RESPONSE SURFACE DESIGNS AND ANALYSIS FOR BI-RANDOMIZATION ERROR STRUCTURES

by
Jennifer J. Davison

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

DOCTOR OF PHILOSOPHY
in
Statistics

APPROVED:

Raymond H. Myers, Chair

Marvin Lentner, Chair

Jesse C. Arnold

Klaus Hinkelmann

Marion R. Reynolds, Jr.

July 5, 1995
Blacksburg, Virginia

RESPONSE SURFACE DESIGNS AND ANALYSIS FOR BI-RANDOMIZATION ERROR STRUCTURES

by
Jennifer J. Davison

Raymond H. Myers, Chair Marvin Lentner, Chair
Department of Statistics
Virginia Polytechnic Institute and State University

ABSTRACT

Cost control, resource availability, or difficulty in performing complete randomizations may dictate the necessity to run response surface experiments in a bi-randomization error control format of which the split plot design is a special case. A bi-randomization scheme allows for certain factor levels to be applied at random to large experimental units with the remaining factor levels randomly applied to nested smaller units. For example, in the dual response surface approach to robust parameter design, process mean and variance models are formulated to aid in designing products to be "robust" to uncontrollable system influences called noise variables. In model development, noise variables are assumed to be controllable in the laboratory, but due to their random nature they may be costly and/or difficult to control. This suggests the need for a bi-randomization scheme in which the noise variables constitute the levels applied to the larger experimental units.

For the bi-randomization situation, two types of bi-randomization designs are explored along with their respective analyses and various error variance estimation procedures. The efficiency of common response surface designs are also examined in the presence of this alternative error structure to determine the necessity of design modifications to better accommodate the error structure. General recommendations for efficient designs and practical analysis methods are outlined.

ACKNOWLEDGEMENTS

I wish to extend my sincere gratitude to my family and friends for their support and encouragement throughout this endeavor. I would especially like to thank my parents, Jim and Gail Davison, and my brothers and sister, Tom, Mike, and Diane, for their unending love, support, and patience throughout my years of schooling.

In addition, I wish to thank my advisors, Dr. Raymond H. Myers and Dr. Marvin Lentner for their guidance, expertise, and encouragement. Most of all I would like to thank them for believing in me.

Finally, I wish to thank my husband, Dr. William C. Letsinger, II, for his love and friendship which are invaluable to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

§1.1 Response Surface Methodology

In 1951, Box and Wilson published a landmark paper in the area of industrial experimentation which ignited development in what has become response surface methodology (RSM). Response surface analysis originates from the assumption that there exists a response, y, which is a function of a set of controllable inputs called design variables, denoted by $\underline{x}$, i.e.

$$y=f(\underline{x},\underline{\theta}). \tag{1.1.1}$$

While $f(\underline{x})$ is typically unknown and may in fact be a complex function, within a given region of interest it is assumed that it can be adequately approximated by a linear polynomial model in the design variables, i.e.

$$y \approx h(\underline{x},\underline{\beta}). \tag{1.1.2}$$

Using this model assumption in 1.1.2, statisticians and engineers developed methods for not only response optimization of 1.1.1 with respect to $\underline{x}$, but also for the exploration of the response surface itself which is invaluable knowledge in industrial and experimental applications. Through the years the variety of methods, tools, and techniques used to accomplish these goals has grown. This collection of RSM tools can be broken into three categories corresponding to the

three basic steps in RSM: (i) regression model building, (ii) region seeking, and (iii) response optimization/region exploration.

The first step revolves around the formulation of appropriate designs for the estimation of the polynomial model, 1.1.2. Unlike regression methods where the design matrices are observational, RSM has the unique characteristic that the designs are controllable by the experimenter. Because of this, a multitude of designs have been developed. Included for first order modeling are two-level full and fractional factorials, Plackett-Burman designs [Plackett & Burman, 1946] and orthogonal two-level designs [Box & Wilson, 1951]. In the presence of second order models, designs such as the central composite [Box & Wilson, 1951], Box-Behnken [Box & Behnken, 1960], small composite [Hartley, 1959], hybrid [Roquemore, 1976], and variations on the 3^k factorials [Hoke, 1974] were introduced. The goal of these researchers was to formulate designs which provide desirable properties for the estimation of 1.1.2. See Box & Draper [1987], Myers [1976], Khuri & Cornell [1987], and Myers and Montgomery [1995] for further discussion of these designs and their properties. After choosing an appropriate design for estimating 1.1.2, the next step in the RSM procedure is region seeking through sequential experimentation.

The concept of region seeking arose out of a need to consider alternative regions of interest for modeling. See Box & Wilson [1951]. It allows for a movement across the response surface to other regions which may in turn contain the location of optimum response. This location is of primary importance in RSM. Region seeking methods provide a "protection" against the experimenter's initial selection of the region of optimum response. Included in the region seeking techniques are the steepest ascent/descent methods, simplex search method, and method of parallel tangents. Additional information on these procedures can be found in Davies [1954], Myers [1976], Myers & Khuri [1979], Khuri & Cornell [1987], and Box & Draper [1987]. In addition to region movement, sequential region seeking also incorporates design augmentation for the detection of curvature within a given region and thus is an indicator of the necessity for second order augmentation and alteration of 1.1.2. After the final experimental region is

determined, internal investigation can then begin within that region.

As mentioned previously, exploration within the region of optimum accomplishes two primary goals: the obtainment of an optimum response and, in addition, an understanding of the response surface or system. Through the use of methods such as canonical analysis [Box & Wilson, 1951] for the determination of the existence of an optimum and response contour plots, location of optimum can be selected. Unfortunately, the choice is not always clear. In these situations, ridge analysis, a constrained optimization procedure, may be used. See Hoerl [1959], Draper [1963], and Khuri & Cornell [1987] for details and examples. While determination of the optimal response location is important, this collection of tools provides an even more valuable insight into the nature of the response surface. Hill and Hunter [1966] referred to it as an "elucidation of an underlying mechanism" (p. 571).

Recent development in RSM center around the concept of creating "optimal" designs for use in the model building step. The origins of choosing a "best" design with respect to a given property dates back to Kiefer [1958, 1959] and Kiefer & Wolfowitz [1959, 1960]. Many criteria for design selection have been created included among them D optimality which chooses the design which minimizes the generalized variance of the estimated model coefficients, Q optimality which addressed minimization of an "average" prediction variance across a specified region, and G optimality which looks to minimize the maximum prediction variance over a region. These types of criteria are appropriately called the "alphabetic optimality" criteria. Refer to Silvey [1980] and Atkinson [1982] surveys of the various criteria.

Improvements and new techniques are continually being added to the collection of RSM tools in an attempt to find a useful "marriage" of regression modeling and design of experiments. It is an area of research which finds itself nested within many industrial applications and has proven to be beneficial in process understanding and optimization.

§1.2 Assumptions and Analysis

The foundation of RSM, as mentioned earlier, is that there exists a response of interest, y, which is a function of design variables, $\underline{x}$, which may be either discrete or continuous. Also mentioned was the fact that this function is often unknown and complicated, but it is assumed due to Taylor series expansion that on a region of interest, $\mathfrak{R}$, $f(\underline{x})$ can be approximated by a linear polynomial model in the design variables,

$$\underline{y}=X\underline{\beta}+\underline{\epsilon} \tag{1.2.1}$$

where $\underline{y}$ is an $n*1$ vector of responses, X is an $n*(p+1)$ known model matrix, $\underline{\beta}$ is a $(p+1)*1$ vector of unknown parameters, and $\underline{\epsilon}$ is an $n*1$ vector of unknown errors. Any order polynomial may be used in X, but typically the restriction is to first or second order models. In addition, we often assume

$$\underline{\epsilon} \sim N(0, \sigma^2 I). \tag{1.2.2}$$

Under this error assumption maximum likelihood estimation (MLE) of 1.2.1 is obtained equivalently by using the method of ordinary least squares (OLS), i.e.

$$\hat{\underline{\beta}}=(X'X)^{-1}X'\underline{y} \tag{1.2.3}$$

and

$$\mathrm{Var}(\hat{\underline{\beta}})=\sigma^2(X'X)^{-1}. \tag{1.2.4}$$

At this point, estimation of σ^2 can be done and variable screening using t-tests can be performed on 1.2.1 to obtain a final fitted model. Response surface analysis techniques can then be employed as described in §1.1.

The model matrix, X, found in 1.2.1 is controllable through design by the experimenter. Careful selection of that design, denoted by $\mathfrak{D}$, is essential to provide a good approximation to $f(\underline{x})$. Based on the assumption that ϵ_i iid $N(0,\sigma^2)$ a pool of "useful" designs have been developed. A few of these were mentioned in

§1.1 for first and second order polynomial models. The 3^k full factorial and central composite design (ccd) are two second order designs which are among the more universally used for the exploration of second order RSM models.

§1.3 3^k Factorial and Central Composite Design Examples

To illustrate these specific second order RSM designs, consider two design variables, x_1 and x_2, i.e. $y=f(x_1,x_2)$. For the 3^2 design

$$\mathfrak{D}= \begin{array}{c} \begin{array}{cc} x_1 & x_2 \end{array} \\ \left[\begin{array}{cc} 1 & 1 \\ 1 & 0 \\ 1 & -1 \\ 0 & 1 \\ 0 & 0 \\ 0 & -1 \\ -1 & 1 \\ -1 & 0 \\ -1 & -1 \end{array}\right] \end{array}$$

Note that the range of extreme possible values on each variable is used to appropriately center and scale the largest value on each variable to 1 and the lowest value to -1. This centering and scaling technique is common in RSM and will be used throughout this dissertation. Notice the 3^2 RSM design provides an attractive balanced pairing of the unique levels of each design variable.

The central composite design [Box & Wilson, 1951], however, does not preserve this pairing relationship. It instead has the following form:

$$
\mathfrak{D}=\begin{array}{c} \begin{matrix} x_1 & x_2 \end{matrix} \\ \left[\begin{array}{cc} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \\ \alpha & 0 \\ -\alpha & 0 \\ 0 & \alpha \\ 0 & -\alpha \\ 0 & 0 \end{array}\right] \end{array} \begin{array}{l} \text{(i)} \\ \\ \\ \\ \text{---} \\ \\ \text{(ii)} \\ \\ \text{---} \\ \text{(iii)} \end{array}
$$

The ccd is comprised of three unique portions: (i) a full or fractional two-level factorial, (ii) an axial portion to introduce additional levels for second order modeling, and finally (iii) center run(s). Within the axial portion, α is referred to as the axial value and is chosen to obtain desirable design properties. The number of center runs, n_0, is also chosen to achieve desirable properties. Note that the ccd requires at least one center run to prevent singularity of X in equation 1.2.1.

In this specific example, both the 3^2 and ccd have a total of 9 design points. As the number of variables grows, however, the ccd maintains a smaller design size than that of the 3^k, thus making it an attractive alternative when design size is a concern. For example, for a four variable situation , a ccd ≥ 27 total runs whereas the 3^4 has a total of 81 runs .

Much research effort has been expended in the formulation of treatment designs such as the 3^k and ccd which provide overall "nice" properties in RSM. However, in the development of these designs little attention has been given to the role of the error control structure. The assumption of a completely randomized (CRD) (or randomized complete block (RCBD)) error control structure has been universally accepted in a majority of RSM research and

applications. Unfortunately, this assumption may not always be valid.

§1.4 **Randomization Difficulty - Example**

A situation which could complicate the use of a CRD structure arises when one or more of the design variables, $\underline{x}$, are difficult and/or costly to control. For example, consider an opalescent glass manufacturing process in which variables such as kiln temperature, percentage(%) sand content in the glass mixture, and melting/heating time in the kiln are of interest in relation to the overall quality of the final sheet of glass. Percentage sand and melting time are two easily manipulated design variables in experimentation; kiln temperature, however, although controllable, can be difficult and costly to operate in order to accommodate a completely randomized structure.

The kiln is a room sized oven with 8-10 smaller ovens located within it. Although there are separate ovens, all ovens must be heated to the same kiln temperature. After running a single treatment combination of temperature, % sand, and melting time (say from a ccd) in one of the ovens, it takes around five days for the kiln to cool down enough to be able to scrape out the remaining glass residue and prepare the oven for the next design run. It then takes another few days to fire the kiln back up to temperature for the next randomly chosen treatment combination. Even if a separate oven was utilized for each design run, the manipulation of temperature can be a difficult and time consuming process. Because of this, a CRD in terms of kiln temperature is not feasible.

A more realistic approach to this situation would involve heating the kiln to a randomly chosen temperature and running all corresponding treatment combinations of % sand and melting time in the available ovens of the kiln. Once completed, the kiln can be cooled and prepared for the next temperature until all desired temperatures are run. The entire experiment would remain costly and time consuming, but considerable savings would be obtained over the use of a CRD. This is just one of many possible situations in which a CRD is not economical or easily implemented. For example, in the RSM field of robust

parameter design (RPD) this problem would arise frequently in the manipulation of a subset of design variables called noise variables.

§1.5 Application to Robust Parameter Design

In the early 1980's, Taguchi [Taguchi [1986], [1987]] sparked interest in the consideration of variability in product design. His philosophy focused on the precept that poor product quality is in part due to inconsistency in product performance. The idea is to design a system or product whose inputs called control design variables produce desired mean responses with minimal variability. The variability to be minimized is that introduced through fluctuations in levels of factors called noise factors which in the system or in use of the product in the field are assumed uncontrollable. The name given to this philosophy is robust parameter design (RPD).

The choice of optimum control inputs involves designing an experiment to allow for the presence of noise variability. In the laboratory setting, however, for the purpose of model building, noise variables must be assumed controllable so they may be incorporated into the design set-up along with the control variables and also in the modeling of the system. The control*noise interactions found in the overall model provide the key to controlling process variability through settings of the controls. Analysis of the noise/control experiment through either signal-noise ratios [Taguchi [1986], [1987], Taguchi & Wu [1980], [1985], and Kackar [1985]] or the dual response surface approach [Vining & Myers [1990] and Myers, Khuri & Vining [1992]], provide operating conditions which optimize the mean response while providing a robustness to changing noise variables. For a general discussion of Taguchi's RPD see the panel discussion edited by Nair [1992].

In RPD, the assumption concerning the ability to control noise variables in the laboratory for modeling may in some cases be a bold assumption. By the very nature of some noise variables, they will be difficult and/or costly to control. For example, consider environmental conditions such as humidity or pressure, velocity

of machine parts, or once again fluctuations in oven or kiln temperature. Completely randomized designs in these cases may not be efficient options.

§1.6 Split Plot Application in RPD

To alleviate the complications associated with running CRDs in the presence of hard to control variables, a solution lies in using alternative error control structures. Box and Jones [1992] recognized the natural application of the split plot design to RPD experiments in which the noise variables are hard/costly to control.

A split plot design is a factorial design structure with randomization restrictions. It is characterized by two particular attributes. First, there are two types of experimental units (EU): whole-plots and sub-plots whereas a CRD has only one type of EU. In addition, the split plot design also involves two separate randomization procedures corresponding to the random assignment of factor levels to each type of EU. The first factor whose levels are randomly applied to large EUs, the whole-plots, is referred to as the whole-plot factor. This initial randomization generates an error component called the whole-plot error which measures natural variability among the whole-plot EUs. After the assignment of this factor, each whole-plot is subdivided into smaller sub-plot EUs. Within each whole-plot, all levels of the second factor of interest, the sub-plot factor, are then randomly assigned to sub-plots. Each whole-plot, therefore, receives all levels of the sub-plot factor but only one level of the whole-plot factor. Randomization within whole-plots generates a second error component which now accounts for the variability among sub-plots within a whole-plot. For the origin of the split plot design, see Yates [1935], [1937], Lentner and Bishop [1993], and Hinkelmann & Kempthorne [1994].

The most natural application of the split plot design to RPD corresponds to the use of combinations of noise variables as the levels of the whole-plot factor with combinations of control variables constituting the levels of the sub-plot factor. Box and Jones [1992] noted the influences of this design/error structure on

RPD in a treatment comparison setting. Use of the split plot design in RPD considerably simplifies experimentation due to the minimal manipulation of noise variable levels; however, a price is paid for this simplification in the precision of estimation of the noise effects. Through the concept of uniformity trials, it can be easily shown that the whole-plot effects (noise) are estimated less precisely in the split plot arrangement than in the completely randomized format. The sub-plot effects (controls and control*noise interactions), on the other hand, are actually estimated more precisely in the split plot design than in the CRD. In other words, let EMS correspond to the expected mean square, and if E_{wp}=EMS(whole-plot error) for a split plot design, E_{sp}=EMS(sub-plot error) for a split plot design, and E_{crd}=(CRD error) for a CRD, then

$$E_{sp}<E_{crd}<E_{wp}. \qquad (1.6.1)$$

The sacrifice comes in the loss of precision in estimation of the whole-plot effects. Box and Jones also discussed other split plot scenarios applied to RPD including the use of the control variables for the whole-plot factor and strip block designs.

While Box and Jones saw the logical application of the split plot design in RPD in a treatment comparison set-up, there exists a more universal need for this type of structure in general response surface methods. This would entail using designs such as the split plot in regression applications instead of in a treatment comparison setting.

§1.7 **Bi-Randomization Error Structure**

A general class of designs will be introduced called bi-randomization designs (BRDs), of which the split plot design is a special case. These designs can provide useful and economical alternatives to CRDs in the presence of hard/costly to control variables. The implications of the BR error structures must, however, be examined. Modification of regression model building, region seeking procedures, and of course optimization techniques may be necessary to accommodate the new format. In this dissertation, the implications of the bi-

randomization error design on model formulation and analysis will be explored along with an investigation into design optimality under the new error control design.

CHAPTER 2

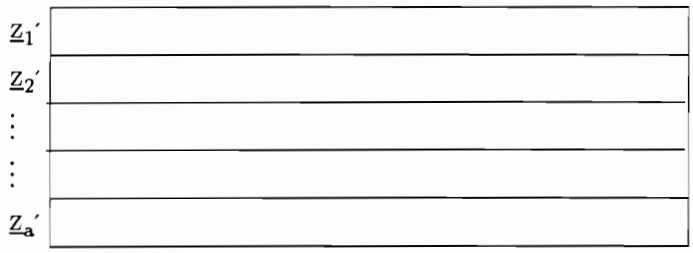
BI-RANDOMIZATION ERROR STRUCTURE

§2.1 Definition and Design Set-up

The concept of the bi-randomization error structure borrows from the randomization structure of the split plot error control design which was introduced in §1.6. The formation of a bi-randomization design (BRD) begins with a chosen response surface design, $\mathfrak{D}$, such as the 2^k full or fractional factorial for first order modeling, or the ccd or 3^k factorial design for second order models. From the set of original design variables that form $\mathfrak{D}$ the subset of difficult and/or costly to control variables are then identified and are denoted by $z_1, z_2, \ldots, z_z$ or $\underline{z}$. These variables will be referred to as whole-plot variables. The remaining design variables are then assumed to fall within the class of more easily controlled variables called sub-plot variables and will be denoted by $x_1, x_2, \ldots, x_x$ or $\underline{x}$. Using this notation, there are a total of $(z+x)$ design variables of interest. The response surface design is assumed to have a total of n design runs where each combination $(\underline{z}',\underline{x}')$ is an individual design run consisting of a treatment combination of $z_1, z_2, \ldots, z_z$ and $x_1, x_2, \ldots, x_x$, respectively.

The next step in the bi-randomization procedure is to identify the "a" unique treatment combinations of the $\underline{z}'=(z_1, z_2, \ldots, z_z)$ found in the original design and denote them as $\underline{z}_1', \underline{z}_2', \ldots, \underline{z}_a'$. These treatment combinations constitute the "a" levels of what will be referred to as the whole-plot factor.

These unique levels are randomly assigned to a set of large sized experimental units called whole-plots, i.e.



This initial randomization of whole-plot variable levels to the whole-plot EUs corresponds to the first randomization procedure in the bi-randomization structure and produces the first variance component, σ_δ^2, called the whole-plot error variance which measures the natural variation among these large EUs.

After the initial randomization, the original treatment combinations of $(\underline{z}',\underline{x}')$ found in $\mathfrak{D}$ are used to identify the $\underline{x}'$ combinations that occur with each unique $\underline{z}_i'$. These $\underline{x}'$ combinations make up the levels of the second BRD factor, the sub-plot factor. The larger EUs are subdivided into an appropriate number of smaller units called sub-plots. The number of sub-plots needed depends directly upon the number of $\underline{x}'$ associated with each unique $\underline{z}_i'$ in the original design. Whole-plots need not have the same number of sub-plots. Let b_i denote the number of sub-plots within the i^{th} whole-plot. Then, within the i^{th} whole-plot the corresponding $\underline{x}'$ treatment combinations are randomly assigned to the sub-plot EUs. The final BRD has the following form:

$\underline{z}_1'$	$\underline{x}_{11}'$	$\underline{x}_{12}'$	$\cdots$	$\underline{x}_{1b_1}'$
$\underline{z}_2'$	$\underline{x}_{21}'$	$\underline{x}_{22}'$	$\cdots$	$\underline{x}_{2b_2}'$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
$\underline{z}_a'$	$\underline{x}_{a1}'$	$\underline{x}_{a2}'$	$\cdots$	$\underline{x}_{ab_a}'$

where

$\underline{z}_i'$ corresponds to the whole-plot factor level (a specific $\underline{z}$ treatment combination) applied to the i^{th} whole-plot EU, i=1,2,...a

and

$\underline{x}_{ij}'$ corresponds to the sub-plot factor level (a specific $\underline{x}$ treatment combination) applied to the j^{th} sub-plot EU within the i^{th} whole-plot EU, i=1,2,...,a j=1,2,...,b_i.

The original design, $\mathfrak{D}$, consists of design variable treatment combinations ($\underline{z}_1'$, $\underline{x}_{11}'$), ($\underline{z}_1'$, $\underline{x}_{12}'$), . . ., ($\underline{z}_a'$, $\underline{x}_{ab_a}'$). Thus, the treatment design does not get altered, only the method in which it is run. Note that terminology similar to that used in the split plot design is incorporated into the BRD set-up due to the relationship between the designs.

The second randomization procedure which randomizes sub-plot factor levels to EUs within a whole-plot introduces a dependent relationship among responses in the same whole-plot, i.e. observations in the same whole-plot are correlated. This stems from the fact that the whole-plot EU receives only one application of a level of the whole-plot factor combination while receiving independent applications of the levels of the sub-plot factor. Throughout this dissertation it will be assumed that the covariance between two observations within the same whole-plot remains constant across the sub-plots. This second randomization procedure also introduces a second variance component, σ_ϵ^2, which will be referred to as the sub-plot error variance. It measures the variation among the sub-plot EUs along with residual variation.

The description just given applies to a general bi-randomization design, but this general class of designs separates into two BRD categories called crossed and non-crossed BRDs. The classification into each category depends upon the resulting structure of the levels of the sub-plot factor that are randomized within each whole-plot.

§2.2 Crossed BRD - 3^k Example

For illustration of the crossed BRD, recall the 3^2 full factorial design given in §1.3. For this example, assume that variable x_1 is a costly and/or difficult to control variable such as temperature, humidity, etc., i.e. a whole-plot variable, while x_2 is more easily controlled and, thus, considered a sub-plot variable. Thus, x_1 is our "z" variable, and x_2 is our "x" variable.

In the original 3^2 design, z (or x_1) has three unique levels identified as 1, 0, and -1. These comprise the levels of the whole-plot factor which are randomly applied to whole-plot EUs, i.e.

1	
0	
-1	

Note that in the 3^2 factorial, given on pg. 5, each level of z (or x_1) is crossed with all the levels of the other variable x (or x_2). For this reason, all the unique levels of x coded as 1, 0, -1 are randomized within each whole-plot to the sub-plot experimental units to complete the bi-randomized structure, i.e.

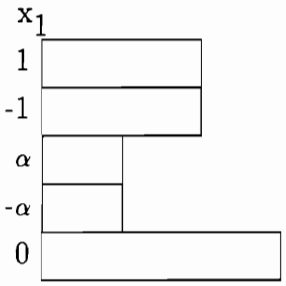
z	x		
1	1	0	-1
0	1	0	-1
-1	1	0	-1

The overall composition of the 3^2 factorial is preserved, but now a dependent relationship exists among the observations within each of the following groups: {(1,1), (1,0), (1,-1)}, {(0,1), (0,0), (0,-1)}, {(-1,1), (-1,0), (-1,-1)} . Also notice that each whole-plot has the same number of sub-plot EUs and identical

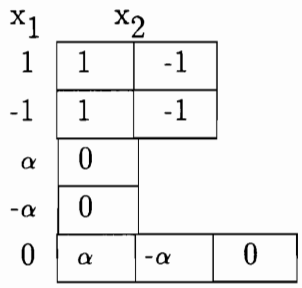
applications of the levels of x. This example falls within the class of crossed BRDs because of these very characteristics. A crossed BRD has each level of the sub-plot factor "crossed" with each level of the whole-plot factor.

§2.3 **Non-crossed BRD - CCD Example**

Now recall the 2 variable central composite response surface design also given in §1.3. Once again, for illustration purposes assume that x_1 is the whole-plot factor while x_2 is the sub-plot factor. In this example, the whole-plot factor now has five unique levels: $-\alpha$, -1, 0 , 1, α which are assigned at random to the whole-plot EUs, i.e.



The corresponding levels of x_2 for each level of x_1 found in the ccd are then randomly applied to the respective sub-plot EU to give the resulting BRD:



Once again, the overall structure of the design remains. Unlike in the previous 3^2 factorial example, however, this BRD no longer has the same number of sub-plots within each whole-plot, nor are the same levels of x_2 found within each whole-plot

unit. The bi-randomization ccd no longer has the crossing of both factor levels. This type of BRD will be called a non-crossed BRD because each level of the sub-plot factor is no longer "crossed" with the all the levels of the whole-plot factor. Even in the presence of this non-crossed structure and varying number of sub-plots within a whole-plot, the assumption of constant covariance between any two observations within the same whole-plot will be maintained.

§2.4 **Model and Estimation**

After having introduced informally through example the two types of bi-randomization designs, attention must now return to the general class of BRDs and how the presence of the new error control scheme and the dependent relationship among observations within whole-plots affect standard response surface modeling and estimation. Under the bi-randomization error structure (either crossed or non-crossed), the general RSM model 1.2.1 can be rewritten as follows:

$$\underline{y}=\beta_0\underline{1} + Z\underline{\gamma} + X\underline{\alpha} + Z'\Delta X + \underline{\delta} + \underline{\epsilon} \qquad (2.4.1)$$

where

$\underline{y}$ is an n∗1 vector of observed responses, y_{ij}

Z is n∗k model matrix for the whole-plot variables

X is n∗l model matrix for the sub-plot variables

β_0 is unknown intercept parameter

$\underline{\gamma}$ is a k∗1 vector of unknown whole-plot variable parameters including whole-plot interactions and higher order polynomials

$\underline{\alpha}$ is a l∗1 vector of unknown sub-plot parameter including sub-plot interactions and higher order polynomials

Δ is a k∗l matrix of unknown whole-plot∗sub-plot interaction parameters

$\underline{\delta} = \{\delta_i\}$ i=1,2,...,a

$\underline{\epsilon} = \{\epsilon_{ij}\}$ i=1,2,...,a; j=1,2,...,b_i

δ_i iid $N(0, \sigma_\delta^2)$; ϵ_{ij} iid $N(0, \sigma_\epsilon^2)$; $Cov(\delta_i, \epsilon_{ij}) = 0 \ \forall$ i,j.

The error term δ_i corresponds to the error generated from being in the i^{th} whole-

plot EU, while ϵ_{ij} is the error associated with the j^{th} sub-plot unit within the i^{th} whole-plot.

Model 2.4.1 is similar to 1.2.1 in its linear form, but 2.4.1 breaks up the general model matrix into portions attributed to both the whole-plot and sub-plot design variables and their interactions. The set of $\underline{\gamma}$ parameters will be referred to as the whole-plot parameters whereas $\underline{\alpha}$ and $\triangle$ together will be collectively termed the sub-plot parameters.

Model 2.4.1 also has restrictions imposed upon it by the bi-randomization structure. Since combinations of whole-plot variables constitute levels of the whole-plot factor, the model in the whole-plot terms, $\underline{\gamma}'\underline{z}_i^d$, must have less than (a-1) terms. The restriction comes from the "a" whole-plots only providing (a-1) degrees of freedom for the modeling of whole-plot effects. The remaining (n-a) degrees of freedom are then available for the modeling of sub-plot effects consisting of all sub-plot terms and whole-plot*sub-plot interactions.

Since the y_{ij}'s are no longer iid under the bi-randomization error scheme, the vector of responses, $\underline{y}$, now has the following distribution:

$$\underline{y} \sim N(\underline{\mu}, V)$$

$$\text{where } \underline{\mu} = \beta_0\underline{1} + Z\underline{\gamma} + X\underline{\alpha} + Z'\triangle X$$

and

$$V=\begin{bmatrix} T_1 & 0 & \cdots & 0 \\ 0 & T_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T_a \end{bmatrix}$$

where

$$T_i=\sigma_\epsilon^2 I_{(b_i*b_i)} + \sigma_\delta^2 \underline{1}_{(b_i*1)}\,\underline{1}'_{(1*b_i)} \quad i=1,2,\ldots,a.$$

T_i corresponds directly to the i^{th} whole-plot EU. The matrix V is block diagonal due to the independence of observations between whole-plots. Within a whole-plot, however, unlike in the CRD the observations are no longer independent, i.e. if y_{ij} is the j^{th} observation within the i^{th} whole-plot then

$$\begin{aligned} &Cov(y_{ij}, y_{jk})=0 \quad \forall\, i \neq j \\ &Cov(y_{ij}, y_{ik})=\sigma_\delta^2 \quad \forall\, i. \end{aligned}$$

The non-diagonal structure of V (and each T_i respectively) alters the maximum likelihood equation for the estimation of model 2.4.1 from the OLS expression that was used in Chapter 1 for a CRD to the generalized least squares (GLS) equation. If model 2.4.1 is rewritten as

$$\underline{y}=X^*\underline{\beta} + \underline{\delta} + \underline{\varepsilon} \tag{2.4.2}$$

where $X^*\underline{\beta} = \beta_0\underline{1} + Z\underline{\gamma} + X\underline{\alpha} + Z'\Delta X$

then

$$\hat{\underline{\beta}} = (X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1}\underline{y} \tag{2.4.3}$$

and

$$Var(\hat{\underline{\beta}}) = (X^{*\prime}V^{-1}X^*)^{-1}. \tag{2.4.4}$$

§2.5 Efficiency/Power Comparison for CRD vs BRD

As was mentioned previously in §1.6 the convenience and economic advantages associated with using a BRD in certain experimental situations in place of a CRD does not come without a price. The price is paid in the precision of estimating the whole-plot parameters in model 2.4.1. For a specific design with the experiment conducted as a BRD, the whole-plot parameters are estimated with less precision than if the same design were run as a CRD. This is due to the fewer number of independent whole-plot EUs found in a BRD as compared to a CRD's "n" independent units. However, the sub-plot parameter estimates actually increase in precision in the presence of a BRD structure (as compared to

a CRD) due to the application of the $\underline{x}_{ij}$'s within the same whole-plot units . The contrasting precisions illustrate the same relationship expressed in equation 1.6.1. Thus, the benefits obtained in experimentation by using the BRD are hindered by a sacrifice in the precision of whole-plot parameter estimation.

To illustrate this precision trade-off in the BRD, consider a five variable experiment consisting of three whole-plot variables, z_1,z_2,z_3, each at two levels (-1,1) and two sub-plot variables, x_1 and x_2, where x_1 has levels (-1,1) and x_2 (-1,0,1). The response surface design is a 2^3*2*3 full factorial. Now assume that this same design is run under two different error control procedures, completely randomized and bi-randomization.

For the CRD, experimental data for all 48 treatment combinations was generated from the following known response model:

$$y_i=10.13 + 1.16z_1 - 0.91z_2 + 1.97z_3 + 2.05x_1 + 0.79x_2 + 0.98x_1x_2 + 0.83x_2^2 + 1.23z_1x_1 - 0.99x_1z_2 + 0.46x_1z_3 - 0.92x_2z_1 + 0.78x_2z_2 + 0.86x_2z_3 + 0.67\,x_2^2z_2 + \theta_i \qquad (2.5.1)$$
$$i=1,2,...,48$$

where θ_i iid $N(0, \sigma_\theta^2=13)$.

Under the bi-randomization scheme, the eight level combinations of the whole-plot variables, z_1,z_2,z_3, make up the levels of the whole-plot factor while the six level combinations of x_1 and x_2 constitute the levels of the sub-plot factor. Due to the factorial structure of the design, this example falls within the class of the crossed BRDs. The true model for generating experimental data for this error control design remains the same as 2.5.1 above except that an additional error term must be incorporated into the model to accommodate the BR structure:

$$y_{ij}=10.13 + 1.16z_1 - 0.91z_2 + 1.97z_3 + 2.05x_1 + 0.79x_2 + 0.98x_1x_2 + 0.83x_2^2 + 1.23z_1x_1 - 0.99x_1z_2 + 0.46x_1z_3 - 0.92x_2z_1 + 0.78x_2z_2 + 0.86x_2z_3 + 0.67\,x_2^2z_2 + \delta_i + \epsilon_{ij} \qquad (2.5.2)$$
$$i=1,2,...,8; \quad j=1,2,...,6$$

where δ_i iid N(0,8) and ϵ_{ij} iid N(0.5).

In both error schemes, the variance of a single observation is held equal to 13: σ_θ^2 for CRD, $\sigma_\delta^2+\sigma_\epsilon^2$ for BRD.

Assuming that all of the respective variance parameters are known for each error situation, the standard errors for the estimated model parameters for each error design can be calculated using the following formulae:

CRD (OLS estimation): $\mathrm{Var}(\hat{\beta})=(X^{*\prime}X^*)^{-1}\sigma_\theta^2$
BRD (GLS estimation): $\mathrm{Var}(\hat{\beta})=(X^{*\prime}V^{-1}X^*)^{-1}$ where $V=f(\sigma_\delta^2, \sigma_\epsilon^2)$.

The resulting standard errors of the estimated coefficients for each error design are given in Table 2.5.1.

Table 2.5.1 Standard Errors of CRD and BRD Estimated Model Coefficients

Coefficient	CRD	BRD
Intercept	.90	1.15
z_1	.52	1.05
z_2	.90	1.15
z_3	.52	1.05
x_1	.52	.32
x_2	.64	.40
x_1x_2	.64	.40
x_2^2	1.10	.68
x_1z_1	.52	.32
x_1z_2	.52	.32
x_1z_3	.52	.32
x_2z_1	.64	.40
x_2z_2	.64	.40
x_2z_3	.64	.40
$x_2^2z_2$	1.10	.68

Table 2.5.1 clearly displays the precision trade-off referred to earlier in this section. For the whole-plot model terms z_1, z_2, z_3, there is a significant loss of precision in parameter estimation in the BRD. In estimating the coefficients of z_1 and z_3, the use of the BRD almost doubles from the CRD standard error whereas there are only moderate increases for the other whole-plot coefficients. The remaining model parameters, the sub-plot parameters, illustrate the gain in precision found in the use of the BRD. In the BRD, observations within a given whole-plot can be viewed as subsampling on that whole-plot EU. With this idea, it can be more easily seen why the sub-plot estimation increases in precision for the BRD. This example illustrates that if an experimenter chooses to utilize the BRD error structure for whatever purpose, he must be willing to relinquish some precision in estimating the whole-plot parameters. If the whole-plot parameters are of primary interest in the response surface analysis, a BRD may not be a suitable alternative. A note must be made that this variation in precision in not unique to the BRD and CRD comparison. Any two different error control structures for the same design will produce different degrees of precision. In this dissertation, however, the BRD vs CRD comparison will be the only focus.

The variations in precision between the two error control designs has a direct impact on the power of each design to detect non-zero model parameters, i.e. the power associated with rejecting H_0

$$H_0: \ \beta_j=0$$
$$H_1: \ \beta_j \neq 0$$

given H_1 is true.

