runs over the completely randomized version. In fact, this is a general result for MWP method BBD's for $\text{D}(1, k)$, $k \in \{2, 3, 4\}$.

The form of $\mathbf{K}$ for a MWP Method BBD is

$$\mathbf{K} = \begin{bmatrix} n_0 & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{m}_\mathbf{0}' \\ \mathbf{0} & n_\mathrm{W}\mathbf{I}_\mathrm{D} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{u}_\mathrm{W} & \mathbf{0} & n_\mathrm{W}\mathbf{I}_\mathrm{Q} & \mathbf{0} & \mathbf{M}_\mathrm{W} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Based on this form of $\mathbf{K}$ and Equation 4.2, the conditions that must be satisfied are

$$\begin{aligned} \mathbf{1}_{n_i}\mathbf{1}_{n_i}'\mathbf{S}_{\mathrm{Q}_i} &= \mathbf{1}_{n_i}\mathbf{m}_\mathbf{0}' + \mathbf{W}_{\mathrm{Q}_i}\mathbf{M}_\mathrm{W} \\ n_i\mathbf{1}_n &= n_0\mathbf{1}_{n_i} + \mathbf{W}_{\mathrm{Q}_i}\mathbf{u}_\mathrm{W} \qquad \forall\, i. \end{aligned} \tag{4.9}$$

The first condition is unchanged from the balanced case, and the second is similar to the VKM unbalanced case, except that there are no conditions on the subplot pure quadratic terms. Recall the form of $\mathbf{u}_\mathrm{W}$ is $\lambda\mathbf{1}_w$, and with $w = 1$ we have $\mathbf{u}_\mathrm{W}$ as a scalar equal to λ. We rewrite Equation 4.9 for $w = 1$ as

$$n_i - n_0 = \lambda\mathbf{W}_{\mathrm{Q}_{ij}} \quad \forall\, i, j\ ,$$

where $\mathbf{W}_{\mathrm{Q}_i}$ is a $n \times 1$ vector for $w = 1$ of the form $z_1^2\mathbf{1}_{n_i}$. Therefore, we only need to consider the two possible values of z_i^2 as 0 and 1. Table 4.10 provides the two cases of whole-plot types. For $z_1^2 = 0$, we see that $n_0 = n_1 = 2k(k-1) + n_c$. For $z_1^2 = 1$, we find

$$\begin{aligned} \lambda &= n_2 - n_0 \\ \lambda &= 2k - [2k(k-1) + n_c] \\ \lambda &= -\left(2k^2 - 4k + n_c\right)\ . \end{aligned}$$

Table 4.10: MWP Method BBD Whole-Plot Types.

Case	$\mathbf{W}_{\mathrm{Q}_{ij}}$	n_i
1	0	$n_1 = 2k(k-1) + n_c$
2	1	$n_2 = 2k$

Using these expressions, we can define $\mathbf{K}$ for an unbalanced MWP BBD when $w = 1$. For the design in Table 4.11, $\mathbf{K}$ is defined by $n_0 = 14$, $\mathbf{u}_\mathrm{W} = -8$, $\mathbf{m}_\mathbf{0}' = 8\mathbf{1}_k$, and $\mathbf{M}_\mathrm{W} = -6\mathbf{1}_k'$.

A summary table of the unbalanced MWP method BBD's is provided in Table 4.12. To achieve the minimum number of whole-plots, these designs require a large number of subplot runs per whole-plot, similar to the VKM method BBD's.

Table 4.11: MWP Method - Box-Behnken Design for D(1, 3).

Whole-Plot	z_1	x_1	x_2	x_3
1	-1	-1	0	0
	-1	1	0	0
	-1	0	-1	0
	-1	0	1	0
	-1	0	0	-1
	-1	0	0	1
2	1	-1	0	0
	1	1	0	0
	1	0	-1	0
	1	0	1	0
	1	0	0	-1
	1	0	0	1
3	0	0	-1	-1
	0	0	-1	1
	0	0	1	-1
	0	0	1	1
	0	-1	-1	0
	0	-1	1	0
	0	1	-1	0
	0	1	1	0
	0	-1	0	-1
	0	-1	0	1
	0	1	0	-1
	0	1	0	1
	0	0	0	0
	0	0	0	0

Table 4.12: Summary of Unbalanced MWP Method Box-Behnken Designs.

Factors				
w	k	m	n_i	N
1	1			
1	2	3	6,4	14
1	3	3	14,6	26
1	4	3	26,8	42

4.6 Minimal Point Designs

Typically, the cost of a split-plot experiment is dominated by the number of whole-plots, which dictates the number of times the HTC factors are reset. However, in some experimental situations it is desirable to simultaneously minimize the overall design size as well. In this case, saturated or near-saturated designs could be considered. As in the CRD case, we recommend the use of the CCD or BBD, unless they are cost prohibitive.

Goos (2002) highlights that all saturated designs are equivalent. For a saturated design, we have N design points to estimate a model with p parameters where $N = p$. Therefore, the model matrix, $\mathbf{X}$ is a square full rank matrix. The OLS estimate is

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{OLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{X}'^{-1}(\mathbf{X}')^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{X}^{-1}\mathbf{y}.\end{aligned}$$

The GLS estimate is

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{GLS} &= (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} \\ &= \mathbf{X}^{-1}\boldsymbol{\Sigma}(\mathbf{X}')^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} \\ &= \mathbf{X}^{-1}\mathbf{y},\end{aligned}$$

which is the same as the OLS estimate. Even though all saturated design are equivalent, some are more suitable for augmentation to provide pure-error estimates of the variance components and lack-of-fit degrees of freedom while maintaining the equivalence property.

The parent designs in this section are common families of minimal point RSM designs for the completely randomized case developed by Notz (1982), Hoke (1974), Box and Draper (1974), and Roquemore (1976). A summary of the features of these designs in the completely randomized context is provided in Myers and Montgomery (2002).

In the previous sections, design construction was based on an analytical derivation of the form of the $\mathbf{K}$ matrix to satisfy the equivalence condition for a particular design structure. As previously mentioned, Equation 3.3 can also be used to numerically test the equivalence of any design without knowledge of $\mathbf{K}$.

If the design is found to be equivalent, then the form of $\mathbf{K}$ can be computed directly from Equation 3.4. While it is not necessary to compute $\mathbf{K}$ to verify equivalence, it does provide insight into augmentation strategies that maintain the equivalence property. For example, we can determine if the design can be augmented by adding subplot centers one-by-one or a whole-plot containing all factors at their center level. This is especially important since these designs do not include subplot replicates.

An interesting feature of these designs are non-orthogonal subplot designs. As previously stated, orthogonality is not a necessary condition, although it is desirable. These examples demonstrate the flexibility in generating equivalent estimation designs. In addition, to obtain

Table 4.13: Notz Saturated Design for D(1, 3).

Whole-Plot	z_1	x_1	x_2	x_3
1	-1	1	1	-1
	-1	1	-1	1
	-1	-1	1	1
	-1	1	1	1
	-1	-1	-1	-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
	1	0	0	0
3	0	1	0	0
	0	0	1	0
	0	0	0	1

minimal point split-plot designs, we have not constrained $n_{min} = 2$, allowing whole-plots of size one that are typically undesirable for HTC factors.

Consider a saturated Notz design for D(1, 3) in Table 4.13, where equivalent estimation was verified by Equation 3.3. This design can be augmented to provide 1 degree of freedom for lack-of-fit, by including a subplot center run to whole-plot number 3, or by adding a subplot run to whole-plot number 2 with all of the subplot factors at their low level, $\begin{bmatrix} -1 & -1 & -1 \end{bmatrix}$. Either of these augmentations will maintain the equivalence property of the design. In other cases, when the augmentation is not intuitive, a $\mathcal{D}$-optimal augmentation strategy was applied to add the minimum number of design points while maintaining the equivalence property. Table 4.14 contains a summary of Notz split-plot designs.

Consider a Hoke saturated design for D(1, 2) in Table 4.15, based on a Hoke D2 parent design. While the Hoke D1 through D3 designs are saturated, the D2 was chosen due to high $\mathcal{D}$-efficiency in the CRD context. A near-saturated design for the same case is illustrated in Table 4.16, based on a Hoke D6 design. Table 4.17 contains a summary of Hoke split-plot designs.

A saturated Box and Draper design for the D(2, 3) case is shown in Table 4.18. While this design offers a small design for studying 5 factors, two of which are HTC, the design has the unattractive feature of 5 whole-plots of size one. In addition, the unusual factor levels may be difficult to set in some experimental applications. Table 4.19 contains a summary of Box and Draper split-plot designs.

Table 4.14: Summary of Equivalent Estimation Notz Designs.

Factors		**Saturated**			**Augmented**			
w	k	m	n_i	N	m	n_i	N	df_{LoF}
1	1	3	3,2,1	6	3	3,2	7	1
1	2	3	4,2	10	3	5,4,2	11	1
1	3	3	7,5,3	15	3	7,5,4	16	1
1	4	3	11,8,2	21	3	11,9,2	22	1
2	1	7	2,1	10	7	2,1	11	1
2	2	7	3,2,1	15	7	4,3,2,1	16	1
2	3	7	6,4,1	21	7	6,5,4,1	22	1
2	4	7	9,5,2,1	28	7	9,5,2,1	30	2
3	1	12	2,1	15	12	2,1	20	5
3	2	12	2,1	21	12	3,2,1	22	1
3	3	12	4,3,2,1	28				
3	4							

Table 4.15: Hoke Saturated Design for D(1, 2).

Whole-Plot	z_1	x_1	x_2
1	-1	-1	-1
	-1	1	1
	-1	1	-1
	-1	-1	1
	-1	0	0
2	1	1	-1
	1	-1	1
	1	-1	-1
3	0	-1	0
	0	0	-1

Table 4.16: Hoke Near-Saturated Design for D(1, 2).

Whole-Plot	z_1	x_1	x_2
1	-1	-1	-1
	-1	1	1
	-1	1	-1
	-1	-1	1
	-1	0	0
2	1	1	-1
	1	-1	1
	1	-1	-1
	1	1	0
	1	0	1
3	0	-1	0
	0	0	-1
	0	1	1

Table 4.17: Summary of Equivalent Estimation Hoke Designs.

