To the Graduate Council:

I am submitting herewith a dissertation written by David Joseph Edwards entitled "Design and Analysis of Screening Experiments Assuming Effect Sparsity". I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Business Administration.

Robert W. Mee, Major Professor

We have read this dissertation
and recommend its acceptance:

Mary G. Leitnaker

Russell L. Zaretzki

Myong K. Jeong

Accepted for the Council:

Carolyn R. Hodges,

Vice Provost and Dean of the
Graduate School

(Original Signatures are on file with official student records.)

Design and Analysis of Screening Experiments Assuming Effect Sparsity

A Dissertation
Presented for the
Doctor of Philosophy Degree
The University of Tennessee, Knoxville

David Joseph Edwards
August 2008

Copyright © 2008 by David Joseph Edwards.
All rights reserved.

Dedication

This dissertation is dedicated in honor of:

Elizabeth, Joseph, LaVerne, Kevin, Arlett, and Lillie

and in memory of:

Wesley and Christine.

Acknowledgments

I would like to thank my advisor, Dr. Robert Mee, for the many hours you have spent teaching me how to be a better researcher and educator. Thank you for your guidance, patience, persistence, and attention to details. You have helped to provide considerable growth and insight for my professional career in academia. Thank you for always keeping your door open.

Thanks also to the other members of my dissertation committee. I would like to thank Dr. Mary Leitnaker for teaching me that there is more to industrial statistics than just a set of methods. Your sense for the needs of practitioners has inspired me to always think of them when developing new methods. Thanks to Dr. Russell Zaretzki for being both a great teacher and colleague. I have enjoyed our conversations on topics ranging from research, work, to life in general. Thanks also to Dr. Myong Jeong for bringing your thoughts and insights based on your expertise in both statistics and industrial engineering.

Thanks to my office mates, Artin Armagan and Maria Weese, for their friendship. Without all of the conversations, encouragements, and laughs, I would have been done with this dissertation six months ago. Finally, I want to thank all of the faculty and staff in the Department of Statistics, Operations, and Management Science for your help and support over the years. This has truly been an adventure.

Abstract

Many initial experiments for industrial and engineering applications employ screening designs to determine which of possibly many factors are significant. These screening designs are usually a highly fractionated factorial or a Plackett-Burman design that focus on main effects and provide limited information for interactions. To help simplify the analysis of these experiments, it is customary to assume that only a few of the effects are actually important; this assumption is known as 'effect sparsity'. This dissertation will explore both design and analysis aspects of screening experiments assuming effect sparsity.

In 1989, Russell Lenth proposed a method for analyzing unreplicated factorials that has become popular due to its simplicity and satisfactory power relative to alternative methods. We propose and illustrate the use of p-values, estimated by simulation, for Lenth t-statistics. This approach is recommended for its versatility. Whereas tabulated critical values are restricted to the case of uncorrelated estimates, we illustrate the use of p-values for both orthogonal and nonorthogonal designs. For cases where there is limited replication, we suggest computing t-statistics and p-values using an estimator that combines the pure error mean square with a modified Lenth's pseudo standard error.

Supersaturated designs (SSDs) are designs that examine more factors than runs available. SSDs were introduced to handle situations in which a large number of factors are of interest but runs are expensive or time-consuming. We begin by assessing

the null model performance of SSDs when using all-subsets and forward selection regression. The propensity for model selection criteria to overfit is highlighted. We subsequently propose a strategy for analyzing SSDs that combines all-subsets regression and permutation tests. The methods are illustrated for several examples.

In contrast to the usual sequential nature of response surface methods (RSM), recent literature has proposed both screening and response surface exploration using only one three-level design. This approach is named "one-step RSM". We discuss and illustrate two shortcomings of the current one-step RSM designs and analysis. Subsequently, we propose a new class of three-level designs and an analysis strategy unique to these designs that will address these shortcomings and aid the user in being appropriately advised as to factor importance. We illustrate the designs and analysis with simulated and real data.