For both the CRD and BRD, power is calculated for any model coefficient, β_j, by the formula

$$\text{Power}=P[Z>Z_{1-\frac{\alpha}{2}} - \frac{\beta_j}{\sqrt{\text{Var}\ (\hat{\beta}_j)}}] + P[Z<-Z_{1-\frac{\alpha}{2}} - \frac{\beta_j}{\sqrt{\text{Var}\ (\hat{\beta}_j)}}] . \qquad (2.5.3)$$

Differences in powers for the two error designs are linked to the $\mathrm{Var}(\hat{\beta}_j)$ found in expression 2.5.3. Assuming α=.05, comparison of powers for the BRD and CRD are given in Table 2.5.2.

Table 2.5.2 Power for Coefficient Detection for CRD and BRD

Coefficient	CRD	BRD
Intercept	1	1
z_1	.61	.20
z_2	.17	.12
z_3	.97	.47
x_1	.98	1
x_2	.24	.52
x_1x_2	.34	.70
x_2^2	.12	.23
x_1z_1	.66	.97
x_1z_2	.48	.87
x_1z_3	.14	.30
x_2z_1	.30	.64
x_2z_2	.23	.51
x_2z_3	.27	.59
$x_2^2z_2$	.10	.17

These powers provide additional insight into the implications of using a BRD as an alternative to the CRD. For the BRD, the power of detection for whole-plot parameters is reduced by the dual randomization scheme indicating that only more significantly sized whole-plot coefficients will be detected. The CRD has greater power for the detection of each whole-plot parameter. On the other hand, the BRD has a greater sensitivity to find differences from zero for the sub-plot parameters than the CRD.

Because of the imbalance in power found in the BRD, an experimenter must be aware that by implementing the BRD error structure in the experiment

he is sacrificing his ability for detection of significant whole-plot effects. The consequences include possible elimination of non-zero whole-plot terms from the fitted response surface model.

§2.6 Implications of Improper BRD Analysis

The impact of the bi-randomization scheme on power of coefficient detection provides the first indication of the consequences of incorrectly analyzing BRD experimental results as if they were obtained from a CRD. Unfortunately, when proper analysis tools are unavailable, some experimenters may resort to using CRD techniques (OLS methods described in Chapter 1) to estimate and edit a response surface model even though the data were obtained from a BRD. Ignoring the dependent relationship among observations within a whole-plot for the BRD (thus, viewing it as a CRD) can have drastic effects on the final fitted RSM model. The implications include an incorrectly fitted model which in turn can lead to the wrong choice of optimum location and provide misleading information about the nature of the response surface. One must keep in mind that up to this point there is no obvious procedure for the estimation of response surface coefficients and standard errors under the BRD scheme when V is not known.

For example, consider the five variable BRD experiment discussed in §2.5 and assume that the experiment was run in a bi-randomization error format. The experimental results, $\underline{y}$, are simulated using a random error mechanism from model 2.5.2 under the error assumptions of δ_i iid N(0,8) and ϵ_{ij} iid N(0,5). Using the MLE equation 2.4.3 and $\underline{y}$, the fitted model is

$$\begin{aligned}\hat{y}_{ij} = 8.53 &+ 2.13z_1 - 0.99z_2 + 1.62z_3 + 2.03x_1 + 1.10x_2 + 0.48x_1x_2 \\ &+ 1.96x_2^2 + 0.99z_1x_1 - 0.63x_1z_2 + 0.99x_1z_3 - 0.94x_2z_1 + 0.33x_2z_2 \\ &+ 1.09x_2z_3 + 0.97\, x_2^2z_2\end{aligned} \quad (2.6.1)$$

with standard errors of estimated coefficients as given in Table 2.5.1.

Now, suppose that the experimenter erroneously ignores the dependency among responses in a whole-plot and analyzes the BRD experimental results as CRD results. This would likely be due to his lack of knowledge of V and, thus, an unavailability of proper analysis tools. In this situation, the fitted model is identical to that of 2.6.1. This equivalency of OLS and GLS for BRDs will be explored more fully in Chapter 3. Thus, estimating the model with the incorrect MLE equation does not alter the fitted model, but it does generate incorrect standard errors calculations. Consider the CRD column of standard errors in Table 2.5.1. Note these standard errors are correct only if the design were truly run as a CRD, but in this case it was not.

The ramifications of analyzing the data as a CRD lies , in this example, not in the estimated model, but becomes evident during the model editing stage. Using the standard errors given in Table 2.5.1, the following p-values were obtained for the proper BRD analysis and the improper CRD approach:

Table 2.6.1 P-values on Estimated Coefficients for CRD and BRD Analyses

Coefficient	CRD	BRD
Intercept	0.0000*	0.0000*
z_1	0.0000*	0.021*
z_2	0.1359	0.194
z_3	0.0009*	0.062
x_1	0.0000*	0.0000*
x_2	0.418	0.0027*
x_1x_2	0.2236	0.1112
x_2^2	0.038*	0.002*
x_1z_1	0.029*	0.001*
x_1z_2	0.113	0.025*
x_1z_3	0.028*	0.001*
x_2z_1	0.071	0.009*
x_2z_2	0.302	0.202
x_2z_3	0.044*	0.002*
$x_2^2z_2$	0.192	0.079

*indicates term to be left in final model for the respective analysis.

The final fitted model for each analysis is:

CRD:

$$\hat{y}_i = 8.53 + 2.13z_1 + 1.62z_3 + 2.03x_1 + 1.10x_2 + 1.96x_2^2 + 0.99z_1x_1 + 0.99x_1z_3 + 1.09x_2z_3 \quad (2.6.2)$$

BRD:

$$\hat{y}_{ij} = 8.53 + 2.13z_1 + 2.03x_1 + 1.10x_2 + 1.96x_2^2 + 0.99z_1x_1 - 0.63x_1z_2 + 0.99x_1z_3 - 0.94x_2z_1 + 1.09x_2z_3 \quad (2.6.3)$$

The OLS analysis of the BRD data incorrectly finds x_1z_2 and x_2z_1 to be insignificant whereas it erroneously keeps z_3 as significant in the fitted model. These discrepancies are based entirely on the use of incorrect standard errors.

In terms of selection of optimum location for the RSM model, using fitted model 2.6.2 can lead to the less than optimal location due to the elimination/inclusion of significant/insignificant parameter estimates. Considerable variation from optimal location can occur especially if significant/insignificant second order terms are incorrectly deleted/included in the model.

This example illustrates as with all error control designs, the proper accommodation of the randomization structure is essential to insure accurate results for the given experimental data. The BRD error structure is certainly no exception.

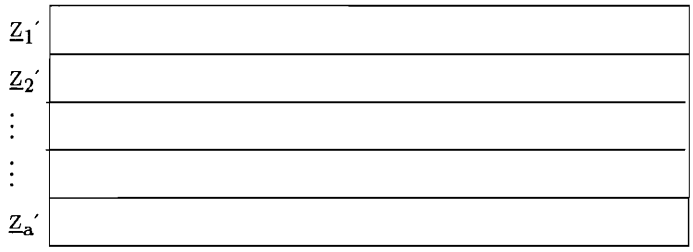
CHAPTER 3

CROSSED BI-RANDOMIZATION DESIGNS

Response surface analysis with the use of bi-randomization designs requires that many changes be made in traditional response surface analyses. In the previous chapter, the general bi-randomization design set-up and analysis using GLS was outlined. In this chapter, a more thorough investigation will be done into the analysis of the crossed bi-randomization design. Estimation techniques for both model coefficients and error variances will be derived along with the development of correct model screening procedures. Note that this chapter will deal only with the crossed BRD. The next chapter will explore the non-crossed design with alternative analyses developed to accommodate the non-crossed BRD structure.

§3.1 **Definition and Design Set-up**

In §2.2 the general structure of a crossed BRD was illustrated using a 3^2 full factorial response surface design. Formally, experimentation under a crossed BRD structure begins as was described in Chapter 2 with the identification of the classes of whole-plot and sub-plot variables of interest in the chosen response surface design. The level combinations of the whole-plot variables found in the treatment design constitute the "a" levels of the whole-plot factor and once again will be denoted $\underline{z}_1'$, $\underline{z}_2'$, . . ., $\underline{z}_a'$. These levels are randomized to whole-plot EUs as described in the general BRD situation, i.e.



Similarly, the "b" unique level combinations of the sub-plot variables in the design form the levels of the BRD sub-plot factor. These levels are denoted $\underline{x}_1'$, $\underline{x}_2'$, . . ., $\underline{x}_b'$. Unlike in the general BRD, however, the same "b" levels are associated with each level combination $\underline{z}_i'$ (i=1,...,a) in the original response surface design. There is a "crossing" of whole-plot and sub-plot factor levels. Thus, each whole-plot EU is subdivided into "b" sub-plots, and each of the "b" levels of the sub-plot factor are randomly applied to these units within a given whole-plot. The final crossed BRD has the following treatment and error control structure:

$\underline{z}_1'$	$\underline{x}_1'$	$\underline{x}_2'$	$\cdots$	$\underline{x}_b'$
$\underline{z}_2'$	$\underline{x}_1'$	$\underline{x}_2'$	$\cdots$	$\underline{x}_b'$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
$\underline{z}_a'$	$\underline{x}_1'$	$\underline{x}_2'$	$\cdots$	$\underline{x}_b'$

Recall that the general BRD definition given in Chapter 2 did not require the same sub-plot factor levels to be randomized within each whole-plot.

By examining the crossed structure above, one may recognize this design to be familiar. It has the same treatment and randomization structure as the split plot treatment comparison design. The major difference between the split plot and crossed bi-randomization designs lies in the model and analysis application. The bi-randomization design discussed here is intended for use in regression model building techniques found in RSM whereas the split plot design has its roots

primarily in treatment comparisons. Because of this, the crossed BRD has no requirement of replication whereas the split plot design requires replication in order to compute estimates of the error variances for testing purposes. The split plot design typically involves replication of the entire structure given above although each replication need not be identical.

In bi-randomization experimentation certain classes of traditional response surface designs fall naturally into the crossed BRD definition. Any full factorial design including the more utilized two and three-level designs are always classified as crossed under a bi-randomization error structure. This is due directly to the "crossing" of levels which is used to generate these designs in the first place. Fractional factorials, however, do not always fit the crossed BRD criteria. Certain fractions are indeed "crossed" BRDs, while others are actually classified as "non-crossed" bi-randomization designs.

Model 2.4.1 remains essentially unchanged under the crossed treatment structure. However, $\underline{x}_{ij}'$ can no longer be defined as the $\underline{x}$ treatment combination applied to the j^{th} sub-plot within the i^{th} whole-plot but instead is the j^{th} unique $\underline{x}$ combination applied within the i^{th} whole-plot where i=1,2,...a and j=1,2,...,b not b_j, as before.. The n*1 vector of responses, $\underline{y}$, has the distribution:

$$\underline{y} \sim N(\underline{\mu}, V)$$

$$\text{where } \underline{\mu} = \beta_0 \underline{1} + Z\underline{\gamma} + X\underline{\alpha} + Z' \triangle X$$

and

$$V = \begin{bmatrix} T & 0 & \cdots & 0 \\ 0 & T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T \end{bmatrix}$$

where

$$T = \sigma_\epsilon^2 I_{(b*b)} + \sigma_\delta^2 \underline{1}_{(b*1)} \underline{1}'_{(1*b)}.$$

§3.2 Coefficient Estimation and Equivalency of Ordinary and Generalized Least Squares

As was mentioned in §2.4, if model 2.4.1 is rewritten as

$$\underline{y}=X^*\underline{\beta} + \underline{\delta} + \underline{\epsilon} \tag{3.2.1}$$

where $X^*\underline{\beta} = \beta_0\underline{1} + Z\underline{\gamma} + X\underline{\alpha} + Z'\triangle X$

then the maximum likelihood estimator (MLE) of $\hat{\underline{\beta}}$ with V known is given by

$$\hat{\underline{\beta}} = (X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1}\underline{y} \tag{3.2.2}$$

and

$$\mathrm{Var}(\hat{\underline{\beta}}) = (X^{*\prime}V^{-1}X^*)^{-1}. \tag{3.2.3}$$

Hence, estimation of coefficients is complicated by the unknown error variances, ${\sigma_\delta}^2$ and ${\sigma_\epsilon}^2$, which are components of the V matrix. Thus, for discussion of the analysis, estimators for these variances must be derived and calculated. Fortunately, however, due to the special treatment structure of the crossed BRD model matrix and the structure of V, this estimation task is greatly simplified through the equivalence of ordinary and generalized least squares estimating equations.

Theorem 3.2.1: Let $\mathfrak{D}$ be a design with a crossed treatment structure,

$\underline{z}_1'$	$\underline{x}_1'$	$\underline{x}_2'$	$\cdots$	$\underline{x}_b'$
$\underline{z}_2'$	$\underline{x}_1'$	$\underline{x}_2'$	$\cdots$	$\underline{x}_b'$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
$\underline{z}_a'$	$\underline{x}_1'$	$\underline{x}_2'$	$\cdots$	$\underline{x}_b'$

(3.2.4)

which is run under a bi-randomization experimental format. Here $\underline{z}_i'$ is a combination of whole-plot variables and $\underline{x}_j'$ a combination of sub-plot variables.

Let $\underline{y}=X^*\underline{\beta} + \underline{\epsilon}^*$ where $\underline{\epsilon}^*=\underline{\delta}+\underline{\epsilon}$; $\underline{\epsilon}^* \sim N(\underline{0}, V)$ where

$$V=\begin{bmatrix} T & 0 & \cdots & 0 \\ 0 & T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T \end{bmatrix}$$

and

$$T=\sigma_\epsilon^2 I_{(b*b)} + \sigma_\delta^2 \underline{1}_{(b*1)}\, \underline{1}'_{(1*b)}.$$

Then for a polynomial model of any order,

$$(X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1}\underline{y}=(X^{*\prime}X^*)^{-1}X^{*\prime}\underline{y}$$

(OLS=GLS) provided the following hold:

1. the intercept is included in the model (or all sub-plot variable model terms sum to zero over each whole-plot)

2. for each whole-plot*sub-plot variable interaction term included in the model, the respective whole-plot term is also found in the model (or the sum over each whole-plot of the levels of sub-plot variable term found in the interaction equals zero).

Proof: The proof of this theorem will utilize the following theorem and corollary which may be found in Graybill [1976]:

> Theorem: Consider the general linear model $\underline{y}=X\underline{\beta} + \underline{\epsilon}$, $\underline{\epsilon} \sim N(\underline{0}, \Sigma)$. The UMVU estimator of $\underline{\beta}$ is given by $(X'X)^{-1}X'\underline{y}$ if and only if there exists a p*p nonsingular matrix F such that
>
> $$\Sigma X=XF$$
>
> or equivalently,

$$X'\Sigma^{-1}=GX' \text{ where } G=F^{-1}.$$

Corollary: Consider the general linear model $\underline{y}=X\underline{\beta}+\underline{\epsilon}$, $\underline{\epsilon}\sim N(\underline{0},\Sigma)$ where X includes the intercept. If $\Sigma=\sigma^2(1-\rho)I+\sigma^2\rho J$ where $J=\underline{1}\,\underline{1}'$ and $\frac{-1}{n-1}<\rho<1$ then

$$(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underline{y}=(X'X)^{-1}X'\underline{y}\ .$$

The BRD model matrix in the original theorem can be partitioned as $X^{*'}=[X_1'|X_2'|...|X_a']$ where X_i' is a (p+1)*b matrix that corresponds directly to the i^{th} whole-plot. In addition,

$$V^{-1}=\begin{bmatrix} T^{-1} & 0 & \cdots & 0 \\ 0 & T^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T^{-1} \end{bmatrix}.$$

Using these expressions, $(X^{*'}V^{-1}X^{*})^{-1}X^{*'}V^{-1}$ can then be written as

$$\left[\sum_{i=1}^{a}[X_i'T^{-1}X_i]\right]^{-1}[X_1'T^{-1}\mid X_2'T^{-1}\mid ...\mid X_a'T^{-1}]. \tag{3.2.5}$$

Each T found in V can be reexpressed as

$$T=(\sigma_\delta^2+\sigma_\epsilon^2)[1-\frac{\sigma_\delta^2}{\sigma_\epsilon^2+\sigma_\delta^2}]I+\ (\sigma_\delta^2+\sigma_\epsilon^2)[\frac{\sigma_\delta^2}{\sigma_\epsilon^2+\sigma_\delta^2}]J.$$

Thus, using the above corollary, expression 3.2.5 becomes

$$\left[\sum_{i=1}^{a}[G_iX_i'X_i]\right]^{-1}[G_1X_1'\mid G_2X_2'\mid ...\mid G_aX_a'] \tag{3.2.6}$$

where $X_i'T^{-1}=G_iX_i'$ where G_i is a (p+1)*(p+1) matrix.

For the crossed bi-randomization structure of 3.2.4,

$$G_i=\begin{bmatrix} 1\text{-}cb & \vdots & \underline{0}' & \vdots & \underline{0}' & \vdots & \underline{0}' \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \underline{0} & \vdots & (1\text{-}cb)I_{k*k} & \vdots & 0 & \vdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -cs_1 & & & \vdots & & \vdots & \\ -cs_2 & \vdots & & \vdots & & \vdots & \\ \vdots & \vdots & 0 & \vdots & I_{l*l} & \vdots & 0 \\ -cs_l & \vdots & & \vdots & & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \underline{0} & \vdots & K & \vdots & 0 & \vdots & I_{p-k-l} \end{bmatrix} \qquad (3.2.7)$$

where $c=\dfrac{\sigma_\delta^2}{b\sigma_\delta^2+\sigma_\epsilon^2}$

b=number of sub-plots within a whole-plot

$s_m=\sum_{\text{whole-plot}} x_m$ where x_m is m^{th} sub-plot model term

K=matrix which weights the appropriate whole-plot term found in the respective whole-plot*sub-plot interaction by a constant, $-c\sum_{wp} x_m$ or $(-cs_m)$ where x_m is the sub-plot term found in the interaction.

Under the crossed structure of 3.2.4, each s_m remains constant across whole-plots. Thus, $G_i=G\ \forall\ i$, and 3.2.5 is now equal to

$$\left[\sum_{i=1}^{a} G[X_i'X_i]\right]^{-1}G[X_1' \mid X_2' \mid \ldots \mid X_a']$$

$$= (X^{*\prime}X^*)^{-1}G^{-1}GX^{*\prime}$$

$$= (X^{*\prime}X^*)^{-1}X^{*\prime}.$$

Therefore, for any order model under crossed BR structure, the OLS estimation equations for the model coefficients, $\hat{\beta} = (X^{*\prime}X^*)^{-1}X^{*\prime}\underline{y}$ are equivalent to the GLS equations,

$$\hat{\underline{\beta}} = (X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1}\underline{y}\ .\ \square$$

The importance of the above result lies in the fact that knowledge of the error variances, ${\sigma_\delta}^2$ and ${\sigma_\epsilon}^2$, is no longer essential for model estimation. However, even though V does not influence model estimation in the crossed BRDs, the following equality still holds for the model of 3.2.1:

$$\mathrm{Var}(\hat{\underline{\beta}}) = (\mathrm{X}^*\mathrm{V}^{-1}\mathrm{X}^*)^{-1}$$

This implies the necessity of either knowing V or properly estimating V for variable screening and model editing of the response surface model. Since it is uncommon to actually know V, estimation techniques must be developed. Estimation of V is obtained through derivations of estimators of each individual error variance, ${\sigma_\delta}^2$ and ${\sigma_\epsilon}^2$. These can then be substituted into V to form $\widehat{\mathrm{V}}$.

§3.3 Estimation of Error Variances, ${\sigma_\epsilon}^2$ and ${\sigma_\delta}^2$

Two methods will be presented to accomplish the estimation of the error variances in the crossed BRD. In the end each method produces identical estimators for both variance components, but each incorporates a different approach to the bi-randomization design and can provide some additional insight into the bi-randomization structure. The first method will be called the two model approach, while the second will be referred to as the lack of fit approach.

§3.3.1 Method 1-Two Model Approach

The first method for deriving estimators of the error variances focuses on the basic structure of the bi-randomization design and uses two models to reflect the different randomization schemes.

Model 1:

The first model originates from the consideration of whole-plots as the primary experimental units. Recall that the whole-plot error variance, ${\sigma_\delta}^2$, arises from the natural variability among these units. In this setting, the observations

taken on sub-plots within each whole-plot may be viewed as "subsampling" and thus, observations within each whole-plot are averaged to create an overall whole-plot response. The crossed bi-randomization response surface model now takes the form:

$$\bar{y}_{i.}=\beta_0 + \underline{\gamma}'\underline{z}_i^d + \underline{\alpha}'\,\underline{\bar{x}}_{.}^{\,d} + \underline{z}_i'^d \Delta\, \underline{\bar{x}}_{.}^{\,d} + \delta_i + \bar{\epsilon}_{i.} \qquad (3.3.1)$$

where $\bar{y}_{i.}$ is average response on i^{th} whole-plot, $\underline{z}_i^d$ is the model power vector [Myers, 1976] associated with the i^{th} whole-plot combination, and $\underline{x}_j^d$ is the power vector associated with the j^{th} sub-plot variable combination. Here d reflects the order of the model. Recall, by the structure of the crossed BRD identical levels of the sub-plot factor are found within each whole-plot, thus making the average of the $\underline{x}_j^d$'s within each whole-plot, $\underline{\bar{x}}_{.}^{\,d}$, a constant. As a result, 3.3.1 can be written as

$$\bar{y}_{i.}=\beta_0^* + (\underline{\gamma}' + \underline{\bar{x}}_{.}'^d \Delta')\underline{z}_i^d + \delta_i + \bar{\epsilon}_{i.} \qquad (3.3.2)$$

where $\beta_0^* = \beta_0 + \underline{\alpha}'\underline{\bar{x}}_{.}^{\,d}$

or equivalently,

$$\bar{y}_{i.}=\underline{\gamma}^{*\prime}\underline{z}_i^{+d} + \delta_i + \bar{\epsilon}_{i.} \qquad (3.3.3)$$

where $\underline{\gamma}^{*\prime}= [\beta_0^*, \underline{\gamma}' + \underline{\bar{x}}_{.}'^d \Delta\,']$ and $\underline{z}_i^{+\prime d} = [1, \underline{z}_i'^d]$.

Based upon 3.3.3, the distribution of the a*1 vector of $\bar{y}_{i.}$ is given by

$$\underline{\bar{y}} \sim N(Z^+\underline{\gamma}^*, \frac{b\sigma_\delta^2 + \sigma_\epsilon^2}{b} I).$$

where $\underline{\bar{y}} = \begin{bmatrix} \bar{y}_{1.} \\ \vdots \\ \bar{y}_{a.} \end{bmatrix}$ and Z^+ is the model matrix associated with $\underline{z}_i^{+\prime d}$, i=1,...,a.

The crossed BRD response surface model has been transformed into a model with iid normal errors. Based upon this error and model assumption, the MLE for $\underset{\sim}{\gamma}^*$ can be found using OLS equations, Graybill [1976]. Model 3.3.2 was not developed with the intention for use in model estimation of 3.2.1. See equation 3.2.2 for its estimation. Under these assumptions, an unbiased estimator for the error variance of 3.3.3,

$$\frac{b\sigma_\delta^2 + \sigma_\epsilon^2}{b},$$

is given by

$$\frac{\underset{\sim}{Y}'[I\text{-}Z^+(Z^{+\prime}Z^+)^{-1}Z^{+\prime}]\,\underset{\sim}{Y}}{a\text{-}(k+1)} = MSE1 \qquad (3.3.4)$$

where a=number of whole-plots and k=number of whole-plot terms modeled not including the intercept. Because of the normality of errors in 3.3.3, expression 3.3.4 is a maximum likelihood estimator adjusted to achieve unbiasedness.

Model 2:

The second model is formulated by viewing the sub-plots as the primary experimental units in the design. In this particular setting, the whole-plots are blocks used to account for the error introduced by whole-plot differences. The model is now

$$y_{ij} = \beta_0+\beta_1 w_{1,i} +\beta_2 w_{2,i}+\ldots+\beta_{a-1} w_{a-1,i}+\underset{\sim}{\alpha}'\underset{\sim}{x}_j^d+\underset{\sim}{z}_i'^d \triangle \underset{\sim}{x}_j^d + \epsilon_{ij} \qquad (3.3.5)$$
$$i=1,2,\ldots,a;\ j=1,2,\ldots,b$$

where $w_{ri}=\begin{cases} 1 \text{ if } y_{ij} \text{ in } i^{th} \text{ whole-plot} \\ 0 \text{ otherwise} \end{cases}$ r=1,2,...,a-1.

The w_{ri} component represents the indicator variable associated with the i^{th} whole-plot or "block". The inclusion of the blocks is necessary to model the appropriate sum of squares and degrees of freedom associated with the modeled whole-plot terms and error found in the original response surface model, 3.2.1. This ensures

the proper remaining degrees of freedom and sum of squares for estimation of the sub-plot error variance. A simplified form of 3.3.5 is

$$\underline{y}=X^{**}\underline{\theta} + \underline{\epsilon} \tag{3.3.6}$$

where $\underline{\theta}$ is an (a+(p-k))*1 vector of unknown coefficients and X^{**} is an (n*(a+(p−k))) model matrix of 0's and 1's for blocks, $\underline{x}_j^d$'s, and $\underline{z}_i^d$'s. Based upon 3.3.6, an unbiased estimator for σ_ϵ^2 is

$$\frac{\underline{y}'[I-X^{**}(X^{**\prime}X^{**})^{-1}X^{**\prime}]\,\underline{y}}{n-p^*} = MSE2 \tag{3.3.7}$$

where p^*=a+(p-k). Finally, an unbiased estimator for σ_δ^2 can be found by taking a linear combination of the two previous estimators and creating a synthesized mean square:

$$\hat{\sigma}_\delta^2 = \frac{(bMSE1-MSE2)}{b}\,. \tag{3.3.8}$$

§3.3.2 Properties of Estimated Error Variances

Now that estimators have been derived for the respective error variances, properties for these estimators can be investigated. The first property of unbiasedness of the estimators has already been discussed. In addition to unbiasedness, variance properties can also be examined. The variances can be derived by invoking distributional properties of the mean squares. For MSE1 in (3.3.4),

$$\frac{(a-(k+1))*MSE1}{\frac{\sigma_\epsilon^2 + b\sigma_\delta^2}{b}} \sim \chi^2_{(a-(k+1))}\,.$$

Consequently,

$$Var[MSE1] = Var\left[\frac{\hat{\sigma_\epsilon^2 + b\sigma_\delta^2}}{b}\right] = \frac{2[\sigma_\epsilon^2 + b\sigma_\delta^2]^2}{b^2*(a-(k+1))}\,. \tag{3.3.9}$$

Similarly, for the estimator of σ_ϵ^2,

$$\frac{(\text{n-p}^*)*\text{MSE2}}{\sigma_\epsilon^2} \sim \chi^2_{(\text{n-p}^*)}$$

and

$$\text{Var[MSE2]} = \text{Var}[\hat{\sigma}_\epsilon^2] = 2\frac{[\sigma_\epsilon^2]^2}{(\text{n-p}^*)}. \tag{3.3.10}$$

The estimator for σ_δ^2, unlike the other two estimators, is a linear combination of MSE1 and MSE2 and, thus, a linear combination of two independent chi-square random variables. That is,

$$\hat{\sigma}_\delta^2 = \frac{\sigma_\epsilon^2 + \text{b}\sigma_\delta^2}{\frac{\text{b}}{\text{a-(k+1)}}}\left[\frac{\text{a-(k+1)}*\text{MSE1}}{\sigma_\epsilon^2 + \text{b}\sigma_\delta^2}\right] - \frac{\sigma_\epsilon^2}{\text{b(n-p}^*)}\left[\frac{(\text{n-p}^*)\text{MSE2}}{\sigma_\epsilon^2}\right]$$

$$= c_1\chi^2_{(\text{a-(k+1)})} + c_2\chi^2_{(\text{n-p}^*)}.$$

An approximate distribution for the estimator of σ_δ^2 can be found using Satterthwaite's procedure [Satterthwaite , 1946].

As a result,

$$\nu\frac{\hat{\sigma}_\delta^2}{\sigma_\delta^2} \sim \text{approximately } \chi^2_{(\nu)} \text{ where } \nu = \frac{\left[\frac{\sigma_\epsilon^2 + \text{b}\sigma_\delta^2}{\text{b}} - \frac{\sigma_\epsilon^2}{\text{b}}\right]^2}{\frac{\left[\frac{\sigma_\epsilon^2 + \text{b}\sigma_\delta^2}{\text{b}}\right]^2}{\text{a-(k+1)}} + \frac{\left[\frac{\sigma_\epsilon^2}{\text{b}}\right]^2}{\text{n-p}^*}}.$$

Using this approximate distribution,

$$\text{Var}[\hat{\sigma}_\delta^2] = 2\frac{\left[\sigma_\delta^2\right]^2}{\nu}. \tag{3.3.11}$$

§3.3.3 Method 2-Lack of Fit Approach

Throughout this discussion of bi-randomization designs, it is assumed that, unlike the split plot designs, there is no replication of whole-plot factor levels. Recall, the primary purpose of the bi-randomization design is to prevent the application and reapplication of the same levels of the difficult/costly to control factor combinations. Because of this, the whole-plot error variance must be derived exclusively from lack of fit contributed by whole-plot variable terms not included in the model. For estimation of the sub-plot variance component, both lack of fit sum of squares from sub-plot variables and whole-plot*sub-plot interactions and any available replication sum of squares will be pooled. To determine the appropriate sum of squares for each respective error, a model slightly altered from the response surface model 2.4.1 must be considered. The new model is composed of the original model terms but is augmented with higher order whole-plot variable terms that account for the remaining (a-(k+1)) degrees of freedom available for whole-plots. The model becomes

$$y_{ij} = \beta_0 + \underline{\gamma}'\underline{z}_i^{\,d} + \underline{\rho}'\underline{z}_i^{\,*d} + \underline{\alpha}'\underline{x}_j^{\,d} + \underline{z}_i^{\,\prime d} \triangle \underline{x}_j^{\,d} + \epsilon_{ij} \qquad (3.3.12)$$
$$i=1,2...,a;\ j=1,2,...,b$$

where ϵ_{ij} iid $N(0,\ \sigma_\epsilon^{\,2})$ and $\underline{z}^{*d}$ is an (a-(k+1))*1 vector of whole-plot terms not originally modeled, thus saturating the available whole-plot degrees of freedom. The term $\underline{\rho}'\underline{z}_i^{\,*d}$ can be viewed as a model of δ_i. The saturation of whole-plot degrees of freedom in the model accomplishes the same goal as the modeling of blocks earlier in Method 1. It removes the degrees of freedom and sum of squares necessary to leave a correct lack of fit/replication remainder for the calculation of the sub-plot error.

The first error variance combination, $\sigma_\epsilon^{\,2} + b\sigma_\delta^{\,2}$, is estimated by summing the type 1 lack of fit sum of squares for the augmented whole-plot terms and dividing by the respective a-(k+1) degrees of freedom. Based on model 3.3.12 the lack of fit sum of squares can be written as

$$SS(\underline{\varrho}|\underline{\gamma})=SS(\underline{\varrho},\underline{\gamma})\text{-}SS(\underline{\gamma})$$
$$=\underline{y}'[T[T'T]^{-1}T']\underline{y} - \underline{y}'[Z[Z'Z]^{-1}Z']\underline{y} \qquad (3.3.13)$$

where Z^* is the model matrix associated with $\underline{\varrho}'\underline{z}_i^{*d}$ and $T=[Z|Z^*]$.
Since every design point within a whole-plot has the same $\underline{z}_i$ and $\underline{z}_i^*$ values, expression 3.3.13 is equivalent to

$$b\overline{\underline{y}}'[T_r[T_r'T_r]^{-1}T_r']\overline{\underline{y}} - b\overline{\underline{y}}'[Z_r[Z_r'Z_r]^{-1}Z_r']\overline{\underline{y}} \qquad (3.3.14)$$

where T_r and Z_r are reduced versions of T and R with the replicated whole-plot combinations having been removed, leaving only "a" unique whole-plot factor levels. The second term in expression 3.3.14 is equivalent to b*SS(Model1) found in Method 1 above. In addition, the first term simplifies to $b\overline{\underline{y}}'\overline{\underline{y}}$ or b*(SSTotal) from Model 1 in Method 1 since $[T_r[T_r'T_r]^{-1}T_r'] = I$. Therefore, equation 3.3.14 is equivalent to b*(SSE1) from Method 1 and consequently,

$$E\left[\frac{SS(\underline{\varrho}|\underline{\gamma})}{a\text{-}k\text{-}1}\right] = \sigma_\epsilon^2 + b\sigma_\delta^2 .$$

Similarly, $\hat{\sigma}_\epsilon^2$ is found by dividing the remaining lack of fit sum of squares of non modeled sub-plot and whole-plot by sub-plot terms from 3.3.12 and any replication sum of squares by the remaining n-p^* df. This corresponds directly to the derivation of Model 2 in Method 1 since $\beta_0+\underline{\gamma}'\underline{z}_i^d + \underline{\varrho}'\underline{z}_i^{*d}$ models the same sum of squares as the w_{ri}s. A synthesized mean square then appropriately produces an estimator of the whole-plot error variance, σ_δ^2. This method provides identical estimators to those described in the two model approach with the same variances as derived earlier.

Once $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_\delta^2$ are calculated from the experimental data, these estimators can be substituted into V to create an appropriate estimator, $\hat{V}$. The matrix $\hat{V}$ can then in turn be used to estimate $Var(\hat{\underline{\beta}})$, 3.2.3, i.e.

$$\hat{Var}(\hat{\underline{\beta}})=(X^{*'}\hat{V}^{-1}X^*)^{-1}. \qquad (3.3.15)$$

§3.4 Variable Screening

Methods have been given for both the estimation of model parameters and the bi-randomization error variances in the crossed BRD. The next step in RSM analysis is model editing. Ratios of estimated model coefficients and their respective standard errors obtained from $(X^{*'}\hat{V}^{-1}X^{*})^{-1}$ generate appropriate t-tests. One must be careful, however, in the degrees of freedom associated with these tests because of the presence of two error variance components. Expressions for the variances of the estimated model coefficients as found below indicate the degrees of freedom necessary for proper testing.

§3.4.1 Variances of Estimated Regression Coefficients

Assuming V is known, individual variance and covariance expressions can be derived from $(X^{*'}V^{-1}X^{*})^{-1}$ for both whole-plot and sub-plot coefficients. Based upon these, proper tests can be developed. We begin with variances of estimated whole-plot coefficients.

For a given estimated whole-plot coefficient, $\hat{\gamma}_k$, its variance can be expressed as

$$\mathrm{Var}[\hat{\gamma}_k]=c_{k,k}[\sigma_\epsilon^2 + b\sigma_\delta^2] + \sum_{m}[c_{k,km}(\sigma_\delta^2[\sum_{wp} x_m])] \qquad (3.4.1)$$

where x_m = m^{th} sub-plot model term which interacts with k^{th} whole-plot model term

$c_{k,k}$= k^{th} diagonal element of $[X^{*'}X^{*}]^{-1}$ corresponding to γ_k

$c_{k,km}$= off-diagonal element of $[X^{*'}X^{*}]^{-1}$ corresponding to the the k^{th} whole-plot term and the interaction of the k^{th} whole-plot term and the m^{th} sub-plot term.

Note: the intercept is considered a whole-plot coefficient and pure sub-plot terms are viewed as interactions with the intercept for calculation purposes of 3.4.1.

Now consider an estimated sub-plot term coefficient, $\hat{\alpha}_m$. Its respective variance is given by

$$\mathrm{Var}[\hat{\alpha}_m]=c_{m,m}\sigma_\epsilon^2 \tag{3.4.2}$$

where $c_{m,m}$ = m^{th} diagonal element of $[X^{*\prime}X^*]^{-1}$ corresponding to α_m.