Factors		**Saturated (D2)**			**Augmented (D6)**			
w	k	m	n_i	N	m	n_i	N	df_{LoF}
1	1							
1	2	3	5,3,2	10	3	5,3	13	3
1	3	3	6,3	15	3	9,6,4	19	4
1	4	3	9,8,4	21	3	12,9,5	26	5
2	1	7	2,1	10				
2	2	7	3,2,1	15				
2	3	7	4,3,1	21				
2	4							
3	1	12	2,1	15				
3	2	12	2,1	21				
3	3							
3	4							

Table 4.18: Box and Draper Saturated Design for D$(2, 3)$.

Whole-Plot	z_1	z_2	x_1	x_2	x_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	0.5335	0.5335
	-1	-1	0.5335	-1	0.5335
	-1	-1	0.5335	0.5335	-1
2	-1	0.5335	-1	-1	0.5335
	-1	0.5335	-1	0.5335	-1
	-1	0.5335	0.5335	-1	-1
3	-1	1	-1	-1	-1
	-0.8108	1	1	1	1
	0.5335	-1	-1	-1	0.5335
4	0.5335	-1	-1	0.5335	-1
5	0.5335	-1	0.5335	-1	-1
6	0.5335	0.5335	-1	-1	-1
7	1	-1	-1	-1	-1
8	1	-0.8108	1	1	1
9	1	1	-0.8108	1	1
	1	1	1	-0.8108	1
	1	1	1	1	-0.8108

Table 4.19: Summary of Equivalent Estimation Box and Draper Designs.

Factors		**Saturated**		
w	k	m	n_i	N
1	1			
1	2	4	4,3,2,1	10
1	3	4	7,4,3,1	15
1	4	4	11,5,4,1	21
2	1	9	2,1	10
2	2	9	4,2,1	15
2	3	9	7,3,1	21
2	4			
3	1	14	2,1	15
3	2	14	4,2,1	21
3	3			
3	4			

Finally, consider a hybrid design for the D$(2, 4)$ case in Table 4.20 built from a D628A. Similar to the Box and Draper design, it features a small size for studying 6 factors, however it also contains whole-plots of size one. In both of these cases, if the HTC factors are truly hard-to-change, then we might consider replicating the subplot designs in the whole-plots of size one, thereby providing degrees of freedom for a pure-error estimate of subplot variance. The resulting designs are still saturated, in terms of zero lack-of-fit degrees of freedom (df_{LoF}), and retain the equivalence property. Based on the construction of a hybrid design, the unusual factor levels are isolated to x_4. Table 4.21 contains a summary of hybrid split-plot designs.

The ability to numerically verify that a design provides equivalent estimates using Equation 3.3 has proven useful in exploring other families of RSM designs that do not provide a consistent analytical form of $\mathbf{K}$, for example the small composite design.

Table 4.20: Hybrid Saturated Design for $D(2,4)$.

Whole-Plot	z_1	z_2	x_1	x_2	x_3	x_4
1	-2	0	0	0	0	-1.15
2	-1	-1	-1	-1	-1	0.58
	-1	-1	1	1	-1	0.58
	-1	-1	1	-1	1	0.58
	-1	-1	-1	1	1	0.58
3	-1	1	1	-1	-1	0.58
	-1	1	-1	1	-1	0.58
	-1	1	-1	-1	1	0.58
	-1	1	1	1	1	0.58
4	0	-2	0	0	0	-1.15
5	0	0	0	0	0	2.31
	0	0	-2	0	0	-1.15
	0	0	2	0	0	-1.15
	0	0	0	-2	0	-1.15
	0	0	0	2	0	-1.15
	0	0	0	0	-2	-1.15
	0	0	0	0	2	-1.15
	0	0	0	0	0	0
6	0	2	0	0	0	-1.15
7	1	-1	1	-1	-1	0.58
	1	-1	-1	1	-1	0.58
	1	-1	-1	-1	1	0.58
	1	-1	1	1	1	0.58
8	1	1	-1	-1	-1	0.58
	1	1	1	1	-1	0.58
	1	1	1	-1	1	0.58
	1	1	-1	1	1	0.58
9	2	0	0	0	0	-1.15

Table 4.21: Summary of Equivalent Estimation Hybrid Designs.

Factors					
w	k	m	n_i	N	df_{LoF}
1	1				
1	2				
1	3	5	6,4,1	16	1
1	4				
2	1				
2	2	9	4,2,1	16	1
2	3				
2	4	9	8,4,1	28	0
3	1				
3	2				
3	3	15	6,2,1	28	0
3	4	15	8,4,1	46	10

Chapter 5

Robustness of OLS Estimates to Near-Equivalence

In the previous chapters, we proposed several classes of equivalent designs. Even with this broad catalog of designs, there are cases where an equivalent design is not known to exist or the available equivalent designs are not suitable. Under these circumstances, we might entertain designs that are logical extensions from classical completely randomized designs, but do not satisfy the conditions for equivalence. We refer to these designs as *near-equivalent* and expect them to occur frequently in practice. In addition, other types of near-equivalent designs may occur caused by missing subplot runs and set-point errors, however we leave these other cases for future research. In this chapter, we consider the following examples of near-equivalent designs.

- VKM construction designs that are unbalanced; not satisfying the conditions derived for equivalent unbalanced designs. For example, a central composite design for D(w,3) with all of the subplot axial subplot runs contained in a single whole-plot. An example design with one whole-plot of size 6 and the remainder of size 4 design is shown for the $D(1, 3)$ case in Table 5.1. Recall from Section 3.4.1 and 4.3, that this construction is not permissible to achieve equivalence. In Table 5.1, the subplot pure quadratic terms remain at the subplot level, thereby increasing their precision. While we could reduce the number of whole-plots in this design by choosing $n_i = 8$, we use $n_i = 4$ to avoid excessive degrees of freedom for estimating subplot error variance.

- A VKM construction near-equivalent Box-Behnken design. The equivalent balanced version discussed in Section 3.4.1 for D(w,3) is large in terms of subplot runs due to its pseudo-crossed construction. A balanced near-equivalent BBD for $D(1, 3)$ is shown in Table 5.2, constructed as a logical extension of a CRD. This design is an attractive alternative in terms of the number of subplot runs per whole-plot, however the number of whole-plots has increased significantly. This design serves as a useful simulation case; however, it may have limited practical application.

Table 5.1: D$(1, 3)$ VKM Near-Equivalent CCD, One Whole-Plot of Size 6.

Whole-Plot	z_1	x_1	x_2	x_3	Whole-Plot	z_1	x_1	x_2	x_3
1	-1	1	-1	-1	5	-1	0	0	0
	-1	-1	1	-1		-1	0	0	0
	-1	-1	-1	1		-1	0	0	0
	-1	1	1	1		-1	0	0	0
2	1	1	-1	-1	6	1	0	0	0
	1	-1	1	-1		1	0	0	0
	1	-1	-1	1		1	0	0	0
	1	1	1	1		1	0	0	0
3	-1	-1	-1	-1	7	0	-1	0	0
	-1	1	1	-1		0	1	0	0
	-1	1	-1	1		0	0	-1	0
	-1	-1	1	1		0	0	1	0
4	1	-1	-1	-1		0	0	0	-1
	1	1	1	-1		0	0	0	1
	1	1	-1	1	8	0	0	0	0
	1	-1	1	1		0	0	0	0
						0	0	0	0
						0	0	0	0

When a near-equivalent design is executed, the practitioner must choose the appropriate method of analysis. OLS estimation enjoys the advantages of ease of computation and robustness to model assumptions; however, it does not account for the correlation structure of the observations. The OLS estimates are no longer BLUE (best linear unbiased). Moreover, neither are the GLS estimates when they are based on estimated variance components. Therefore, if an equivalent estimation design is not employed or the variance components are estimated, then a best linear unbiased estimator is not available for split-plot analysis.

By simulation, we compare these estimation methods and consider if it remains advantageous to use OLS, rather than moving to a more complex procedure such as GLS. A recommendation to use OLS depends on the degree of departure from the equivalence condition, and therefore the choice is design dependent. In addition, we acknowledge that there are cases when the use of OLS is not acceptable or advantageous.

Both estimation methods provide unbiased parameter estimates, assuming that the model is correct. As a result, we can restrict our attention to a comparison of the variance of the parameter estimates. Typically, we expect the OLS estimates to have a larger variance than the GLS estimates. The question we address in the simulation study is whether the increase

Table 5.2: D(1, 3) VKM Near-Equivalent BBD.

Whole-Plot	z_1	x_1	x_2	x_3	Whole-Plot	z_1	x_1	x_2	x_3
1	-1	-1	0	0	6	1	0	-1	0
	-1	1	0	0		1	0	1	0
	-1	0	-1	0		1	0	0	-1
	-1	0	1	0		1	0	0	1
2	-1	-1	0	0	7	0	-1	0	-1
	-1	1	0	0		0	-1	0	1
	-1	0	0	-1		0	1	-1	0
	-1	0	0	1		0	1	1	0
3	-1	0	-1	0	8	0	0	-1	-1
	-1	0	1	0		0	0	-1	1
	-1	0	0	-1		0	-1	1	0
	-1	0	0	1		0	1	1	0
4	1	-1	0	0	9	0	0	-1	-1
	1	1	0	0		0	0	1	-1
	1	0	-1	0		0	-1	0	1
	1	0	1	0		0	1	0	1
5	1	-1	0	0	10	0	0	0	0
	1	1	0	0		0	0	0	0
	1	0	0	-1		0	0	0	0
	1	0	0	1		0	0	0	0

in variance is meaningful from a practical perspective and outweighs the benefits of OLS estimation as discussed in Section 3.1.

In all of the near-equivalent designs cases we have considered, pure-error estimates of the variance components are available enabling us to construct an unbiased estimate of the variance-covariance matrix as proposed by Vining, Kowalski, and Montgomery (2005). Consequently, we consider GLS estimates based on pure-error, leaving residual based estimates of the variance components, as implemented in REML, for future research.

The remainder of this chapter is organized as follows. First, we develop the machinery necessary to quantify the performance of the estimation methods and measure the departure from the equivalence property. The parameters of the simulation are presented and we discuss the results. In the final section, we provide a description of the simulation algorithm and diagnostics procedures performed to support the simulation results.

5.1 Performance of Estimation Methods

To compare the performance of OLS and GLS estimation methods, we adopt the approach of Kowalski, Cornell, and Vining (2002) where they investigated the performance of estimation methods in the context of mixture experiments with process variables executed in a split-plot mode.