Contents

1 Introduction **1**

2 Empirically Determined p-Values for Lenth t-Statistics **10**
2.1 Introduction and Motivation 10
2.2 Simulations to Compute p-values 14
2.2.1 Example 1: An Unreplicated Orthogonal Design 16
2.3 The Impact of Correlated Parameter Estimates 17
2.3.1 Example 2: A 2^4 with One Observation Missing 20
2.3.2 Example 3: Retchshaffner Design for Seven Factors 22
2.4 Pure Error Degrees of Freedom 25
2.4.1 Example 4: A Fractional Factorial with Center Point Replication 32
2.5 Discussion 35

3 Supersaturated Designs: Are Our Results Significant? **38**
3.1 Introduction and Motivation 38
3.2 A Strategy for Analyzing Supersaturated Designs 50
3.2.1 More on All-Subsets vs. Forward Selection 50
3.2.2 Permutation Tests 55
3.2.3 Proposed Analysis Strategy 58
3.3 Examples 62
3.3.1 Williams Data 62
3.3.2 A Bayesian D-Optimal SSD 73
3.3.3 AIDS Data 78
3.4 Discussion 81

4 A New Class of Designs for Factor Screening and Response Surface Exploration **85**
4.1 Introduction and Motivation 85
4.2 New Three-Level Designs for Screening and Response Surface Exploration 96
4.2.1 Motivation 96
4.2.2 Resolution III Subsets of Four Runs (Fractional Box-Behnken Designs) 102

4.2.3 Other Designs 119
4.3 Analysis Strategy 121
4.3.1 Analysis Approach 1 125
4.3.2 Analysis Approach 2 153
4.4 Discussion 159

5 Concluding Remarks **164**

Bibliography **171**

Vita **179**

List of Tables

2.1 Example from Hicks and Turner [40] 16
2.2 Example 1: Lenth t Statistics and Comparison of P-value Approximations 17
2.3 $\alpha = 0.05$ Critical Values and IER for m Equicorrelated Contrasts . . 19
2.4 Example 2 - a 2^4 with one missing observation 21
2.5 Example 2 Estimates and P-value Comparisons 21
2.6 Example 3 - Retchshaffner's Nonorthogonal Design 23
2.7 Example 3 Estimates and Empirical P-values for Saturated Model . . 24
2.8 Power Simulation [$m = 7$,$\pi^* = 2.\bar{3}/(2.\bar{3} + df_{PE})$, $\omega^* = 2.\bar{3}/(2.\bar{3} + 5df_{PE})$] 27
2.9 Power Simulation [$m = 15$,$\pi^* = 5/(5 + df_{PE})$, $\omega^* = 5/(5 + 5df_{PE})$] . . 28
2.10 Power Simulation [$m = 31$,$\pi^* = 10.\bar{3}/(10.\bar{3} + df_{PE})$, $\omega^* = 10.\bar{3}/(10.\bar{3} + 5df_{PE})$] 29
2.11 Power Simulation with Effects of Various Sizes [$\pi^* = (m/3)/((m/3) + df_{PE})$, $\omega^* = (m/3)/((m/3) + 5df_{PE})$] 31
2.12 Example 4- 2^{5-1} Design with 3 Center Point Runs 33
2.13 Example 4 Comparison of P-values based on MSPE, PSE, and CSE_π 34