§3.4.2 Covariances Between Estimated Regression Coefficients

Let $\hat{\gamma}_k$ and $\hat{\gamma}_l$ correspond to the k^{th} and l^{th} whole-plot term's estimated coefficients. The covariance between these two estimates is

$$\mathrm{Cov}[\hat{\gamma}_k\ ,\ \hat{\gamma}_l]=\ c_{k,l}[\sigma_\epsilon^2+b\sigma_\delta^2]+\sum_m[c_{k,lm}(\sigma_\delta^2[\sum_{wp}x_m])] \tag{3.4.3}$$

where x_m = m^{th} sub-plot model term which interacts with l^{th} whole-plot model term

$c_{k,l}$ = off-diagonal element of $[X^{*\prime}X^*]^{-1}$ corresponding to γ_k and γ_l

$c_{k,lm}$= off-diagonal element of $[X^{*\prime}X^*]^{-1}$ corresponding to the the k^{th} whole-plot term and the interaction of the l^{th} whole-plot term and the m^{th} sub-plot term.

Similarly, let $\hat{\alpha}_m$ and $\hat{\alpha}_n$ correspond to the m^{th} and n^{th} sub-plot term's estimated coefficients. The covariance between these two estimates is given by

$$\mathrm{Cov}[\hat{\alpha}_m\ ,\ \hat{\alpha}_n]=\ c_{m,n}[\sigma_\epsilon^2] \tag{3.4.4}$$

where $c_{m,n}$ = off-diagonal element of $[X^{*\prime}X^*]^{-1}$ corresponding to α_m and α_n.

Finally, let $\hat{\gamma}_k$ and $\hat{\alpha}_m$ correspond to the k^{th} whole-plot term's estimated coefficient and the m^{th} sub-plot term's estimated coefficient, respectively. The covariance between these two estimators is

$$\mathrm{Cov}[\hat{\gamma}_k\ ,\ \hat{\alpha}_m]=\ c_{m,n}[\sigma_\epsilon^2] \tag{3.4.4}$$

where $c_{m,n}$ = off-diagonal element of $[X^*'X^*]^{-1}$ corresponding to α_m and α_n.

Notice, the expression for the variance of an estimated whole-plot parameter is not a function of purely $(\sigma_\epsilon^2 + b\sigma_\delta^2)$ unless interacting sub-plot terms sum to zero within each whole-plot, i.e. $\sum_{wp} x_m = 0$. If the sub-plot terms which interact with the k^{th} whole-plot term are not orthogonal to that whole-plot term within a whole-plot, then a synthesized mean square must be used in the denominator of the t-test on the whole-plot parameter, Satterthwaite's procedure employed, and degrees of freedom altered.

§3.4.3 Testing on Regression Coefficients

Based upon the variances of the derived estimators of the error variances given in §3.3.2 and the expressions for the variances of the estimated coefficients, appropriate t-tests can be formulated for model editing. We first consider tests for whole-plot coefficients.

An appropriate test of hypothesis for significance of a whole-plot model coefficient,

$$H_0: \ \gamma_k = 0 \ \text{ vs. } \ H_1: \gamma_k \neq 0,$$

is derived from the ratio of a standard normal random variable and an independent χ^2. Using 3.4.1,

$$\frac{\hat{\gamma}_k - 0}{\left[c_{k,k}[\sigma_\epsilon^2 + b\sigma_\delta^2] + \sum_m [c_{k,km}(\sigma_\delta^2[\sum_{wp} x_m])]\right]^{\frac{1}{2}}} \sim N(0,1). \qquad (3.4.6)$$

Now let $s^2 = c_{k,k}[bMSE1] + \sum_m [c_{k,km}(\frac{bMSE1 - MSE2}{b}[\sum_{wp} x_m])]$

$$= (r+t)bMSE1\text{-}tMSE2$$

where $r=c_{k,k}$ and $t=\sum_{m}\frac{1}{b}[c_{k,km}[\sum_{wp}x_m]]$. It follows that

$$E[s^2]=c_{k,k}[\sigma_\epsilon^2+b\sigma_\delta^2]+\sum_{m}[c_{k,km}(\sigma_\delta^2[\sum_{wp}x_m])].$$

Using the χ^2 distributions developed for MSE1 and MSE2 in §3.3, Satterthwaite's approximation defines

$$\frac{\rho s^2}{c_{k,k}[\sigma_\epsilon^2+b\sigma_\delta^2]+\sum_{m}[c_{k,km}(\sigma_\delta^2[\sum_{wp}x_m])]} \dot{\sim} \chi^2_{(\nu)} \qquad (3.4.7)$$

$$\text{where } \nu=\frac{\left[(r+t)(\sigma_\epsilon^2+b\sigma_\delta^2)\text{-}t\sigma_\epsilon^2\right]^2}{\frac{\left[(r+t)(\sigma_\epsilon^2+b\sigma_\delta^2)\right]^2}{a\text{-}(k+1)}+\frac{\left[\frac{\sigma_\epsilon^2}{b}\right]^2}{n\text{-}p^*}}.$$

Finally, an appropriate t-test can be formulated from expressions 3.4.6 and 3.4.7. Under the assumption of H_0 true

$$\frac{\hat{\gamma}_k}{s} \dot{\sim} t_{(\nu)} \text{ where } s=\sqrt{s^2}. \qquad (3.4.8)$$

An estimate of ν may be found be replacing unknown error variances with their respective estimates.

To test the hypothesis H_0: $\alpha_m = 0$ vs. H_1: $\alpha_m \neq 0$ where α_m corresponds to a sub-plot coefficient, we take advantage of the fact that under H_0

$$\frac{\hat{\alpha}_m}{[MSE2\ c_{m,m}]^{\frac{1}{2}}} \sim t_{(n-p^*)}. \qquad (3.4.9)$$

Note, Satterthwaite's approximation is unnecessary for testing sub-plot coefficients because the corresponding variances are functions of only σ_ϵ^2.

To maintain the validity of these t-tests the numerator and denominator must be independent. First consider the testing of a sub-plot coefficient, i.e. 3..4.9. Instead of considering just one parameter, consider the entire vector of sub-plot coefficients, $\underline{\alpha}$. If the model matrix X^* is partitioned into $[X_1|X_2]$ where X_1 is the model matrix associated with the whole-plot variables and X_2 is the matrix associated with the sub-plot variables then

$$\hat{\underline{\alpha}}=C\underline{y} \tag{3.4.10}$$

where C=

$$\left[(X_2'X_2)^{-1}X_2'X_1[(X_1'X_1)^{-1}-X_1'X_2(X_2'X_2)^{-1}X_2'X_1]^{-1}|[X_2'X_2-X_2'X_1(X_1'X_1)^{-1}X_1'X_2]^{-1}X_2'\right]X^{*'}$$

Similarly, the denominator, can be written as $\underline{y}'A\underline{y}$ with

$$A= c[I-X^{**}(X^{**'}X^{**})^{-1}X^{**'}] \tag{3.4.11}$$

where c is a constant and X^{**} is the model matrix X^* augmented with nonmodeled whole-plot terms to saturate whole-plot degrees of freedom.

In order for 3.4.10 and 3.4.11 to be independent,

$$CVA=0$$

where V=var($\underline{y}$).

For X^{**},

$$X^{**'}[X^{**}(X^{**'}X^{**})^{-1}X^{**'}]=X^{**'}$$

and thus,

$$X^{**'}[I-X^{**}(X^{**'}X^{**})^{-1}X^{**'}]=0. \tag{3.4.12}$$

Now partition X^{**} into $[X^*|Z_s]$ where Z_s is the matrix of whole-plot terms

saturating the whole-plot degrees of freedom. From 3.4.12,

$$X^{*\prime}[I-X^{**}(X^{**\prime}X^{**})^{-1}X^{**\prime}]=0. \qquad (3.4.13)$$

Thus,

$$CVA=0 \text{ since } X^{*\prime}V=G^{*}X^{*\prime}$$

and independence holds. A similar argument may be used to verify the independence for the whole-plot t-tests.

Based on the variable screening results, a final response surface model for the crossed BRD may be fit, and the other various RSM techniques may then be used to achieve the final goals of region exploration and optimization.

CHAPTER 4

NON-CROSSED BI-RANDOMIZATION DESIGNS

In the previous chapter, the first class of bi-randomization designs, the crossed BRDs, was defined and estimation procedures were outlined for both the model coefficients and the error variances. Based upon these estimators, techniques were developed for proper variable screening tests. In this chapter, the second class of bi-randomization designs, the non-crossed BRDs, will be defined along with respective estimation methods and analysis techniques. Unfortunately, for this class of designs, the equivalence of OLS and GLS which simplified the crossed BRD analysis in Chapter 3 does not universally hold for the non-crossed designs. Thus, the investigation into estimation and analysis will not be as straightforward as that described in Chapter 3.

§4.1 Definition and Design Set-up

In Chapter 2, the example of a two variable central composite design illustrated the concept of a non-crossed bi-randomization design. The design is created by randomizing the sub-plot factor levels associated with each whole-plot factor level in the given response surface design within the respective whole-plots giving the following BRD form:

$\underline{z}_1'$	$\underline{x}_{1(1)}'$	$\underline{x}_{1(2)}'$	$\cdots$	$\underline{x}_{1(b_1)}'$
$\underline{z}_2'$	$\underline{x}_{2(1)}'$	$\underline{x}_{2(2)}'$	$\cdots$	$\underline{x}_{2(b_2)}'$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
$\underline{z}_a'$	$\underline{x}_{a(1)}'$	$\underline{x}_{a(2)}'$	$\cdots$	$\underline{x}_{a(b_a)}'$

where

$\underline{z}_i'$ corresponds to the whole-plot factor level ($\underline{z}$ treatment combination) applied to the i^{th} whole-plot EU, i=1,2,...a.

and

$\underline{x}_{i(j)}'$ corresponds to the j^{th} sub-plot factor level nested within the i^{th} whole-plot EU, i=1,2,...,a j=1,2,...,b_i.

The nesting notation reflects that in the non-crossed BRD, there is the possibility of a nesting of certain sub-plot factor levels within a single whole-plot. Typically, for standard response surface designs, there is at least partial nesting of sub-plot factor levels within whole-plots. This characteristic distinguishes the non-crossed BRD from the crossed BRD. Recall, the crossed BRD had the requirement that every level of the sub-plot factor must be randomized within each whole-plot. The non-crossed design does not have this "crossing" treatment restriction, thus, giving rise to the terminology non-crossed.

If an experiment is run under a bi-randomization format, and the design has a non-crossed treatment structure, the standard response surface model 2.4.1 takes the following form:

$$y_{ij} = \beta_0 + \underline{\gamma}'\underline{z}_i^d + \underline{\alpha}'\underline{x}_{i(j)}^d + \underline{z}_i^{'d} \triangle \underline{x}_{i(j)}^d + \delta_i + \epsilon_{ij} \qquad (4.1.1)$$
$$i=1,2,...,a \text{ and } j=1,2,...,b_i.$$

The distribution of the vector of responses remains

$$\underline{y} \sim N(\underline{\mu}, V)$$

$$\text{with } \underline{\mu} = \beta_0\underline{1} + Z\underline{\gamma} + X\underline{\alpha} + Z' \triangle X$$

and

$$V=\begin{bmatrix} T_1 & 0 & \cdots & 0 \\ 0 & T_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T_a \end{bmatrix}$$

where

$$T_i=\sigma_\epsilon{}^2 I_{(b_i*b_i)} +\sigma_\delta{}^2 \underline{1}_{(b_i*1)}\, \underline{1}'_{(1*b_i)} \quad i=1,2,\ldots,a.$$

The only difference between models 4.1.1 and 2.4.1 is the reflection of nesting of sub-plot factor levels within a whole-plot. The assumptions on the errors remain the same including the independence of observations across whole-plots. In Chapter 2 the universal assumption of constant covariance between two observations within the same whole-plot was made. For the non-crossed BRD this assumption will still be maintained even though each whole-plot may no longer be subdivided into the same number of sub-plot units.

For the non-crossed BRDs, maximum likelihood estimation is still achieved through the use of the GLS equations. If model 4.1.1 is rewritten as

$$\underline{y}=X^*\underline{\beta} + \underline{\delta} + \underline{\epsilon} \tag{4.1.2}$$

where $X^*\underline{\beta} = \beta_0\underline{1} + Z\underline{\gamma} + X\underline{\alpha} + Z'\triangle X$

then the MLE of $\hat{\underline{\beta}}$ is

$$\hat{\underline{\beta}} = (X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1}\underline{y} \tag{4.1.3}$$

and

$$\mathrm{Var}(\hat{\underline{\beta}}) = (X^{*\prime}V^{-1}X^*)^{-1}. \tag{4.1.4}$$

Under bi-randomization experimentation, classes of workable standard

response surface designs fall naturally into the categories of crossed and non-crossed BRDs. Full factorial designs, as was mentioned earlier, are always crossed BRDs whereas only certain fractional factorials are "crossed". The choice of the fraction can determine into which class of BRD the design falls. For those that are crossed BRDs, the estimation and analysis techniques necessary are described in Chapter 3.

§4.2 Two-level Fractional Factorial Designs

Two-level fractional factorials which do fall into the class of non-crossed BRDs compose a broad class of the standard first order non-crossed BRDs. Model estimation and analysis techniques for these non-crossed BRDs will be considered separately from the second order non-crossed BRDs due to a simplification invoked by the first order structure.

First consider the class of two-level orthogonal fractional factorial designs. In reference to both crossed and non-crossed fractions, care should be taken in the choice of fraction in order to prevent the inflation of the naturally smaller variances associated with estimation of sub-plot model parameters in a BRD. To prevent unnecessary inflation in variances, fractions must be chosen to ensure that important sub-plot effects including pure sub-plot terms and sub-plot*whole-plot interactions do not become aliased with less important, but yet poorly estimated, whole-plot term parameters. The following theorem addresses this issue.

Theorem 4.2.1:

To minimize the variance of estimated sub-plot model parameters in the class of bi-randomization designs, two-level fractional factorials with levels coded to -1 and 1 should be chosen such that within each whole-plot, the sub-plot effect contrasts sum to zero.

Proof:

Consider any fraction of a 2^k full factorial design arranged to be run in a bi-randomization format with "a" whole-plots and b_i sub-plots within the i^{th} whole-plot. The contrast of any sub-plot effect can be written as

$$\underline{c}_{n*1}=\begin{bmatrix} \underline{c}_1 \\ \underline{c}_2 \\ \vdots \\ \underline{c}_a \end{bmatrix}$$

where $\underline{c}_i$ is the b_i*1 vector corresponding to the portion of the sub-plot effect found within the i^{th} whole-plot. The vector of responses, $\underline{y}$, can also be expressed as

$$\underline{y}_{n*1}=\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_a \end{bmatrix}$$

where $\underline{y}_i$ is a b_i*1 vector of responses for the i^{th} whole-plot unit.

The coefficient for any sub-plot term is then estimated by

$$\tfrac{1}{n}\,\underline{c}'\underline{y} = \tfrac{1}{n}[\,\underline{c}_1'\underline{y}_1 + \ldots + \underline{c}_a'\underline{y}_a]\,.$$

Thus,

$$\mathrm{Var}[\tfrac{1}{n}\,\underline{c}'\underline{y}] = \tfrac{1}{n^2}[\,\underline{c}_1'T_1\underline{c}_1 + \ldots + \underline{c}_a'T_a\underline{c}_a]$$

$$= \tfrac{1}{n^2}[\underline{c}_1'(\sigma_\epsilon^2 I_{(b_1*b_1)}+\sigma_\delta^2 \underline{1}_{(b_1*1)}\,\underline{1}'_{(1*b_1)})\underline{c}_1+\ldots+\underline{c}_a'(\sigma_\epsilon^2 I_{(b_a*b_a)}+\sigma_\delta^2 \underline{1}_{(b_a*1)}\,\underline{1}'_{(1*b_a)})\underline{c}_a]$$

$$=\frac{1}{n^2}[(\sigma_\epsilon{}^2 b_1+\sigma_\delta{}^2(\sum_{j=1}^{b_1} c_{1j})^2)+\ldots+(\sigma_\epsilon{}^2 b_a+\sigma_\delta{}^2(\sum_{j=1}^{b_a} c_{aj})^2)]. \qquad (4.2.1)$$

To minimize expression 4.2.1 with respect to design, set $(\sum_{j=1}^{b_i} c_{ij})^2=0$ for all i which is equivalent to setting $\sum_{j=i}^{b_i}(c_{ij})=0$ for all i. Thus, choose the fractional factorial design such that each portion of a sub-plot contrast sums to zero within each respective whole-plot. □

For example, consider the following $\frac{1}{2}$ fraction of a 2^4 factorial written in the form of a BRD:

whole-plot	z_1	z_2	x_1	x_2	x_1	x_2
1	1	1	1	1	-1	-1
2	1	-1	1	-1	-1	1
3	-1	1	1	-1	-1	1
4	-1	-1	1	1	-1	-1

This is a non-crossed BRD due to differing (x_1, x_2) combinations within whole-plots 1,4 and 2,3. Also notice that within each whole-plot, $\sum z_i x_j = 0$. This holds true for any sub-plot model term and sub-plot*whole-plot model term combination. Since these conditions hold, Theorem 4.2.1 applies, and this fractional factorial provides minimal variance for estimation of sub-plot model coefficients.

In addition to minimizing variances, choosing the fraction according to the theorem above also has an impact on estimation of model 4.1.2. Just as in all crossed bi-randomization designs, non-crossed fractional factorial designs chosen such that sub-plot contrast portions within a whole-plot sum to zero maintain an equality between the OLS and GLS model estimating equations. Recall in the proof of theorem 3.2.1 for the crossed BRDs, equality was maintained as long as

G_i=G for all i. For a two-level fractional factorial chosen as described above, each whole-plot has the same number of sub-plots so T_i=T for all i and within each whole-plot the sum over the sub-plot contrasts remains constant, i.e. s_m= k (s_m= 0 if levels are coded to -1 and 1). Thus, using this information and by examining the G_i matrix found in 3.2.7, it is evident that G_i=G for two-level fractional factorials as described here. This implies by the proof of theorem 3.2.1 that OLS=GLS for these non-crossed fractional factorial BRDs.

Other first order non-crossed bi-randomization designs besides the two-level regular fractional factorial designs also appear to maintain the same equivalency of OLS and GLS. Altered fractions were examined along with designs in which treatment combinations were arbitrarily deleted. In each case, the design had a non-crossed treatment structure, but yet OLS=GLS. Investigations of these results were conducted, and an attempt was made to adapt the proof of theorem 3.2.1 to these situations. Unfortunately, a direct adaptation was not found nor has any other explanation for the estimation equivalency for the first order non-crossed altered fractional factorial BRDs. The key, however, must lie in the first order nature of the model. Further research must be done on these designs before a universal statement can be made about these equalities for first order non-crossed BRDs.

Based upon this equivalency of OLS and GLS, estimation and analysis of first order non-crossed BRDs and models is now simplified. The MLE of Model 4.1.2 is given by

$$\hat{\underline{\beta}} = (X^{*\prime}V^{-1}X^{*})^{-1}X^{*\prime}V^{-1}\underline{y} = (X^{*\prime}X^{*})^{-1}X^{*\prime}\underline{y}$$

and

$$\mathrm{Var}(\hat{\underline{\beta}}) = (X^{*\prime}V^{-1}X^{*})^{-1}. \qquad (4.2.2)$$

Information of V, or σ_δ^2 and σ_ϵ^2, is not essential for model estimation, but is necessary for estimation of 4.2.2.

The methods described in Chapter 3 (the lack of fit method and the two model method) must be adapted to reflect the non-crossed structure of the first order designs. Due to the lack of identical sub-plot factor levels within each whole-plot, the lack of fit method will be used for the theoretical development.

§4.3 Estimation of V/Analysis for First Order Non-crossed BRDs

As described in §3.3, estimators for the whole-plot and sub-plot error variances in first order non-crossed BRDs will be primarily functions of lack of fit sum of squares. To produce the appropriate sum of squares and degrees of freedom, Model 4.1.2 is rewritten as

$$y_{ij} = \beta_0+\underline{\gamma}'\underline{z}_i^{d}+\underline{\varrho}'\underline{z}_i^{*d}+\underline{\alpha}'\underline{x}_{i(j)}^{d}+\underline{z}_i'^{d} \triangle \underline{x}_{i(j)}^{d} + \epsilon_{ij} \qquad (4.3.1)$$
$$i=1,2...,a;\ j=1,2,...,b_i$$

where ϵ_{ij} iid $N(0, \sigma_\epsilon^2)$ and $\underline{z}^{*d}$ is an (a-(k+1))*1 vector of nonmodeled whole-plot terms saturating the available whole-plot degrees of freedom. Recall, $\underline{\varrho}'\underline{z}_i^{*d}$ can be viewed as a model of δ_i. The whole-plot lack of fit sum of squares is given by

$$SS(\underline{\varrho}|\underline{\gamma})=SS(\underline{\varrho},\underline{\gamma})-SS(\underline{\gamma}) \qquad (4.3.2)$$

as in the crossed case.

Using 4.3.2,

$$E\left[\frac{SS(\underline{\varrho}|\underline{\gamma})}{a-(k+1)}\right]=E[MSE_{\text{whole-plot lof}}]$$

$$=\sigma_\epsilon^2+\left[\frac{\left[\sum_{i=1}^{a} b_i-\text{trace}[Z(Z'Z)^{-1}Z'B^*]\right]\sigma_\delta^2}{a-(k+1)}\right] \qquad (4.3.3)$$

where Z is the model matrix associated with the modeled whole-plot parameters, $\underline{\gamma}$, and

$$B^{*}=\begin{bmatrix} b_1 I_{(b_1)} & & & \\ & b_2 I_{(b_2)} & & \\ & & \ddots & \\ & & & b_a I_{(b_a)} \end{bmatrix} .$$

Once whole-plot lack of fit is computed, sub-plot lack of fit/replication sum of squares can then be calculated for model 4.3.1 and

$$E\left[\frac{SSE_{\text{sub-plot lof}}}{\sum_{i=1}^{a} b_i - p^{*}}\right] = E[MSE_{\text{sub-plot lof}}]=\sigma_{\epsilon}^{2} \qquad (4.3.4)$$

where $\sum_{i=1}^{a} b_i$ = total design size and p^{*} corresponds to the total number of parameters found in model 4.3.1.

Combining the expected mean squares of 4.3.3 and 4.3.4 the following estimators are derived:

$$\hat{\sigma}_{\epsilon}^{2} = MSE_{\text{sub-plot lof}} \qquad (4.3.5)$$

and

$$\hat{\sigma}_{\delta}^{2} = \frac{[MSE_{\text{whole-plot lof}}-MSE_{\text{sub-plot lof}}]*(a-(k+1))}{\sum_{i=1}^{a} b_i - \text{trace}[Z(Z'Z)^{-1}Z'B^{*}]} . \qquad (4.3.6)$$

For first order non-crossed designs in which $b_i=b \ \forall \ i$, the denominator in expression 4.3.6 simplifies to $b*(a-(k+1))$ which in turn reduces 4.3.6 to

$$\hat{\sigma}_{\delta}^{2} = \frac{[MSE_{\text{whole-plot lof}}-MSE_{\text{sub-plot lof}}]}{b} . \qquad (4.3.7)$$

Note this estimator is identical to that found in Chapter 3 for the crossed bi-randomization design. If $b_i=b$ holds, model editing for the first order designs may

be performed using the same tests outlined previously in Chapter 3. Refer to Appendix A for distributions which may be used for the formulation of t-tests when the number of sub-plots within each whole-plot is not the same. The development of these distributions utilize Cholesky's decomposition which is formally given in §4.4.

Using the estimators from 4.3.5 and 4.3.6, an estimator for V may be formed and substituted into 4.2.2 to obtain the respective estimator for $\mathrm{Var}(\underline{\hat{\beta}})$.

§4.4 Second Order Non-crossed Bi-Randomization Designs

While first order non-crossed BRDs maintain equivalency between the OLS and GLS model estimation equations, this same equivalency does not universally hold for second order non-crossed BRDs such as the ccd. For model 4.1.2 in these cases, the MLE of $\underline{\hat{\beta}}$ remains

$$\underline{\hat{\beta}} = (X^{*'}V^{-1}X^{*})^{-1}X^{*'}V^{-1}\underline{y} \tag{4.4.1}$$

and

$$\mathrm{Var}(\underline{\hat{\beta}}) = (X^{*'}V^{-1}X^{*})^{-1}. \tag{4.4.2}$$

Unlike the first order case, both the model parameter estimates and their respective variances are functions of the unknown error variances, ${\sigma_\delta}^2$ and ${\sigma_\epsilon}^2$. Thus, before model estimation can be done using 4.4.1 a reasonable estimator for V must be determined.

An interesting and useful reparametrization of both 4.4.1 and 4.4.2 which may simplify this task and others that follow may be found using Cholesky's decomposition. See Graybill [1976]. Cholesky's decomposition makes use of the square root of a positive definite matrix such as V found in 4.4.1. Using this theorem and the assumption of a positive definite V, V^{-1} can be rewritten as

$$V^{-1}=V^{\frac{-1}{2}}V^{\frac{-1}{2}}. \tag{4.4.3}$$

It follows that

$$X^{*\prime}V^{-1}X^{*} = X^{*\prime}V^{\frac{-1}{2}}V^{\frac{-1}{2}}X^{*} = X_t^{*\prime}X_t^{*}. \tag{4.4.4}$$

Consequently, using 4.4.4

$$\hat{\underline{\beta}} = (X_t^{*\prime}X_t^{*})^{-1}X_t^{*\prime}\underline{y}^{*} \tag{4.4.5}$$

where $\underline{y}^{*}=V^{\frac{-1}{2}}\underline{y}$ and $X_t^{*}=V^{\frac{-1}{2}}X^{*}$.

As a result,

$$\mathrm{Var}(\hat{\underline{\beta}}) = (X_t^{*\prime}X_t^{*})^{-1}. \tag{4.4.6}$$

The matrix X_t^{*} can be viewed as the new model matrix. For the derivation of $V^{\frac{-1}{2}}$ for a specific example and to obtain an idea of the structure of $V^{\frac{-1}{2}}$ refer to Appendix B. The matrix $V^{\frac{-1}{2}}$ is a function of

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

Thus, this reparametrization is still a function of the two unknown variance parameters.

By either parametrization, 4.4.1 or 4.4.5, estimators of the error variances, and thus V, are essential to proceed with the desired response surface analysis. Estimators given for the first order non-crossed BRDs in §4.3, however, must be modified to reflect the dependence on V of the estimated parameter coefficients. A more complex approach must be taken for estimation. Four methods are presented for estimation in second order non-crossed BRDs.

§4.5 **Estimation of V for Second Order Non-crossed BRDs**

As indicated in the previous section, the estimation procedures for V in first order cases and in crossed situations were derived using the estimators for $\underline{\beta}$ given by the MLE=GLS=OLS expression. In the second order non-crossed case, however, the MLE of $\underline{\beta}$ is no longer equal to OLS and is a function of V. Thus in

order to estimate the parameters and their variances, initial estimators must be found for the error variances and updated through an iterative scheme. This approach is the basis for a majority of the error variance estimation procedures presented.

§4.5.1 Method 1: Ordinary Least Squares

The first method proposed can be thought of as the "ignorance" approach. The name is appropriate due to the fact that this method chooses to ignore the non-independent relationship among observations within a given whole-plot in bi-randomization designs and makes the assumption of universal independence. Thus, the analysis for the second order non-crossed designs reduces to using OLS for estimation. i.e., for model 4.1.2

$$\hat{\underline{\beta}} = (X^{*\prime}X^{*})^{-1}X^{*\prime}\underline{y}. \tag{4.5.1}$$

At first glance, "ignorance" in this procedure may appear to have dual meanings. Primarily, "ignorance" refers to the ignoring of the dependent nature of the observations, but, in addition, ignorance can also have the negative connotation of a simplified analysis that may produce less than "optimal" results. Even in the presence of a bi-randomization error structure, the estimating equations in 4.5.1 (OLS) produce unbiased estimators for $\underline{\beta}$. The sacrifice in using the OLS estimators comes from the increased variation associated with these estimators. However, recall that in the presence of a BRD structure, a majority of error variance estimation is based upon lack of fit sum of squares. In addition, recall that this discussion involves standard second order response surface designs in which there are in most cases minimal degrees of freedom for lack of fit. As a result, there may not be sufficient information available to provide "good" estimates for V. In other words, the researcher may in some situations be better served by not trying to estimate V for use in model estimation than using poor estimates and inducing more variability into the methodology. The OLS method, thus, does have a valid basis for consideration as a bi-randomization estimation procedure.

§4.5.2 Method 2: 1 Step Iterated Re-weighted Least Squares (1IRLS)

The second estimation procedure incorporates OLS estimators into the GLS expressions. In 1 step IRLS, $\underline{\beta}$ is initially estimated using expression 4.5.1 (OLS). From these estimated coefficients, estimators for ${\sigma_\delta}^2$ and ${\sigma_\epsilon}^2$ are then calculated using expressions 4.3.5 and 4.3.6, i.e. from OLS expected mean square formulae. Using these estimators for the error variances, $\hat{V}$ can be created and the model fitted by

$$\hat{\underline{\beta}} = (X^{*'}\hat{V}^{-1}X^{*})^{-1}X^{*'}\hat{V}^{-1}\underline{y}. \qquad (4.5.2)$$

§4.5.3 Method 3: Iterated Re-weighted Least Squares (IRLS)

The first step in the IRLS procedure is to use the OLS E(MS) to provide an initial estimator for V for use in the GLS expressions just as described in Method 2. Method 3 then extends this idea. Consider the $\hat{V}$ created in Method 2 as the initial estimator of V and call it $\hat{V}_0$. Thus, 4.5.2 becomes

$$\hat{\underline{\beta}}_0 = (X^{*'}\hat{V}_0^{-1}X^{*})^{-1}X^{*'}\hat{V}_0^{-1}\underline{y}\,. \qquad (4.5.3)$$

By using 4.5.3 and E(MS) formulae derived from the GLS expressions, estimators of ${\sigma_\delta}^2$ and ${\sigma_\delta}^2$ can be updated with the estimators that follow. Consider a second order model in the form of 4.3.1. The same assumptions hold including the modeling of δ_i by $\underline{\varrho}'\underline{z}_i^{*d}$ to saturate the whole-plot portion of the model with

$$\text{MSE}_{\text{whole-plot lof}} = \left[\frac{\text{SS}(\underline{\varrho}|\underline{\alpha})}{\text{a-(k+1)}}\right]$$

and

$$\text{MSE}_{\text{sub-plot lof}} = \frac{y'[I\text{-}X^{**}(X^{**'}X^{**})^{-1}X^{**'}]y}{n\text{-}p^{*}}.$$

Here X^{**} is the model matrix for 4.3.1 including saturated whole-plot terms.

The updated estimators are then given by

$$\hat{\hat{\sigma}}_{\epsilon}^{2} = \frac{[MSE_{\text{sub-plot lof}}][n-p^{*}]}{tr[\hat{V}_0^{-1} - \hat{V}_0^{-1}X^{**}(X^{**'}\hat{V}_0^{-1}X^{**})^{-1}X^{**'}\hat{V}_0^{-1}]} \tag{4.5.4}$$

and

$$\hat{\hat{\sigma}}_{\delta}^{2} = \tag{4.5.5}$$

$$\frac{[MSE_{\text{wp lof}}]*(a-(k+1))-[\hat{\hat{\sigma}}_{\epsilon}^{2}]tr[\hat{V}_0^{-1}[Z^{**}(Z^{**'}\hat{V}_0^{-1}Z^{**})^{-1}Z^{**'}-Z(Z'\hat{V}_0^{-1}Z)^{-1}Z']\hat{V}_0^{-1}]}{tr[\hat{V}_0^{-1}[Z^{**}(Z^{**'}\hat{V}_0^{-1}Z^{**})^{-1}Z^{**'}-Z(Z'\hat{V}_0^{-1}Z)^{-1}Z']\hat{V}_0^{-1}O]}$$

where $J_i = \underline{1}_{(b_i*b_i)}\underline{1}'_{(b_i*b_i)}$, Z^{**} is the saturated whole-plot model matrix, and

$$O = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_a \end{bmatrix} .$$

Using the new estimates from 4.5.4 and 4.5.5, $\hat{V}_0$ can be updated to $\hat{V}_1$ and $\hat{\underline{\beta}}_0$ to $\hat{\underline{\beta}}_1$, where

$$\hat{\underline{\beta}}_1 = (X^{*'}\hat{V}_1^{-1}X^{*})^{-1}X^{*'}\hat{V}_1^{-1}\underline{Y} . \tag{4.5.6}$$

The vector $\hat{\underline{\beta}}_1$ is then used to update and create $\hat{V}_2$, and this cyclic procedure continues until convergence is reached in $\hat{\underline{\beta}}$. Once convergence is reached, the final $\hat{V}_i$ is used in the GLS equations to estimate the model.

§4.5.4 Method 4: Restricted Maximum Likelihood

The final procedure presented for error variance estimation is based on the

concept of maximum likelihood estimation (MLE). In the bi-randomization situation presented, the assumption is made that errors are normally distributed with $\mathrm{Var}(\underline{\delta}+\underline{\epsilon})=V$ which is a function of σ_δ^2 and σ_ϵ^2 . Thus, for a BRD there are a total of three parameters of interest to be estimated, $\underline{\beta}$, σ_δ^2, and σ_ϵ^2. For now, the primary focus is on the variance components and for this reason, instead of the traditional maximum likelihood methods, restricted maximum likelihood (REML) methods will be used. REML is similar to ML in concept with the distinction being that REML uses the likelihood of a transformation of $\underline{y}$. This transformation is based on the vector of residuals and eliminates $\underline{\beta}$ from the likelihood expression, thus, leaving only the variance components for estimation. In addition, REML corrects for the wrong degrees of freedom usually associated with ML estimators for treatment comparison analyses such as ANOVA. REML was first developed by Anderson and Bancroft [1952] and Russell and Bradley [1958]. Extensions and further developments are given by W.A. Thompson [1962] and Patterson and R. Thompson [1971,1974]. For further reference on REML and its relationship to MLE, consult Searle, Casella, and McCulloch [1992].

For determination of REML estimators in a bi-randomization experiment, consider a second order response surface design (conducted as a bi-randomization experiment) and the response model given by 4.1.2,

$$\underline{y}=X^*\underline{\beta} + \underline{\delta} + \underline{\epsilon}$$

$$\text{where } \underline{y} \sim N(X^*\underline{\beta}, V).$$

Let the $n*n$ matrix $K=[I-X^*(X^{*\prime}X^*)^{-1}X^{*\prime}]$ and $K\underline{y}$ be the $n*1$ transformed response vector. The new responses have the following distribution:

$$K\underline{y} \sim N(0, KVK). \qquad (4.5.7)$$

Note the distribution in 4.5.7 is not a function of $\underline{\beta}$. Based on 4.5.7,
-2(log likelihood) of $K\underline{y}$ (the objective function for REML) can be rewritten as

$$-2\log L= C+\log|V|+ \underline{r}'V^{-1}\underline{r} + \log|X^{*\prime}V^{-1}X^*| \qquad (4.5.8)$$

where C is a constant and $\underline{r}=\underline{y}-X^*[X^*V^{-1}X^*]^{-1}X^*V^{-1}\underline{y}$.

See Appendix C for the development of 4.5.8. Note that the log-likelihood expression given in 4.5.8 is a function of only $V=f(\sigma_\delta^2, \sigma_\epsilon^2)$. Derivatives of 4.5.8 are then taken with respect to the parameters to produce the REML equations. The derivatives are outlined in Appendix D along with the comparative MLE derivatives. These equations are non-linear in σ_δ^2 and σ_ϵ^2 so an iterative procedure such as Newton-Raphson algorithm along with a stopping criterion must be used to find solutions.