Referring to the model definition in Section 1.3, we review the parameter estimates and their associated variance. In the simulation context, the variance components are known, and therefore the true, or BLUE, variance-covariance matrix of the parameter estimates is

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{BLUE}\right) = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1},$$

and the variance-covariance matrix of $\hat{\boldsymbol{\beta}}_{OLS}$ is

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{OLS}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

The estimated finite sample variance-covariance matrix provided by Kowalski, Cornell, and Vining (2002) for generalized least squares (GLS) parameter estimates is

$$\widehat{\mathrm{Var}}\left(\hat{\boldsymbol{\beta}}_{GLS}\right) = (\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1},$$

where $\widehat{\boldsymbol{\Sigma}}$ is the estimated variance-covariance matrix of the observations is given by,

$$\widehat{\boldsymbol{\Sigma}} = \hat{\sigma}_{\epsilon}^2\mathbf{I} + \hat{\sigma}_{\delta}^2\mathbf{J}_b.$$

Note that the variance-covariance matrix of the parameter estimates is a $p\times p$ matrix, making comparison cumbersome. Even if we restrict our attention to the variance of the parameter estimates, we still require p comparisons. This approach could be useful if specific parameters were of more interest than others, which is consistent with a split-plot experiment where we implicitly consider the subplot factors to be of more importance that the whole-plot factors. Furthermore, in a robust product design context, we are primarily interested in the whole-plot by subplot interactions. In the absence of specific information about the parameters, we choose to collapse the variance-covariance matrix to a single summary value that provides the generalized variance by taking the determinant, thereby providing a measure of the size of the joint confidence ellipsoids associated with the parameter estimates. We define the relative efficiencies as

$$\eta_{\,OLS/BLUE} = \frac{\left|\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{OLS}\right)\right|}{\left|\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{BLUE}\right)\right|}$$

$$\eta_{GLS/BLUE} = \frac{\left|\widehat{\text{Var}}\left(\hat{\boldsymbol{\beta}}_{GLS}\right)\right|}{\left|\text{Var}\left(\hat{\boldsymbol{\beta}}_{BLUE}\right)\right|}$$

where the subscript on η denotes the comparison performed.

5.2 Measures of Near Equivalence

To quantify the departure from equivalence, we define the following measures of near equivalence. From the necessary and sufficient condition for equivalence from Equation 3.3 we have,

$$\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{J}_b\mathbf{X} = \mathbf{J}_b\mathbf{X}.$$

The left-hand side can be expressed in terms of the HAT matrix, $\mathbf{H}$, as

$$\mathbf{H}\mathbf{J}_b\mathbf{X} = \mathbf{J}_b\mathbf{X}.$$

Rearranging this expression yields

$$(\mathbf{I} - \mathbf{H})\mathbf{J}_b\mathbf{X} = \mathbf{0}.$$

We define the difference matrix $\mathbf{D}$ as

$$\mathbf{D} = (\mathbf{I} - \mathbf{H})\mathbf{J}_b\mathbf{X}.$$

Notice that $\mathbf{D}$ is not a function of the variance components; either estimated or assumed. However, $\mathbf{D}$ is model dependent, and we consider a complete second-order model in the simulation.

We require a measure of the size, or norm, of matrix $\mathbf{D}$. Note that this matrix has dimension $N \times p$, and therefore the usual measures of size based on functions of the eigenvalues are not well defined. However, we still desire an approach consistent with this philosophy. Note that when the design is equivalent, $\mathbf{D}$ is a matrix of zeroes. The farther the design departs from the equivalence property, the more non-zero entries there are in $\mathbf{D}$.

We perform a singular value decomposition (SVD) of $\mathbf{D}$ as follows. Since $\mathbf{D}$ is an $N \times p$ matrix with $N \geq p$, then $\mathbf{D}$ can be expressed as

$$\mathbf{D} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}'$$

where $\mathbf{U}$ is an $N \times p$ matrix with orthogonal columns, $\boldsymbol{\Delta}$ is a $p \times p$ diagonal matrix with non-negative elements, and $\mathbf{V}$ is a $p \times p$ orthogonal matrix. We pre-multiply this equation by $\mathbf{D}'$ and simplify, we obtain

$$\begin{aligned}\mathbf{D}'\mathbf{D} &= (\mathbf{U}\boldsymbol{\Delta}\mathbf{V}')'\,\mathbf{U}\boldsymbol{\Delta}\mathbf{V}' \\ \mathbf{D}'\mathbf{D} &= \mathbf{V}\boldsymbol{\Delta}^2\mathbf{V}'.\end{aligned}$$

Post-multiplying by $\mathbf{V}$, we obtain,

$$\mathbf{D}'\mathbf{D}\mathbf{V} = \mathbf{V}\boldsymbol{\Delta}^2.$$

By expressing this in terms of the columns of $\mathbf{V}$ and the elements of $\boldsymbol{\Delta}$ we see that

$$\mathbf{D}'\mathbf{D}\mathbf{v}_j = \mathbf{v}_j\delta_j^2$$

where $\mathbf{v}_j$ is the j^{th} column of $\mathbf{V}$ and δ_j^2 is the j^{th} diagonal of $\boldsymbol{\Delta}$. By inspection, we see that δ_j^2 is the j^{th} eigenvalue and $\mathbf{v}_j$ is the corresponding eigenvector of $\mathbf{D}'\mathbf{D}$. By definition, δ is a singular value of $\mathbf{D}$, and $\lambda = \delta^2$ is an eigenvalue of $\mathbf{D}'\mathbf{D}$. Golub and Van Loan (1998) define the p-norm family of matrix norms that are applicable to rectangular matrices (see also Hadi (1996) and Searle (1992)). A common matrix norm is based on the largest singular value, known as the 2-norm of the matrix. This norm corresponds to the square root of the largest eigenvalue of $\mathbf{D}'\mathbf{D}$ denoted by,

$$\left\|(\mathbf{I}-\mathbf{H})\mathbf{J}_b\mathbf{X}\right\|_2.$$

This quantity is similar in nature to E-optimality since it is a function of the largest eigenvalue of $\mathbf{D}'\mathbf{D}$. Extending this concept, we entertain other measures based on functions of the non-zero eigenvalues of $\mathbf{D}'\mathbf{D}$. Similar to D-optimality and A-optimality, we consider the product and sum of the eigenvalues, respectively. Let λ_i for $i = 1, \ldots, z$ represent the z non-zero eigenvalues of $\mathbf{D}'\mathbf{D}$. We define the following measures of departure from equivalence as

$$\begin{aligned}\mathrm{E}_{\mathrm{norm}} &= \max_{i \leq z} \lambda_i = (2 - \mathrm{norm})^2 \\ \mathrm{D}_{\mathrm{norm}} &= \prod_{i=1}^{z} \lambda_i \\ \mathrm{A}_{\mathrm{norm}} &= \sum_{i=1}^{z} \lambda_i\,.\end{aligned}$$

5.3 Simulation Parameters

In this section, we provide an overview of the simulation, with a more detailed discussion of the implementation provided in Section 5.5.1. Rather than specifying a range of values for $\mathbf{D}$, we construct a near equivalent design and then compute $\mathbf{D}$. The four near-equivalent designs considered include unbalanced CCD's for $\mathrm{D}(w, 3), w \in \{1, 2, 3\}$ and a BBD for $\mathrm{D}(1, 3)$.

In the CRD case, the variance of the observations serves as a scale factor of the variance-covariance matrix of the parameter estimates and therefore does not need to be considered in the comparison of designs. However in the split-plot case, the variance-covariance matrix of the parameter estimates is a function of the two variance components. As a result, we consider a range of values for the variance components in the simulation study. It is convenient to express $\mathbf{\Sigma}$ as (see Equation 1.3)

$$\mathbf{\Sigma} = \sigma_\epsilon^2 \left[\mathbf{I} + d\mathbf{J}_b\right].$$

Recall, that d is the ratio of the whole-plot to subplot error variance, $\sigma_\delta^2/\sigma_\epsilon^2$. As d increases in magnitude, the correlation among the observations within a whole-plot becomes stronger. Conversely, as d decreases the correlation becomes weaker and the variance-covariance matrix approaches the completely randomized case, $\mathbf{\Sigma} = \sigma_\epsilon^2\mathbf{I}$. Without loss of generality, we let $\sigma_\epsilon^2 = 1$ and thus $d = \sigma_\delta^2$; the whole-plot error variance. Then, $\mathbf{\Sigma}$ becomes

$$\mathbf{\Sigma} = \mathbf{I} + d\mathbf{J}_b.$$

The values of d were chosen to be $\left\{\frac{1}{2}, 1, 2, 5, 10\right\}$; based on practical experience.

Based on a near-equivalent design model matrix, $\mathbf{X}$, the vector of model parameters, $\boldsymbol{\beta}$, and the ratio of the variance components, d, we can simulate observations, $\mathbf{y}$. However, since our variance component estimates are all based on pure-error, we only simulate the replicated observations. Moreover, the estimated variance components are model independent, requiring no assumption of the parameter values in $\boldsymbol{\beta}$.

To form our pure-error estimate of the whole-plot error variance, we include whole-plot replicates in the design. We consider 3, 4, 5, and 6 replicates. In most cases, 6 whole-plot replicates would not feasible in practice due to the HTC nature of the factors, however we consider this case in the simulation to provide additional information in studying the effect of the number of whole-plot replicates.

From each trial of the simulation, we estimate the variance-covariance matrix of the parameter estimates and compute its determinant. Then, the average determinant over all of the simulation trials is computed, which is used to compute the relative efficiency. In this study, 100,000 trials are used for each simulation case.

Table 5.3: Norms of the Difference Matrix, **D**, for Simulation Designs.

	Central Composite			Box-Behnken		
	D(1,3)	D(2,3)	D(3,3)	D(1,3)		
WP_{rep}	E_{norm}	E_{norm}	E_{norm}	E_{norm}	D_{norm}	A_{norm}
3	19.69	15.26	12.03	35.00	1225.00	70.00
4	21.33	17.07	13.84	35.00	1225.00	70.00
5	22.46	18.37	15.21	35.00	1225.00	70.00
6	23.27	19.36	16.28	35.00	1225.00	70.00

Table 5.4: Relative Efficiencies for D(1,3) VKM Near-Equivalent CCD.