3.1 Plackett-Burman design with n=12, k=11, and two SSDs with $n_{SS} = 6$ and $k_{SS} = 10$ 40
3.2 Unbalanced Supersaturated Design to Maximize Probability of Identifying Active Factors 41
3.3 Data from Williams [94] 46
3.4 Analysis of Full Williams Data 63
3.5 Full Williams Data - All-Subsets 64
3.6 Full Williams Data - Global Model Test 65
3.7 Full Williams Data - ΔAIC_c 65
3.8 Analysis of Supersaturated Subset of Williams Data 1 - Global Test . 67
3.9 Analysis of Supersaturated Subset of Williams Data 1 - ΔAIC_cs . . . 68
3.10 Analysis of Supersaturated Subset of Williams Data 1 - Partial Tests 68
3.11 Analysis of Supersaturated Subset of Williams Data 2 - Global Test . 70
3.12 Analysis of Supersaturated Subset of Williams Data 2 - ΔAIC_cs . . . 71
3.13 Analysis of Supersaturated Subset of Williams Data 2 - Partial Tests 71
3.14 Analysis of Supersaturated Subset of Williams Data 3 - Global Test . 72
3.15 Bayesian D-Optimal Design 74

3.16 Analysis of Bayesian D-Optimal SSD - Global Test 76
3.17 Analysis of Bayesian D-Optimal SSD - ΔAIC_cs 77
3.18 Analysis of Bayesian D-Optimal Design - Partial Tests 78
3.19 Analysis of AIDS Data - Global Test 80

4.1 Example from CW 88
4.2 Simulated Example using CW's Design - Pure Quadratic Model . . . 90
4.3 Simulated Example using CW's Design - Second-Order Model 90
4.4 18-run OA from XCW 94
4.5 Example 3 Results 94
4.6 First Bayesian D-optimal Search using OPTEX 100
4.7 Box-Behnken Design for 6 factors 101
4.8 Subset of Second Bayesian D-optimal Search using OPTEX 102
4.9 Bayesian D-optimal Design for 7 factors in 29 runs 103
4.10 Aliasing Structure of new 29-run design in 7 factors 105
4.11 Aliasing Structure of 8-factor, 33-run design (Cyclic IBD Construction) 108
4.12 Summary of New Three-Level Designs 108
4.13 Eligible Projections and D-Efficiencies of Table 4.12 Designs 110
4.14 Eligible Projections and D-Efficiencies of XCW's Recommended 18-run OA 111
4.15 Eligible Projections and D-Efficiencies of XCW's Recommended 27-run OA 112
4.16 Eligible Projections and D-Efficiencies of D-optimal Designs 113
4.17 Simulation Results (Scenario 1) 116
4.18 Simulation Results (Scenario 2) 118
4.19 41-run 10 factor design with Resolution II Blocks 120
4.20 Critical Values for Analysis Approach 1 for Table 4.12 Designs 128
4.21 Design and Response for Example 4 130
4.22 Aliasing of Example 4 Design 131
4.23 Example 4 - Analysis Stage 1 132
4.24 Second-Order Analysis Results for Example 4 133
4.25 Variance of Pseudo-Estimates - Example 4 135
4.26 Design and Response for Example 5 137
4.27 Aliasing of Example 5 Design 138
4.28 Example 5 - Analysis Strategy Step 1 139
4.29 Second-Order Analysis Results for Example 5 140
4.30 Variance of Pseudo-Estimates - Example 5 142
4.31 Aliasing of Example 6 Design 144
4.32 Example 6 - Analysis Strategy Step 1 145
4.33 Second-Order Analysis Results for Example 6 147
4.34 Design and Response for Example 7 149
4.35 Aliasing of Example 7 Design 149
4.36 Example 7 - Analysis Stage 1 150

4.37 Second-Order Analysis Results for Example 7 152
4.38 Critical Values for Approach 2 154
4.39 Example 8 - Example 5 Approach 2 155
4.40 Example 9 - Example 6 Approach 2 157
4.41 Example 10 - Example 7 Approach 2 159

List of Figures

2.1 Comparison of T-Distributions with Lenth's T 13
2.2 Upper Tail Comparison of Distributions 14

3.1 Mean AIC_c vs. Number of Factors- Null Model (Random Models) . . 53
3.2 Mean AIC_c vs. Number of Factors -Null Model (All-Subsets and Forward Selection) 53
3.3 R^2 vs. Number of Factors - Null Model 55
3.4 Histogram of Untransformed AIDS Data 79