The computational aspect of determining REML estimators for the respective error variances in a bi-randomization experiment was accomplished through the use of the PROC MIXED procedure in SAS. PROC MIXED was developed to fit mixed linear models, i.e. with fixed and random effects, and provide appropriate testing procedures, etc. by allowing more than one source of variation. This is a useful extension of the PROC GLM procedure in SAS which allows for only one source of variation, and thus, testing statements are required to conduct proper tests in mixed model situations. The PROC MIXED procedure may be found in Chapter 16 of SAS Technical Report P-229, SAS/STAT Software: Changes and Enhancements, Release 6.07, [1992].

The PROC MIXED procedure was not developed for direct application to situations such as the response surface bi-randomization experiment described in this dissertation. The discrepancy arises from the fact that in the bi-randomization design a portion (k+1 degrees of freedom) of the a degrees of freedom associated with random effects must be modeled as fixed. A few modifications, therefore, were necessary to accommodate the additional fixed model terms. See Appendix E for the specific details and modified program code.

Once the REML estimates are found, $\hat{V}$ can be formed along with

$$\hat{\underline{\beta}} = (X^*\hat{V}^{-1}X^*)^{-1}X^*\hat{V}^{-1}\underline{y}. \qquad (4.5.9)$$

Example:

To illustrate these estimation methods, consider a four variable (z_1 z_2 x_1 x_2) ccd run under a bi-randomization experimental format with σ_δ^2=16 and σ_ϵ^2=4. The response model of interest involves the following terms: intercept, z_1, z_2, z_1z_2, z_1^2, x_1, x_2, x_1x_2, x_1^2, x_2^2, x_1z_1, x_1z_2. The response data was simulated for this bi-randomization situation assuming normality of errors. Based on the simulated data the following estimators for σ_δ^2 and σ_ϵ^2 were calculated using the presented methods. The results are given in Table 4.5.1.

Table 4.5.1 Estimators for Error Variances-Example

	1IRLS	IRLS	REML
σ_δ^2	61.1	95.5	29.8
σ_ϵ^2	4.6	4.6	4.6

The effect of the differing estimates on coefficient estimates are given in Table 4.5.2.

Table 4.5.2 Coefficient Estimates-Example

	1IRLS	IRLS	REML	OLS
β_0	31.94	31.94	31.95	32.76
z_1	8.06	8.08	8.00	6.56
z_2	5.17	5.19	5.13	3.90
z_1^2	6.26	6.26	6.25	5.64
x_1^2	11.01	11.00	11.04	11.41
x_2^2	-1.93	-1.94	-1.9	-1.5

The estimation of the whole-plot error variance varies widely with respect to these methods while the sub-plot error variance has greater estimation consistency. Interestingly, even in the presence of the diverse $\widehat{\sigma_\delta}^2$, the $\underline{\beta}$ estimates for each method are similar indicating a robustness to the estimated value of $\sigma_\delta{}^2$. This will become more evident in the comparison of error variance estimators in §4.6.

Four methods have now been presented for the purpose of providing estimators for $\sigma_\delta{}^2$ and $\sigma_\epsilon{}^2$ and, consequently, V in bi-randomization experimental designs. Interest centers on which of these methods provides "better" variance estimators. This decision can be based on the properties of the various estimators. Due to the iterative nature of these estimators and specifically the complexion of the IRLS estimators, however, direct properties of these estimators are unable to be easily derived. Simulation, therefore, must be utilized for discrimination purposes.

§4.6 **Comparison of Estimators**

The comparison of the error variance estimation methods through simulation has a dual purpose. The first is to provide information concerning the performance of the estimators in various design/model situations. The central composite and hybrid designs will be used along with two different models. The first model will involve only sub-plot quadratic terms and maintain a first order model in the whole-plot terms. The second model will then incorporate second order whole-plot model terms. The differing models were considered because the bi-randomization structure primarily affects the estimation of quadratic terms in the model. The influence of adding quadratic terms to the model on the estimation procedures can then be explored. The second purpose of the study is to investigate the small sample properties of these estimators. It was already mentioned that due to the iterative nature of the estimation the distributions of these error variance estimators are unknown, and thus, so is $\mathrm{Var}(\widehat{\underline{\beta}})$. Through simulation, an approximation to this small sample unknown $\mathrm{Var}(\widehat{\underline{\beta}})$ can be compared to the asymptotic result to assess whether if in future investigations the asymptotic result may be used for simplification.

For a bi-randomization response surface experiment, if V were known the MLE of the model parameters, $\underline{\beta}$, would be given by

$$\hat{\underline{\beta}} = (X^{*'}V^{-1}X^{*})^{-1}X^{*'}V^{-1}\underline{y} \tag{4.6.1}$$

and

$$\mathrm{Var}(\hat{\underline{\beta}}) = (X^{*'}V^{-1}X^{*})^{-1}. \tag{4.6.2}$$

If $\hat{V}$ is a "good" estimator of V, then

$$\hat{\underline{\beta}} = (X^{*'}\hat{V}^{-1}X^{*})^{-1}X^{*'}\hat{V}^{-1}\underline{y} \tag{4.6.3}$$

should be a very close approximation to 4.6.1, and in turn the variance of the estimator in 4.6.3 should be close to 4.6.2. An exact expression for this variance is unknown due to unknown distributions of the error variance estimators. One way to quantify the "closeness" of this approximation is to compare the resulting variance-covariance matrix of 4.6.3 to that of 4.6.2 through their respective traces and determinants. Note that the $\mathrm{Var}(\hat{\underline{\beta}})$ given in 4.6.2 is also the asymptotic variance-covariance matrix of the estimator in 4.6.3. This will discussed in more detail later. Also note that simulation was necessary to obtain small sample numerical values for the variance of 4.6.3.

Using a simulation based on 2000 iterations, $\mathrm{Var}(\hat{\underline{\beta}})$ was generated for each of the estimation procedures outlined for a given design/model combination. The trace and determinant of each simulated matrix were then able to be calculated. These values were then compared to the $\mathrm{trace}(X^{*'}V^{-1}X^{*})^{-1}$ and $\det(X^{*'}V^{-1}X^{*})^{-1}$, the asymptotic values.

Four types of bi-randomization experimental designs were investigated: (i)

a modified four variable ccd with α=2, (ii) a modified four variable hybrid 416a, (iii) standard four variable ccd with α=2, and (iv) standard hybrid 416a. Designs (i) and (ii) were used for modeling a second order model in which the whole-plot terms were first order and the sub-plot terms second order. Modified versions of standard second order designs were used to reflect a first order nature in the whole-plot terms. Designs (iii) and (iv) were then used to model a second order response surface model in which both whole-plot and sub-plot terms involved second order terms. The standard RS designs were used in this situation.

§4.6.1 Model 1: First Order Whole-plot/Second Order Sub-plot

The first design\model combination investigated involves four variables: two whole-plot variables, z_1 and z_2, and two sub-plot variables, x_1 and x_2. The response surface model used to simulate the data (and to be fit) is

$$y_{ij}=34.45+7.24z_1+5.91z_2+4.89x_1-1.34x_2+11.06x_1^2+6.66x_2^2 \\ -5.91x_1z_1+11.07x_1z_2+\delta_i+\epsilon_{ij}. \qquad (4.6.5)$$

In addition, the assumption is made that the $\text{var}(y_{ij})=\sigma_\delta^2+\sigma_\epsilon^2=12$, but $\frac{\sigma_\delta^2}{\sigma_\epsilon^2}=d$ is allowed to vary. The value d measures variation among whole-plots as compared to variation within whole-plots. A value of d=0 implies that observations within a whole-plot are independent, i.e. as if a CRD were implemented. As $d\rightarrow\infty$, the strength of the dependent relationship among observations within a whole-plot grows in relation to the sub-plot error variance, i.e. the covariance increases while the variance remains constant. As d rises, the differences between the completely randomized and bi-randomization design situations become more severe.

The following modified standard ccd and hybrid 416a designs were used for generation of data and analysis:

ccd

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
0	0	2	0
0	0	-2	0
0	0	0	2
0	0	0	-2
0	0	0	0
0	0	0	0
0	0	0	0

hybrid

1	1	1	.6444
1	1	-1	.6444
-1	1	1	.6444
-1	1	-1	.6444
1	-1	1	.6444
1	-1	-1	.6444
-1	-1	1	.6444
-1	-1	-1	.6444
0	0	1.6853	-.9075
0	0	-1.6853	-.9075
0	0	0	-.9075
0	0	0	-.9075
0	0	0	-.9075
0	0	0	1.7844
0	0	0	-1.4945

The modifications of these designs better accommodate the first order nature of the whole-plot variables and allow for the presence of fewer whole-plot EUs. While fewer whole-plot EUs may be beneficial, at the same time this modification may be impairing the designs. The following results were obtained from the simulation.

The first table displays the determinant of the asymptotic $\mathrm{Var}(\underline{\hat{\beta}})$ along with the simulated determinants for various combinations of ${\sigma_\delta}^2$ and ${\sigma_\epsilon}^2$ for each estimation method.

Table 4.6.1 Asymptotic/Simulated Det[Var($\hat{\beta}$)] for CCD/Model1 for Various Estimation Methods

d	ASYM	OLS	1IRLS	IRLS	REML	ASY(Sim)
.2	0.0165	0.0147	2.9892	0.6111	0.0145	0.0144
.5	0.0105	0.0111	0.2559	0.0436	0.0107	0.0105
1	0.0037	0.0042	0.0114	0.0139	0.0038	0.0038
2	0.00057	0.00067	0.01139	0.00428	0.00054	0.00057
3	0.00013	0.00015	0.00098	0.00037	0.00011	0.00011
5	0.000014	0.000022	0.000059	0.000040	0.000015	0.000015

Table 4.6.2 Asymptotic/Simulated Det[Var($\hat{\beta}$)] for Hybrid 416A/Model1 for Various Estimation Methods

d	ASYM	OLS	1IRLS	IRLS	REML	ASY(Sim)
.2	5.96	5.21	11.27	10.44	8.66	4.98
.5	4.51	4.45	6.06	5.72	5.21	3.53
1	1.29	1.80	1.77	1.70	1.69	1.84
2	0.20	0.41	0.27	0.26	0.27	0.19
3	0.054	0.156	0.075	0.073	0.073	0.054
5	0.0053	0.020	0.0070	0.0068	0.0066	0.0051

Note: Asy(Sim) corresponds to a simulation of Var($\hat{\beta}$) using the known V. This is primarily a check to ensure a large enough simulation is being conducted. If the simulation is indeed large enough, Asy(Sim) values should be close to the asymptotic values.

Tables 4.6.1 and 4.6.2 show interesting results. The most obvious is the performance of the OLS estimators for small to moderate values of d. In the ccd, the OLS estimation procedure even performs well for the larger d values. This is primarily due to the fact that for these specific designs and this model, few model

coefficient estimates are being affected by the presence of V. The OLS procedure will not fare as well for the second model in which more terms are influenced by V. For the ccd, it appears that over all values of d the REML estimators appear to be the "best" in terms of closeness to $|(X^*V^{-1}X^*)^{-1}|$. IRLS also appears to do well for larger values of d while 1IRLS is only competitive for the even larger values of d. In the case of the hybrid design, however, all three methods (1IRLS, IRLS, and REML) appear to be acceptable and very similar alternatives except for very small values of d. More discrimination between these methods for the determinant criterion will be found in Model 2.

The second discriminatory value for the methods was the trace of simulated $\mathrm{Var}(\widehat{\underline{\beta}})$. Instead of using a table form as above, these results are displayed graphically in Figures 4.6.1 and 4.6.2. Results similar to those displayed for the determinant are found.

Figure 4.6.1 shows the simulated trace values for the ccd. Like the case of the determinant, the OLS estimator provides a very competitive procedure for all values of d. REML estimators also give results close to their asymptotic values. The iterative estimators, 1IRLS and IRLS, however, require larger values of d (close to 3) before approaching their asymptotic trace. IRLS requires moderate d's whereas 1IRLS requires even larger values of d to "settle" down. It appears so far that 1IRLS may not be a satisfactory estimation method.

Figure 4.6.2 illustrates the simulated traces for the hybrid design. Unlike Figure 4.6.1, this graph shows the OLS method for estimation to be very unsatisfactory for moderate to large values of d. The OLS trace quickly diverges from the set of other estimators as d grows. Once again, this may differ from the results of the ccd due to the fact that the hybrid has more terms affected by the bi-randomization experimental set-up. The indication is that as more terms become affected by V, the worse the OLS estimation procedure will perform. The other variance estimation procedures all generate criterion values close to the asymptotic trace with REML and IRLS being the best of the set.

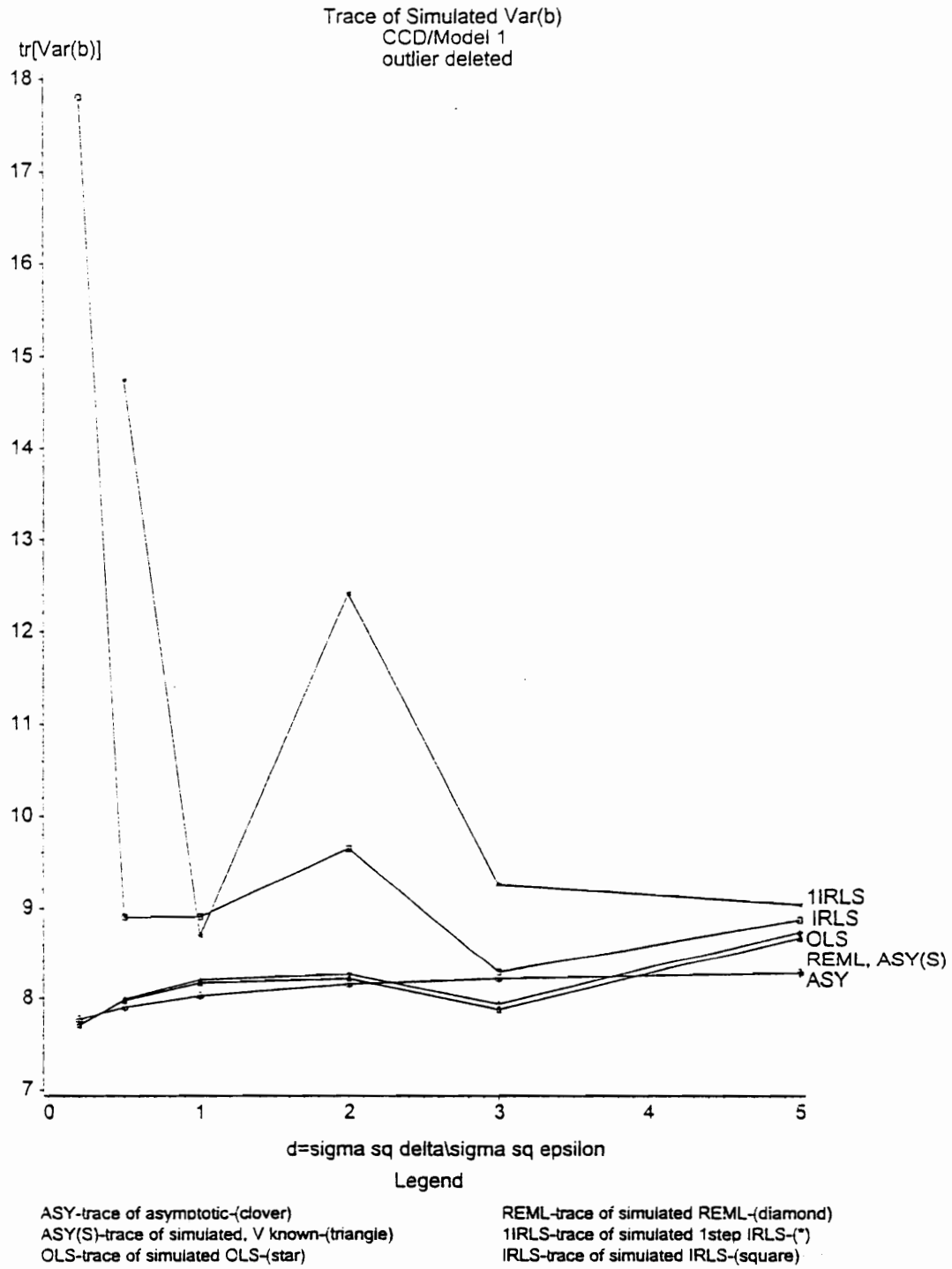


Figure 4.6.1 Asymptotic/Simulated Trace[Var($\hat{\beta}$)] for CCD/Model1 for Various Estimation Methods

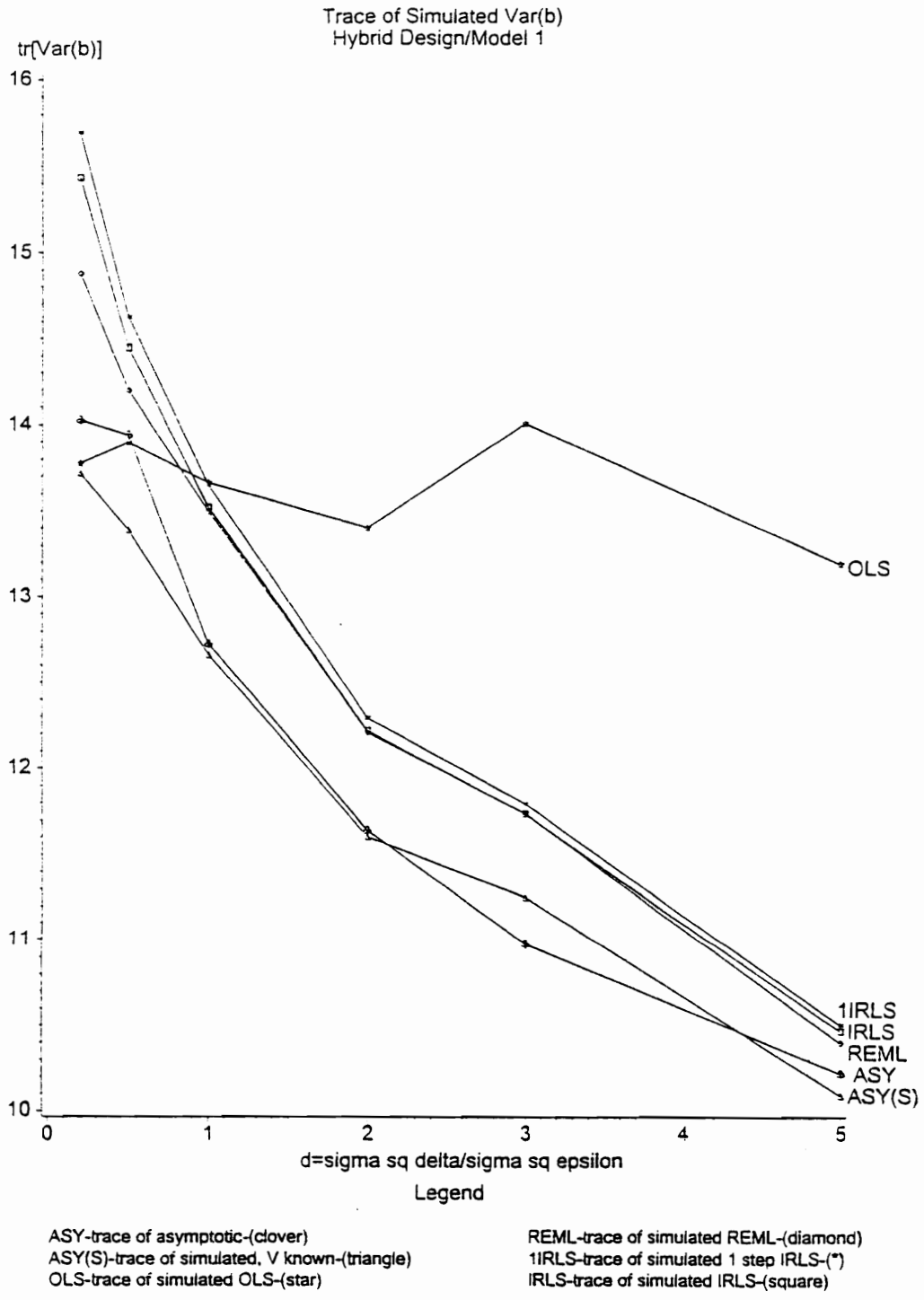


Figure 4.6.2 Asymptotic/Simulated Trace[Var($\hat{\underline{\beta}}$)] for Hybrid 416A/Model1 for Various Estimation Methods

In terms of determining the "best" method among those presented for estimating V, the plots do not point to an overall winner. For the ccd, OLS and REML both proved to be excellent choices across all levels of d with IRLS entering the picture at moderate levels of d. Estimation through OLS should clearly not be ruled out for the ccd with limited second order terms. For the case of the smaller hybrid design, all methods apart from OLS were very close to each other in terms of their determinant and trace values. Discrimination among methods for this model/design situation would be difficult.

The choice of the "best" estimation procedure is clearly a function of design, but with respect to an overall winner, this author would be forced to select the REML estimation method for this model situation. For both designs, the asymptotic results provided close approximations to the simulated REML trace and determinant values. This will be further supported later in the chapter through the standard errors associated with the simulation. For large values of d, however, the IRLS procedure can also be considered as an alternative.

§4.6.2 Model 2: Second Order Whole-plot/Second Order Sub-plot

Now consider an alternative situation in which the response surface model used to generate the data and to be fit involves an additional second order term in the whole-plot variables and has a larger error variance sum. The model is

$$y_{ij}=34.45+7.24z_1+5.91z_2+4.92z_1z_2+4.67z_1^2+4.89x_1-1.34x_2+4.95x_1x_2 \\ +11.06x_1^2-2.10x_2^2-5.91x_1z_1-2.91x_1z_2+\delta_i+\epsilon_{ij}. \qquad (4.6.5)$$

with

$$\text{var}(y_{ij})=\sigma_\delta^2+\sigma_\epsilon^2=20 \text{ and } \frac{\sigma_\delta^2}{\sigma_\epsilon^2}=d.$$

The larger sum for the error variances will provide additional information on the influence of the actual variances and d on the estimation methods.

The standard versions of the ccd and the hybrid design were used:

ccd

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	0	0	0
-2	0	0	0
0	2	0	0
0	-2	0	0
0	0	2	0
0	0	-2	0
0	0	0	2
0	0	0	-2
0	0	0	0

hybrid

1	1	1	.6444
1	1	-1	.6444
-1	1	1	.6444
-1	1	-1	.6444
1	-1	1	.6444
1	-1	-1	.6444
-1	-1	1	.6444
-1	-1	-1	.6444
1.6853	0	0	-.9075
-1.6853	0	0	-.9075
0	1.6853	0	-.9075
0	-1.6853	0	-.9075
0	0	1.6853	-.9075
0	0	-1.6853	-.9075
0	0	0	-1.4945
0	0	0	1.7844
0	0	0	0

Three center runs were included in each design.

The simulation results are given in Tables 4.6.3 and 4.6.4 for the determinant criterion and Figures 4.6.3 and 4.6.4 for the trace criterion.

Table 4.6.3 Asymptotic/Simulated Det[Var($\hat{\beta}$)] for CCD/Model2 for Various Estimation Methods

d	ASYM	OLS	1IRLS	IRLS	REML	ASY(Sim)
.11	0.549	0.510	0.942	2.535	0.626	0.487
.43	0.344	0.450	0.483	0.445	0.370	0.305
1	0.077	0.154	0.078	0.082	0.079	0.067
2.3	0.00407	0.01490	0.00378	0.00385	0.00391	0.00355
4	0.000311	0.00181	0.000305	0.000307	0.000318	0.000311

Table 4.6.4 Asymptotic/Simulated Det[Var($\hat{\beta}$)] for Hybrid 416A/Model2 for Various Estimation Methods

d	ASYM	OLS	1IRLS	IRLS	REML	ASY(Sim)
.11	674.8	722.8	3515.5	1704.8	906.6	669.9
.43	292.5	412.8	53051	589.3	418.6	296.5
1	57.0	95.1	66.6	70.3	65.2	47.8
2.3	3.0	9.9	3.8	3.8	3.8	2.8
4	0.24	1.32	0.26	0.26	0.31	0.21

Results similar to those of the first model are found for this situation. The OLS method of estimation, however, does not perform as well with the addition of the whole-plot quadratic term. In fact, even in the case of the ccd, it is only competitive with the other methods for small values of d in which V more closely resembles the CRD variance-covariance matrix. As d rises, it quickly becomes an unsatisfactory option while according to the determinants the methods of REML and IRLS both prove to be more promising. The 1IRLS estimation method is also competitive for the ccd, but in the case of the hybrid design it has trouble with unstable determinant values due to outliers or "wild" results that occasionally occur. It requires much larger values of d before it becomes stable.

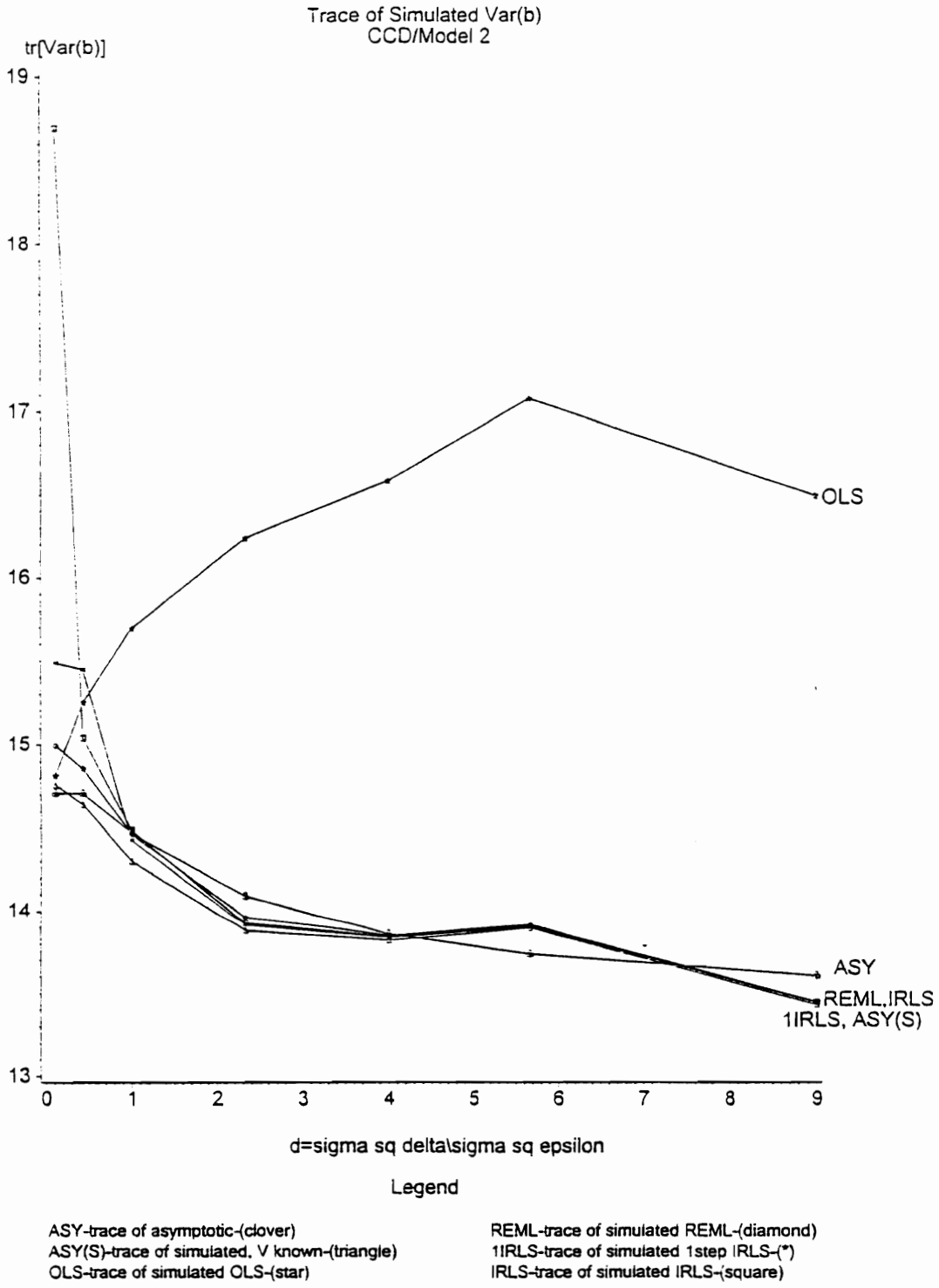


Figure 4.6.3 Asymptotic/Simulated Trace[Var($\hat{\beta}$)] for CCD/Model2 for Various Estimation Methods

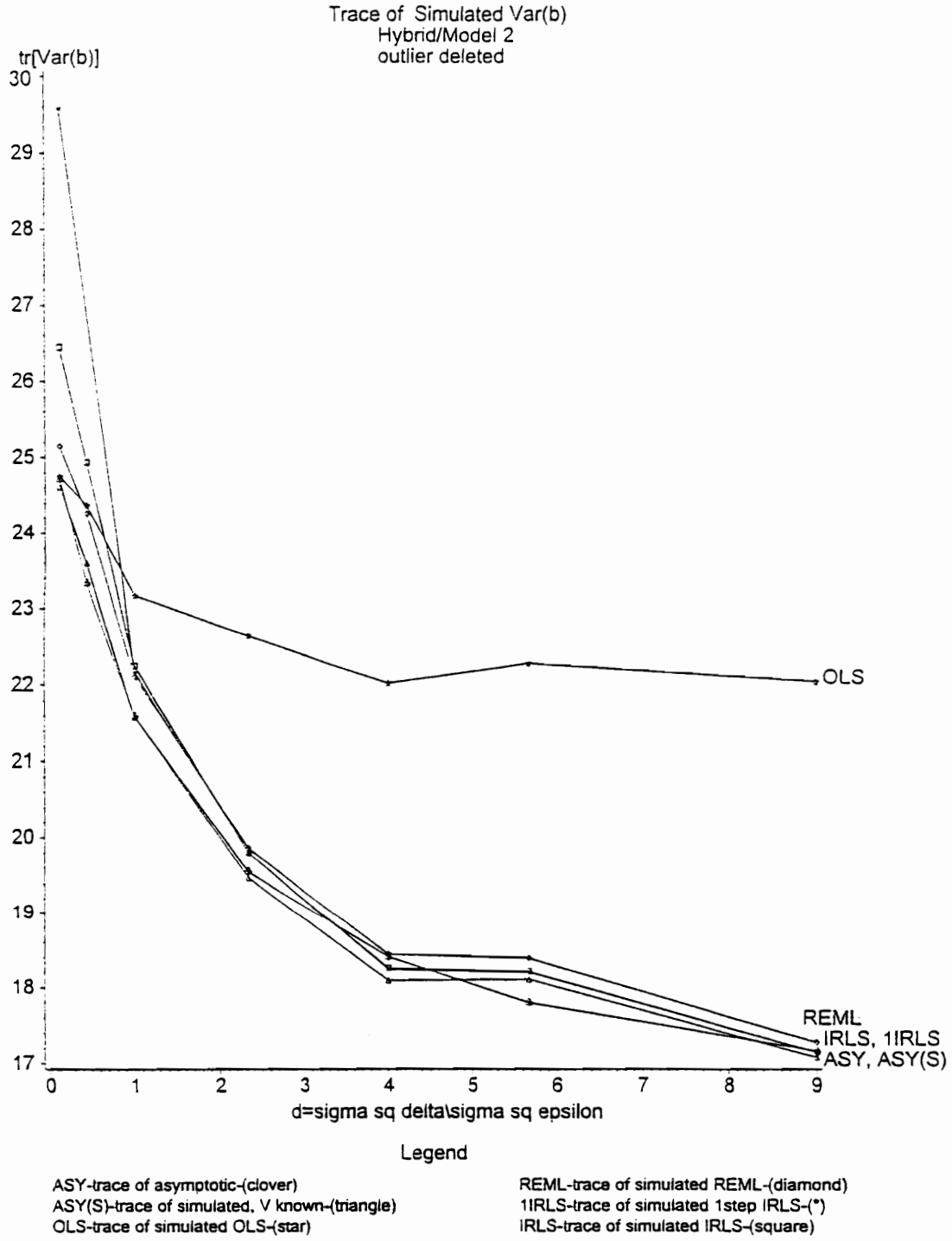


Figure 4.6.4 Asymptotic/Simulated Trace[Var($\widehat{\beta}$)] for Hybrid 416A/Model2 for Various Estimation Methods

Examination of the trace plots (with the outlier in the case of the hybrid design for d=.43 for 1IRLS deleted) given in Figures 4.6.3 and 4.6.4 shows some contrasts to the conclusions drawn for Model 1. Recall for Model 1, the OLS method of estimation of V proved, especially for the ccd, to be competitive with the other methods and also competitive for the hybrid for d<1. For Model 2, however, as was also illustrated by the determinant criterion, OLS estimation can only be considered competitive for very small d and definitely should be considered inefficient for the larger d values. As was conjectured earlier, the performance of OLS can be most likely attributed to the fact that more terms in Model 2 are affected by V. The other estimation methods provide values that are close to the trace of the asymptotic Var($\hat{\underline{\beta}}$), especially for large d. Since all the 1IRLS, IRLS, and REML methods are very competitive for the larger d, the choice of an overall winner for this model situation must rely on the performance at the smaller values of d. The REML method gives results that are closest to the asymptotic trace in both designs for the smaller d values, and hence is the preferable procedure for this model situation. Once again, this will become more evident through the examination of the simulation errors. One note of caution, however, concerning the REML estimators. Through this simulation, the REML on occasion would produce "wild" estimators for σ_δ^2. The experimenter must be aware of the possibility of such estimates when using this method.

An additional investigation was conducted on the same Model 2 scenario as given above with the exception that now $\sigma_\delta^2+\sigma_\epsilon^2=50$. Similar results to those found using the sum=20 were obtained indicating that the magnitude of the variance of a single observation does not influence the characteristics of the variance estimating procedures. The only real observable difference came in the trace of Var($\hat{\underline{\beta}}$) plots. For the larger sum of the error variances, it took longer for the simulated traces for all the methods to converge to the asymptotic trace, especially the 1IRLS traces. For larger errors, 1IRLS should not be considered as a good alternative for the hybrid variance component estimation.

One of the primary goals of this study was to determine the relationship between the simulated trace and determinant of Var($\hat{\underline{\beta}}$) for the various estimation

procedures and small designs to the respective asymptotic values. Throughout this discussion, the relationship has been described in terms of "closeness" to the asymptotic results. The simulation standard errors given for various combinations of model, designs, d, and variance sum for each estimation method can help quantify the meanings of "close".

Table 4.6.5 Simulation Standard Errors

	d	1IRLS	IRLS	REML	OLS	ASYM
Model1-hybrid						
$\sigma_\delta^2+\sigma_\epsilon^2$=12:	2.0					
Trace avg		12.23	12.15	12.09	13.42	11.64
se		0.144	0.143	0.157	0.191	-
Det avg		0.265	0.256	0.252	0.406	0.204
se		0.015	0.014	0.014	0.014	-
Model2-ccd						
$\sigma_\delta^2+\sigma_\epsilon^2$=12:	0.43					
Trace avg		15.19	15.63	14.97	15.35	14.71
se		0.154	1.04	.099	.084	-
Det avg		0.469	0.591	0.452	0.478	0.549
se		0.022	0.25	0.098	0.029	-
Model2-hybrid						
$\sigma_\delta^2+\sigma_\epsilon^2$=12:	0.43					
Trace avg		132.4	24.60	23.86	24.04	23.34
se		145.66	0.31	0.275	0.24	-
Det avg		20433.8	525.46	367.32	370.0	292.54
se		27060.9	76.47	43.7	42.48	-

Model2-hybrid						
$\sigma_\delta^2+\sigma_\epsilon^2$=12:	2.3					
Trace	avg	19.91	19.94	20.00	22.75	19.55
	se	0.182	0.187	0.188	0.267	-
Det	avg	3.75	3.87	3.99	10.62	3.00
	se	0.295	0.305	0.321	1.119	-

Model2-ccd						
$\sigma_\delta^2+\sigma_\epsilon^2$=50:	4.0					
Trace	avg	35.15	34.79	34.76	41.80	34.65
	se	0.78	0.35	0.35	0.72	-
Det	avg	21.15	20.37	20.02	107.6	18.51
	se	2.75	1.77	1.54	15.10	-

Model2-hybrid						
$\sigma_\delta^2+\sigma_\epsilon^2$=50:	4.0					
Trace	avg	47.3	46.74	47.10	56.22	46.01
	se	0.88	0.62	0.67	0.50	-
Det	avg	18553.2	16836	19155.9	83497.3	14140.6
	se	2653.4	979.05	1493.6	4429.6	-

The magnitude of the standard errors first of all clearly suggests that a set of 2000 iterations is sufficient in size. In fact, in some cases, 2000 iterations may perhaps be too many. The table helps illustrate the closeness of asymptotic values to the resulting simulated trace and determinant values. For the smaller values of d, the asymptotic values, on the most part, consistently fall within two standard errors of the OLS average criteria indicating a "closeness" of values. As d rises, however, this is no longer true. Similarly, the asymptotic criteria also consistently fall within approximately two standard errors for the REML estimators when one uses the ccd. Even in the case of the hybrid when the REML (or even IRLS) does not always fall within $\pm$ two standard errors of the asymptotic result for the trace, the asymptotic result is truly a close

approximation, and the "small" standard errors may suggest misleading conclusions. This "closeness" of values also holds for 1IRLS and IRLS for the larger d values.