	$WP_{rep} = 3$			$WP_{rep} = 4$		
d	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$
0.5	1.020	1.007	1.012	1.019	1.005	1.013
1	1.027	1.007	1.020	1.025	1.004	1.021
2	1.032	1.005	1.027	1.030	1.003	1.028
5	1.036	1.003	1.033	1.034	1.001	1.033
10	1.038	1.002	1.036	1.036	1.000	1.035

	$WP_{rep} = 5$			$WP_{rep} = 6$		
d	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$
0.5	1.017	1.004	1.013	1.017	1.003	1.013
1	1.024	1.003	1.021	1.023	1.002	1.020
2	1.029	1.002	1.027	1.027	1.001	1.026
5	1.032	1.000	1.032	1.030	1.000	1.030
10	1.033	1.000	1.033	1.032	1.000	1.032

5.4 Simulation Results

Table 5.3 summarizes the norms of the difference matrix **D**, where WP_{rep} is the number of replicated whole-plots. For the CCD's, there is one non-zero eigenvalue of $\mathbf{D}'\mathbf{D}$, therefore all three norms are equivalent. For the BBD's, we see that the norms do not change as function WP_{rep}, indicating that the replicated whole-plots do not influence the departure from equivalence.

The simulation results for each design are summarized in Tables 5.4, 5.5, 5.6, and 5.7. Figure 5.1 is a plot of the relative efficiencies contained in Table 5.4 and illustrates how the variance of the parameter estimates increases with d. Similar plots for the other VKM CCD's are consistent and are not reported. For the near-equivalent BBD, we can see the rapid increase in the variance of the parameter estimates shown Figure 5.2.

Table 5.5: Relative Efficiencies for D(2, 3) VKM Near-Equivalent CCD.

	$\mathrm{WP}_{rep} = 3$			$\mathrm{WP}_{rep} = 4$		
d	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$
0.5	1.021	1.008	1.013	1.021	1.006	1.015
1	1.029	1.007	1.021	1.028	1.005	1.023
2	1.034	1.005	1.028	1.034	1.003	1.031
5	1.038	1.003	1.035	1.038	1.001	1.037
10	1.040	1.002	1.038	1.040	1.001	1.039

	$\mathrm{WP}_{rep} = 5$			$\mathrm{WP}_{rep} = 6$		
d	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$
0.5	1.020	1.005	1.015	1.020	1.004	1.016
1	1.028	1.004	1.024	1.027	1.003	1.025
2	1.033	1.002	1.032	1.033	1.001	1.031
5	1.038	1.001	1.037	1.037	1.000	1.036
10	1.039	1.000	1.039	1.038	1.000	1.038

Table 5.6: Relative Efficiencies for D(3, 3) VKM Near-Equivalent CCD.

	$\mathrm{WP}_{rep} = 3$			$\mathrm{WP}_{rep} = 4$		
d	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$
0.5	1.020	1.007	1.012	1.020	1.006	1.014
1	1.027	1.007	1.020	1.028	1.005	1.023
2	1.032	1.005	1.027	1.034	1.003	1.030
5	1.036	1.003	1.033	1.038	1.001	1.036
10	1.038	1.002	1.036	1.039	1.001	1.039

	$\mathrm{WP}_{rep} = 5$			$\mathrm{WP}_{rep} = 6$		
d	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$
0.5	1.021	1.005	1.015	1.021	1.005	1.016
1	1.029	1.004	1.025	1.029	1.004	1.025
2	1.034	1.002	1.032	1.034	1.002	1.032
5	1.038	1.001	1.038	1.038	1.001	1.038
10	1.040	1.000	1.040	1.040	1.000	1.040

Table 5.7: Relative Efficiencies for D(1, 3) VKM Near-Equivalent BBD.

	$\text{WP}_{rep} = 3$			$\text{WP}_{rep} = 4$		
d	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$
0.5	1.596	1.249	1.278	1.596	1.202	1.327
1	2.663	1.470	1.811	2.663	1.335	1.995
2	5.781	1.830	3.159	5.781	1.498	3.860
5	22.673	2.760	8.216	22.673	1.741	13.022
10	75.806	4.144	18.293	75.806	1.965	38.579

	$\text{WP}_{rep} = 5$			$\text{WP}_{rep} = 6$		
d	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$	$\eta_{OLS/BLUE}$	$\eta_{GLS/BLUE}$	$\eta_{OLS/GLS}$
0.5	1.596	1.170	1.364	1.596	1.145	1.394
1	2.663	1.252	2.127	2.663	1.194	2.231
2	5.781	1.316	4.392	5.781	1.209	4.780
5	22.673	1.350	16.798	22.673	1.178	19.250
10	75.806	1.342	56.479	75.806	1.145	66.185

Figure 5.3 illustrates the relationship between the matrix norms and the relative efficiencies of the OLS estimates to the BLUE for the central composite designs. Note that all three norms are equivalent for these designs. Since the norms of the BBD are constant for all cases, a similar plot is not useful.

We now summarize our observations from the simulation results and draw conclusions.

- For the CCD's, we see that $\eta_{OLS/GLS} > 1$ for all cases, indicating that there is an increase in variance of the OLS estimates over the GLS estimates. However, the difference is quite small, with a maximum improvement of 4% by using GLS. In contrast, the performance increase of GLS is quite large for the BBD. This indicates that there is little loss of efficiency in using OLS for the CCD's, however OLS would not be recommended for this particular BBD. In addition, we see that the appropriateness of using OLS for these CCD's is not dependent on the number of whole-plot factors, due to a consistent construction method.

- The value of $\eta_{OLS/GLS}$ is proportional to d. Beyond $d = 5$, there is very little additional increase. Furthermore, we see that $\eta_{OLS/BLUE}$ is also proportional to d, since as $d \to 0$ the design approaches a CRD favoring the OLS estimates. Alternatively, $\eta_{GLS/BLUE}$ is inversely proportional to d.

- For the CCD's, as we increase our degrees of freedom to estimate whole-plot variance, there is no meaningful change in the relative efficiencies, indicating that $\text{WP}_{rep} = 3$ is sufficient. We caution that it is sufficient in terms of the determinant of the variance-covariance matrix, however increasing the number of whole-plot replicates provides higher power to detect whole-plot level effects.

- For the BBD, $\eta_{OLS/BLUE}$ is constant as a function of WP_{rep}, which is consistent with the norm of $\mathbf{D}$. There is improved performance of $\eta_{GLS/BLUE}$ with more whole-plot replicates, which we would expect due to the increase in the degrees of freedom available for the estimates of the variance components.

- For the CCD's, the norm of the difference matrix increases with number of whole-plot replicates, as shown in Table 5.3. From Figure 5.3, we see that the relative efficiency is a function of the variance ratio, d, and the matrix norm, E_{norm}. Unfortunately, the relationship between the norm and the relative efficiency is not monotonic. This finding does not seem to support our intuition that the larger the norm, the larger the relative efficiency. A model of this relationship is not considered useful due to the small increase in variance.

- For the BBD, there is no relationship between the norm of the difference matrix and the relative efficiency since the norms are constant for all design configurations. However, the norms are larger than those of the CCD, indicating a further departure from equivalence which is consistent with the performance of the BBD.

Overall, we conclude that for the proposed near-equivalent CCD's the performance difference in the estimation methods is negligible from a practical perspective, making either method a viable candidate for analysis. This demonstrates that for certain designs the OLS estimator is robust to the near-equivalence. Conversely, for the BBD the use of OLS would be inappropriate.

5.5 Simulation Algorithm and Diagnostics Procedures

5.5.1 Algorithm Description

In this section, we discuss the implementation details of the simulation algorithm. The computer code is provided in Appendix A.

1. We begin with a base design, without whole-plot replicates, and having the designated whole-plot to be replicated at the end of the design. The typical choice is the whole-plot containing all factors simultaneously at their center, or zero, level. We denote the number of subplot runs in this whole-plot as n_{Wr}. In addition, we specify the number of whole-plot replicates, WP_{rep}. The base design is denoted as $\text{D}_b(w, k, m, N)$.

2. The base design is augmented to provide the specified number of whole-plot replicates, with the parameters of

$$\mathrm{D}_a(w,\ k,\ m + \mathrm{WP}_{rep} - 1,\ N + n_{Wr}(\mathrm{WP}_{rep} - 1)).$$

   Note that one of the whole-plot replicates is already contained in the base design.

3. From the augmented design, we generate $\mathbf{X}$ for a complete second-order model and $\mathbf{J}_b$. The difference matrix, $\mathbf{D}$ is constructed and the norms are computed.

4. We then compute the known variance-covariance matrix of the observations as

$$\mathbf{\Sigma} = \mathbf{I} + d\mathbf{J}_b.$$

   The BLUE, or true, covariance matrix of the parameter estimates is calculated as

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{BLUE}\right) = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1},$$

   and the determinant as

$$\left|\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{BLUE}\right)\right| = \left|(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\right|.$$

   In addition, the variance-covariance matrix of the OLS estimates is computed by,

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{OLS}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

   and the determinant as

$$\left|\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{OLS}\right)\right| = \left|(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right|.$$

5. The design is inspected to identify whole-plots that contain subplot replicates, and the structure of the subplot replicates. For example, in a whole-plot with $n = 4$, there could be 4 replicates of the same subplot design point, or 2 pairs of replicates of 2 unique subplot design points. In the first case, the subplot degrees of freedom, $df_{SP} = 3$, while in the second case $df_{SP} = 2$. We adhere to this subplot replicate structure when computing the estimate of subplot error variance.

6. Observations are simulated for each whole-plot that contains subplot replicates. Note that we assume that the replicated whole-plot contains subplot replicates, which is true for all of the designs we considered. We assume that the observations have a normal distribution with a variance-covariance structure in Equation 1.3. Since our sample variance calculations are based on exact replicates, we can assume the mean of the observations to be zero, highlighting that the simulation is model independent. A vector of observations, $\mathbf{y}$, is randomly generated for each whole-plot from a multivariate

normal distribution with a mean of $\mathbf{0}$ and a within whole-plot variance-covariance matrix defined as follows

$$\mathbf{\Sigma}_{WP} = \mathbf{I}_{n_i} + d\mathbf{1}_{n_i}\mathbf{1}'_{n_i},$$

where n_i is the size of the whole-plot.

7. To obtain a pure-error estimate of the subplot variance we calculate the sum-of-squares (SS) from each set of subplot replicates and then pool them together as SS_{total}. The estimated subplot variance is computed as

$$S^2_{\text{SP}} = \frac{SS_{total}}{df_{\text{SP}_{total}}}$$

Since the $\text{E}(S^2_{\text{SP}}) = \sigma^2_\epsilon$, then $\hat{\sigma}^2_\epsilon = S^2_{\text{SP}}$ is an unbiased estimate of σ^2_ϵ with $df_{\text{SP}_{total}}$ degrees of freedom.