4.1 Aliasing Structure of XCW 18-run OA in 7 factors 104
4.2 Aliasing Structure of XCW 27-run OA in 7 factors 104
4.3 Example 4 - Analysis Stage 1 131
4.4 Example 5 - Analysis Stage 1 138
4.5 Example 6 - Analysis Stage 1 146
4.6 Example 7 - Analysis Stage 1 151

Chapter 1

Introduction

Experiments are performed by researchers in many different fields of study ranging from agriculture and the life sciences to business and engineering with the common objective of utilizing experimental data to obtain useful information about some system, process, organization, etc. under study. Given such widespread experimentation, statistical design and analysis of experiments has become a vital tool and essential topic in the graduate level statistics curriculum. In fact, its importance is easily recognized by its representation in the statistical literature. Textbooks such as Montgomery [72], Wu and Hamada [96], Box, Hunter, and Hunter [15], Ryan [88], Dean and Voss [26], and Lentner and Bishop [54] have all been developed for the purpose of introducing students and practitioners with the essential methods with which to go about designing, conducting, and analyzing experiments.

According to Montgomery [72], the "statistical design of experiments refers to the process of planning an experiment so that appropriate data that can be analyzed by statistical methods will be collected, resulting in valid and objective conclusions. The statistical approach to experimental design is necessary if we wish to draw meaningful conclusions from the data. When the problem involves data that are subject to experimental errors, statistical methodology is the only objective approach to analysis. Thus, there are two aspects to any experimental problem: the design of the

experiment and the statistical analyses of the data. These two subjects are closely related because the method of analysis depends directly on the design employed."

This dissertation will explore both design and analysis aspects of experimentation in the context of screening experiments assuming that only a small portion of the effects are active (effect sparsity). Box and Meyer [16] stated that typically only about 20% of the contrasts are active. Our intention is to propose and illustrate useful methods that can be easily implemented by practitioners seeking to appropriately carry out the design and analysis of an experiment. Rather than narrow ourselves to a single system or process, we maintain a general setting in order to allow for a broader focus. In fact, our examples will come from different areas of study. Throughout, we will assume the usual linear model $\mathbf{y} = \mathbf{X}\beta + \epsilon$, $\epsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ where $\mathbf{y}$ is the $n \times 1$ vector of responses and $\mathbf{X}$ is the $n \times p$ model matrix of a given design. Let b be the least squares estimate of β. Then, $b = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $\mathbb{V}(b) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, assuming $\mathbf{X}$ is of full rank. We will investigate the non-full rank case in Chapter 3.

Chapter 2 of this dissertation explores the analysis of saturated designs (i.e. designs in which $rank(\mathbf{X})$=n using the method of Lenth [53]). Many initial experiments for industrial and engineering applications employ screening designs to determine which of possibly many factors are significant. Due to their economical run sizes, unreplicated two-level factorial or fractional factorial designs are used extensively. However, due to the lack of information about the error variance, identification of active contrasts from such designs is not a straightforward task using the tools of classical statistical inference.

Lenth [53] proposed a method for analyzing unreplicated factorials by computing a pseudo standard error (PSE) that may be used in conjunction with normal effects plots to alleviate some of the plot's subjective nature. Since its introduction, Lenth's method has become a popular tool for analyzing such designs due to its simplicity

and satisfactory power, relative to alternative methods. Software packages, such as Minitab and JMP, have incorporated Lenth's method.

Subsequent authors found that empirically-determined percentiles from the null distribution of Lenth t-statistics were preferred to using percentiles from an approximating Student t distribution. Ye and Hamada's (2000) tables of critical values for Lenth's method are often utilized to determine statistical significance. In Chapter 2, we propose and illustrate the use of empirically-determined p-values for Lenth t-statistics. As will be seen, these p-values can be easily computed via Monte Carlo simulation and should make the analysis of unreplicated factorials more user-friendly. This approach is recommended for its versatility, since tabulated critical values are designed to handle only certain situations. Our illustrative examples include both orthogonal and non-orthogonal designs.