From this investigation into the variance estimating procedures for the specific cases of the central composite and hybrid designs and the given second order models, some recommendations may be made with respect to the "best" estimating procedure for V. The OLS procedure remarkably proved to be a useful alternative for second order models with limited quadratic terms and for designs such as the ccd whose model terms are minimally influenced by the bi-randomization structure. For complete or near complete second order models and designs such as the hybrid which are more sensitive to the bi-randomization structure, the OLS procedure only performs well for values of $d<1$ as expected due to the closer resemblance to a completely randomized scenario.

With respect to the other methods and more importantly, for "stronger" bi-randomization situations ($d>1$), the REML and IRLS estimating procedures are quite competitive and give results that are close to the asymptotic values for both design/model combinations. The method of REML, however, is better able to handle estimation in cases with small d values, making it the "best" method over all considered designs, models, and d. If the experimenter should have some prior information on the relationship of the error variances, $\sigma_\delta{}^2$ and $\sigma_\epsilon{}^2$, and believes that the ratio is small, the REML or even the OLS methods are the useful estimation alternatives. If on the other hand, the belief is that the ratio is large, he/she would be best served using either the REML or IRLS estimating procedures. For more extreme values of d, the 1IRLS cannot be ruled out for certain designs.

This investigation is by no means an exhaustive study. It only provides the experimenter with some reference from which to make a decision. Before a general recommendation can be made for the pool of response surface designs and models, a more extensive and complete simulation study must be performed.

In addition to studying competitive estimation procedures for V, a useful relationship between the traces and determinants of the simulated $\text{Var}(\hat{\underline{\beta}})$ for these small response surface designs and the asymptotic trace and determinant was discovered. For moderate to large values of d (sometimes even for smaller d), the trace and determinant criteria for $\text{Var}(\hat{\underline{\beta}})$ for both the ccd and hybrid designs were remarkably consistent with the asymptotic $\text{Var}(\hat{\underline{\beta}})$. As a result, it is safe to make the assumption that for designs such as the ccd and hybrid, $\text{Var}(\hat{\underline{\beta}}) \approx (X^*V^{-1}X^*)^{-1}$. This closeness of small design and asymptotic results is quite useful due to the fact that for small designs such as these investigated the $\text{Var}(\hat{\underline{\beta}})$ is unknown due the iterative nature of the estimators of V. Thus, the asymptotic $\text{Var}(\hat{\underline{\beta}})$ may be used as a close approximation to $\text{Var}(\hat{\underline{\beta}})$ for the standard response surface designs.

§4.7 **Asymptotic Properties of Estimators of Coefficients**

Under the normality assumption and known variance-covariance matrix V, the MLE for $\hat{\underline{\beta}}$ is given by $(X^*V^{-1}X^*)^{-1}X^*V^{-1}\underline{y}$. The matrix V, however, is rarely known and thus, must be estimated using one of the procedures described. Once estimated, $\hat{V}$ is substituted into the MLE expression to give

$$\hat{\underline{\beta}} = (X^*\hat{V}^{-1}X^*)^{-1}X^*\hat{V}^{-1}\underline{y}. \qquad (4.7.1)$$

The REML estimators for σ_δ^2 and σ_ϵ^2 are consistent estimators as are all MLEs. Similarly, estimators for σ_δ^2 and σ_ϵ^2 using the 1IRLS or IRLS methods are also consistent assuming the sample size grows in both whole-plots units and sub-plot units alike. As the sample size grows for both units, more degrees of freedom become available for lack of fit and/or replication sum of squares. Assuming lack of fit accurately measures whole-plot and sub-plot error, as the sample size grows large, the estimators converge in probability to the parameters, i.e.

$$\hat{\sigma}_\delta^2 \xrightarrow{\mathcal{P}} \sigma_\delta^2$$

and

$$\hat{\sigma}_\epsilon^2 \xrightarrow{\mathcal{P}} \sigma_\epsilon^2.$$

Based on the consistency of $\hat{V}$, the asymptotic variance-covariance matrix for the estimator discussed is given by

$$\mathrm{Var}(\hat{\underline{\beta}}) = (X^{*'}V^{-1}X^{*})^{-1}. \qquad (4.7.2)$$

From the earlier exploration in §4.6, $\mathrm{Var}(\hat{\underline{\beta}})$ for small designs such as the central composite and hybrid is close (at least in terms of trace and determinant) to that of the the asymptotic result. Thus, for model editing purposes the estimated asymptotic variance-covariance matrix may be used as a good approximation to the true by unknown matrix, i.e.

$$\hat{\mathrm{Var}}(\hat{\underline{\beta}}) = (X^{*'}\hat{V}^{-1}X^{*})^{-1}. \qquad (4.7.3)$$

The diagonal elements of 4.7.3 can then be used as denominators values in coefficient testing procedures.

The close approximation of $\mathrm{Var}(\hat{\underline{\beta}})$ to the asymptotic variance for small designs will be utilized extensively in the next chapter which addresses the efficiency of standard second order response surface designs under bi-randomization error structures. The asymptotic variance of the model estimators will be used in the optimality criteria expressions to circumvent the necessity for extensive simulation for comparison of the small designs.

CHAPTER 5

DESIGN OPTIMALITY

A unique characteristic of RSM is that the experimenter actually has the ability to choose the design he or she wishes for experimentation. As was mentioned in Chapter 1, this characteristic spawned an area of research called design optimality. Design optimality aids the experimenter in his choice of design by selecting a collection of design points that provides an "overall" best design with respect to a prespecified criterion. Common criteria used in optimal design selection include D and Q. This letter notation has given rise to the reference of alphabetic optimality criteria, and note that D and Q are just two out of many available criteria.

The D and Q criteria are direct functions of the $\text{Var}(\hat{\underline{\beta}})$. Thus, in the presence of a completely randomized structure they are functions of $(X'X)^{-1}$ which is a function purely of design. Under a bi-randomization error control format, however, these criteria are functions of $(X'V^{-1}X)^{-1}$ assuming V is known. The nuisance parameters, ${\sigma_\delta}^2$ and ${\sigma_\epsilon}^2$, are nested within the criteria. While many efficient designs have been developed for the completely randomized error structure, the natural question is whether these designs perform well under a bi-randomization structure. In other words, standard response surface designs may require, for each criterion, modification to better accommodate the dependent relationship among observations within a whole-plot found in a BRD.

The investigation into design efficiency under bi-randomization error structures will focus on altering standard completely randomized response surface designs. Under the bi-randomization structure, the overall composition of the standard CR design in terms of sample size and other properties will remain unchanged, but based on the resulting V matrix, the levels of the design variables will be chosen to gain efficiency. These resulting levels can then be compared to the standard levels used in the CRDs. Both crossed and non-crossed BRDs will be considered.

§5.1 **Design Optimality for Crossed BRD**

Design optimality will be first explored for crossed BRDs. Both the D and Q criteria will be considered.

§5.1.1 D Optimality

The D optimality criterion for design selection is the most widely used and accepted (most likely due to its simplicity) from the alphabetic collection. It originated with Kiefer and Wolfowitz in 1959 and continues to be a popular choice among experimenters for design selection. This criterion concentrates on choosing the design, denoted $\mathfrak{D}$, that minimizes the generalized variance associated with estimating model coefficients. For standard RSM, D optimality solves the criterion

$$\min_{\mathfrak{D}} |(X^{*\prime}X^{*})^{-1}| \qquad (5.1.1)$$

where X^{*} is the model matrix for the CRD. Equation 5.1.1 is derived from $\mathrm{Var}(\hat{\beta})=\sigma_{\theta}^{2}[X^{*\prime}X^{*}]^{-1}$.

For a BRD, the D optimality criterion must reflect the dependent covariance structure of the observations and thus, becomes

$$\min_{\mathfrak{D}} |(X^{*\prime}V^{-1}X^{*})^{-1}|. \qquad (5.1.2)$$

Expression 5.1.2 can also be written as

$$\max_{\mathfrak{D}}|(X^{*\prime}V^{-1}X^{*})|. \tag{5.1.3}$$

The variations in precision of coefficient estimation in a BRD suggests intuitively that their D optimal designs may differ from D optimal CRDs to compensate for the unequal variances and nonzero covariances.

Recall from Chapter 3 that for a crossed BRD with a model of any order, the ordinary least squares and generalized least squares equations were proven to be equal based upon the existence of a G matrix such that

$$X^{*\prime}V^{-1}=GX^{*\prime}.$$

Using this equality, 5.1.3 becomes

$$\max_{\mathfrak{D}}|GX^{*\prime}X^{*}|=\max_{\mathfrak{D}}|G|\ |X^{*\prime}X^{*}|. \tag{5.1.4}$$

The matrix G is a (p+1)*(p+1) lower triangular matrix and therefore,

$$|G|=\prod_{i=1}^{p+1} g_i=\left[\frac{(1\text{-}cb)}{\sigma_\epsilon^{\,2}}\right]^{k+1}.$$

Since $c=\dfrac{\sigma_\delta^{\,2}}{\sigma_\epsilon^{\,2}+b\sigma_\delta^{\,2}}$ and b are both constants, 5.1.4 is equivalent to choosing $\mathfrak{D}$ to

$$\max_{\mathfrak{D}}|(X^{*\prime}X^{*})|.$$

Consequently, the D optimal BRD is equal to the D optimal CRD. No modifications are necessary.

§5.1.2 D_s Optimality for Robust Parameter Design

Often the entire vector of parameters is not of primary interest in an analysis; instead, only a subset of parameters may be the focus. For this

situation, a design should be chosen to concentrate on the estimation of this subgroup. Kiefer [1961] and then Karlin and Studden [1965] modified the D optimality criterion found in §5.1.1 to reflect this change of focus. This criterion is called D_s optimality, and it chooses the design, $\mathfrak{D}$, that minimizes the determinant of the respective submatrix of $\mathrm{Var}(\hat{\underline{\beta}})$ corresponding to the subset of desired model terms. Suppose the model matrix X^* is partitioned into $[X_1|\ X_2]$ where X_2 contains the subset of model terms of interest. The D_s optimality criterion for a CRD is given by

$$\min_{\mathfrak{D}} |\ [X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2]^{-1}|. \tag{5.1.5}$$

This particular criterion has a natural application in robust parameter design. Recall, the driving force behind controllable process variability through design are the modeled control*noise interactions. Designs should, thus, be chosen for experimentation which provide precise information about these model terms. This can be accomplished using D_s optimality on the subset of control*noise interactions. Alteration of 5.1.5 to incorporate the covariance structure of a crossed BRD is based on the variance-covariance expressions derived for estimated coefficients in §3.4.1. Assuming the noise variables form the levels of the whole-plot variable and the control variables are the sub-plot variables, the D_s optimality criterion for the noise*control interactions (sub-plot terms) becomes

$$\min_{\mathfrak{D}} \left| \begin{bmatrix} c_{m,m} & c_{m,m+1} & \cdots & c_{m,t} \\ c_{m+1,m} & \cdots & \cdots & c_{m+1,t} \\ \vdots & \cdots & \ddots & \vdots \\ c_{m,t} & \cdots & c_{t,t-1} & c_{t,t} \end{bmatrix} \cdot \sigma_\epsilon^2 \right| \tag{5.1.6}$$

where $c_{i,j}$ is the i^{th} row and j^{th} column entry of $[X^{*\prime}X^*]^{-1}$. This determinant is proportional to the respective noise*control interaction submatrix of $\sigma_\theta^2[X^{*\prime}X^*]^{-1}$ from the CRD analysis. Thus, if sub-plot terms are of primary interest, the D_s optimal design is invariant to the error control structure.

When a subset combination of whole-plot and sub-plot model terms are of interest in a BRD, the D_s optimality criterion does not easily transform into a function of $[X^{*\prime}X^{*}]$ as it did in 5.1.6. Intuitively, however, by examining the different components of the submatrix determinant, it appears, at least in the first order situation, to have the same goals as the minimization of the same criterion for a CRD. The goals are to push variable levels out to extremes and to make model term contrasts orthogonal. See Myers [1976] and Box and Draper [1987].

§5.1.3 Q Optimality

The Q optimality criterion is another useful tool for choosing a "good" design to implement. Unlike D optimality, Q optimality focuses on choosing the design, $\mathfrak{D}$, which provides an "overall" minimum prediction variance throughout the researcher's region of interest. "Overall" typically refers to a type of average prediction variance. For a CRD with model $\underline{y}=X\underline{\beta}+\underline{\theta}$ where $\underline{\theta} \sim N(0,\sigma_\theta{}^2 I)$, the prediction variance at the point $\underline{x}_0$ is given by

$$\mathrm{Var}[\hat{\underline{y}}]=\underline{x}_0{}^{\prime d}[X'X]^{-1}\underline{x}_0{}^{d}\cdot\sigma_\theta{}^2 \qquad (5.1.7)$$

where $\underline{x}_0{}^d$ is the model vector for the given point. The Q criterion does not attempt to minimize the prediction variance at each $\underline{x}_0$ but instead minimize the prediction variance average over all $\underline{x}_0$ in the region of interest. Formally, the Q criterion is given by the expression

$$\min_{\mathfrak{D}} \frac{N}{k}\int_{\mathfrak{R}} \underline{x}^{\prime d}[X'X]^{-1}\underline{x}^{d}\, d\underline{x} \qquad (5.1.8)$$

where N is the total sample size, $\mathfrak{R}$ is the region of interest, and k is the volume of that region.

Now consider a crossed bi-randomization design with model $\underline{y}=X^{*}\underline{\beta}+\underline{\epsilon}^{*}$ where $\underline{\epsilon}^{*}=\underline{\delta}+\underline{\epsilon} \sim N(0,V)$. The prediction variance at a given point $\underline{x}_0$ is now given by

$$\mathrm{Var}[\hat{\underline{y}}]=\underline{x}_0^{'d}[X^{*'}V^{-1}X^*]^{-1}\underline{x}_0^d \tag{5.1.9}$$

and consequently, the Q criterion becomes

$$\min_{\mathfrak{D}} \frac{N}{k}\int_{\mathfrak{R}} \underline{x}^{'d}[X^{*'}V^{-1}X^*]^{-1}\underline{x}^d \, d\underline{x} \tag{5.1.10}$$

or equivalently,

$$\min_{\mathfrak{D}} \text{trace}\left[\frac{1}{k}\int_{\mathfrak{R}} \underline{x}^d \, \underline{x}^{'d} d\underline{x}\right]\cdot\left[N[X^{*'}V^{-1}X^*]^{-1}\right]. \tag{5.1.11}$$

Notice that 5.1.8 for the CRD is not a function of the error variance. For the BRD, however, the error variances are nested within the criterion given by 5.1.10 and 5.1.11.

Since the Q criterion is model driven, in order to see the influence of V on design choice a specific model must be selected. Consider a two variable, one whole-plot, z, and one sub-plot variable, x, crossed BRD with a first order model involving the intercept, z, x, xz. Assuming the region is symmetric and using the variance and covariance expressions derived in Chapter 3 for the crossed BRD coefficient estimates, 5.1.11 becomes

$$\begin{aligned} &N[c_{b_0,b_0}[\sigma_\epsilon{}^2+b\sigma_\delta{}^2]+c_{b_0,x}[\sigma_\delta{}^2]\sum_{wp} x + c_{b_0,xz}[\sigma_\delta{}^2]\sum_{wp} xz + \tfrac{1}{3}\, c_{z,z}[\sigma_\epsilon{}^2+b\sigma_\delta{}^2] \\ &+\tfrac{1}{3}c_{z,x}[\sigma_\delta{}^2]\sum_{wp} x + \tfrac{1}{3}\, c_{x,x}[\sigma_\epsilon{}^2] + \tfrac{1}{9}\, c_{xz,xz}[\sigma_\epsilon{}^2] \end{aligned} \tag{5.1.12}$$

where

$c_{i,i}$ = diagonal element of $[X^{*'}X^*]^{-1}$ associated with i model term

c_{b_0,b_0} = diagonal element of $[X^{*'}X^*]^{-1}$ associated with intercept

$c_{i,j}$ = off-diagonal element of $[X^{*'}X^*]^{-1}$ associated with ij interaction model term.

To minimize 5.1.12 with respect to design, the diagonal and off-diagonal elements of $[X^{*'}X^*]^{-1}$ should be minimized. In addition, the sub-plot terms should be

orthogonal to whole-plots. These two goals can be achieved simultaneously by spreading the levels of each variable to its extremes and making model term contrasts orthogonal. These are the same goals used to create first order Q optimal designs under a completely randomized structure. Thus, the presence of V has no influence on first order crossed BRDs optimal design levels.

For first order crossed BRDs, all three alphabetic criteria failed to be influenced by the BRD's dependent covariance structure. In addition, D optimality was not influenced for a crossed BRD of any order. Because of the preservation of the equality of the OLS and GLS equations in the two-level non-crossed fractional factorial BRDs, they too are unaffected in terms of optimality by the presence of the V matrix. Second order crossed and non-crossed BRDs will be considered in the next section.

§5.2 **Design Optimality for Second Order Response Surface Designs**

Exploration into design optimality for second order crossed and non-crossed BRDs is more complex than that presented in the previous section for first order designs. Recall for second order non-crossed BRDs, the equivalency of the OLS and GLS estimating equations does not universally hold. This implies that $X^{*'}V^{-1}$ cannot be rewritten as $GX^{*'}$ which was the basis of the design optimality investigation in §5.1. Because this equivalency does not always hold a different approach to design optimality for second order BRDs must be taken. Instead of arbitrarily trying to "improve" the efficiency of standard second order response surface designs run under a bi-randomization error structure, the design optimality investigation will begin with a comparison of efficiencies among both crossed and non-crossed second order standard RS designs. The efficiencies of the second order designs will be compared for varying strengths of the dependency among observations within a whole-plot. This investigation will provide an insight into desirable design characteristics for second order bi-randomization experimentation.

A collection of twelve standard second order response surface designs run

under a bi-randomization experimental structure will be compared. Included in the collection are Box-Behnken, hybrid, 3^k full factorial, and central composite designs. Variations on each design constitute the twelve total designs.

Two efficiency measures will be used for comparison. Each is based upon the variance of the estimated vector of model coefficients, but recall, in Chapter 4 that this variance, although unknown, is well approximated by the asymptotic var($\hat{\underline{\beta}}$) expression. The measures will therefore be functions of $(X^{*'}V^{-1}X^{*})^{-1}$. The first measure is based upon the D optimality criterion. For each of the twelve BRDs,

$$\frac{\left|(X^{*'}V^{-1}X^{*})\right|^{\frac{1}{p+1}}}{N} \qquad (5.2.1)$$

will be calculated for various combinations of σ_δ^2 and σ_ϵ^2. The $\text{var}(y_{ij})=\sigma_\delta^2 + \sigma_\epsilon^2$ will be held fixed to one, but the values of the error variances will individually be varied to reflect different "strengths" for the bi-randomization structure. The value of 5.2.1 provides a per observation measure of the D optimality criterion. The overall goal is to maximize this expression. In addition, comparisons will be made in terms of an approximate Q optimality criterion. Recall that the Q optimality criterion involves an integrated prediction variance over a region of interest. An approximation utilizing the asymptotic var($\hat{\underline{\beta}}$) is given by

$$N\left[\frac{\sum_{\underline{x}_0 \text{ in } \mathcal{R}} \underline{x}_0^{d'}(X^{*'}V^{-1}X^{*})^{-1}\underline{x}^{d}_{0}}{K}\right] \qquad (5.2.2)$$

where $\underline{x}_0$ is a point in the design region of interest, $\underline{x}_0^d$ is the respective model vector for that design point, and K is the total number of $\underline{x}_0$ combinations used in the sum. Once again, the error variances will be held to a fixed sum, but their ratio, d, will be allowed to vary. A sizable grid search over $\mathcal{R}$ is used to calculate 5.2.2. Unlike expression 5.2.1, the goal is to minimize the 5.2.2 values. For

evaluation purposes of both criteria, V is known.

To be able to fairly compare the twelve designs with respect to these measures, each design will be scaled to the same region of interest. This was done by scaling each design such that the outermost design point for each design is on the same radius. Note that when the designs are presented they will be in the unscaled, but more familiar notation. All calculations, however, are based on the scaled versions.

Comparisons among the set of second order BRDs will be performed for four different variable combinations. The variable combinations included are: (i) one whole-plot variable/two sub-plot variables, (ii) one whole-plot variable/three sub-plot variables, (iii) two whole-plot variables/one sub-plot variable, and (iv) two whole-plot variables/two sub-plot variables. Later results will indicate that the number of whole-plot variables influences the resulting efficient designs. For this dissertation, situations with only one and two whole-plot variables will be examined for three and four variable designs. For each variable combination, a second order model including all first order interactions and quadratic terms will be used.

§5.2.1 One Whole-plot Variable/Two Sub-plot Variables

The first variable combination has one whole-plot variable, z_1, and two sub-plot variables denoted x_1 and x_2. The second order model under investigation is

$$E[y_{ij}]=\beta_0+\beta_1 z_1+\beta_2 z_1^2+\beta_3 x_1+\beta_4 x_2+\beta_5 x_1 x_2+\beta_6 x_1^2+\beta_7 x_2^2+\beta_8 x_1 z_1+\beta_9 x_2 z_1.$$

As previously mentioned, twelve standard RS designs run under a bi-randomization error structure will be examined with respect to these variables and this model. The designs include a 3^3 full factorial; 3 variable standard (full factorial portion) ccds and α=1, 1.682, and 2; 3 variable small composites (a ccd with a half fraction in the factorial portion) and α=1, 1.682, and 2; 3 variable

Box-Behnken designs with 1 and 3 center runs; hybrid 310 designs with 0 and 1 center runs; and finally, a hybrid 311b design. The respective design matrices are given in Appendix F. For each design, the D and Q criterion measures were evaluated at ten levels of $d=\sigma_\delta^2/\sigma_\epsilon^2$ ranging from d=0 to d=30. The results for the D criterion comparison are given in Figures 5.2.1 and 5.2.2 while the Q criterion results are found in Figures 5.2.3 and 5.2.4. In each pair of figures, the first plot shows a range on d from 0 to 1 whereas the second plot displays the entire range of d. Also note that the side axis corresponds to the D criterion in Figures 5.2.1 and 5.2.2 whereas the bottom axis displays the $d=\sigma_\delta^2/\sigma_\epsilon^2$ values.

While examining Figures 5.2.1 and 5.2.2 one must remember that the goal is to maximize the D criterion value. Beginning with Figure 5.2.1, for d=0, (a CRD), the standard designs cluster together with competitive D values except for the small composite design with α=1 which is known not to be an efficient completely randomized design. The full factorial design ranks low for d=0 due to its larger sample size. As d begins to slowly grow, the efficiency ranking of the designs begin to change. Efficiency ranking refers to how the given design's criterion values compare to the those of other designs. The D value for the 3^3 full factorial initially declines as d grows from 0 to 0.25 indicating an initial loss in efficiency as compared to the other designs under the stronger bi-randomization structure. After the initial drop, the full factorial design does slowly rise in D value but never regains a competitive efficiency ranking. On the contrary, over the same range the Box-Behnken design with 3 center runs, the standard ccd with α=1.682, the standard ccd with α=2, and the hybrid 310 with 1 center run all appear to be improving in terms of their efficiency rankings as d gets larger. Figure 5.2.2 illustrates the continuation of that trend. As d becomes larger indicating a stronger dependency within each whole-plot, the most efficient class of designs for the bi-randomization structures contains the ccds with α=1.682 and 2, with the Box-Behnken designs with 1 and 3 center running close behind. According to the per observation D criterion, these designs are the most efficient under a fairly strong bi-randomization error structure. The hybrid designs, even though they are competitive for small d, become inefficient alternatives as d increases. The small composite designs and the full factorial design also proved themselves to be inefficient over a broad range of d.

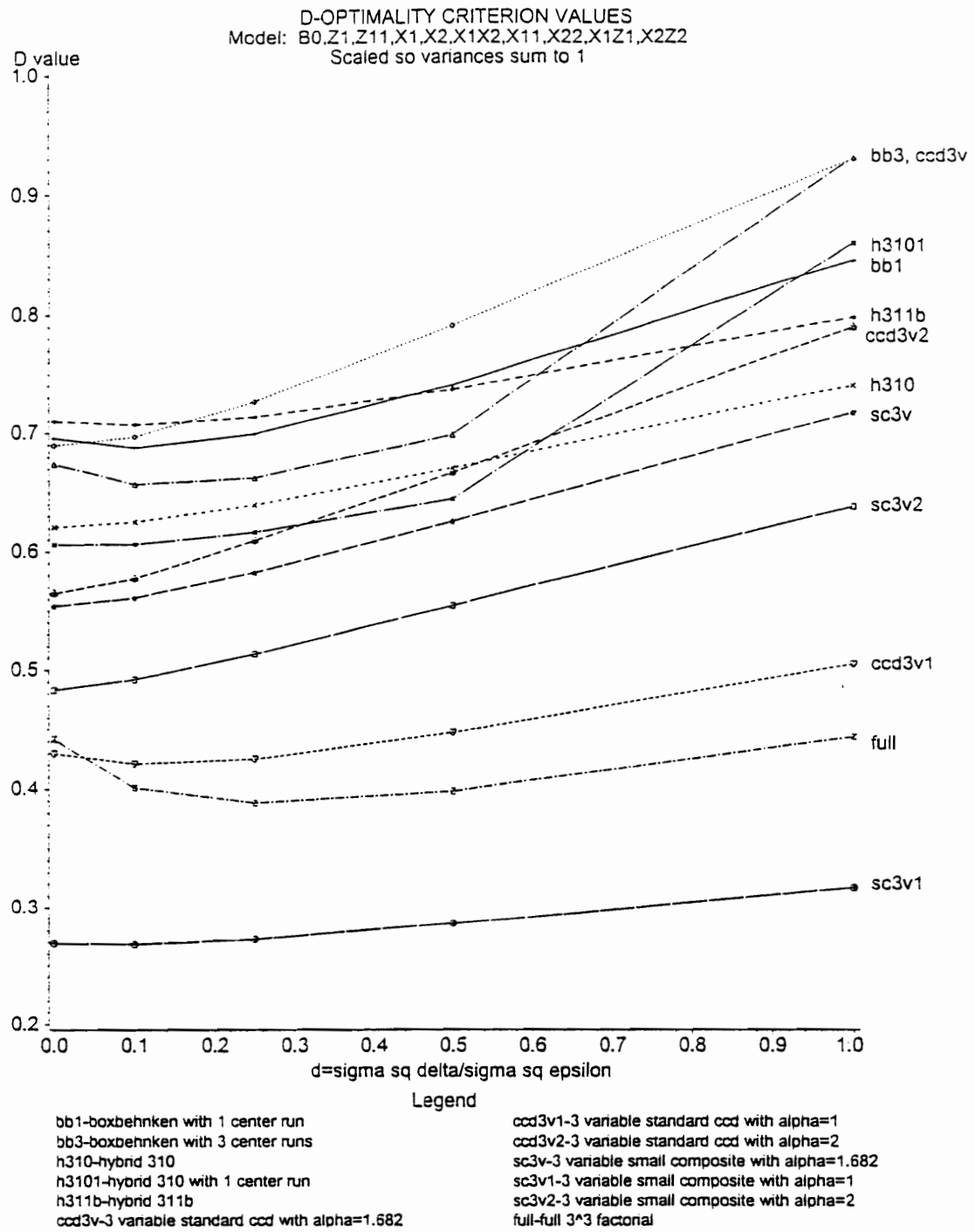


Figure 5.2.1 D Criterion Values for One Whole-plot/Two Sub-plot Variables d=0 to 1

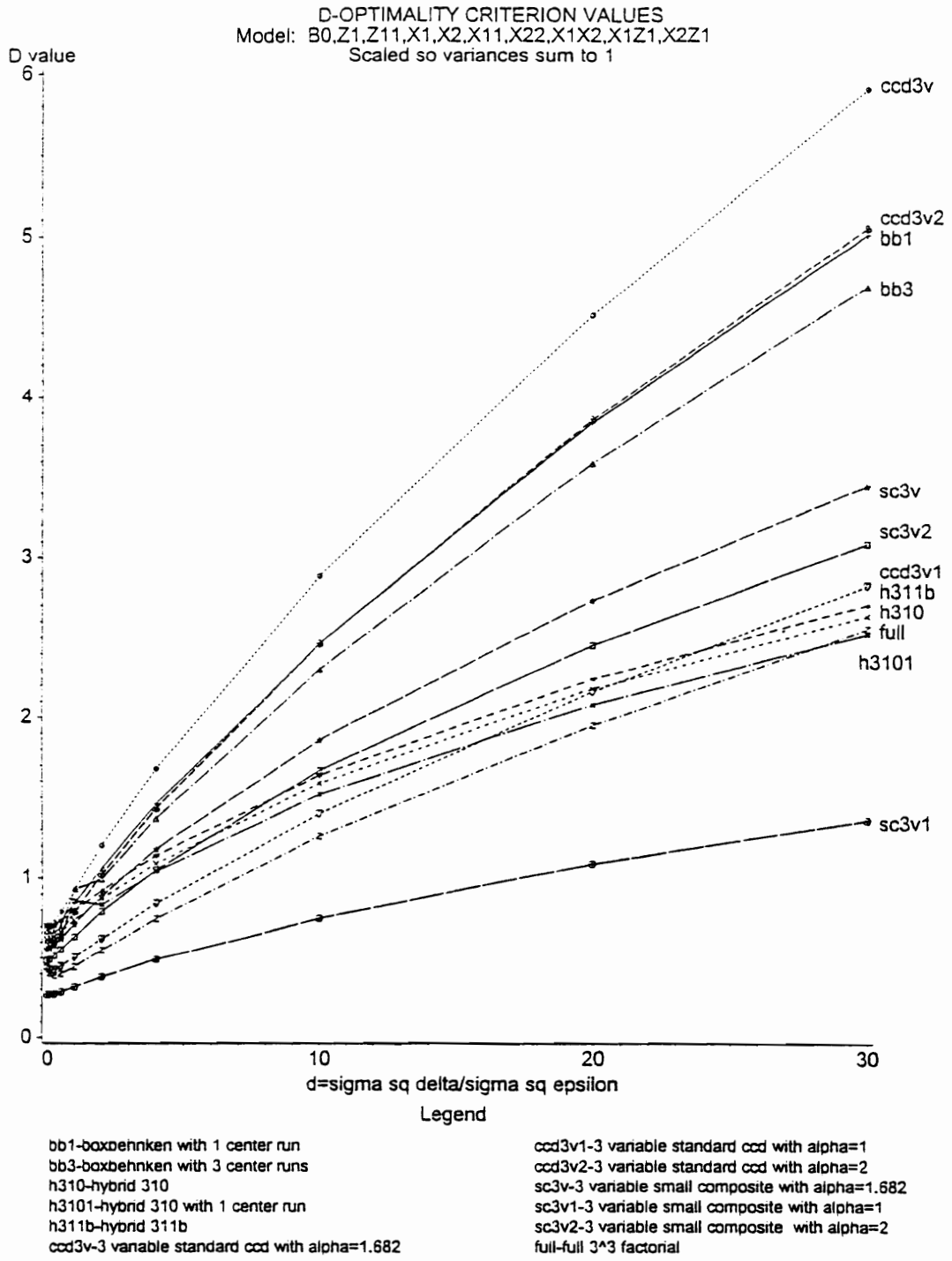


Figure 5.2.2 D Criterion Values for One Whole-plot/Two Sub-plot Variables d=0 to 30

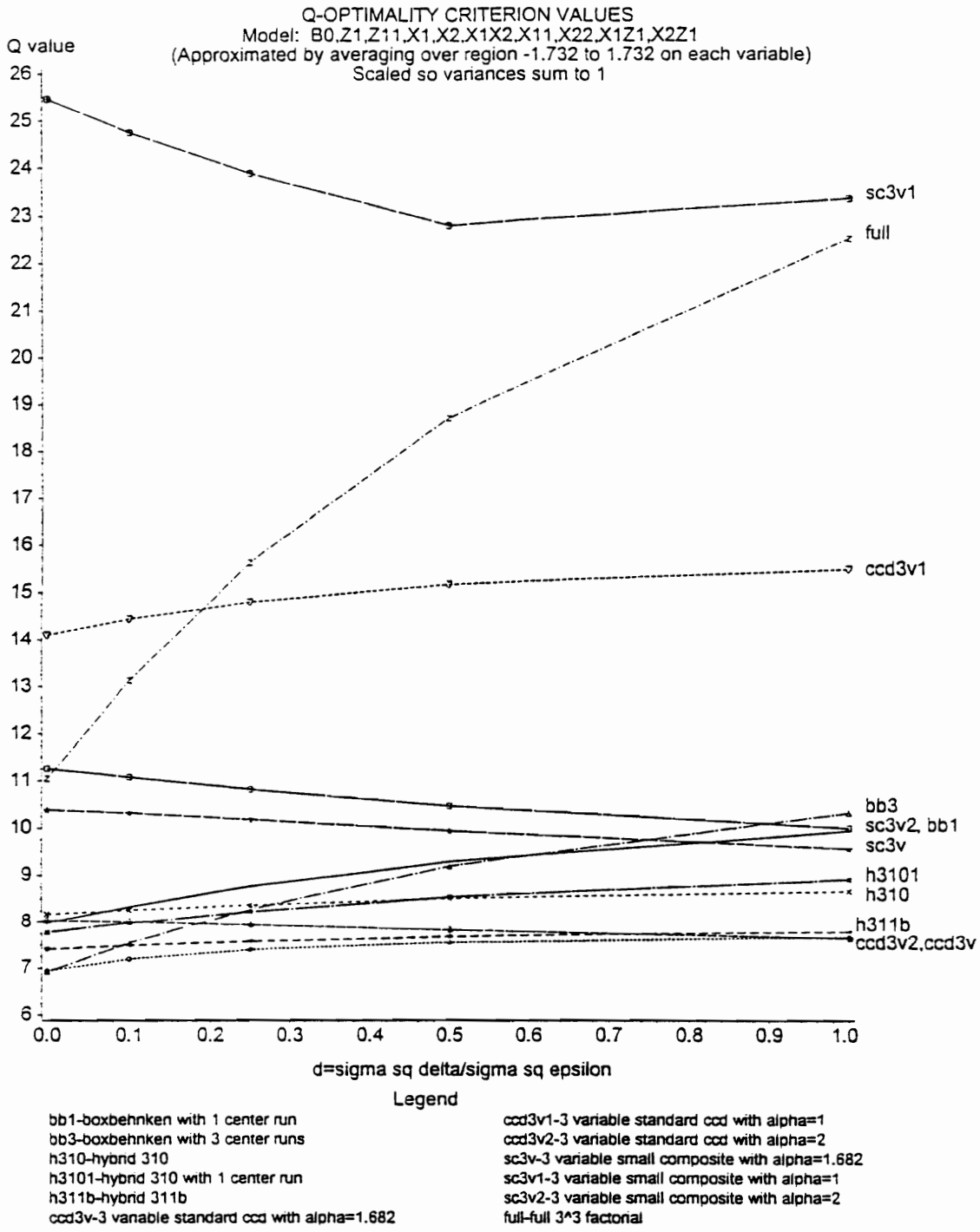


Figure 5.2.3 Q Criterion Values for One Whole-plot/Two Sub-plot Variables d=0 to 1

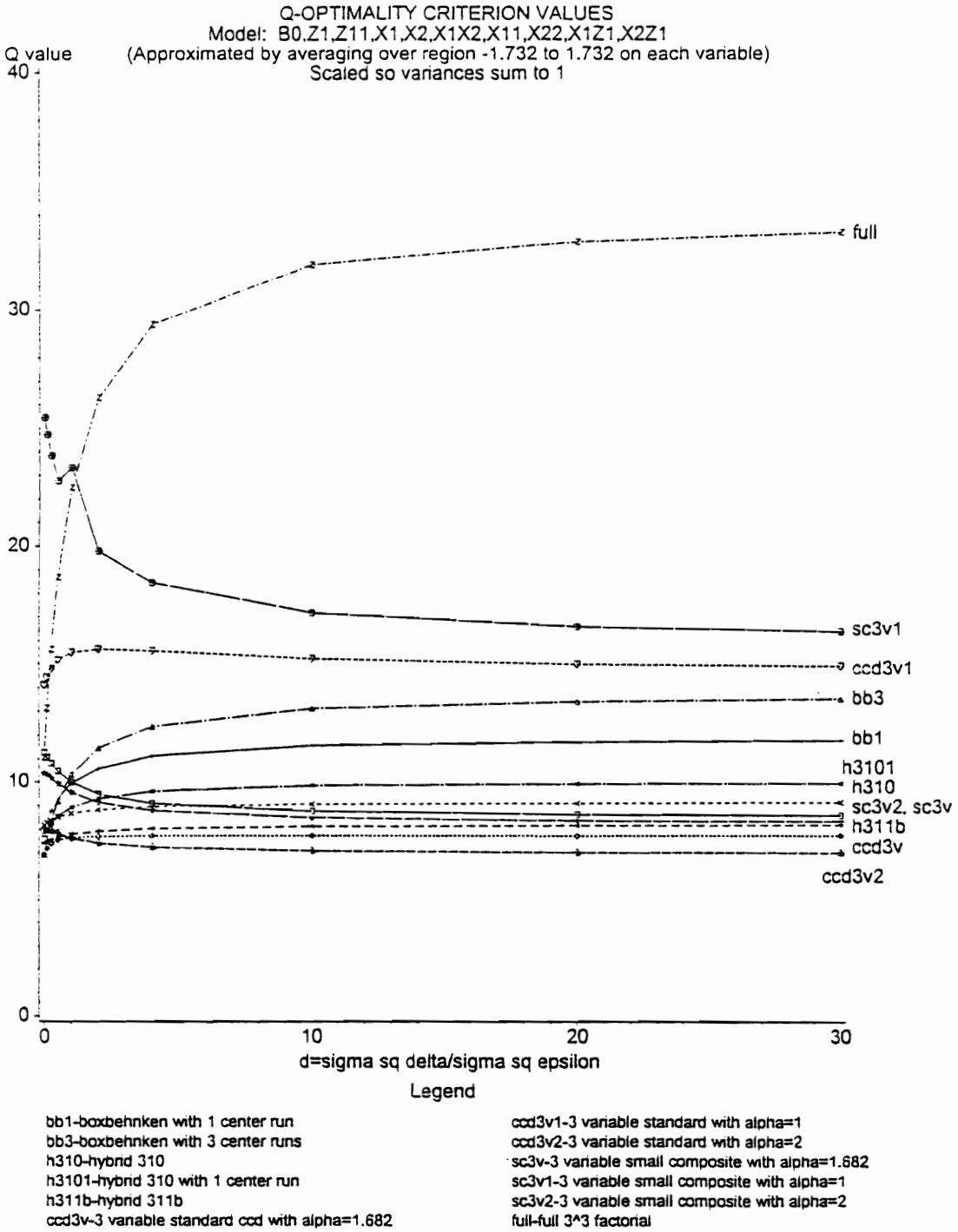


Figure 5.2.4 Q Criterion Values for One Whole-plot/Two Sub-plot Variables d=0 to 30

An alternative ranking of efficient designs according to the Q criterion is found in Figures 5.2.3 and 5.2.4. For these plots, recall the goal is to minimize the Q value. Once again, for d=0 the designs cluster together except for the inefficient ccd and small composite design with α=1. As d rises, the 3^3 full factorial design's Q value quickly begins to rise. At a small d=0.1 value, it can no longer be considered among the competitors of efficient designs. According to the Q criterion, the 3^3 full factorial is very inefficient under most degrees of bi-randomization structure. Even if design size were not a consideration, it would still not be an acceptable option from an efficiency standpoint. The other designs appear to have fairly stable Q values across d=0 to 1 except for both Box-Behnken designs which like the 3^3 also lose efficiency and continue to lose until about d=10. Note that the hybrid designs fare much better in efficiency with respect to the Q criterion than the D criterion. As d continues to rise, the standard ccds with α=1.682 and 2 again prove themselves to be the most efficient under these error control structures along with selected hybrid and small composite designs. When examining these plots, remember that the var(y_{ij}) has been scaled to be equal to one. If a different scaling were used, the differences in Q values would be more exaggerated.