8. To obtain a pure-error estimate of the whole-plot variance we let $\overline{Y}_{i\cdot}$ be the sample mean of the values in the i^{th} whole-plot replicate, and $\overline{Y}_{\cdot\cdot}$ the overall sample mean of all replicated whole-plots, and define S^2_{WP} as

$$S^2_{\text{WP}} = \frac{\sum_{i=1}^{\text{WP}_{rep}} \left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2}{\text{WP}_{rep} - 1}.$$

Since the $\text{E}(S^2_{\text{WP}}) = \sigma^2_\delta + \frac{1}{n_{Wr}}\sigma^2_\epsilon$, using the method of moments we obtain an unbiased estimator of σ^2_δ as

$$\hat{\sigma}^2_\delta = S^2_{\text{WP}} - \frac{1}{n_{Wr}}\hat{\sigma}^2_\epsilon$$

with $(\text{WP}_{rep} - 1)$ degrees of freedom. Note that if $\hat{\sigma}^2_\delta$ is negative, we bias the estimator and set $\hat{\sigma}^2_\delta = 0$.

9. We form the estimated variance-covariance matrix of the observations as

$$\widehat{\mathbf{\Sigma}} = \hat{\sigma}^2_\epsilon\mathbf{I} + \hat{\sigma}^2_\delta\mathbf{J}_b.$$

Then, we compute the variance-covariance matrix of the GLS parameter estimates as

$$\widehat{\text{Var}}\left(\hat{\boldsymbol{\beta}}_{GLS}\right) = (\mathbf{X}'\widehat{\mathbf{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{X}(\mathbf{X}'\widehat{\mathbf{\Sigma}}^{-1}\mathbf{X})^{-1},$$

and compute the det $\left|\widehat{\text{Var}}\left(\hat{\boldsymbol{\beta}}_{GLS}\right)\right|$.

10. Steps 6 - 9 are replicated 100,000 times for each value of WP_{rep} and d resulting in 20 combinations of parameters for each design. A running sum of the determinant is maintained to compute the average determinant over all of the simulation trials denoted by

$$\overline{\left|\widehat{\text{Var}}\left(\hat{\boldsymbol{\beta}}_{GLS}\right)\right|}.$$

In addition, a running sum of the squared values of the determinant is maintained to compute the sample variance of the simulated trials.

11. From the average determinant we compute the relative efficiency, η, as

$$\eta_{\,OLS/BLUE} = \frac{\left|\widehat{\text{Var}}\left(\hat{\boldsymbol{\beta}}_{OLS}\right)\right|}{\left|\text{Var}\left(\hat{\boldsymbol{\beta}}_{BLUE}\right)\right|}$$

$$\eta_{\,GLS/BLUE} = \frac{\overline{\left|\widehat{\text{Var}}\left(\hat{\boldsymbol{\beta}}_{GLS}\right)\right|}}{\left|\text{Var}\left(\hat{\boldsymbol{\beta}}_{BLUE}\right)\right|}$$

$$\eta_{\,OLS/GLS} = \frac{\left|\text{Var}\left(\hat{\boldsymbol{\beta}}_{OLS}\right)\right|}{\overline{\left|\widehat{\text{Var}}\left(\hat{\boldsymbol{\beta}}_{GLS}\right)\right|}}$$

where the subscript on η denotes the comparison performed.

5.5.2 Diagnostics

Various diagnostics have been performed to check the simulation code for errors and to validate the algorithm. Several of them are discussed.

1. An equivalent design was run through the simulation to verify,

$$\left|\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{OLS}\right)\right| = \overline{\left|\widehat{\mathrm{Var}}\left(\hat{\boldsymbol{\beta}}_{GLS}\right)\right|}$$

and therefore $\eta_{OLS/BLUE} = \eta_{GLS/BLUE}$, and thus $\eta_{OLS/GLS} = 1$.

2. To investigate the distributions of the simulated values, the individual samples of the determinant of the variance-covariance matrix of the GLS parameter estimates and the estimated values of σ_ϵ^2 and σ_δ^2 were retained. The design in Table 5.1 with $\mathrm{WP}_{rep} = 3$ and $d = 1$, implying that $\sigma_\epsilon^2 = \sigma_\delta^2 = 1$, was used as the test case. The values of the determinants and standard deviation, for 100,000 trials are

$$\begin{aligned} \det{}_{BLUE} &= 9.9830 \times 10^{-15} \\ \det{}_{OLS} &= 1.0253 \times 10^{-14} \\ \mathrm{avgdet}_{GLS} &= 1.0052 \times 10^{-14} \\ \mathrm{stdev}_{GLS} &= 1.0564 \times 10^{-16}. \end{aligned}$$

Histograms of the individual observations for each variable are provided in Figure 5.4. There is a larger number of extreme values of the determinant on the right tail of the distribution. The maximum value of the GLS determinant is 1.0253×10^{-14}, which is equal to the OLS determinant. An investigation into the samples, reveals that as $\hat{\sigma}_\delta^2$ approaches zero, the value of the GLS determinant increases exponentially. This relationship is consistent with our expectations, since $\hat{\sigma}_\delta^2 = 0$ is the completely randomized case and produces $\widehat{\boldsymbol{\Sigma}} = \hat{\sigma}_\epsilon^2\mathbf{I}$. In this case, the variance of the OLS and GLS estimates are equal. The large number of values at the maximum, is due to the biasing of $\hat{\sigma}_\delta^2$ if we compute a negative estimate.

In addition, we see that $\hat{\sigma}_\epsilon^2$ appears to follow a chi-squared distribution with 15 degrees of freedom (df), and the sample mean is 1.0703. This distribution of $\hat{\sigma}_\delta^2$ is also consistent with a chi-squared distribution based on 2 df, and the sample mean is 1.0193. Both of these distributions and average values are consistent with our expectations.

3. To determine if 100,000 runs is sufficient, we use the sample standard deviation and compute the coefficient of variation of the average value of the GLS determinant. The standard deviation of the mean is

$$\frac{1.0564 \times 10^{-16}}{\sqrt{100,000}} = 3.3330 \times 10^{-19}.$$

This results in a coefficient of variation equal to

$$\frac{3.3330 \times 10^{-19}}{1.0052 \times 10^{-14}} \times 100 = 0.0033\%.$$

This value is typical of the all cases involving the VKM method CCD's, while some cases for the near-equivalent BBD are as high as 1.09%. Based on the high-precision of these results, we could consider running a smaller simulation study in future research.

Figure 5.1: Relative Efficiencies for D(1, 3) VKM Near-Equivalent CCD.

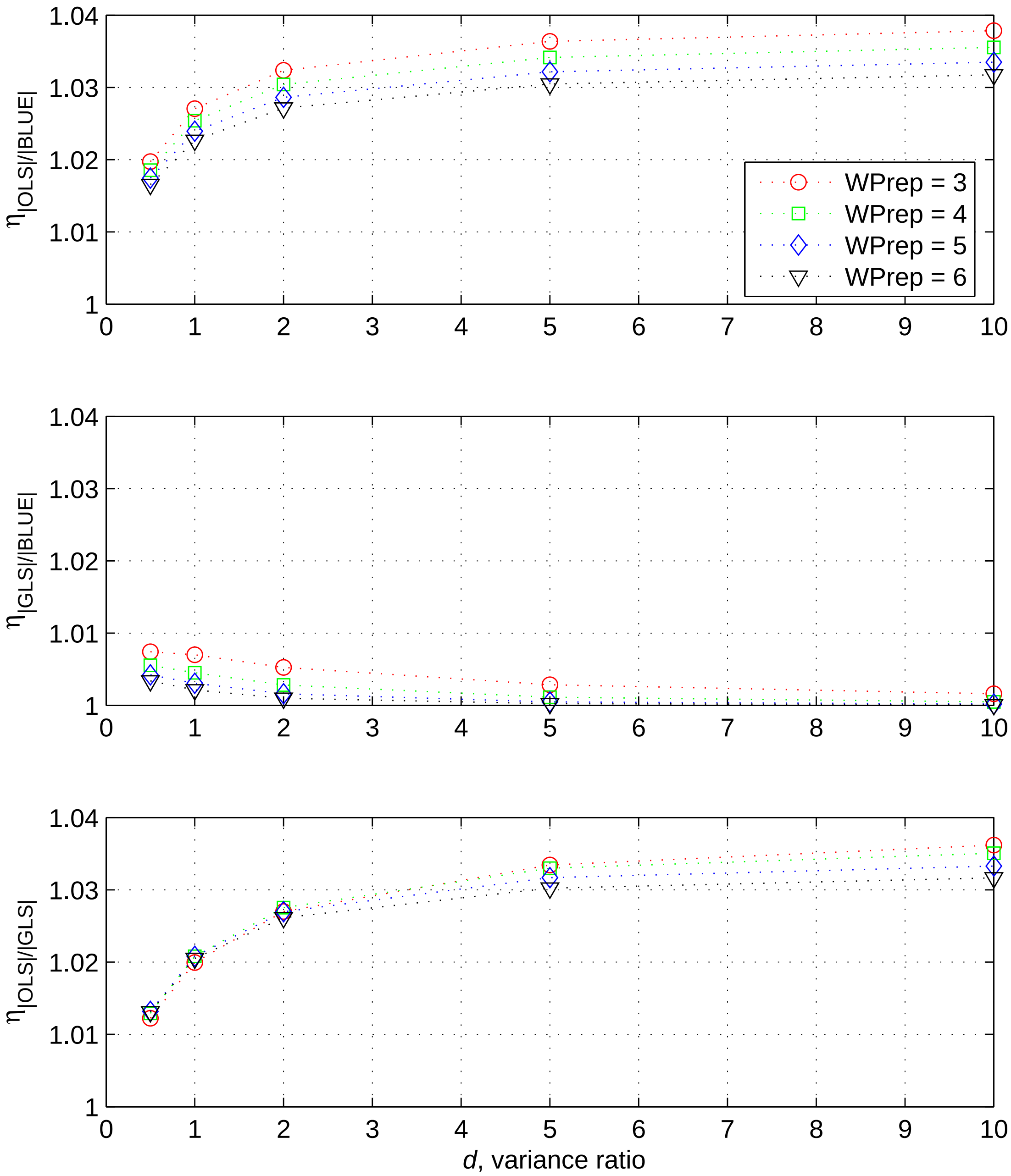


Figure 5.2: Relative Efficiencies for D(1, 3) VKM Near-Equivalent BBD.