Furthermore, we propose a simple method of combining Lenth's PSE with a pure error variance estimator for cases where there is limited replication. Pure error degrees of freedom (df) are obtained by replicating any design point(s). Although one could estimate the error variance with the mean square pure error (MSPE), such an estimator will be imprecise if there is little relication. Assuming effect sparsity, it seems that one should not ignore the information available in Lenth's PSE. Larntz and Whitcomb [51] proposed using a linear combination of MSPE and PSE^2, weighting each by its degrees of freedom (df) (or pseudo df). In conjunction with the use of this combined standard error, we propose a modification to incorporate the model independent information provided by MSPE into the adaptive feature of Lenth's method before computing the PSE. The use of this modified PSE will be shown to make the procedure more robust to violations of effect sparsity.

If the number of factors is very large and/or experimental runs are very expensive, then even the use of the common screening designs can become impractical.

Supersaturated designs (SSDs) are designs that examine more than n-1 factors in n runs. Naturally, SSDs have too few runs to allow for estimation of main effects of all the factors of interest, which is an unavoidable source of ambiguity in analysis. Although it is relatively easy to find literature involving the construction of SSDs, there is much less available that focuses on their analysis. Gilmour [34] states that although methods for analyzing data from SSDs have been proposed, none of them seem very convincing. Therefore, although methods exist for constructing SSDs with good statistical properties, current methods of analysis are not as convincing. Thus, in Chapter 3 we will investigate further the analysis of SSDs.

Whether using forward selection or all-subsets regression, it is common to select models from SSDs that explain a very large percentage of the total variation in a response. The naïve p-values one sees for the selected model can persuade the user that the included factors are clearly active. This has contributed to the poor performance of the stepwise and even the all-subsets procedure in terms of Type I error. The forward selection procedure also suffers from an inability to entertain and compare multiple models of the same size. That is, since the aliasing structure inherent in the SSD can hide real effects or encourage identifying nonactive effects as active, it is common for a stepwise procedure to easily be led astray by the entry of a nonactive effect. Therefore, it has been suggested that all-subsets regression be utilized instead of forward selection.

In Chapter 3, we begin by assessing the null model performance of SSDs when using all-subsets and forward selection regression. The propensity for model selection criteria to overfit is highlighted. We subsequently propose a strategy for analyzing SSDs that combines all-subsets regression and permutation tests. That is, in conjunction with all-subsets regression, we show how permutation procedures may be used to more appropriately select candidate models of interest and ascertain statis-

tical significance for individual coefficients. Also, we will show how the power for detecting active effects decreases as the number of factors in the SSD increases. We will illustrate the method with several real and simulated data sets.

Sir Ronald Fisher once said that 'the best time to design an experiment is after you've done it.' The sequential approach to design and analysis of experiments is utilized, in general, as a way to gain more information on a decreasing number of factors. As mentioned above, the beginning stage of experimentation consists of an initial experiment involving many factors. These screening designs are usually a highly fractionated factorial or a Plackett-Burman design (see Plackett and Burman [84]) that focus on main effects and provide limited information for interactions. Subsequent stages may entertain fewer factors, but provide us with the opportunity to gain more information regarding interaction effects. Nelson, Montgomery, Elias, and Maass [77] discuss and compare several sequential design strategies.

Box [12] discusses the questions that arise after a preliminary experiment has been conducted and how such questions lead us to naturally consider the ideas of sequential (follow-up) experimentation. In particular, Box [12] lists six customary design follow-up strategies. We discuss each of them briefly below and cite some relevant literature.