From this collection of plots, based on both criteria, the standard ccds with α=1.682 and 2 provide the most efficient alternatives for this model under the bi-randomization structure for various levels of d. In addition, this investigation shows for small to large d the 3^3 full factorial design can quickly be eliminated from possible design choices, even if sample size is not a concern. It is simply not efficient under the bi-randomization error scheme. Before continuing with another variable combination, note that for the Q criterion at d=1 the most efficient ccd changes. From d=0 to 1, the standard ccd with an α=1.682 is the most efficient, but for d=1-30 the other standard ccd with α=2 is the most efficient. Under a CRD structure, α=1.732 is optimal. This fluctuation in the most efficient axial value for ccds will be explored later.

§5.2.2 One Whole-plot Variable/Three Sub-plot Variables

The example explored in the previous section will now be extended to incorporate one more sub-plot variable, x_3. The second order model becomes

$$E[y_{ij}]=\beta_0+\beta_1 z_1+\beta_2 z_1^2+\beta_3 x_1+\beta_4 x_2+\beta_5 x_3+\beta_6 x_1 x_2+\beta_7 x_1 x_3+\beta_8 x_2 x_3+$$
$$\beta_9 x_1^2+\beta_{10} x_2^2+\beta_{11} x_3^2+\beta_{12} x_1 z_1+\beta_{13} x_2 z_1+\beta_{14} x_2 z_1.$$

For this scenario a similar set of designs will be compared, just now in terms of four variables. Included are once again a 3^4 full factorial design, standard ccds with α=1,2, and 3, small composite designs with α=1,2, and 3, 4-variable Box-Behnken designs with 1 and 3 center runs, hybrid 416A designs with 0 and 1 center runs, and a hybrid 416C design. The designs may be found in Appendix G. The same analysis as outlined in §5.1.1 for the three variable combination was performed. The resulting D and Q plots are given in Figures 5.2.5-5.2.8.

The set of plots for the D criterion illustrate the same results found for the previous three variable example. The 3^4 full factorial design D values initially decrease as d rises from 0 to 0.25 quickly pushing it out and keeping it out of the category of D efficient designs. The other competitive designs show increases in their criterion values over the same range of d. As d continues to grow, the standard ccd with α=2 and the set of Box-Behnken designs (even after an initial decline in efficiency) are once again the most D efficient designs with respect to the bi-randomization error structure. Note that a large portion of the designs appear to have consistent increases in D values across d=0 to 1. Only the Box-Behnken designs, one hybrid design, and of course, the full factorial design seem to be influenced by the dependency among observations. This can be seen by their non-constant increases/decreases in D. For most designs this influence improves the D efficiency as d grows, but for the full factorial, however, the initial loss is never recovered.

Figures 5.2.7 and 5.2.8 for the Q criterion also produce a similar picture to that found in the earlier scenario. For almost the entire range of d, the full

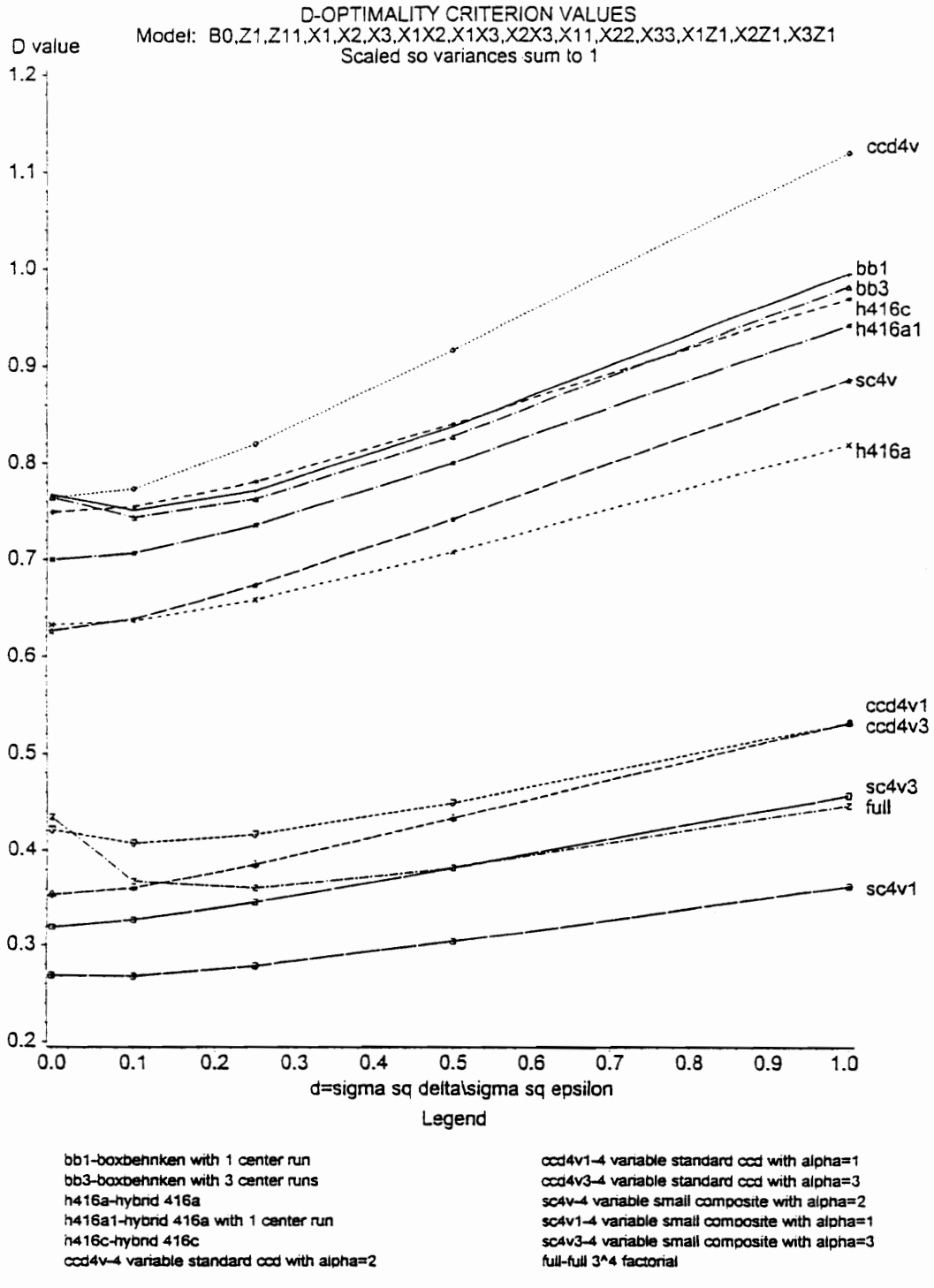


Figure 5.2.5 D Criterion Values for One Whole-plot/Three Sub-plot Variables d=0 to 1

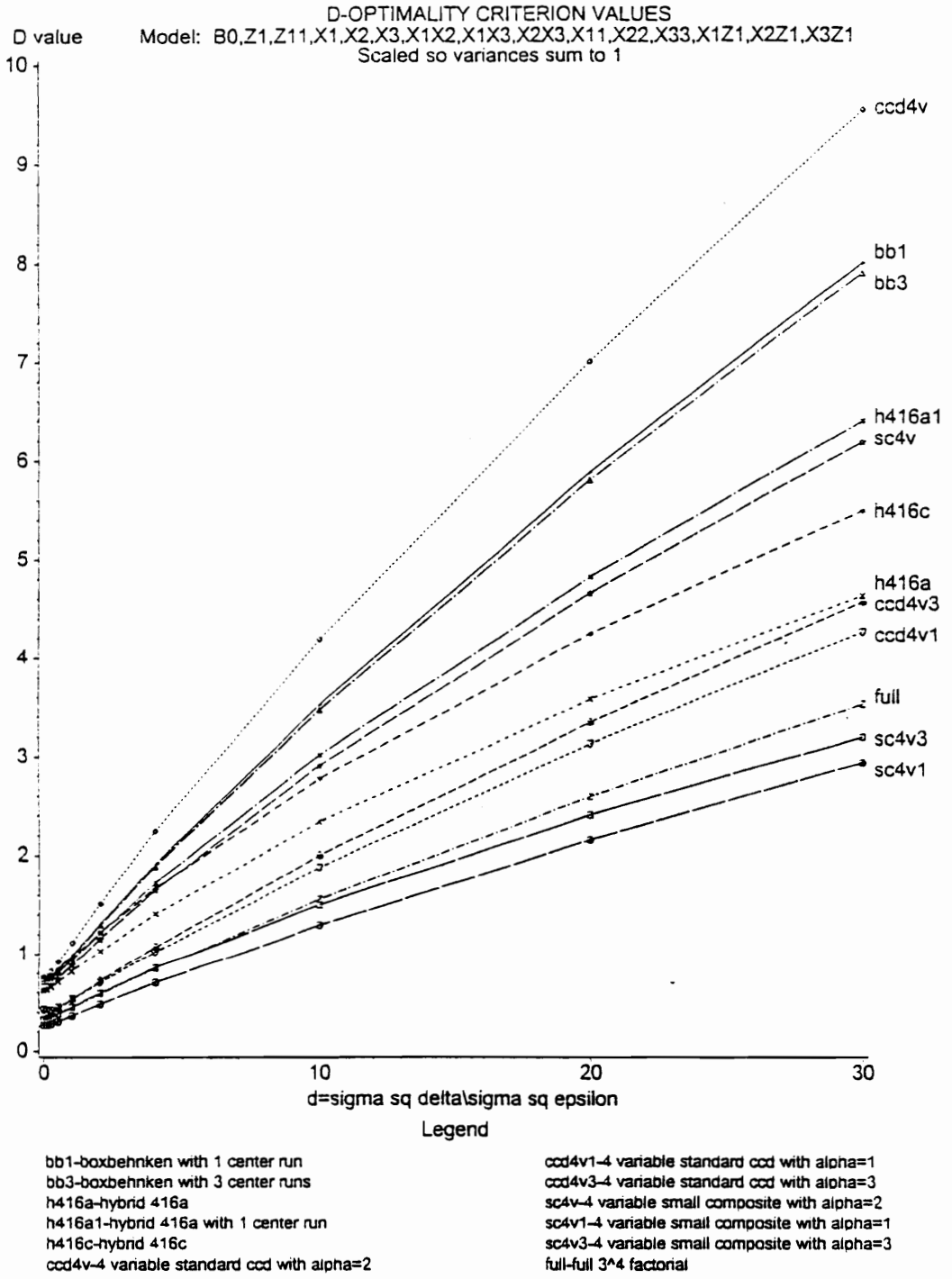


Figure 5.2.6 D Criterion Values for One Whole-plot/Three Sub-plot Variables d=0 to 30

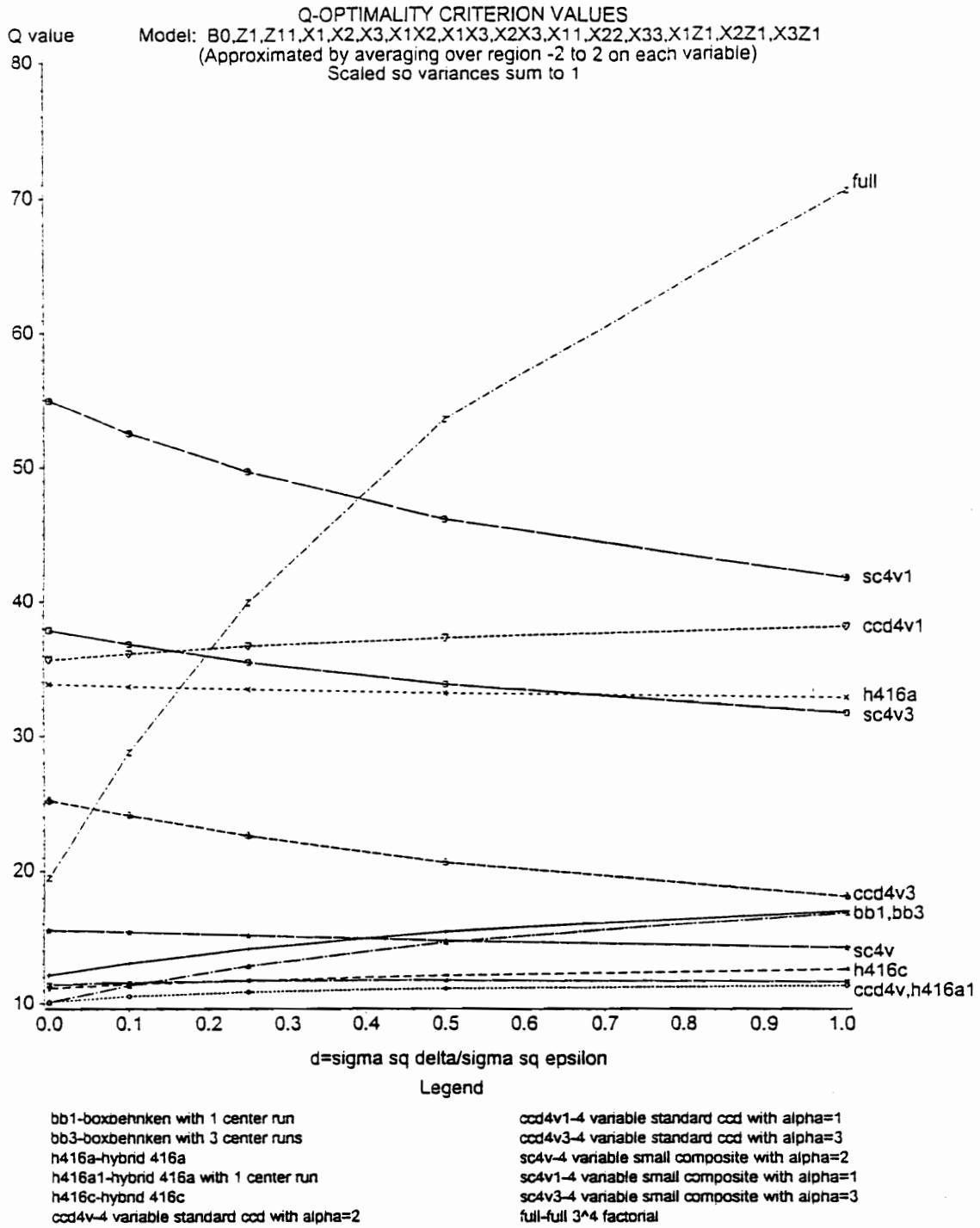


Figure 5.2.7 Q Criterion Values for One Whole-plot/Three Sub-plot Variables d=0 to 1

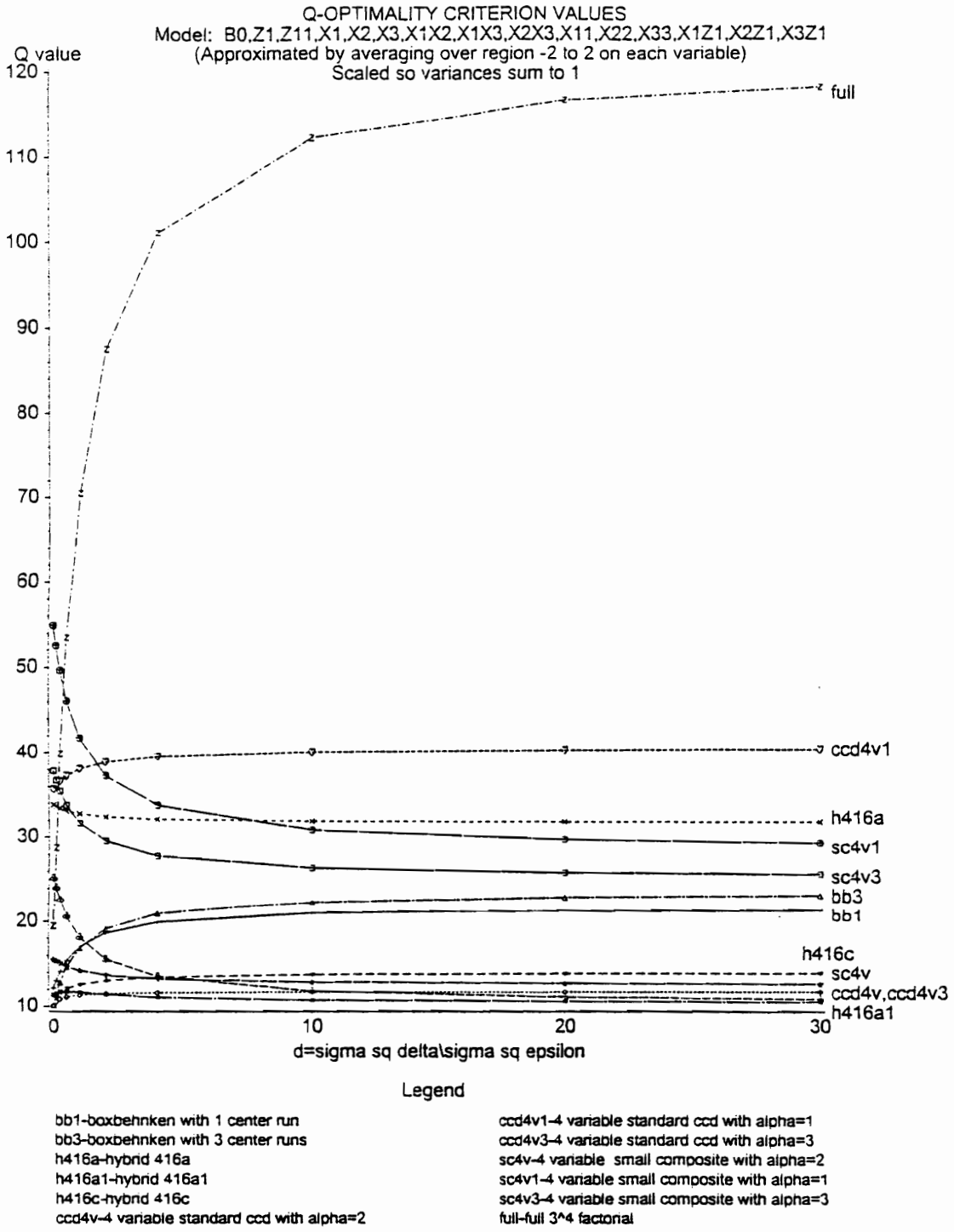


Figure 5.2.8 Q Criterion Values for One Whole-plot/Three Sub-plot Variables d=0 to 30

factorial design is clearly inefficient as compared to the other second order design alternatives. Its structure cannot adequately handle, in terms of average prediction variance, even a weak dependent relationship among observations within a whole-plot. The Box-Behnken designs, as was found for the D criterion, are clearly affected by the dependent covariance structure; however, for the Q criterion, the effect is "negative". While the Box-Behnken designs are good competitors for a CRD or for small d BRD, as d grows they soon become inefficient. One interesting thing to note: that of all the designs considered the full factorial design and the Box-Behnken designs are the ones with only three levels (-1,0,1) for each variable. The ccd with α=1 also falls in that group, but whereas the 3^4 factorial and Box-Behnken designs are competitive for d=0, this ccd is inefficient even for the CRD. Having only three unique levels for the whole-plot variable implies only three whole-plot EUs, as opposed to five for the standard ccd, are required for the BRD. So few units and degrees of freedom for whole-plot estimation may account for the poor performance of these designs under the bi-randomization structure. The full factorial design fares worse than the others due to its larger sample size. It provides information where it is least needed. The most efficient designs in reference to the Q criterion are once again the hybrid designs and the standard ccds with α=2 and 3. Also note that on Figure 5.2.8 at d=10, the Q efficient ccd changes. From d=0 to 10, the standard ccd with α=2 is the most efficient; from d=10 to 30, however, an α=3 ccd is more efficient. Recall that this same fluctuation in the Q efficient ccd was seen in the three variable example. One possible explanation may be that a most efficient axial value exists for each value of d or, perhaps, for ranges of d. These fluctuations in optimal axial values was not observed for the D criterion.

§5.2.3 Q Optimal Axial Values for Standard CCD

While the ccd has already proven itself to be the most efficient design for both criteria under the bi-randomization error control structure, evidence has been found in the preceding examples for the Q criterion that the standard ccds may be able to be modified to even better accommodate the bi-randomization scheme. The evidence suggests that an optimal axial value (different from the standard used in CRDs) may exist for each value of d or perhaps for ranges of d.

Instead of selecting the optimal axial value for a ccd for various bi-randomization situations, the axial values will be fixed to maintain a comparable design region of interest. The factorial values in the design will instead be chosen to minimize the Q criterion. The resulting design apart from scaling is equivalent to what would be found by choosing the "optimal" axial value. The relationship between the factorial and axial values is what is of primary importance. Symmetric factorial levels for each type of variable are assumed. The factorial levels on the whole-plot variables will be denoted $\pm$ g, while the factorial levels on the sub-plot variables will be given by $\pm$ h. The values g and h are then chosen to minimize the Q criterion given in 5.2.2 for the two previous variable/model combinations. The Nelder-Mead simplex program for function minimization is used to obtain the minimizing g/h values. See Nelder and Mead [1965].

First consider the standard ccd from the example in §5.2.1 with one whole-plot variable and two sub-plot variables. The axial value was fixed to 1.732, and the resulting "optimal" factorial levels chosen for a range of d values are given in Table 5.2.1.

Table 5.2.1 Optimal Factorial Values for CCD Run in Bi-randomization Format: One Whole-plot Variable/Two Sub-plot Variables

$d=\frac{\sigma_\delta^2}{\sigma_\epsilon^2}$	g	h	Efficiency
0.0	1.00	1.00	1.00
0.5	0.87	1.06	0.99
1.0	0.79	1.09	0.97
3.0	0.64	1.13	0.91
5.0	0.58	1.15	0.89
8.0	0.53	1.17	0.87
10.0	0.50	1.17	0.86
30.0	0.40	1.19	0.83
50.0	0.35	1.20	0.82

The efficiency column in the Table 5.2.1 compares the Q criterion values for a ccd using factorial values of ± 1 for both g and h to the "optimal ccd" using the tabled g and h values. An efficiency < 1 indicates that the traditional ccd with factorials of ± 1 (used for CRDs) is not as efficient as the ccd using the alternative factorial levels under the given bi-randomization error structure. An obvious trend appears in the optimally chosen factorial levels. As d becomes large, the chosen whole-plot factorial level continually decreases indicating that levels pushed to the extremes are no longer efficient. At d=10, for example, the most efficient factorial location is half way to the extremes. This is a sizable change in factorial level placement. Meanwhile, the chosen sub-plot factorial values remain close to the traditional value of one with a slight increase as d gets larger. For these bi-randomization designs, the use of the alternative factorial values in the ccd provides a more efficient design. For example, for a d=10 the traditional ccd is only 86% as efficient as the alternative ccd.

Similar results are found for the example given in §5.2.2 using one whole-plot and three sub-plot variables. In this case, α was set to 2, the traditional axial value for a CRD. The results are given in Table 5.2.2.

Table 5.2.2 Optimal Factorial Values for CCD Run in Bi-randomization Format: One Whole-plot Variable/Three Sub-plot Variables

$d=\frac{\sigma_\delta^2}{\sigma_\epsilon^2}$	g	h	Efficiency
0.0	1.00	1.00	1.00
0.5	0.89	1.03	0.99
1.0	0.82	1.05	0.98
3.0	0.68	1.09	0.94
5.0	0.61	1.10	0.92
8.0	0.56	1.11	0.90
10.0	0.53	1.11	0.89
30.0	0.42	1.13	0.86
50.0	0.37	1.14	0.85

Once again, the chosen factorial levels on the whole-plot variable is inversely related to d whereas the sub-plot factorial levels remain fairly constant. Overall, the same conclusions for one whole-plot variable can be drawn in the presence of either two or three sub-plot variables.

While the conclusions from this investigation into efficient ccd factorial values for the Q criterion are quite intriguing, there is no obvious intuitive justification for these results. To help understand the necessity of altering the factorial levels on the whole-plot variables, two variable prediction variance surface plots will be examined for the specific case of a ccd with one whole-plot variable and two sub-plot variables.

Recall for a given model and design, prediction variance at a given point, $\underline{x}_0$, for a BRD is given by the expression

$$\underline{x}_0^{'d}(X^{*'}V^{-1}X^{*})^{-1}\underline{x}_0^{d}. \qquad (5.2.3)$$

Using 5.2.3, a surface of prediction variances at design points within the region $\mathfrak{R}$ was generated for the standard ccd with factorial values ± 1 on each variable and $\alpha=1.732$.

Figure 5.2.9 displays the prediction variance surface for any combination of two variables from the ccd run under a completely randomized error control structure, i.e. d=0. The surface is mound shaped with the highest peak at the center (0,0). The design points for the ccd are indicated on the surface. According to design optimality theory, in order to improve this design with respect to average prediction variance, additional design points must be placed at areas of poor prediction variance. In other words, to improve the standard ccd under a completely randomized structure, all that can be done is to possibly place some additional points at the center of the region, (0,0). The ± 1 factorial levels should remain unchanged. This holds true for any combination of variables from the ccd. Thus, the standard ccd is indeed a "good" CRD.