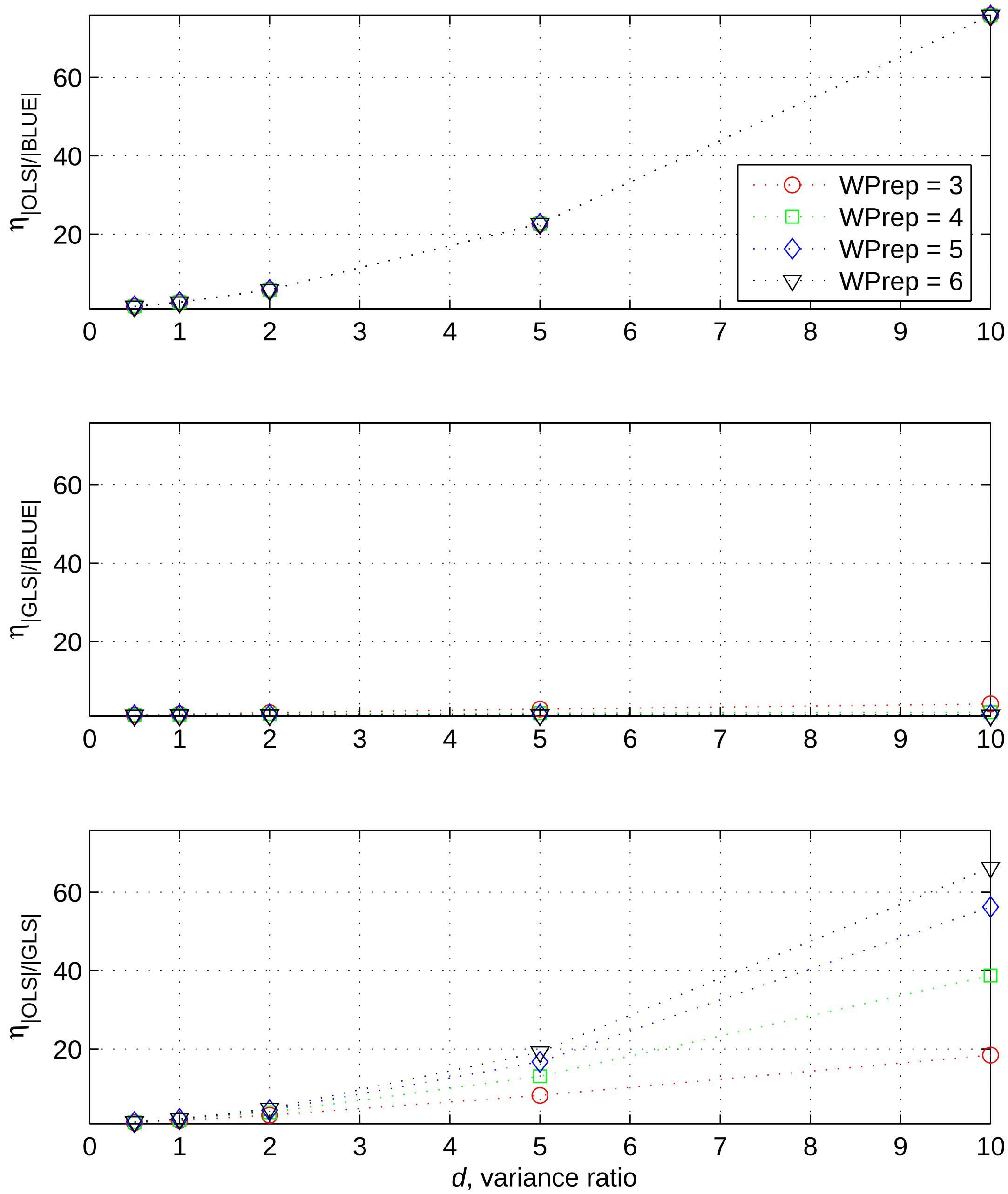


Figure 5.3: Relative Efficiency of OLS Estimator versus the E_{norm}.
(VKM Near-Equivalent Central Composite Designs)

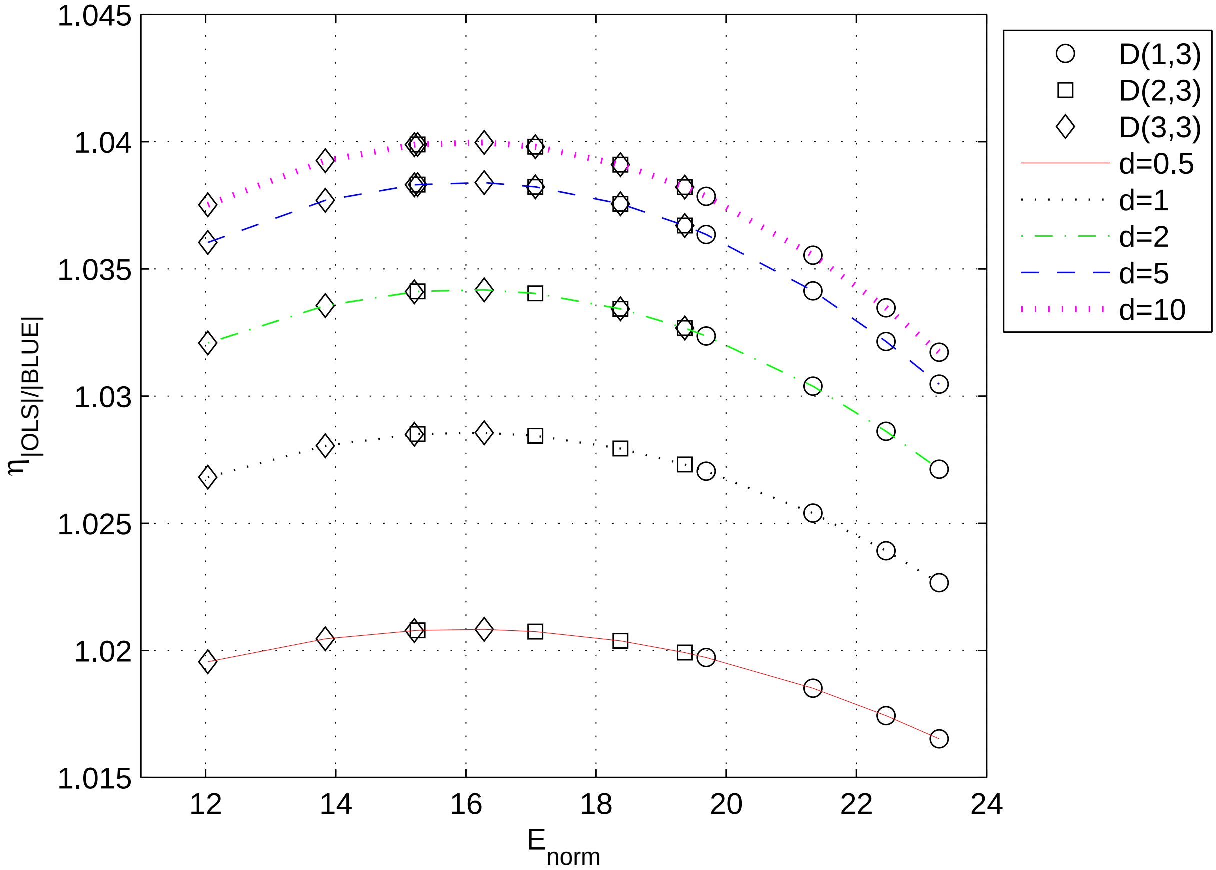
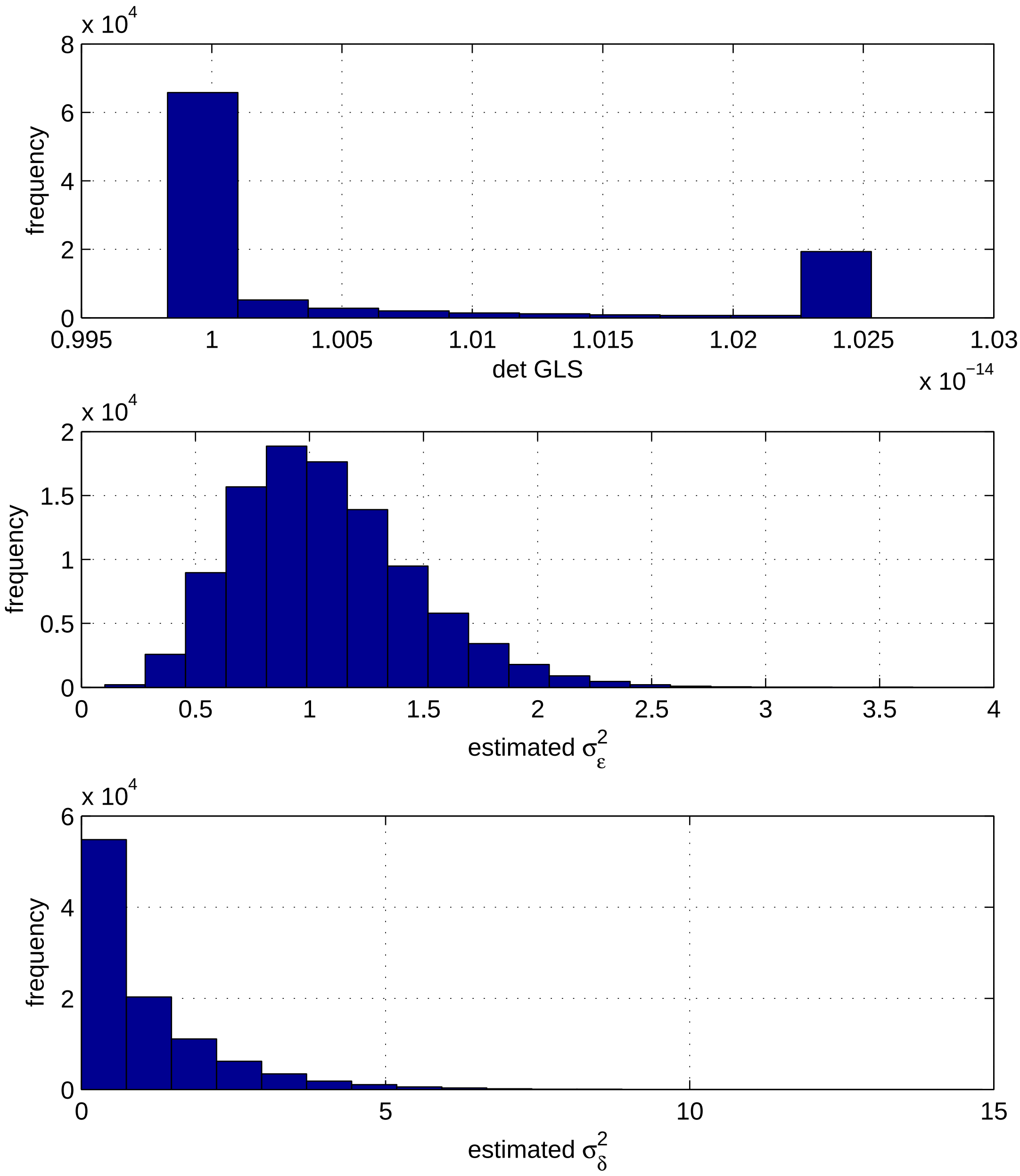


Figure 5.4: Distribution of Simulation Observations.
(D(1, 3) VKM Near-Equivalent CCD, 100,000 runs)



Chapter 6

Concluding Remarks

Restricted randomization presents additional challenges in the design and analysis of response surface experiments. In the introduction we discussed how the development of RSM from classical agricultural research has provided an effective means to conduct industrial and scientific experiments. A natural extension is the use of split-plot designs in cases where some factors are costly-to-change or hard-to-change. In Chapter 1, we provided some examples of industrial applications requiring a split-plot structure as motivation for research in this area. While the RSM community has been reluctant to address the issue of restricted randomization, in recent years there has been an emergence of research interest in the area of industrial split-plot designs, which we discussed in Chapter 2.