1. *Move to a new location.* Suppose the results of an initial experiment suggest that a new region of experimentation is appropriate. It is then appropriate to choose a set of follow-up runs that assist the experimenter in moving to such a region of improvement. One of the most common design augmentation strategies for moving to a new location is the method of steepest ascent (SA). SA was first introduced by Box and Wilson [18] and arises most often within the framework of response surface methods (RSM). RSM is defined by Myers, Montgomery, et. al. [76] as a collection of statistical design and numerical optimization techniques

used to optimize processes and product designs. In particular, sequential design of experiments is inherent within the RSM framework. For more on RSM, see the excellent review articles by Hill and Hunter [41] and Myers [74] and the books Khuri and Cornell [50], Myers and Montgomery [75], and Box and Draper [13]. Simply stated, SA is a gradient search optimization method that has the objective of increasing the value of some response function.

2. *Add another fraction.* Adding another fraction to some initial fractional factorial design may be necessary in order to resolve ambiguities involving aliasing of effects in the initial experiment and/or to increase the precision of factorial effect estimates.

   Foldover is a classic technique used to select follow-up runs. In particular, foldover is often used to increase the resolution of a fractional factorial design from III to IV. These runs are obtained by reversing the signs of one or more factors (columns) in an initial design. Li and Lin [57] consider optimal foldover plans for regular two-level fractional factorial designs of run sizes of 16 and 32 and demonstrate that there are several equivalent ways to generate a particular foldover design. Li, Lin, and Ye [58] determine optimal foldover plans for two-level nonregular designs. The reader should recall that regular fractional factorial designs are those constructed by using interaction contrasts to generate additional factors. Designs that do not possess this property are called nonregular designs (see Wu and Hamada [96] for more). Mee and Peralta [70] address semifolding, which consists of adding another fraction half the size of the original experiment. Meyer, Steinberg, and Box [71] introduce a Bayesian approach to adding runs that help resolve confounding and allow the experimenter to discriminate between competing models.

3. *Rescale the Design.* It may happen that after an initial experiment has been conducted that the experimenter discovers that certain factor levels were too wide or narrow to gain any useful information. Rescaling factors is simple and commonly done. However, there is little literature pertaining to this follow-up strategy.

4. *Drop or Add Factors.* This is another simple and commonly used strategy that is not commonly addressed conceptually in the literature, although it does arise in application articles. Suppose that information is needed regarding a factor that was not included in an initial experiment. The obvious solution would be to incorporate that factor in any remaining experiments. On the other hand, certain factors may be deemed insignificant following an initial experiment and thus, no longer included in a follow-up experiment. Gilmour and Mead [35] investigate designs in which one of the factors is fixed at a particular level after the first experiment is conducted. They show that fixing a factor at a particular level can sometimes allow all important effects in the remaining factors to be estimated in fewer runs than standard designs.

5. *Replicate Runs.* Replicating certain runs provides a means of gaining an estimate of pure error, checking model adequacy, and/or repairing the results of a run because it was believed to be wrongly conducted. For instance, replicating the center point of a design allows one to check for pure quadratic curvature as well as obtain an estimate of pure error, which would not be available in an unreplicated design (see Montgomery [72] for more).

6. *Augment.* In this instance, the word augment is intended to indicate that an initial experiment is indeed in some appropriate region of experimentation and further runs are desired in this region. If an optimal setting of factors is sought,

one may consider augmenting the initial design with 'axial points' and/or center point runs in order to construct a second-order design known as a central composite design (CCD) (especially if curvature is suspected). CCD's are discussed in detail in, for example, Myers and Montgomery [75] and Montgomery [72].