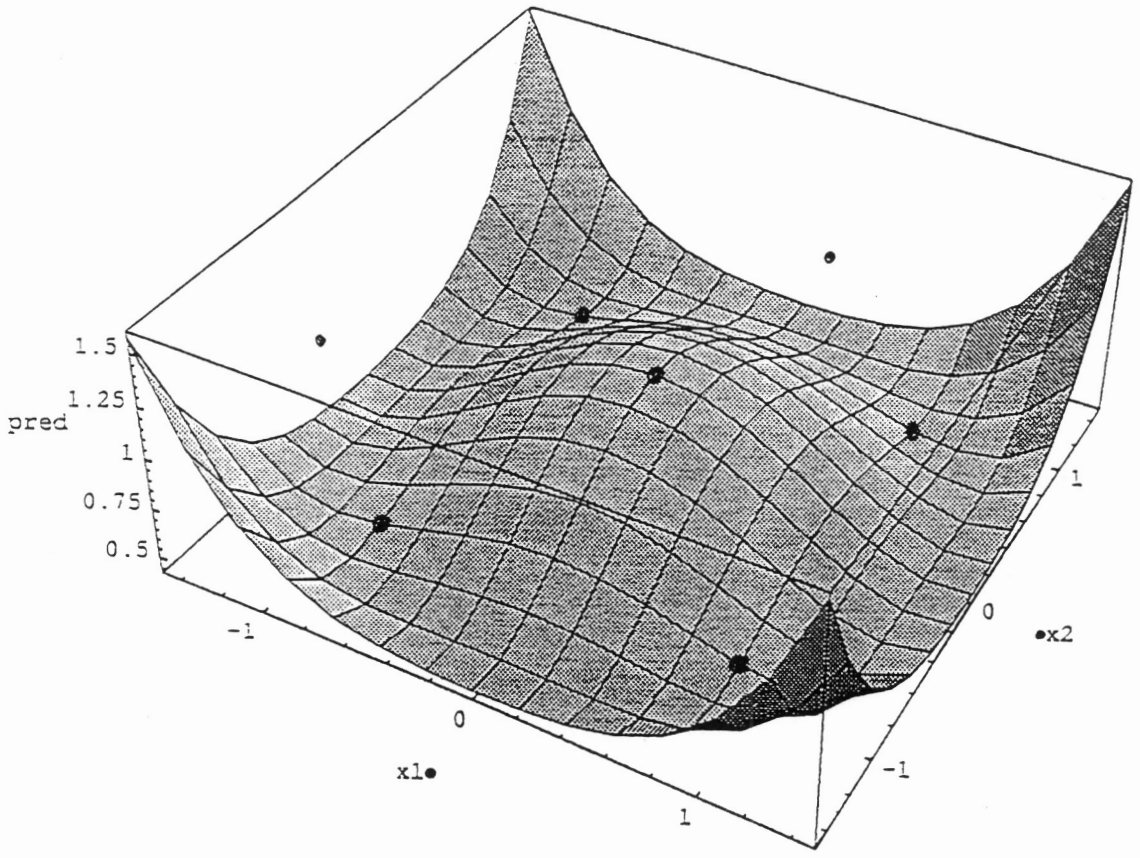


Figure 5.2.9 Prediction Variance Surface for d=0, any two variables

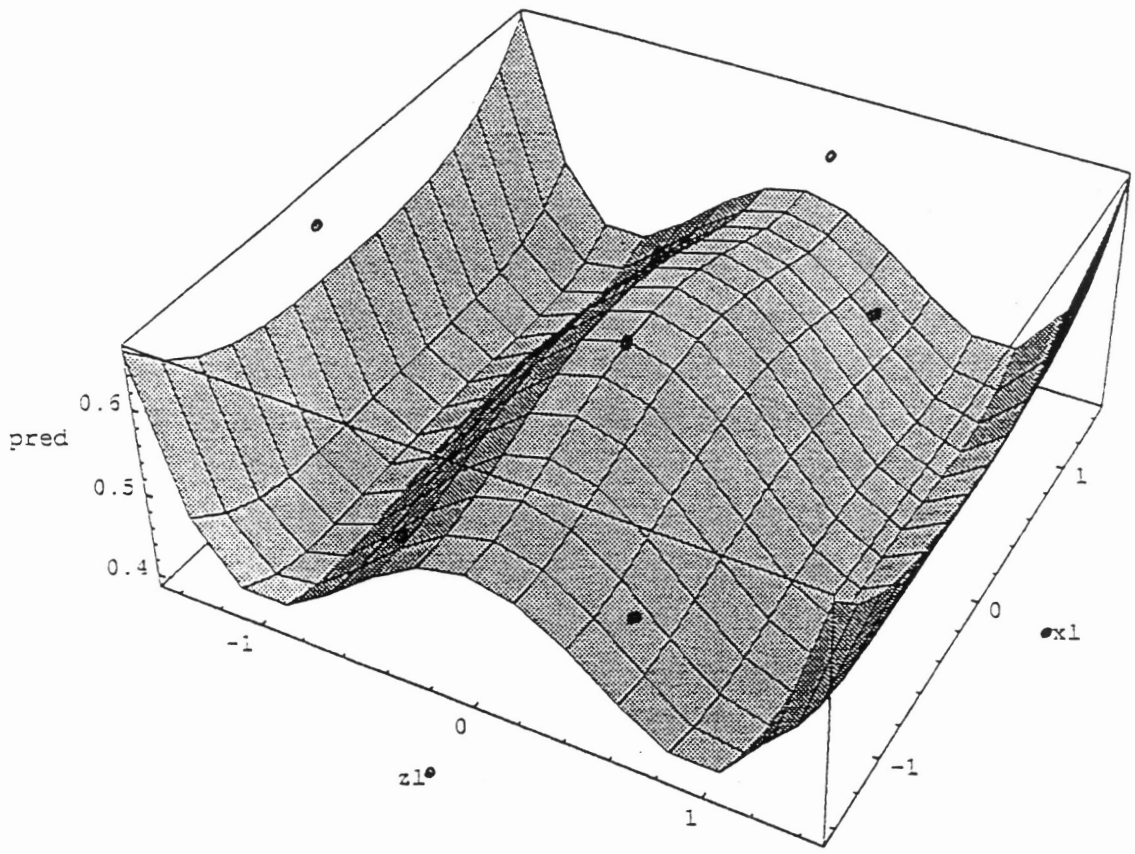


Figure 5.2.10 Prediction Variance Surface for d=10, $z_1 * x_1$ or $z_1 * x_2$

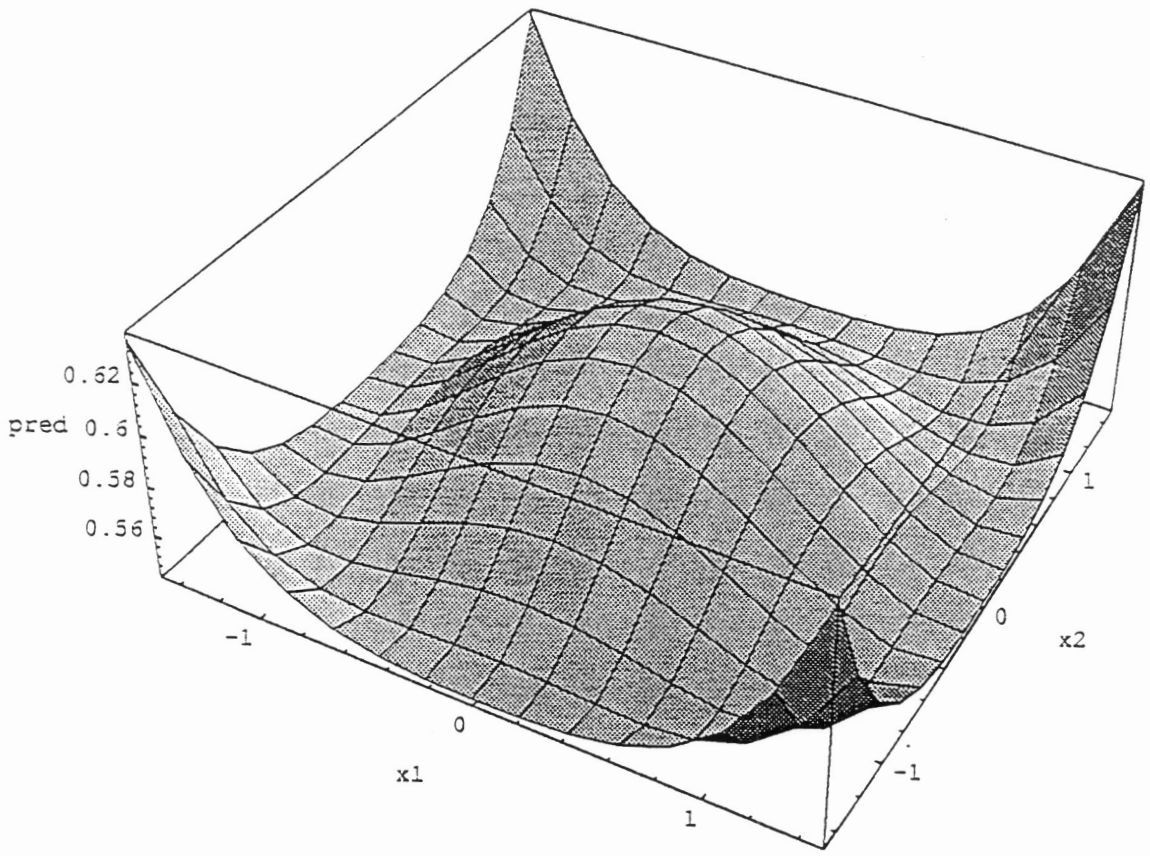


Figure 5.2.11 Prediction Variance Surface for d=10, $x_1 * x_2$

Now consider the same standard ccd but run as a BRD with $d=\sigma_\delta^2/\sigma_\epsilon^2=10$. Figure 5.2.10 displays the prediction variance surface for a whole-plot and sub-plot variable combination. Notice that the surface is no longer mound shaped. It has a ridge of poor prediction variance running along $z_1=0$. The mound in the center is still visible but is dominated by the ridge. According to design optimality theory, in this situation, points need to be placed along that ridge of poor prediction variance. This is exactly what is achieved by literally pushing the whole-plot factorial levels from ± 1 closer to the center. Augmentation of additional smaller factorial levels to the original ccd to better address that ridge of poor prediction variance would not be acceptable due to the fact that adding new whole-plot variable levels introduces additional whole-plots. In this investigation, the number of whole-plots was to remain unchanged. Thus, pushing the factorials toward the center provided the best means for improving the prediction capability of the design. The "optimal" ccd design points are indicated on Figure 5.2.10. It is evident that these new factorial levels more directly target the poor prediction variance. While the whole-plot variable factorial levels change, recall the sub-plot levels remain close to ± 1. According to Figure 5.2.10, it appears that these levels should be pushed to the outer limits of the region to once again best target the prediction variance ridge. Figure 5.2.11, however, provides an explanation for the constant factorial values for the x's. This surface shows prediction variances for the x_1 and x_2 combination. It has the same mound shaped surface found for the completely randomized ccd indicating that according to design optimality theory the factorial levels should remain at ± 1. The benefit obtained by leaving the sub-plot factorials at one outweighs the benefit of pushing them to the edges to better address the variance ridge.

These prediction variance surfaces provide the reasoning based on design optimality theory behind the necessity for modification of whole-plot variable factorial levels for bi-randomization ccds. This is also the first evidence to support the claim that standard response surface designs may indeed require modification for use in bi-randomization experiments. In addition to the prediction variance surfaces, variances of estimated model coefficients for both the traditional and the alternative "optimal" ccds can be helpful.

For the first variable combination studied with one whole-plot and two sub-plot variables, consider the traditional ccd with factorial values=± 1 and also the "Q optimal" ccd using with tabled factorial values for d=10 of g=0.50 and h=1.17. For this design and model, $\mathrm{Var}(\hat{\beta})$ is calculated and an efficiency value computed where

$$\text{efficiency}=\frac{\mathrm{Var}(\hat{\beta}_{\text{optimal}})}{\mathrm{Var}(\hat{\beta}_{\text{trad}})}.$$

When efficiency<1 the alternative ccd has smaller variance associated with its coefficient estimator than the traditional ccd with g/h=1. Efficiencies for the three variable design are given in Table 5.2.3.

Table 5.2.3 Efficiency of Coefficient Estimation for Traditional and Alternative CCD for d=10

Intercept	z_1	z_1^2	x_1	x_2	x_1x_2	x_1^2	x_2^2	x_1z_1	x_2z_1
.71	1.25	.76	.83	.83	.55	1	1	3	3

The efficiencies indicate that the alternative ccd is sacrificing precision in estimation the whole-plot*sub-plot interactions in order to gain needed efficiency in whole-plot coefficient estimation. These interactions are sub-plot terms and recall, that sub-plot model coefficients in a BRD are estimated with greater precision than the whole-plot coefficients. Thus, loss of precision is sub-plot estimation can be easily afforded in order to combat the larger variances in whole-plot parameter estimation. Similar efficiencies and conclusions were obtained for the one whole-plot variable/three sub-plot variables example.

For the optimality investigations conducted thus far, both examples with one whole-plot variable produced similar findings regardless of the number of sub-plot variables. To investigate whether these conclusions are also invariant to the number of whole-plot variables, two more variable combinations will be considered: two whole-plot variables/one sub-plot variable and two whole-plot variables/two sub-plot variables.

§5.2.4 Two Whole-plot Variables/One Sub-plot Variable

In the first combination the whole-plot variables will be denoted z_1, z_2 with the sub-plot variable denoted x_1. The second order model for this investigation is

$$E[y_{ij}]=\beta_0+\beta_1 z_1+\beta_2 z_2+\beta_3 z_1 z_2+\beta_4 z_1^2+\beta_5 z_2^2+\beta_6 x_1+\beta_7 x_1^2+\beta_8 x_1 z_1+\beta_9 x_1 z_2.$$

The same set of twelve standard second order response surface designs will be compared with respect to the D and Q optimality criteria. Since there are a total of three variables, the same designs as those in the one whole-plot/two sub-plot variable situation will be used. The D plot is given in Figures 5.2.12 for d=0-30 and the Q plot in Figures 5.2.13, also for d=0-30.

The number of whole-plot variables seems to have little affect on the D criterion plots. Figure 5.2.12 is similar to those for one whole-plot variable examples (except for scale). The full factorial design, however, appeared to fare better than in the previous examples. There are some differences in terms of the influence of V on the hybrid designs, but the standard ccd with α=1.682 and both Box-Behnken designs are still the most efficient under the bi-randomization error structure.

The Q criterion plot, however, does show variations from the previous examples. The obvious differences involve the Box-Behnken and hybrid designs. In the one whole-plot variable examples, the standard ccds with α=1.682 and 2 were the most efficient designs, while the Box-Behnken designs did not perform well. For this example, however, the Box-Behnken designs are now among the most efficient. The ccds are only fourth and sixth in the ranking of most efficient designs for moderate to large values of d. The set of hybrid designs are also more efficient than the ccds for larger d values. The switch in efficient designs from the one-whole plot variable case are the result of simply altering the role of a variable. Of all the designs and situations considered, until now, the ccd seemed to best accommodate the bi-randomization error structure.

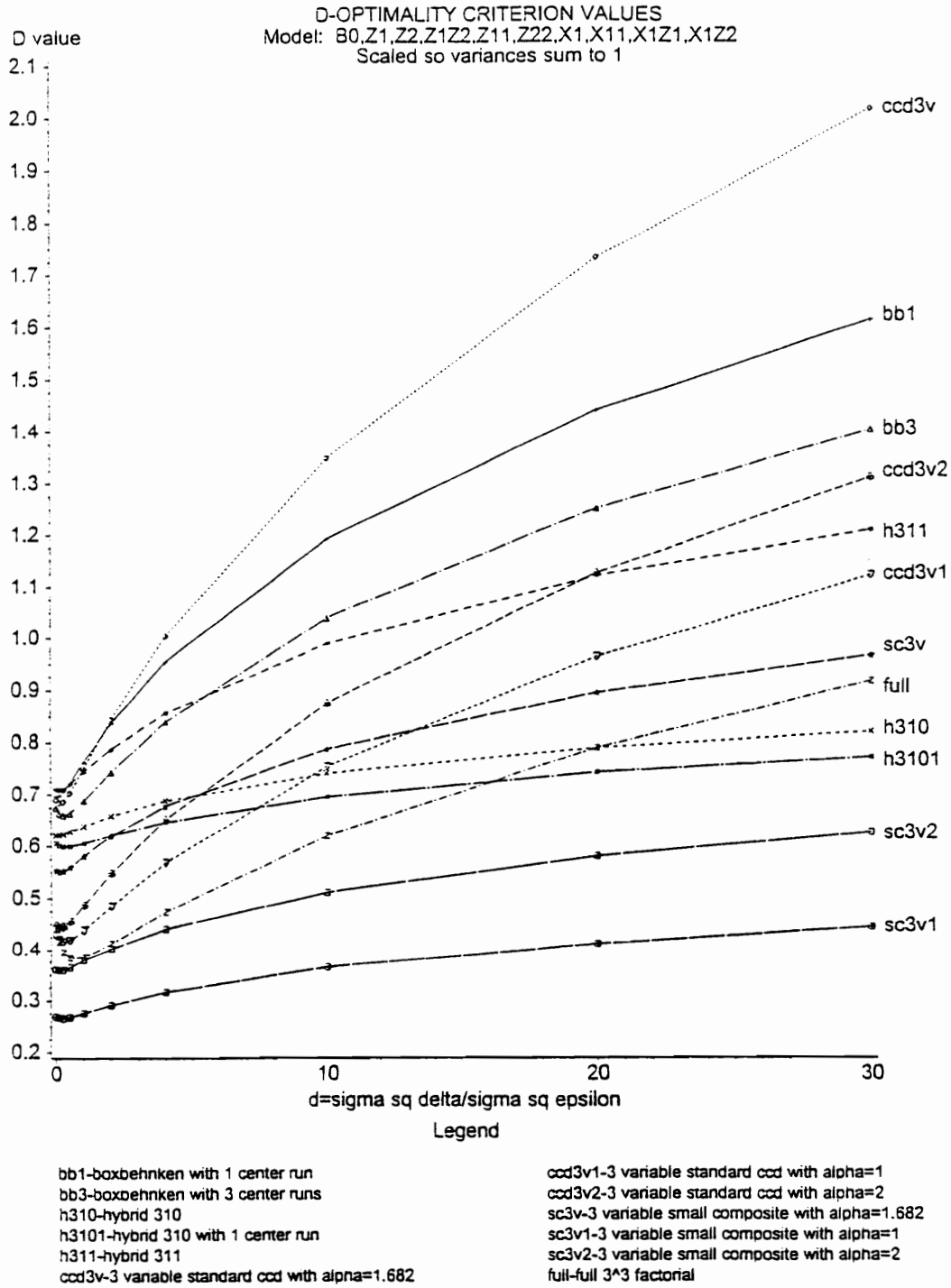


Figure 5.2.12 D Criterion Values for Two Whole-plot/One Sub-plot Variable d=0 to 30

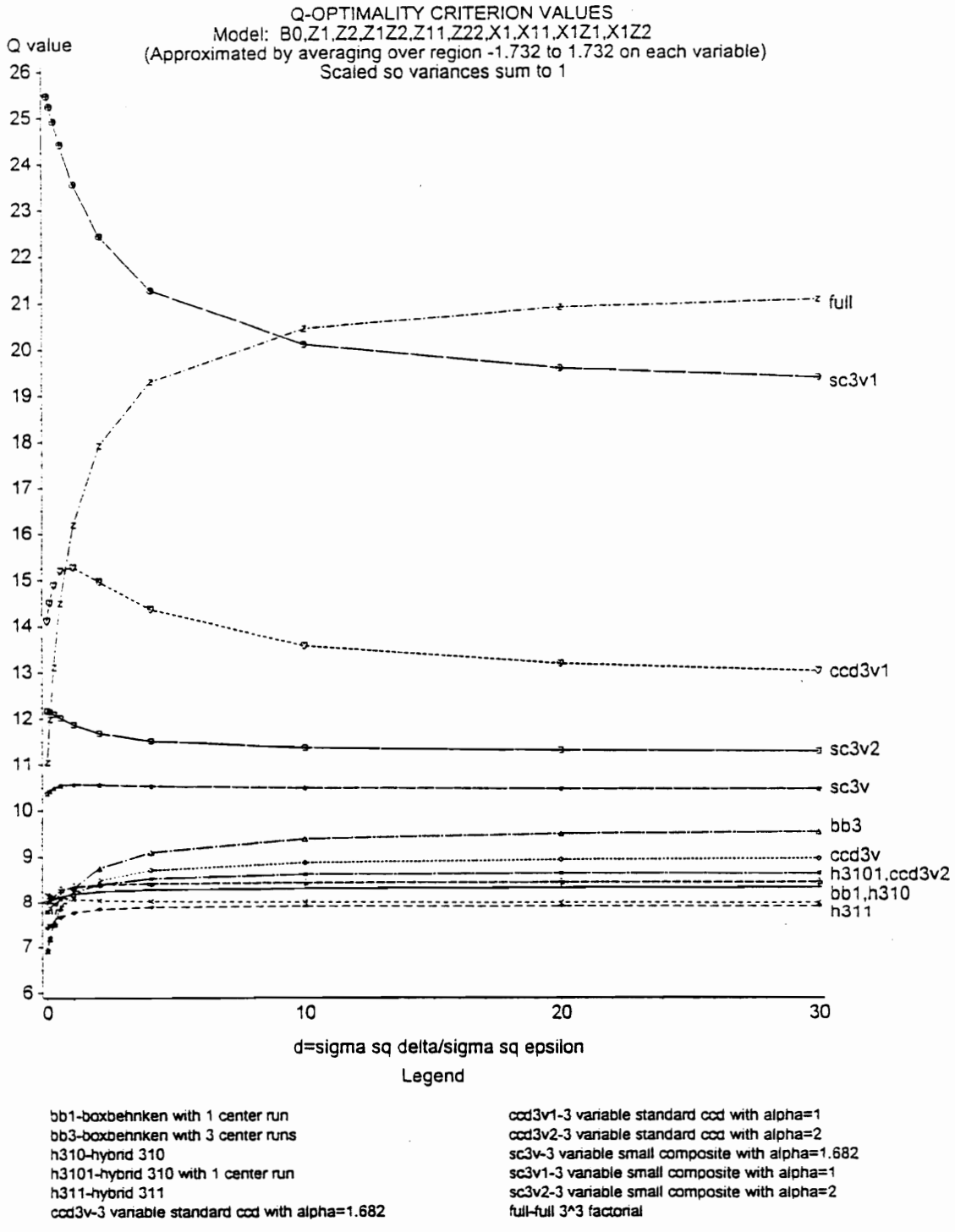


Figure 5.2.13 Q Criterion Values for Two Whole-plot/One Sub-plot Variable d=0 to 30

It is possible from the observed results in the previous investigation that perhaps an alternative ccd may be superior to the Box-Behnken and hybrid designs. Using the Nelder-Mead algorithm optimal factorial values were chosen for the ccd under this bi-randomization structure. For each value of d, the most efficient factorial levels, with respect to the Q criterion, were close to g=1 and h=1. The standard ccd is indeed the best that the ccd can do for average prediction variance under this structure. To better understand why the ccd could not be improved for the two whole-plot variables/one sub-plot variable combination, prediction variance surface plots will again be consulted. Figure 5.2.14 displays the surface for the two whole-plot variables, z_1 and z_2. This surface suggests that the z's factorial levels should remain at ± 1 to best address the "hump" of poor prediction variance. The prediction variance surface for a whole-plot variable and a sub-plot variable combination found in Figure 5.2.15, however, indicates pushing the levels on the whole-plot variables in closer to the center to address its ridge of poor prediction variance.

The two surfaces clearly warrant different whole-plot factorial levels. From the Nelder-Mead results, the benefit of leaving the factorials at ± 1 must outweigh the benefit found in Figure 5.2.15 of using smaller levels. This is easily understood by considering the relationship of the variances of the types of estimated model coefficients. Recall that variances of estimated whole-plot term coefficients are naturally larger than those of estimated sub-plot coefficients for a BRD and also recall that whole-plot*sub-plot interactions are considered sub-plot terms. Figure 5.2.14 is a purely a function of estimated whole-plot coefficient variances while 5.2.15 is a function of estimated sub-plot*whole-plot coefficient variances. Since the whole-plot variances are large, more benefit is found in reducing them by leaving the factorial levels unchanged than what can be obtained by altering them to improve sub-plot variances which are already small. The small sacrifices in the estimation of sub-plot term coefficients by leaving levels at ± 1 is overshadowed by the improvement in whole-plot coefficient estimation.

The same concept can help us understand the better performance of the

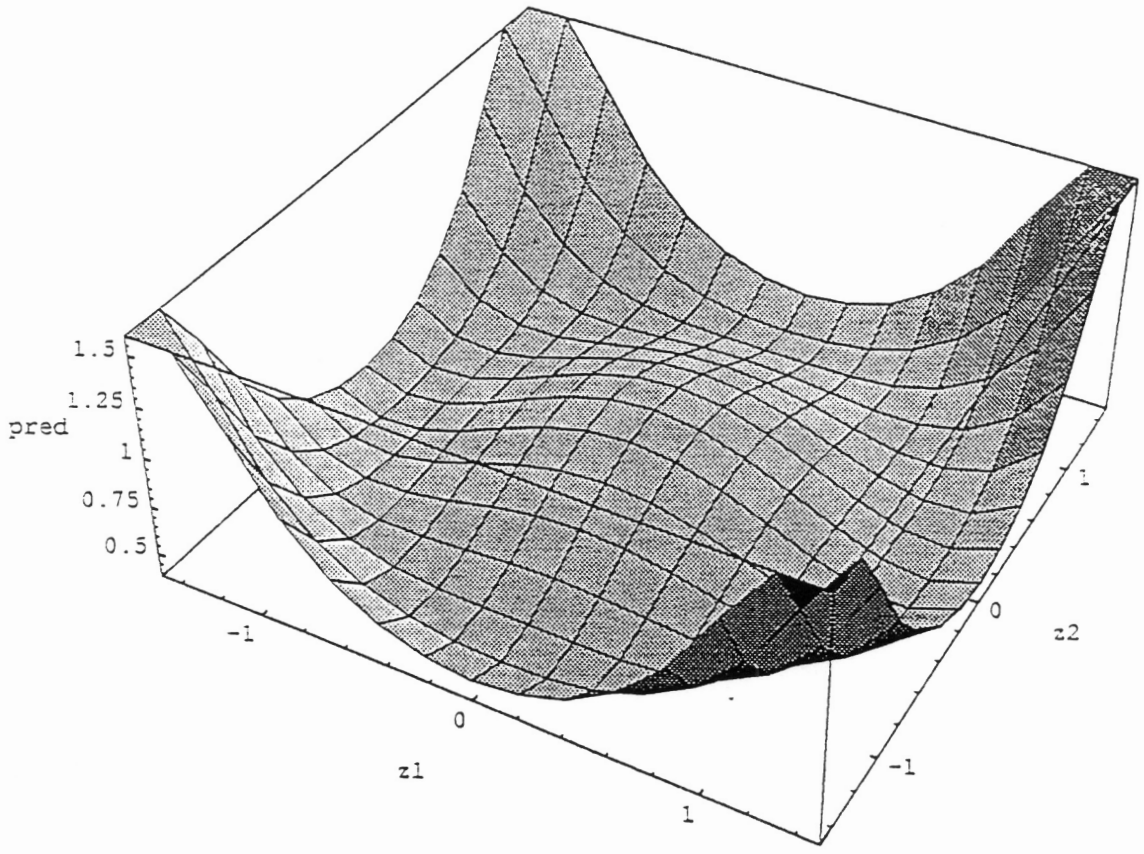


Figure 5.2.14 Prediction Variance Surface for d=10, $z_1 * z_2$

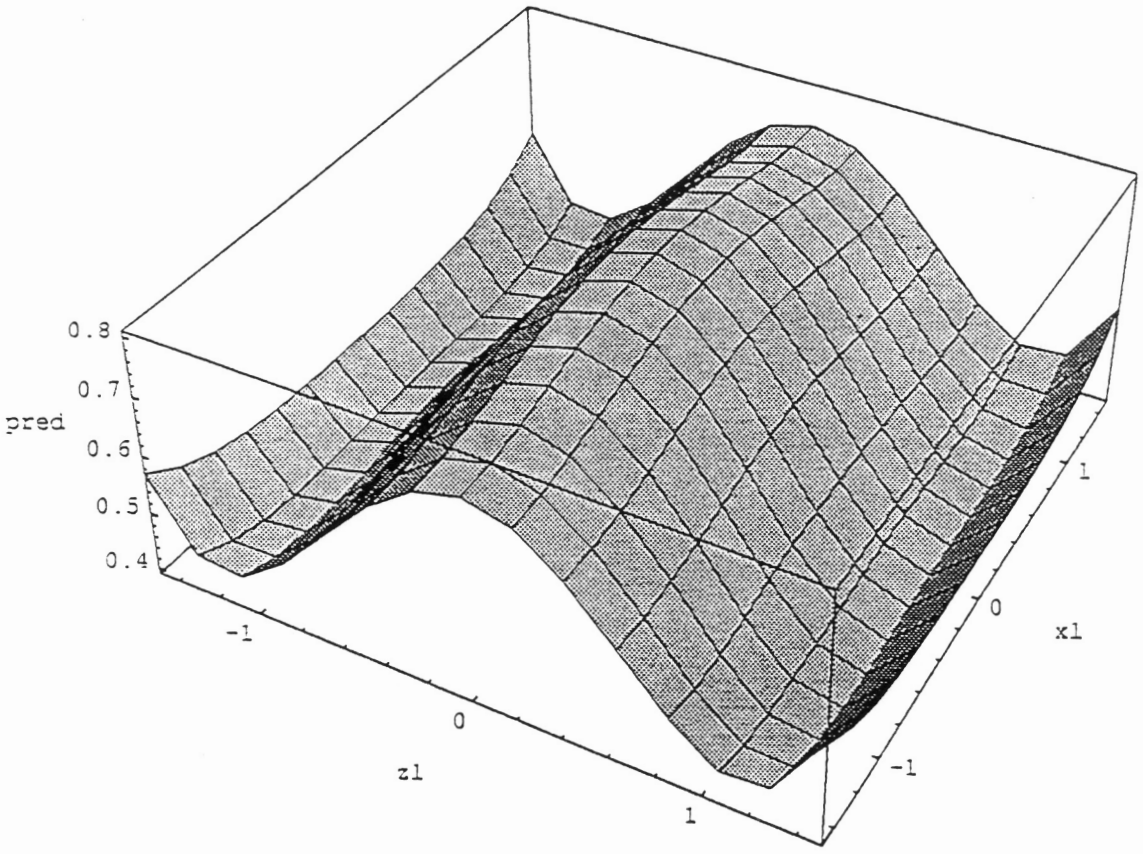


Figure 5.2.15 Prediction Variance Surface for d=10, z_1* x_1

hybrid design over the ccd for this variable combination. Due to the larger variance associated with whole-plot coefficient estimation, design choice in the variables z_1 and z_2 should attempt to minimize their estimation variance. This is best achieved by using a two variable ccd (or perhaps 3^2 factorial) in z_1 and z_2. The design choice in x_1, however, has limited influence on the overall design prediction variance due to the already inherently small variances on the estimated sub-plot coefficients for a BRD. The three variable hybrid design is constructed in just this manner: a two variable ccd in z_1 and z_2 augmented with an additional column of x_1 levels. The standard ccd is also constructed in a similar manner, but design size plays crucial role. The hybrid design obtains this structure with 10 or 11 design points whereas the ccd requires 15 or 16. On a per observation basis, the hybrid design is more efficient.

Efforts were also made using the Nelder-Mead minimization algorithm to improve in terms of the Q criterion both the hybrid and Box-Behnken designs. The same methods that were used to obtain optimal ccd factorial levels were applied to both types of designs. No useful modifications, however, were found.

§5.2.5 Two Whole-plot Variables/Two Sub-plot Variables

The Q criterion results observed for the previous example appear to be heavily dependent on having only one sub-plot variable. To explore the effects of an additional sub-plot variable on the efficiency of these designs, two whole-plot variables, z_1 and z_2, along with two sub-plot variables, x_1 and x_2, will be considered. The assumed model is

$$E[y_{ij}]=\beta_0+\beta_1 z_1+\beta_2 z_2+\beta_3 z_1 z_2+\beta_4 z_1^2+\beta_5 z_2^2+\beta_6 x_1+\beta_7 x_2+\beta_8 x_1 x_2+$$
$$\beta_9 x_1^2+\beta_{10} x_2^2+\beta_{11} x_1 z_1+\beta_{12} x_1 z_2+\beta_{13} x_2 z_1+\beta_{14} x_2 z_2.$$

The same twelve four variable second order designs that were explored in §5.2.2 will be compared with respect to the D and Q criteria. The design performances are given in Figures 5.2.16 and 5.2.17.

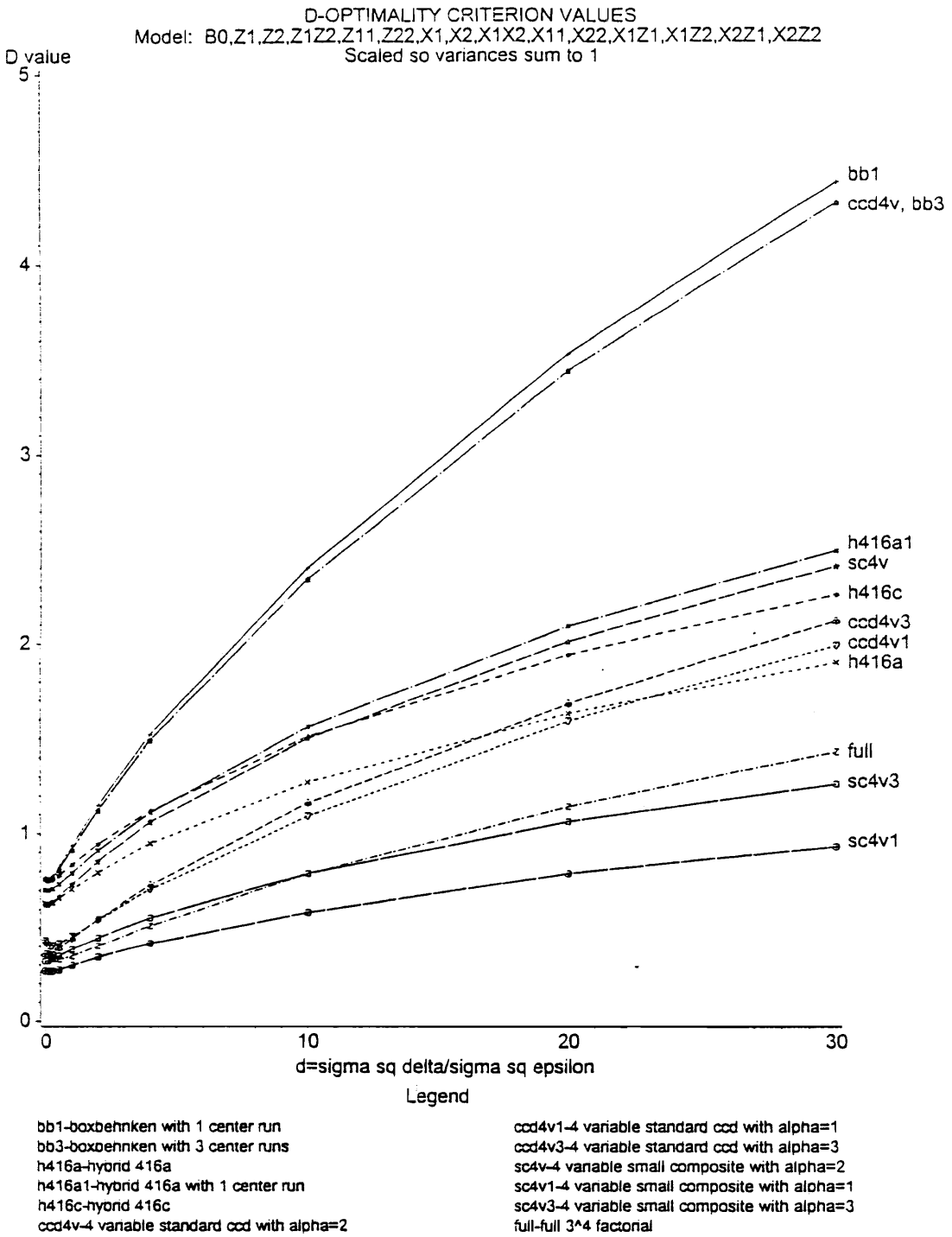


Figure 5.2.16 D Criterion Values for Two Whole-plot/Two Sub-plot Variables d=0 to 30

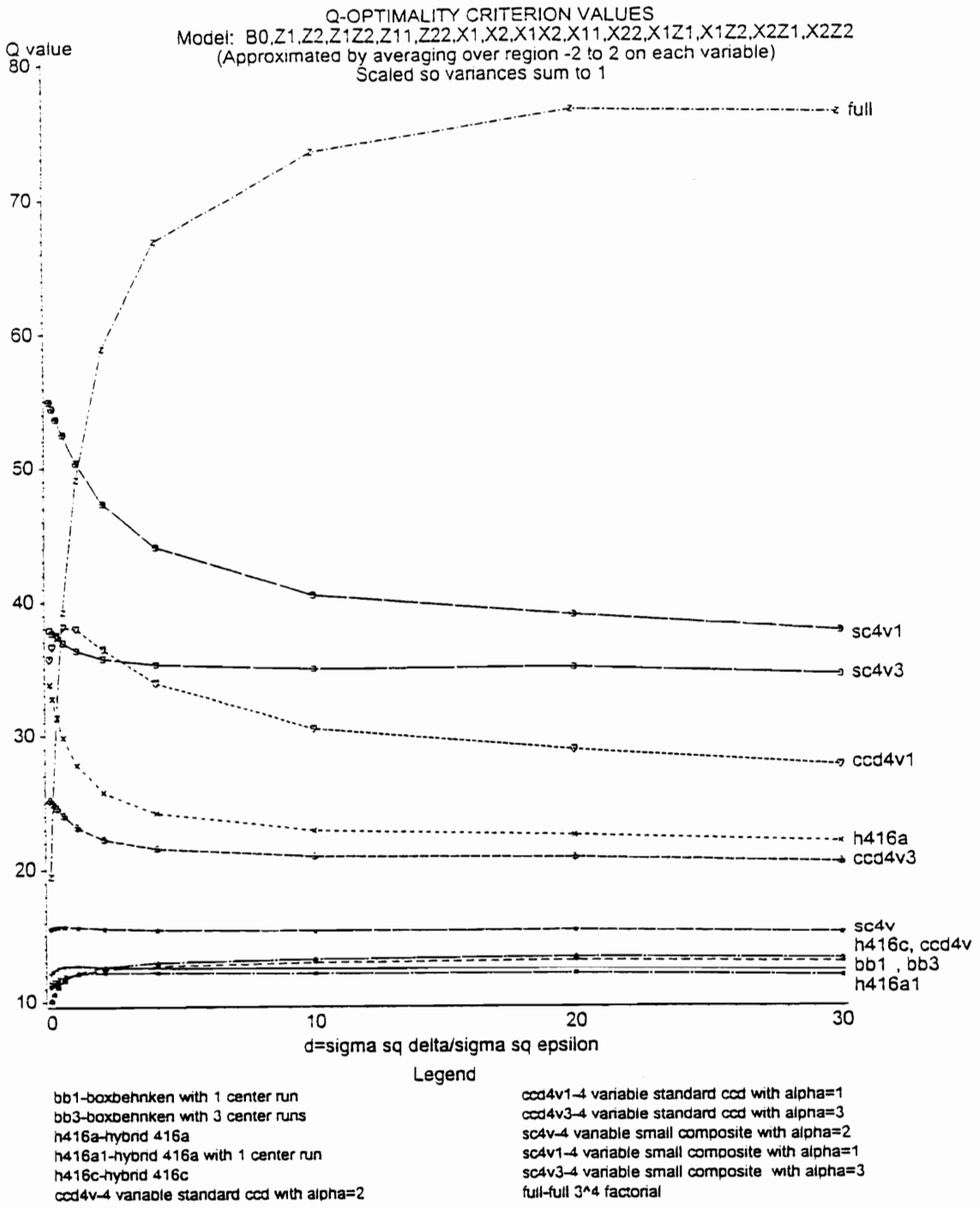


Figure 5.2.17 Q Criterion Values for Two Whole-plot/Two Sub-plot Variables d=0 to 30

Figure 5.2.16 tells the same story for the D criterion that has been told so far. The Box-Behnken designs and the standard ccd with the traditional α value are the most efficient with respect to generalized variance over all values of d. The full factorial design, however, does fare better for larger d than it did in the earlier examples. Unlike the Q criterion, the D criterion appears to be robust to the distribution and roles of the variables. Of greater interest is the more sensitive Q criterion results which can found in Figure 5.2.17. The plots themselves do not show any discrepancies from the results in Figure 5.2.13 for only one sub-plot variable except the expected improved performance of the standard ccd. With the presence of two sub-plot variables, design in the sub-plot variables now plays a more critical role in overall prediction variance due to more sub-plot terms in the model. While the two variable ccd is still desirable in the whole-plot variable, a two variable ccd is also desirable in the sub-plot variables. The reduction in prediction variance gained by using a two variable ccd in the sub-plot variables, however, is countered by the the number of additional design points. The hybrid and Box-Behnken designs satisfactorily estimate the sub-plot coefficients without the added design points found in the ccd. This trade-off results in the designs being equal competitors in terms of average prediction variance. Results such as these indicate that the use of average prediction variance as a discriminating tool may in fact wash out important differences among the designs under bi-randomization experimentation.