The discovery by Vining, Kowalski, and Montgomery (2005) of second-order split-plot designs where the ordinary least squares model estimates are equivalent to the generalized least squares estimates was the impetus for this research effort. Designs that possess the equivalence property enjoy the advantage of best linear unbiased estimates that are robust to the assumption of normality, are not a function of the variance components, and can be easily computed in any statistical software package. In addition, design selection is independent of the variance components and robust to model misspecification. Furthermore, design construction can be based on common families of RSM designs including the central composite and Box-Behnken, which are familiar to practitioners. Moreover, these designs are easily augmented to obtain pure-error estimates of the variance components and are well-suited for sequential experimentation.

In Chapter 3, a generalized proof of the equivalence conditions was presented that enables the development of several systematic design construction techniques and provides the ability to verify numerically that a design provides equivalent estimates. Throughout Chapters 3 and 4 we explored the construction of balanced and unbalanced split-plot versions of the central composite and Box-Behnken designs. In addition, we illustrated the utility of numerical verification in generating $\mathcal{D}$-optimal and minimal point designs including split-plot versions of the Notz, Hoke, Box and Draper, and hybrid designs. As a result, we presented a comprehensive catalog of equivalent estimation designs that minimize the number of whole-plots and feature efficient allocation of experimental resources. The variety of designs in this catalog demonstrates the flexibility in achieving the equivalence condition.

Even with this broad catalog of designs, there are cases when a suitable equivalent design is not available. In Chapter 5, we addressed this practical issue of providing guidance to practitioners on analysis alternatives. By simulation, we compared estimation methods to consider if it remains advantageous to use OLS, rather than moving to a more complex procedure such as GLS. This work serves as a foundation for refining the measures of near-equivalence and considering residual based estimation, as implemented in REML, in the comparison of estimation efficiency.

Throughout this research, we focus on experimentation strategies that give proper consideration to hard-to-change factors and are consistent with design robustness, as defined by Box and Draper (1975). Restricted randomization represents a new facet of RSM, and future research is required to extend classical RSM techniques to accommodate hard-to-change factors.

Bibliography

[1] Addelman, S. (1964). "Some Two-Level Factorial Plans With Split-Plot Confounding". *Technometrics* 6, pp. 253–258.

[2] Bingham, D. and Sitter, R. S. (1999a). "Minimum-Aberration Two-Level Fractional Factorial Split-Plot Designs." *Technometrics* 41, pp. 62–70.

[3] Bingham, D. and Sitter, R. S. (1999b). "Some Theoretical Results for Fractional Factorial Split-Plot Designs." *The Annals of Statistics* 27, pp. 1240–1255.

[4] Bingham, D. and Sitter, R. S. (2001). "Design Issues in Fractional Factorial Split-Plot Experiments." *Journal of Quality Technology* 33, pp. 2–15.

[5] 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.

[6] Bisgaard, S. and Kulahci, M. (2001). "Robust Product Design: Saving Trials with Split-Plot Confounding". *Quality Engineering* 13, pp. 525–530.

[7] Bisgaard, S. and Steinberg, D. M. (1997). "The Design and Analysis of $2^{k-p} \times s$ Prototype Experiments". *Technometrics* 39, pp. 52–62.

[8] Box, G. E. P. (1996). "Split-Plot Experiments". *Quality Engineering* 8, pp. 515–520.

[9] Box, G. E. P. (1999). "Statistics as a Catalyst to Learning by Scientific Method Part II - A Discussion". *Journal of Quality Technology* 31, (1), pp. 16–29.

[10] Box, G. E. P. and Behnken, D. W. (1960). "Some New Three-Level Designs for the Study of Quantitative Variables". *Technometrics* 2, pp. 455–475.

[11] Box, G. E. P. and Draper, N. R. (1974). "Some Minimum Point Designs for Second Order Response Surfaces". *Technometrics* 16, pp. 613-616.

[12] Box, G. E. P. and Draper, N. R. (1975). "Robust Designs". *Biometrika* 62, (2), pp. 347–352.

[13] Box, G. E. P. and Hunter, J. S. (1957). “Multifactor Experimental Designs for Exploring Response Surfaces.” *Annals of Mathematical Statistics* 28, pp. 195–241.

[14] Box, G. E. P. and Jones, S. (1992). “Split-Plot Designs for Robust Product Experimentation.” *Journal of Applied Statistics* 19, pp. 3–26.

[15] Box, G. E. P. and Jones, S. (2001). “Split-Plots for Robust Product and Process Experimentation.” *Quality Engineering* 13, pp. 127–134.

[16] Box, G. E. P. and Wilson, K. B. (1951). “On the Experimental Attainment of Optimum Conditions”. *Journal of the Royal Statistical Society* Ser.B, 13, pp. 195–241.

[17] Cornell, J. A. (1988). “Analyzing Data from Mixture Experiments Containing Process Variables: A Split-Plot Approach”. *Journal of Quality Technology* 20, pp. 2–23.

[18] Cochran W. G. and Cox, G. M. (1957). *Experimental Designs* 2nd edition. John Wiley and Sons, New York, NY.

[19] Draper, N. R. and John, J. A. (1998). “Response Surface Designs where Levels of Some Factors Are Difficult to Change”. *Australian and New Zealand Journal of Statistics* 40, pp. 487–495.

[20] Ganju, J. and Lucas, J. M. (1997). “Bias in Test Statistics when Restrictions in Randomization Are Caused by Factors”. *Communications in Statistics, Part A - Theory and Methods* 26, pp. 47–63.

[21] Ganju, J. and Lucas, J. M. (1999). “Detecting Randomization Restrictions Caused by Factors”. *Journal of Statistical Planning and Inference* 81, pp. 129–140.

[22] Ganju, J. and Lucas, J. M. (2000). “Analysis of Unbalanced Data from an Experiment with Random Block Effects and Unequally Spaced Factor Levels”. *The American Statistician* 54, pp. 5–11.

[23] Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations* 3rd edition. The Johns Hopkins University Press, Baltimore, MD.

[24] Goos, P. (2002). *The Optimal Design of Blocked and Split-Plot Experiments.* Springer-Verlag, New York.

[25] Goos, P. and Vandebroek, M. (2001a). “D-Optimal Response Surface Designs in the Presence of Random Block Effects”. *Computational Statistics and Data Analysis* 37, pp. 433-453.

[26] Goos, P. and Vandebroek, M. (2001b). “Optimal Split-Plot Designs”. *Journal of Quality Technology* 33, pp. 436–450.

[27] Goos, P. and Vandebroek, M. (2003). "D-optimal Split-Plot Designs with Given Numbers and Sizes of Whole Plots". *Technometrics* 45, pp. 235–245.

[28] Goos, P. and Vandebroek, M. (2004). "Outperforming Completely Randomized Designs". *Journal of Quality Technology* 36, pp. 12–26.

[29] Graybill, F. A. (1976). *Theory and Application of the Linear Model.* Duxbury, North Scituate, MA.

[30] Hadi, A. S. (1996). *Matrix Algebra As A Tool.* Duxbury Press.

[31] Hahn, G. J. (1978). "More on Randomization". *Chemical Technology* 8, pp. 164–168.

[32] Hartley, H. O. (1959). "Small Composite Design for Quadratic Response Surfaces." *Biometrika* 15, pp. 611-624.

[33] Hoke, A. T. (1974). "Economical Second Order Designs Based on Irregular Fractions." *Technometrics* 17, pp. 375-384.

[34] Huang, P. ; Chen, D. ; and Voelkel, J. O. (1998). "Minimum-Aberration Two-Level Split-Plot Designs." *Technometrics* 40, pp. 314–326.

[35] 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, pp. 411-421.

[36] Kowalski, S. M. (2002). "24 Run Split-Plot Experiments for Robust Parameter Design". *Journal of Quality Technology* 34, pp. 399–410.

[37] 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.

[38] Kowalski, S. M. and Potcner, K. J. (2003). "How to Recognize A Split-Plot Experiment". *Quality Progress* 36, pp. 60-66.

[39] Letsinger, J. D. ; Myers, R. H. ; and Lentner, M. (1996). "Response Surface Methods for Bi-Randomization Structures". *Journal of Quality Technology* 28, pp. 381–397.

[40] Loeppky, J. L. and Sitter, R. R. (2002). "Analyzing Unreplicated Blocked or Split-Plot Fractional Factorial Designs". *Journal of Quality Technology* 34, pp. 229-243.

[41] Lucas, J. M. and Hazel, M. C. (1997). "Running Experiments With Multiple Error Terms: How an Experiment is Run is Important". *ASQC Technical Conference Transactions*, American Society for Quality Control, Milwaukee, WI, pp. 283–296.

[42] Lucas, J. M. and Ju, H. L. (1992). "Split Plotting and Randomization in Industrial Experiments". *ASQC Quality Congress Transactions*, American Society for Quality Control, Nashville, TN, pp. 374–382.

[43] 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, pp. 1302–1304.

[44] Mee, R. W. and Bates, R. L. (1998). Split-Lot Designs: Experiments for Multistage Batch Processes". *Technometrics* 40, pp. 127–140.

[45] Miller, A. (1997). "Strip-Plot Configurations of Fractional-Factorials". *Technometrics* 39, pp. 153–161.

[46] Montgomery, D. C. (2001). *Design and Analysis of Experiments* 5th edition. John Wiley & Sons, New York, NY.