Cheng and Wu [22] introduce a novel method for exploring a response surface using only one design. As alluded to above, it is common practice in response surface optimization to perform experiments sequentially by beginning with a screening experiment, then moving to a new region using steepest ascent, and finally fitting a response surface model using a second-order design. Cheng and Wu's [22] approach is to use 3-level regular or nonregular designs (as opposed to the usual 2-level designs) in order to first screen a large number of factors and then project from the larger factor space onto a smaller factor space to perform response surface exploration. In particular, step 1 of Cheng and Wu's [22] analysis strategy begins with a main-effects only analysis in which the important factors are identified. The goal of step 2 is to then project the design onto the important factors and fit a second-order model. One of the drawbacks of this method is that, given the usual screening design scenario with many factors in a small number of runs, the initial screening stage is unable to entertain the possibility of two-factor interactions and thus potentially neglect important factors. Another drawback is that the projection onto the factors of interest does not always yield a second-order design.

In Chapter 4, we address these two drawbacks and propose a new class of three-level designs for the dual purpose of screening and response surface exploration. It will be shown that these designs are straightforward to construct, possess a much simpler aliasing structure, and have better projection properties than those currently advocated in the literature for the same purpose. Furthermore, we propose an analysis strategy that better equips the experimenter to consider two-factor interactions in the

screening stage of analysis and thus be more advised as to factor importance before proceeding to the projection stage of analysis.

The final chapter of the dissertation, Chapter 5, provides concluding remarks. Suggestions for future research are referred to in this chapter. The reader should note that the author plans future work in this field during his career and more work is certainly forthcoming.

Chapter 2

Empirically Determined p-Values for Lenth t-Statistics

2.1 Introduction and Motivation

Initial experiments for industrial and engineering applications commonly employ screening designs to determine which of possibly many factors affect the response. Due to their economical run sizes, unreplicated two-level factorial and fractional factorial designs are used extensively. However, due to the lack of information about the error variance, identification of active contrasts from such designs is not a straightforward task using the tools of classical statistical inference.

Lenth [53] proposed a simple, intuitive method that has become popular for analyzing unreplicated factorials. Although the computations are easy to carry out by hand, software packages such as JMP and Minitab have integrated Lenth's method into their output, assisting in its widespread use.

Hamada and Balakrishnan [39] present an excellent and thorough review of methods for analyzing unreplicated factorial experiments. They examine two dozen methods, either as proposed in the literature or with reasonable modifications, and compare them in terms of level and power via an extensive simulation study. The simulation showed that Lenth's method performs acceptably in terms of power; Haaland and

O'Connell [38] reached a similar conclusion. Thus, given the computational simplicity and satisfactory power of Lenth's method, its widespread use is justified. We take a moment now to introduce notation and review Lenth's method.

In this chapter, we assume the linear model,

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \tag{2.1}$$

where $\mathbf{Y}$ is an $n \times 1$ vector of responses, $\mathbf{X}$ is an $n \times (m+1)$ model matrix of full column rank, $\beta = (\beta_0, \beta_1, \beta_2, ..., \beta_m)'$ is a vector of unknown coefficients, and ϵ is an $n \times 1$ vector of independently distributed random variables with mean 0 and unknown variance σ^2. Let $\mathbf{b} = (b_0, b_1, b_2, ..., b_m)' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ be the least squares estimator of β. In order to handle cases where the diagonal elements of $\mathbb{V}(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ are not all equal, for $i = 1, ..., m$, we define the standardized coefficients

$$c_i = \frac{b_i}{\sqrt{v_{ii}}}, \tag{2.2}$$

where v_{ii} is the diagonal element of $\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}$ corresponding to b_i.

Lenth [53] proposed fitting a *saturated* model to analyze unreplicated factorial designs, so that under an assumption of effect sparsity the many negligible estimates could be used to estimate σ. The model (2.1) is saturated if $m+1$ equals the number of distinct rows of $\mathbf{X}$. For unreplicated designs, the saturated X is a square matrix. Each standardized coefficient c_i has mean $\beta_i / v_{ii}^{1/2}$, variance σ^2, and being a linear combination of $\mathbf{Y}$, is approximately normally distributed. If the saturated model (2.1) contains many terms with zero coefficients, then the vector of standardized estimates $(c_i, ..., c_m)$ resembles a sample from a $N(0, \sigma^2)$ distribution, plus perhaps a few outliers.