The addition of the other sub-plot variable in this example naturally led to larger designs for comparison than the two whole-plot/one sub-plot variable case. With larger design sizes and more design points with which to work, the Nelder-Mead minimization algorithm was utilized in hope of finding improvements in the ccd, hybrid, and Box-Behnken designs. The structure of the design remained unchanged, only new design levels were "optimally" selected. No useful level modifications were discovered.

At this point structural modifications of the designs, specifically for the hybrid design, were now considered for this variable combination. The construction of a k variable hybrid design involves a ccd in k-1 variables, with the

levels of the k^{th} variable chosen to obtain desirable moment properties. For the two whole-plot/two sub-plot example the design consisted of a three variable standard ccd augmented with an additional column. Recall from previous work, however, that to minimize overall estimation variance a ccd may only be necessary in the whole-plot variables with the sub-plot variables augmented in an "optimal" (according to the Q criterion) way. By only considering a ccd in the whole-plot variables a small, but yet efficient, alternative design may be found.

One such alternative hybrid as it will be called will have the following structural shell:

z_1	z_2	x_1	x_2
i	i	-g	h
i	i	h	g
i	i	g	-h
i	i	-h	-g
i	-i	g	h
i	-i	h	-g
i	-i	-g	-h
i	-i	-h	g
-i	i	g	h
-i	i	h	-g
-i	i	-g	-h
-i	i	-h	g
-i	-i	-g	h
-i	-i	h	g
-i	-i	g	-h
-i	-i	-h	-g
2	0	0	0
-2	0	0	0
0	2	0	0
0	-2	0	0
0	0	0	0

where $\pm$ i are the factorial levels for the whole-plot variables, and $\pm$g and $\pm$h constitute the unique levels of both sub-plot variables. Values for i, g, and h can then be selected to minimize the Q criterion over a given region of interest for various values of d=$\sigma_\delta^2/\sigma_\epsilon^2$ using the Nelder-Mead algorithm. The results are given in Table 5.2.4.

Table 5.2.4 Optimal i/g/h Values for Two Whole-plot/Two Sub-plot Variables Alternative Hybrid

d	i	g	h	Q value
0	0.87	0.60	1.46	11.7
5	0.94	0.71	1.32	12.2
10	0.94	0.72	1.32	12.2
20	0.94	0.72	1.32	12.2

The alternative 21 point hybrid design does not appear to be influenced by fluctuations in the "strength" of the bi-randomization structure. It is robust over varying levels of d which is a desirable attribute. In comparison to the standard second order designs already studied, at d=10 the original hybrid 416a1 with 17 runs had a Q value of 12.1 while the standard ccd with 27 runs has a Q value=13.2. The 21 run alternative hybrid is a close competitor to the original hybrid and by far surpasses in efficiency the ccd. Traditional hybrid designs are categorized under near saturated second order designs due to their small design size. In some situations, too small a design size may be a concern. For these situations, the alternative hybrid provides a useful option without committing to the larger sized ccd. For d=0, however, the alternative hybrid is not as efficient as the original hybrid and clearly not as efficient as the standard ccd, but it is a good option for a bi-randomization experiment.

The variances for the estimated model coefficients under a bi-randomization structure of d=10 for both the alternative and standard hybrid designs are given below in Table 5.2.5.

The comparison of variances is quite surprising. Since both designs are essentially based on a similar concept, it would seem logical that the variances of estimated model coefficients would be similar. Obviously, that is not the case. The variances of the estimated quadratic whole-plot coefficients are picking up additional precision with the alternative hybrid, but there are sacrifices in intercept and interaction estimation. The sub-plot terms, however, show more obvious differences.

Table 5.2.5 Var($\widehat{\beta}$) for Alternative and Standard Hybrids, d=10

Model Term	Standard Hybrid	Alternative Hybrid
Intercept	0.83	1.00
z_1	0.10	0.09
z_2	0.10	0.09
z_1z_2	0.20	0.30
z_1^2	0.15	0.09
z_2^2	0.15	0.09
x_1	0.006	0.005
x_2	0.015	0.005
x_1x_2	0.01	0.006
x_1^2	0.01	0.12
x_2^2	0.016	0.12
x_1z_1	0.01	0.006
x_1z_2	0.01	0.006
x_2z_1	0.14	0.006
x_2z_2	0.14	0.006

Variance in estimation of the sub-plot quadratic terms, for instance, is approximately twelve times higher in the alternative hybrid than in the standard, but notice, however, that the rest of the sub-plot terms have smaller estimation variance, sometimes much smaller, with the alternative hybrid. The differences in variance of estimation of the whole-plot*sub-plot interactions is drastic. The alternative hybrid redistributes variance of estimation and targets the sub-plot

quadratic terms. While overall these designs produce similar average prediction variances, if only a subset of the model were of interest, a clear distinction between the designs could be made.

For example, consider the context of robust parameter design discussed in Chapter 1. Recall that the goal of RPD is to determine levels of design variables called controls that are robust to fluctuations in uncontrollable variables called noise. The determination of these levels is directly related to the modeled control*noise interactions. If the noise variables are considered the whole-plot variables and the controls the sub-plot variables, these interactions would then be whole-plot*sub-plot interactions. To ensure precise estimates of these interactions and proper choice of the robust control levels under a bi-randomization error structure, the alternative hybrid would clearly be the better design for experimentation. This suggests that if different criteria were of interest, such as subset optimality, more efficient alternative designs to the standard response surface designs may be found in the presence of bi-randomization error control structures.

In this subsection, the D and Q optimality criteria were addressed for second order bi-randomization designs under certain variable/model restrictions. The D optimality results appeared to be robust to the situation under investigation. The Box-Behnken designs along with the standard ccd were the most efficient in each case over the values of d. More effort was spent on investigating the Q criterion and prediction variance due to their more easily understood interpretations than the generalized variance in the D criterion. The Q optimal design within the class of twelve designs studied was sensitive to the given situation. For variable combinations with only one whole-plot variable, the ccds with optimally chosen factorial values and also the hybrid designs were the most efficient. For the two-whole plot variable combinations, however, the ccds no longer were the most efficient. The Box-Behnken designs and the hybrid designs now became competitors with the ccd resurfacing in the two whole-plot/two sub-plot example. Efficient designs for the two whole-plot variable situations were influenced by the number of sub-plot variables present.

This design optimality investigation was limited to special cases, special designs, and special discriminating criteria. These investigations, although limited, have provided evidence to warrant more extensive investigations into optimal designs and alternatives for bi-randomization structures. Additional design properties including the consideration of a fixed number of whole-plots or the need for balanced whole-plots could also be incorporated into the investigations.

CHAPTER 6

FUTURE RESEARCH

The research presented in this dissertation is a beginning exploration into alternative error control structures and their implications for RSM. Two specific types of designs, crossed and non-crossed, within the class of bi-randomization designs were examined. Analysis techniques were developed in each case along with methods for error variance estimation. In the case of second order non-crossed bi-randomization designs, several methods for error variance estimation were compared for specific RS design examples. In addition, the efficiency of a set of standard second order response surface designs run under a bi-randomization error structure was explored. Obvious indications existed that standard designs are quite efficient, but modification may be used to better accommodate the alternative error control structure. Although many topics have been addressed in this dissertation, even more questions and problems remain unanswered and unsolved.

§6.1 **Estimation**

For second order non-crossed BRDs, only a limited investigation was conducted into the four presented estimation procedures. The REML and IRLS methods appear to be quite promising. In addition, it appears that properties of estimators of coefficients are well approximated by the asymptotic results even for small designs. Before more general recommendations may be made, however, a

more thorough investigation must be conducted into OLS, 1IRLS, IRLS, and REML. Alternative estimation methods, including perhaps MLE, should also be explored.

§6.2 Design Optimality

As in the case of the estimation problem, only a restricted study into efficient designs for the alternative error structure was performed. In addition to the design optimality criteria considered, other alternatives should be investigated including D_s optimality which was discussed briefly in reference to the alternative hybrid design in Chapter 5.

This study also examined only specific designs and models to determine efficiency under the bi-randomization error structure. To broaden the study away from given models and designs, computer automation is essential. Computer packages currently exist which aid in the selection of efficient designs but only for completely randomized error structures. In order for the optimality programs to accommodate the bi-randomization error structure, the program must allow for the changing of the structure of V as additional design points are added/deleted within various whole-plots. In addition, they must take into consideration parameters such as sample size, number of whole-plots, and the maximum number of sub-plot units within a whole-plot. Once automation is achieved, more general studies can be conducted which would lead to recommendations for efficient designs given a model and values for the error variances. Limited knowledge of the error variances could then, in turn, result in the development of multistage designs incorporating Bayesian techniques.

§6.3 Non-constant Covariance and Other Error Structures

The bi-randomization error structure studied in this dissertation is only one possible error control structure. Further research must also be conducted in the areas of other alternative error control designs for RSM including the possibility of nonconstant covariance among observations within a given whole-plot for a BRD.

APPENDIX

APPENDIX A: Distributions for Non-Crossed, First Order ($b_i \neq b$) Designs

Consider a non-crossed, first order BRD in which $b_i \neq b \ \forall \ i$ and the response surface model,

$$y_{ij}=\beta_0+\underline{z}_i'^{d}\underline{\gamma}+\underline{x}_{ij}'^{d}\underline{\alpha}+\underline{z}'^{d} \triangle \underline{x}_{ij}^{d}+\delta_i+\epsilon_{ij}. \qquad (A.1)$$

For calculation of appropriate sum of squares, let the whole-plot error be modeled as fixed, i.e.

$$\delta_i=\underline{z}_i^{*'d}\underline{\varrho} \qquad (A.2)$$

where $\underline{\varrho}$ is an a-(k+1)*1 vector of whole-plot terms not modeled in A.1.

Based on A.2, the whole-plot error sum of squares (lack of fit) for A.1 can be written as a function of regression sum of squares,

$$\underline{y}'[T(T'T)^{-1}T]\underline{y}-\underline{y}'[Z[Z'Z]^{-1}Z]\underline{y} \qquad (A.3)$$

where $T=[Z, Z^*]$.

Since whole-plot variable levels are constant within a whole-plot, A.3 can be equivalently written as

$$\overline{\underline{y}}'[B^*T_r(T_r'B^*T_r)^{-1}T_rB^*]\overline{\underline{y}}-\overline{\underline{y}}'[B^*Z_r(Z_r'B^*Z_r)^{-1}Z_rB^*]\overline{\underline{y}} \qquad (A.4)$$

where T_r and Z_r are reduced versions of T and Z and B^* is an a*a diagonal matrix of whole-plot sizes.

Because T_r is an a*a full rank matrix, expression A.4 can be rewritten as

$$\overline{\underline{y}}'[B^*]\overline{\underline{y}}-\overline{\underline{y}}'[B^*Z_r(Z_r'B^*Z_r)^{-1}Z_r'B^*]\overline{\underline{y}}. \qquad (A.5)$$

Let

$$\underline{\bar{y}}^{*} = B^{*\frac{1}{2}}\underline{\bar{y}}$$

with

$$\text{var}(\underline{\bar{y}}^{*})=\sigma_{\epsilon}^{2}I+\sigma_{\delta}^{2}B^{*}=V^{*}.$$

The whole-plot error sum of squares in A.5 can then be expressed as

$$\underline{\bar{y}}^{**\prime}[V^{*-1}]\underline{\bar{y}}^{**}- \underline{\bar{y}}^{**\prime}[V^{*-1}Z_r^{*}(Z_r^{*\prime}V^{*-1}Z_r^{*})^{-1}Z_r^{*\prime}V^{*-1}]\underline{\bar{y}}^{**} \qquad (A.6)$$

where $\underline{\bar{y}}^{**}=V^{*-\frac{1}{2}}\underline{\bar{y}}^{*}$ and $Z_r^{*}=V^{*-\frac{1}{2}}B^{*\frac{1}{2}}Z_r$.

Now let

$$\underline{\bar{y}}^{***}=V^{*\frac{-1}{2}}\underline{\bar{y}}^{**} \quad \text{and} \quad Z_r^{**}=V^{*\frac{-1}{2}}Z_r^{*} \qquad (A.7)$$

where

$$\text{var}(\underline{\bar{y}}^{***})=I$$

and $V^{*\frac{-1}{2}}$ is obtained through Cholesky's decomposition on V^{*-1}.

Based on the transformed model, the whole-plot residual sum of squares can then be written as

$$\underline{\bar{y}}^{***\prime}[I-Z_r^{**}(Z_r^{**\prime}Z_r^{**})^{-1}Z_r^{**\prime}]\underline{\bar{y}}^{***} \qquad (A.8)$$

or

$$\underline{\bar{y}}^{***\prime}[A]\underline{\bar{y}}^{***}.$$

Since AI is idempotent with rank a-(k+1), the whole-plot error sum of squares in A.8 is distributed

$$\chi^{2}_{a-(k+1)}.$$

Similarly, based on model A.1 with δ_i modeled as fixed to account for whole-plot error sum of squares, the sub-plot error sum of squares is distributed

$$\chi^{2}_{n-p^{*}}.$$

Using these distributions, appropriate tests may be developed.

APPENDIX B: Example of Cholesky's Decomposition

Consider a two variable bi-randomization central composite design composed of one whole-plot variable, z, and one sub-plot variable, x, i.e.

$$\mathfrak{D}=\begin{array}{c} \begin{matrix} z & x \end{matrix} \\ \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \\ \alpha & 0 \\ -\alpha & 0 \\ 0 & \alpha \\ 0 & -\alpha \\ 0 & 0 \end{bmatrix} \end{array}.$$

By Cholesky's decomposition, $V^{-1}=V^{-\frac{1}{2}}V^{-\frac{1}{2}}$ where

$$V^{-\frac{1}{2}} = \frac{1}{\sigma_\epsilon}\begin{bmatrix} \sqrt{\frac{d+1}{2d+1}} & (\frac{-d}{t}) & & & & & & & & \\ 0 & \sqrt{\frac{1}{d+1}} & & & & & & & & \\ & & \sqrt{\frac{d+1}{2d+1}} & (\frac{-d}{t}) & & & & & & \\ & & 0 & \sqrt{\frac{1}{d+1}} & & & & & & \\ & & & & \sqrt{\frac{1}{d+1}} & & & & & \\ & & & & & \sqrt{\frac{1}{d+1}} & & & & \\ & & & & & & \sqrt{\frac{2d+1}{3d+1}} & (\frac{-d}{u}) & (\frac{-d}{u}) & \\ & & & & & & 0 & \sqrt{\frac{d+1}{2d+1}} & (\frac{-d}{t}) & \\ & & & & & & 0 & 0 & \sqrt{\frac{1}{d+1}} & \end{bmatrix}$$

and where $d=\frac{\sigma_\delta^2}{\sigma_\epsilon^2}$, $t=\sqrt{(d+1)(2d+1)}$, $u=\sqrt{(2d+1)(3d+1)}$.

APPENDIX C: Derivation of Restricted Maximum Likelihood Score Function

Let $\underline{y}=X^*\underline{\beta} + \underline{\delta} + \underline{\epsilon}$ be the response surface model for a given BRD in which the response vector

$$\underline{y} \sim N(X^*\underline{\beta}, V).$$

The variance-covariance matrix can be written as

$$V=\sigma_\epsilon{}^2 I + \sigma_\delta{}^2 J^*$$

where

$$J^*=\begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_a \end{bmatrix} \text{ such that } J_i=\underline{1}_{(b_i*1)}\,\underline{1}'_{(1*b_i)};\ i=1,2,\ldots,a.$$

Define $K=[I-X^*(X^{*\prime}X^*)^{-1}X^{*\prime}]$ such that

$$K'\underline{y} \sim N(0, K'VK).$$

$$\ell(V|K'\underline{y}) = -\frac{n}{2}\ln(2\pi)-\frac{1}{2}\ln|K'VK|-\frac{1}{2}\underline{y}'K(K'VK)^{-1}K'\underline{y} \qquad (C.1)$$

The matrices $X^*(X^{*\prime}X^*)^{-1}X^{*\prime}$ and $K(K'K)^{-1}K'$ are both symmetric and idempotent. In addition, since $K'X^*=0$,

$$X^*(X^{*\prime}X^*)^{-1}X^{*\prime}K=0 \text{ and } K(K'K)^{-1}K'X^*=0. \qquad (C.2)$$

By C.2,

$$M=I-X^*(X^{*\prime}X^*)^{-1}X^{*\prime} - K(K'K)^{-1}K'$$

is also a symmetric and idempotent matrix.

Using the fact that X^* is an $n*(p+1)$ matrix of rank $p+1$ and
K is an $n*n$ matrix of rank $n-(p+1)$,

$$\begin{aligned} \text{trace}(MM') = \text{tr}(M) &= \text{tr}(I - X^*(X^{*\prime}X^*)^{-1}X^{*\prime} - K(K'K)^{-1}K') \\ &= \text{tr}(I) - \text{tr}(X^*(X^{*\prime}X^*)^{-1}X^{*\prime}) - \text{tr}(K(K'K)^{-1}K') \\ &= \rho(I) - \rho(X^*(X^{*\prime}X^*)^{-1}X^{*\prime}) - \rho(K(K'K)^{-1}K') \\ &= n - (p+1) - \rho(K) \\ &= 0. \end{aligned} \tag{C.3}$$

The result of C.3 implies $M=0$ and thus,

$$I - X^*(X^{*\prime}X^*)^{-1}X^{*\prime} = K(K'K)^{-1}K'. \tag{C.4}$$

Now define $K^* = V^{\frac{1}{2}}K$ and $X_t^* = V^{-\frac{1}{2}}X^*$.

Using the new variables similar identities to the above example can be formed:

$$I - X_t^*(X_t^{*\prime}X_t^*)^{-1}X_t^{*\prime} = [I - V^{-\frac{1}{2}}X^*(X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-\frac{1}{2}}] \tag{C.5}$$

and

$$K^*(K^{*\prime}K^*)^{-1}K^{*\prime} = V^{\frac{1}{2}}K(K'VK)^{-1}K'V^{\frac{1}{2}}. \tag{C.6}$$

By C.2,

$$K^{*\prime}X_t^* = 0 \quad \text{and} \quad X_t^{*\prime}K^* = 0.$$

Based upon the same argument given in C.3,

$$I - X_t^*(X_t^{*\prime}X_t^*)^{-1}X_t^{*\prime} = K^*(K^{*\prime}K^*)^{-1}K^{*\prime}$$

and

$$[I - V^{-\frac{1}{2}}X^*(X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-\frac{1}{2}}] = V^{\frac{1}{2}}K(K'VK)^{-1}K'V^{\frac{1}{2}}. \tag{C.7}$$

Multiplying the left and right sides of C.7 by $V^{-\frac{1}{2}}$,

$$[V^{-1} - V^{-1}X^*(X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1}] = K(K'VK)^{-1}K'. \tag{C.8}$$

An outline of the derivation of the equality in C.8 can be found on p.452 of Searle, Casella, and McCulloch [1992].

By exercise E9.12 on p.361 of the same reference the following equality holds provided C.8 holds and $X^*'K=0$,

$$|V|=|K'VK|*|X^{*'}V^{-1}X^*|^{-1}. \qquad (C.9)$$

Substituting the equalities of C.8 and C.9 into the log likelihood for $K'\underline{y}$, C.1,

$$\mathcal{L}(V|K'\underline{y}) = -\frac{n}{2}\ln(2\pi)-\frac{1}{2}\ln|V|-\frac{1}{2}\ln|X^{*'}V^{-1}X^*| -\frac{1}{2}[V^{-1}-V^{-1}X^*(X^{*'}V^{-1}X^*)^{-1}X^{*'}V^{-1}] \quad .$$

The same development can be done using hierarchical models and marginal likelihoods. See p.323-325 in Searle, Casella, and McCulloch [1992].

APPENDIX D: REML and MLE Likelihood Equations

REML:

Assuming $K'\underline{y} \sim N(0, K'VK)$ where $V=f(\sigma_\delta^2, \sigma_\epsilon^2)$,

$$\ell(V|K'\underline{y}) = -\frac{n}{2}\ln(2\pi)-\frac{1}{2}\ln|K'VK|-\frac{1}{2}\underline{y}'K(K'VK)^{-1}K'\underline{y}. \tag{D.1}$$

Writing $V=\sigma_\epsilon^2 I + \sigma_\delta^2 J^*$, the derivatives of the matrix V with respect to the error variances are

$$\frac{dV}{d\sigma_\epsilon^2} = I \tag{D.2}$$

and

$$\frac{dV}{d\sigma_\delta^2} = J^*. \tag{D.3}$$

The REML equations utilizing the derivatives in D.2 and D.3 are

$$\frac{d\ell}{\sigma_\delta^2}: \quad \text{tr}\,[(V^{-1}-V^{-1}X^*(X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1})J^*] \tag{D.4}$$
$$=\underline{y}'[V^{-1}-V^{-1}X^*(X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1}]J^*[V^{-1}-V^{-1}X^*(X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1}]\underline{y}$$

and

$$\frac{d\ell}{\sigma_\epsilon^2}: \quad \text{tr}\,[(V^{-1}-V^{-1}X^*(X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1})] \tag{D.5}$$
$$=\underline{y}'[V^{-1}-V^{-1}X^*(X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1}][V^{-1}-V^{-1}X^*(X^{*\prime}V^{-1}X^*)^{-1}X^{*\prime}V^{-1}]\underline{y}.$$

MLE:

Assuming $\underline{y} \sim N(X^*\underline{\beta}, V)$ where $V=f(\sigma_\delta^2, \sigma_\epsilon^2)$,

$$\ell(V|\underline{y}) = -\frac{n}{2}\ln(2\pi)-\frac{1}{2}\ln|V|-\frac{1}{2}(\underline{y}-X^*\underline{\beta})'V^{-1}(\underline{y}-X^*\underline{\beta}). \tag{D.6}$$

For this likelihood, the MLE equations are

$$\frac{d\ell}{\underline{\beta}}: \quad X^{*\prime}V^{-1}\underline{y} = X^{*\prime}V^{-1}X^*\underline{\beta} \tag{D.7}$$

$$\frac{d\ell}{\sigma_\delta^2}: \quad \text{tr}\,[V^{-1}J^*]=\frac{1}{2}(\underline{y}-X^*\underline{\beta})'V^{-1}J^*V^{-1}(\underline{y}-X^*\underline{\beta}). \tag{D.8}$$

$$\frac{d\ell}{\sigma_\epsilon^2}: \quad \text{tr}\,[V^{-1}]=\frac{1}{2}(\underline{y}-X^*\underline{\beta})'V^{-1}V^{-1}(\underline{y}-X^*\underline{\beta}). \tag{D.9}$$

APPENDIX E: Proc Mixed Code for REML Estimators for BRD

Denote the response surface model to be fit for a BRD as

$$\underline{y}=X^*\beta+\underline{\delta}+\underline{\epsilon}$$

where

$$\underline{y} \sim N(X^*\beta,V).$$

For a BRD,

$$V=\sigma_\epsilon{}^2I+\sigma_\delta{}^2J^*$$

where J^* is a block diagonal matrix of $J_i=\underline{1}\ \underline{1}'$; i=1,...,a.

To use Proc Mixed for error variance estimation, the whole-plot error, $\underline{\delta}$, must be modeled as random effects, i.e.

$$\underline{\delta}=S\underline{\varrho}$$

where

$$Var(\underline{\varrho})=G \quad \text{and} \quad Var(\underline{\delta})=SGS'.$$

In addition,

$$Var(\underline{\epsilon})=R.$$

For a BRD,

$$R=\sigma_\epsilon{}^2I,$$
$$G=\sigma_\delta{}^2I,$$

and

$$S=\begin{bmatrix} \underline{1} & & \\ & \underline{1} & \\ & & \ddots \\ & & & \underline{1} \end{bmatrix}.$$

Thus,

$$Var(\underline{y})=SGS'+R=\sigma_\epsilon{}^2I+\sigma_\delta{}^2J^*=V.$$

The SAS code is:

```
Proc Mixed Method=REML;
  Class wp;
  Model y=fixed effects (elements of X*);
  Random wp;
```

where

wp is a classification variable defining into which whole-plot each observation falls;

the model statement defines the model matrix, X^*;

and

the random statement defines S.

By default, $G=\sigma_\delta{}^2 I$ and $R=\sigma_\epsilon{}^2 I$.

APPENDIX F: Designs for Three Variable Examples

Central Composite Design		
1	1	1
1	1	-1
1	-1	1
1	-1	-1
-1	1	1
-1	1	-1
-1	-1	1
-1	-1	-1
α	0	0
$-\alpha$	0	0
0	α	0
0	$-\alpha$	0
0	0	α
0	0	$-\alpha$
0	0	0

Small Composite Design		
1	1	1
1	-1	-1
-1	1	-1
-1	-1	1
α	0	0
$-\alpha$	0	0
0	α	0
0	$-\alpha$	0
0	0	α
0	0	$-\alpha$
0	0	0

Box-Behnken Design		
1	1	0
1	-1	0
-1	1	0
-1	-1	0
1	0	1
1	0	-1
-1	0	1
0	1	1
0	1	-1
0	-1	1
0	-1	-1
0	0	0

Hybrid 310		
0	0	1.2906
0	0	-0.1360
-1	-1	0.6386
1	-1	0.6386
-1	1	0.6386
1	1	0.6386
1.1736	0	-0.9273
-1.1736	0	-0.9273
0	1.1736	-0.9273
0	-1.1736	-0.9273
0	0	0

Hybrid 311b		
0	0	$\sqrt{6}$
0	0	$-\sqrt{6}$
0	0	0
-.7507	2.1063	1
2.1063	.7507	1
.7507	-2.1063	1
-2.1063	-.7507	1
.7507	2.1063	1
2.1063	-.7507	1
-.7507	-2.1063	1
-2.1063	.7507	1

Full Factorial
3^3

APPENDIX G: Designs for Four Variable Examples

Central Composite Design				Small Composite Design				Box-Behnken Design			
1	1	1	1	1	1	1	1	1	1	0	0
1	1	1	-1	1	1	1	-1	1	-1	0	0
1	1	-1	1	1	-1	-1	1	-1	1	0	0
1	1	-1	-1	1	-1	-1	-1	-1	-1	0	0
1	-1	1	1	-1	1	-1	1	1	0	1	0
1	-1	1	-1	-1	1	-1	-1	1	0	-1	0
1	-1	-1	1	-1	-1	1	1	-1	0	1	0
1	-1	-1	-1	-1	-1	1	-1	-1	0	-1	0
-1	1	1	1	α	0	0	0	1	0	0	1
-1	1	1	-1	$-\alpha$	0	0	0	1	0	0	-1
-1	1	-1	1	0	α	0	0	-1	0	0	1
-1	1	-1	-1	0	$-\alpha$	0	0	-1	0	0	-1
-1	-1	1	1	0	0	α	0	0	1	1	0
-1	-1	1	-1	0	0	$-\alpha$	0	0	1	-1	0
-1	-1	-1	1	0	0	0	α	0	-1	1	0
-1	-1	-1	-1	0	0	0	$-\alpha$	0	-1	-1	0
α	0	0	0	0	0	0	0	0	1	0	1
$-\alpha$	0	0	0					0	1	0	-1
0	α	0	0					0	-1	0	1
0	$-\alpha$	0	0					0	-1	0	-1
0	0	α	0					0	0	1	1
0	0	$-\alpha$	0					0	0	1	-1
0	0	0	α					0	0	-1	1
0	0	0	$-\alpha$					0	0	-1	-1
0	0	0	0					0	0	0	0

Hybrid 416a				Hybrid 416c				Full Factorial
1	1	1	0.6444	1	1	1	0.5675	3^4
1	1	-1	0.6444	1	1	-1	0.5675	
1	-1	1	0.6444	1	-1	1	0.5675	
1	-1	-1	0.6444	1	-1	-1	0.5675	
-1	1	1	0.6444	-1	1	1	0.5675	
-1	1	-1	0.6444	-1	1	-1	0.5675	
-1	-1	1	0.6444	-1	-1	1	0.5675	
-1	-1	-1	0.6444	-1	-1	-1	0.5675	
1.6853	0	0	-0.9075	1.4697	0	0	-1.0509	
-1.6853	0	0	-0.9075	-1.4697	0	0	-1.0509	
0	1.6853	0	-0.9075	0	1.4697	0	-1.0509	
0	-1.6853	0	-0.9075	0	-1.4697	0	-1.0509	
0	0	1.6853	-0.9075	0	0	1.4697	-1.0509	
0	0	-1.6853	-0.9075	0	0	-1.4697	-1.0509	
0	0	0	1.7844	0	0	0	1.7654	
0	0	0	-1.4945	0	0	0	0	
0	0	0	0					

BIBLIOGRAPHY

Anderson, R.L. and Bancroft, T.L. (1952), *Statistical Theory in Research*, McGraw-Hill: New York.

Atkinson, A.C. (1982), "Developments in the Design of Experiments," *International Statistical Review*, 50, 161-177.

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

Box, G.E.P. and Draper, N.R. (1959), "A Basis for the Selection of A Response Surface Design," *Journal of the American Statistical Association*, 54, 622-654.

---------- (1987), *Empirical Model Building and Response Surfaces*, John Wiley & Sons: New York, N.Y.

Box, G.E.P. and Jones, S. (1992), "Split-plot Designs for Robust Parameter Design Approach," *Journal of Applied Statistics*, 19, 3-26.

Box, G.E.P. and Wilson, K.B. (1951), "On the Experimental Attainment of Optimum Conditions," *Journal of the Royal Statistical Society*, Series B, 13, 1-45.

Cochran, W.G. and Cox, Gertrude M. (1957), *Experimental Designs*, John Wiley & Sons, Inc.: New York.

Cox, D.R. (1958), *Planning of Experiments*, Wiley: New York, N.Y.

Davies, O.L. (1954), *The Design and Analysis of Industrial Experiments*, Hafner Press: New York, N.Y.

Draper, N.R. (1963), "Ridge Analysis of Response Surfaces," *Technometrics*, 5, 469-479.

Graybill, F.A. (1976), *Theory and Application of the Linear Model*, Duxbury Press: MA.

Hartley, H.O. (1959), "Smallest Composite Design for Quadratic Response Surfaces," *Biometrics*, 15, 611-624.

Hill, W.J. and Hunter, W.G. (1966), "A Review of Response Surface Methodology: A Literature Review," *Technometrics*, 8, 571-590.

Hinkelmann, K. and Kempthorne, O. (1994), *Design and Analysis of Experiments: Volume 1. Introduction to Experimental Design.* John Wiley & Sons: New York, N.Y.

Hoerl, A.E. (1959), "Optimum Solution of Many Variables Equations," *Chemical Engineering Progress*, 55, 69-78.

Hoke, A.T. (1974), "Economical Second-Order Designs Based on Irregular Fractions of the 3^n Factorial," *Technometrics*, 16, 375-384.

Kackar, R.N. (1985), "Off-Line Quality Control, Parameter Design, and the Taguchi Method," *Journal of Quality Technology*, 17, 176-209.

Karlin, S. and Studden, W.J. (1965), "Optimal Experimental Designs," *Annals of Mathematical Statistics*, 37, 783.

Kiefer, J. (1958), "On the Nonrandomized Optimality and the Randomized Nonoptimality of Symmetrical Designs," *Annals of Mathematical Statistics*, 29, 675-699.

--------(1959), "Optimum Experimental Designs," *Journal of the Royal Statistical Society*, 21, 272-319.

------- (1961), "Optimum Designs in Regression Problems, II," *Annals of Mathematical Statistics*, 32, 298.

Kiefer, J. and Wolfowitz, J. (1959), "Optimum Designs in Regression Problems," *Annals of Mathematical Statistics*, 30, 271-294.

----------(1960), "The Equivalence of Two Extremum Problems," *Canadian Journal of Mathematics*, 12, 363-366.

Khuri, A.I. and Cornell, J.A. (1987), *Response Surfaces: Design and Analysis*, Marcel Dekker: New York, N.Y.

Lentner, M. and Bishop, T. (1993), *Experimental Design and Analysis*, Valley Book Company: Blacksburg, VA.

Myers, R.H. (1976), *Response Surface Methodology*, Author: Blacksburg, VA.

Myers, R.H. and Khuri, A.I. (1979), "A New Procedure for Steepest Ascent," *Communications in Statistics, Part A-Theory and Methods*, 8, 1359-1376.

Myers, R.H., Khuri, A.I. and Carter, W.H. (1989), "Response Surface Methodology: 1966-1988," *Technometrics*, Vol 31, No. 2.

Myers, R.H., Khuri, A.I. and Vining, G.G. (1992), "Response Surface Alternatives to the Taguchi Robust Parameter Design Approach," *The American Statistician*, 46, 2, 131-139.

Myers, R.H. and Montgomery, D.C. (1995), *Response Surface Design and Analysis in Quality Improvement*, John Wiley & Sons: New York, N.Y.

Nair, V.N., editor, (1992), "Taguchi's Parameter Design: A Panel Discussion," *Technometrics*, 34, 2, 127-161.

Nelder, J.A. and Mead, R. (1965), "A Simplex Method for Function Minimization," *Computer Journal*, 7, 308-313.

Patterson, H.D. and Thompson, R. (1971), "Recovery of Inter-block Information When Block Sizes are Unequal," *Biometrika*, 58, 545-554.

----------(1974), "Maximum Likelihood Estimation of Components of Variance," *Proceedings of Eighth International Biom. Conference*, 17-209.

Plackett, R.L. and Burman, J.P. (1946), "The Design of Multifactorial Experiments," *Biometrika*, 33, 305-325.

PROC MIXED, *SAS Technical Report P-229; SAS/STAT Software: Changes and Enhancements*, Release 6.09, 1992, Chapter 16.

Roquemore, K.G. (1976), "Hybrid Designs for Quadratic Response Surfaces," *Technometrics*, 18, 419-423.

Russell, T.S. and Bradley, R.A. (1958), "One-way Variances in the Two-way Classification," *Biometrika*, 45, 11-129.

Satterthwaite, F.E. (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin 2*, 110-114.

Searle, S.R., Casella, G. and McCulloch, C.E. (1992), *Variance Components*, John Wiley & Sons, N.Y.

Silvey, S.D. (1980), *Optimal Designs*, Chapman and Hall: London.

Taguchi, G. (1986), *Introduction to Quality Engineering*, UNIPUB/ Kraus International:White Plains, N.Y.

---------- (1987), *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost*, UNIPUB/Kraus International: White Plains, N.Y.

Taguchi, G. and Wu, Y. (1980), *Introduction to Off-Line Quality Control*, Central Japan Quality Control Association: Tokyo.

---------- (1985), *Introduction to Off-Line Quality Control*, Central Japan Quality Control Association: Nagoya, Japan.

Thompson, W.A. (1962), "The Problem of Negative Estimates of Variance Components," *Annals of Mathematical Statistics*, 33, 273-289.

Vining, G.G. and Myers, R.H. (1990), "Combining Taguchi and Response Surface Philosophies: A Dual Response Approach," *Journal of Quality Technology*, Vol. 22, 1, 38-45.

Yates, F. (1935), "Complex Experiments," *Journal of the Royal Statistical Society,* Supplement 2, 181-247.

--------- (1937), *The Design and Analysis of Factorial Experiments*, Harpenden: Imp. Bur Soil Sci.

VITA

Jennifer J. Davison, daughter of James H. and Gail L. Davison, was born on April 12, 1968 in Chicago, Illinois. At age four, she and her family moved to Springfield, Virginia. In 1986, she graduated as valedictorian from West Springfield High School. She then attended James Madison University in Harrisonburg, Virginia where she majored in Mathematics and minored in Statistics. In 1990, she graduated Summa Cum Laude and was the recipient of the J. Emmert Ikenberry Mathematics Award. She continued her education in Statistics at Virginia Polytechnic Institute and State University in Blacksburg, Virginia where she received her Masters degree in Statistics in the Fall of 1991 and her Ph.D. in Statistics in the Summer of 1995. On December 10, 1994, she married William C. Letsinger, II.