[47] Myers, R. H. (1999). "Response Surface Methodology-Current Status and Future Directions". *Journal of Quality Technology.* 31, pp. 30-44.

[48] Myers, R. H.; Montgomery, D. C.; Vining, G. G. ; Borror, C. M. ; and Kowalski, S. M. (2004). "Response Surface Methodology: A Retrospective and Literature Survey". *Journal of Quality Technology.* 36, pp. 409–419.

[49] Myers, R. H. and Montgomery, D. C. (2002). *Response Surface Methodology* 2nd edition. John Wiley & Sons, New York, NY.

[50] Notz, W. (1982). "Minimal Point Second Order Designs". *Journal of Statistical Planning and Inference.* 6, pp. 47-58.

[51] Parker, P. A. ; DeLoach, R. (2001) "Response Surface Methods for Force Balance Calibration Modeling". *IEEE 19th International Congress on Instrumentation in Aerospace Simulation Facilities*, Cleveland, Ohio.

[52] Parker, P. A. ; Morton, M. ; Draper, N. R. ; Line, W. P. (2001). " A Single-Vector Force Calibration Method Featuring the Modern Design of Experiments". *American Institute of Aeronautics and Astronautics 39th Aerospace Sciences Meeting* & Exhibit, Reno, Nevada.

[53] Roquemore, K. G. (1976). "Hybrid Designs for Quadratic Response Surfaces". *Technometrics* 18, pp. 419–423.

[54] Schoen, E. D. (1999). "Designing Fractional Two-Level Experiments With Nested Error Structures". *Journal of Applied Statistics* 26, pp. 495–508.

[55] Schoen, E. D. and Wolff, K. (1997). "Design and Analysis of a Fractional $4^1 3^1 2^5$ Split-Plot Experiment". *Journal of Applied Statistics* 24, pp. 409–419.

[56] Searle, S. R. (1982). *Matrix Algebra Useful for Statistics* John Wiley & Sons, New York, NY.

[57] Simpson, J. R. ; Kowalski, S. M. ; and Landman, D. (2004). "Experimentation with Randomization Restrictions: Targeting Practical Implementation". *Quality and Reliability Engineering International.* 20, pp. 481-495.

[58] Taguchi, G. (1987). *System of Experimental Design* Vol. 1, White Plains, NY: Kraus International.

[59] Trinca, L. A. and Gilmour, S. G. (2001). "Multistratum Response Surface Designs". *Technometrics* 43, pp. 25-33.

[60] Vining, G. G. ; Kowalski, S. M. ; and Montgomery, D. C. (2005) "Response Surface Designs Within a Split-Plot Structure". *Journal of Quality Technology* 37, pp. 115-129.

[61] Vining, G. G. and Kowalski, S. M. (2004) "An Overview of Composite Designs Run as Split-Plots". *Fourth Annual European Network for Business and Industrial Statistics Conference*, Copenhagen, Denmark.

[62] Vining, G. G. and Kowalski, S. M. (2005) "Exact Inference for Response Surface Designs Within a Split-Plot Structure". Submitted for publication.

[63] Webb, D. F. ; Lucas, J. M. ; and Borkowski, J. J. (2004). "Factorial Experiments when Factor Levels Are Not Necessarily Reset". *Journal of Quality Technology* 36, pp. 1–11.

[64] Wooding, W. M. (1973). "The Split-Plot Design." *Journal of Quality Technology* 5, pp. 16–33.

Appendix A

Simulation Code

```
function [simout, dGLSv, es2ev, es2dv]=ne_sim(bdstruct,wp_reps,ratio,runs)
%function [simout, dOLSv, dGLSv, se, sd]=ne_sim2(bdstruct,wp_reps,ratio,runs)
% INPUTS
% dstruct contains the base design structure to be simulated
%  base design does not contain WP replicates
%  WP to be replicated must be the last WP in design
% wp_reps is the number of replicated whole-plots
% ratio is WPvar/SPvar = sigma2_delta/sigma2_epsilon
%   we let sigma_epsilon = 1, therefore r = sigma2_delta = WPvar
% runs is the number of simulation trials
%
% OUTPUTS
% simout is a structure with results summary
% dGLSv is a vector of all dGLS observations
% es2ev is a vector of all sigma2_epsilon observations
% es2dv is a vector of all sigma2_delta observations

%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% Initialize Variables %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%
bd = bdstruct;
r = ratio;

%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% Exapand Base Design %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Parameters of Base Design
des = bd.des;
N=bd.N;
m=bd.m;
wv = bd.wpvect;

n_wprep = wv(m);

% Expand Base Design with WP replicates of last WP
lastWP = des(N-n_wprep+1:end,:);
edes = [des; kron(ones(wp_reps-1,1),lastWP)];
em = m + (wp_reps-1);
ewv = [wv kron(ones(1,wp_reps-1),n_wprep)];
% build expanded design structure
d=bd;
d.des = edes;
d.m = em;
d.wpvect = ewv;
[d]=process_design(d,0);

% parameters of expanded design
des = d.des;
X=d.X;
J=d.J;
N=d.N;
w=d.w;
m=d.m;
```

```
n=d.n;
wv = d.wpvect;
I=eye(N);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% generate TRUE variance-covariance matrix  %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
s2e=1; s2d=r;
Strue = s2e*I + s2d*J;
det_true = det(inv(X'*inv(Strue)*X));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% generate OLS variance-covariance matrix and Ratio with True  %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
det_OLS = det(inv(X'*X)*X'*Strue*X*inv(X'*X));
eta_OLS = det_OLS / det_true;

% Initializations
sumdet_GLS = 0;
sumdet2_GLS = 0;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%  Identify SP replicates structure within WPs %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% generate SP replicates count vector
% extract subplot desgins

for i=1:m % number of WPs
    if i == 1;  sb = 1;
    else sb = sb + wv(i-1); end;
    % extract Subplot Design Matrix
    SPD = des(sb:sb + wv(i)-1,w+1:end);
    [dup, wpcv] = unique_rows(SPD);
    % build SP replicate structure by WP
    WPrepstruct(i).n = wv(i);
    WPrepstruct(i).wpcv = wpcv';
end;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% BEGIN Simulation Loop %%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
t0 = clock;
for k = 1: runs;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% generate WHOLE-PLOT and SUBPLOT replicates %%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% extract subplot desgins
% and generate replicates for each WP containing SP reps
tss = 0;     % initalize total sums-of-squares
dfsp = 0;    % initalize subplot rep df
z=0;         % intialize WP replicate counter

for j = 1:m
```

```
    % first check if WP contains SP replicates
    if sum(WPrepstruct(j).wpcv - 1) ~=0
        ni = WPrepstruct(j).n; % number of suplot runs in WP

        %%%  Generate Replicated Runs %%%
        % we assume normally distributed random variables
        % CovMatrix is the within WP covarance structure
        %   (same for all if balanced design)
        CovMatrix = s2e*eye(ni) + s2d*ones(ni,ni);
        % random sample from multivariate normal
        %   mean = 0, within WP var-cov matrix specified above
        yv = mvnrnd(zeros(ni,1),CovMatrix);

        %%%  Calc SS for SP reps %%%
        %length(WPrepstruct(j).wpcv)
        % Loop through each set of SP replicates within each WP
        for i = 1:length(WPrepstruct(j).wpcv)
            fr = sum(WPrepstruct(j).wpcv(1:i-1))+1;
            lr = fr + WPrepstruct(j).wpcv(i) - 1;
            yv(fr:lr);
            ss = var(yv(fr:lr))*(WPrepstruct(j).wpcv(i)-1);
            % accumulate total sums-of-squares of SP reps
            tss = tss + ss;
            dfsp = dfsp + (WPrepstruct(j).wpcv(i)-1);
        end;

        %%%  Calc ybar for WP reps %%%
        % if j is a WP rep, then calc ybar for WP
        % replicated WP must be the last WP in the design
        %    (must also contain replicated SP runs to be in cv)
        if j >= (m + 1 - wp_reps)
            z = z + 1;
            ybar(z) = mean(yv);
        end;
    end;
end;

% compute sample variances
S2_wp = var(ybar);
S2_sp = tss/(dfsp-1);

% apply method of moments to obtain estimated variance components
sigma_hat2_epsilon = S2_sp;
sigma_hat2_delta = S2_wp - 1/n_wprep * S2_sp;
    % sigma_hat2_delta can be negative due to MoM
    % if negative and set to zero
    if sigma_hat2_delta < 0; sigma_hat2_delta =0; end;

% form Sigma_hat
Shat = sigma_hat2_epsilon*I + sigma_hat2_delta*J;

% compute var-cov matrix of beta_hat
iSh = inv(Shat);
A = inv(X'*iSh*X);
```

```
varhat_betahat_GLS = A*X'*iSh*Strue*iSh*X*A;

% compute determinant of var-cov matrix of beta_hat
dGLS = det(varhat_betahat_GLS);

% running sum of the determinant of var-cov matrix of beta_hat
sumdet_GLS = sumdet_GLS + dGLS;

% running sum of det^2 used for variance calc
sumdet2_GLS = sumdet2_GLS + dGLS^2;

%%% Output Sample Observation for Diagnostics %%%
%%% Commented out for regular simulation runs %%%
%  es2ev(k) = sigma_hat2_epsilon;
%  es2dv(k) = sigma_hat2_delta;
%  dGLSv(k) = dGLS;

end;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% END of simulation loop %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% compute average determinant
avgdet_GLS = sumdet_GLS/runs;

% compute standard deviation of determinant values
if runs == 1
    stdevGLS = 0;
else
    stdevGLS = sqrt( (sumdet2_GLS - 1/(runs)*sumdet_GLS^2)/(runs-1) );
end;

% ratio with true beta_hat
eta_GLS = avgdet_GLS / det_true;

% ratio of OLS to GLS
eta_REL = det_OLS / avgdet_GLS;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% Form Simulation Output Results %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
simout.comment = d.comment;
simout.design = d.des;
simout.w = d.w;
simout.k = d.k;
simout.Enorm = d.Enorm;
simout.Dnorm = d.Dnorm;
simout.Anorm = d.Anorm;
simout.runs = runs;
simout.var_ratio = r;
simout.wpreps = wp_reps;
simout.eta_OLS = eta_OLS;
simout.eta_GLS = eta_GLS;
simout.eta_REL = eta_REL;
```

```
simout.det_true = det_true;
simout.det_OLS = det_OLS;
simout.avgdetGLS = avgdet_GLS;
simout.stdevGLS = stdevGLS;

disp(sprintf('elapsed time = %4.2f',etime(clock,t0)));

return;
```