Lenth's method begins by computing for a saturated model

$$s_0 = 1.5 * median\left\{|c_1|, |c_2|, ..., |c_m|\right\} \tag{2.3}$$

and then the pseudo standard error (PSE)

$$\text{PSE} = 1.5 * median_{|c_i|<2.5s_0} |c_i| . \tag{2.4}$$

For $C \sim N(0, \sigma^2)$, the median of $|C|$ equals 0.6745σ, so $1.5{*}median\,|C| = 1.01\sigma$. As Lenth [53] shows, the trimming of extreme c_i's makes the PSE roughly consistent for σ when all of the β_i's equal zero. One obtains Lenth t-statistics by computing

$$t_{Lenth,i} = \frac{c_i}{PSE}. \tag{2.5}$$

Lenth [53] suggested the t-distribution with $m/3$ degrees of freedom (df) as an approximation to the null distribution for (2.5), based on fitting the empirical distribution of PSE^2 by scaled chi-squared distributions. Thus, an effect is declared active if $|t_{Lenth,j}| > t^{\alpha/2}_{m/3}$, where $t^{\alpha/2}_{m/3}$ is the upper $\alpha/2$ percentile of a t-distribution with $m/3$ df. Ye and Hamada [99] report that this approximation does not maintain the individual error rate (IER) at the nominal level; see also Loughin [64]. Subsequent authors found that, for independently distributed c_i's, percentiles from the simulated null distribution of (2.5) were preferred to using percentiles from any approximating Student t distribution. Ye and Hamada [99] present extensive individual and simultaneous test critical values for Lenth's method, obtained by simulation, for many values of m that arise with two-level and three-level orthogonal designs.

No t-distribution provides an appropriate approximation to the null distribution of Lenth t-statistics. Figures 2.1 and 2.2 illustrate this by comparing Lenth's distribution

when $m = 7$ with t-distributions that have 7/3 and 7 df. We obtain similar results for other values of m. One notable aspect of the distribution of Lenth's t-statistic is the discrete probability at $|t| = 2/3$. If an odd number of c_i's are smaller than $2.5s_0$, then PSE=$1.5c_i$ for at least one i, and so (2.5) equals 2/3 for that effect. For example, we have found using simulation that $P(|t| = 2/3) = 0.122$ when m=7 and all true effects are zero. The occurrence of this probability mass is one reason why no t-distribution provides a suitable approximation to Lenth's t-distribution over the entire range of possible t-values.

In this chapter, we illustrate the use of empirically determined p-values for Lenth t-statistics for any saturated linear model. P-values can be computed via Monte Carlo simulation even more easily than can percentiles for $t_{Lenth,j}$, and should aid in making analysis of unreplicated designs more user-friendly. Our examples include both orthogonal and non-orthogonal designs. Following Larntz and Whitcomb [51], we also propose a simple modification of Lenth's method that makes use of a pure error variance estimator for cases where there is limited replication.

The rest of this chapter is organized as follows. First we summarize the steps for computing p-values for Lenth t-statistics from unreplicated designs and illustrate

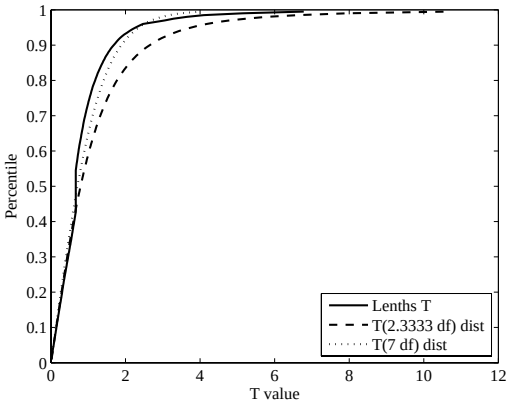

Figure 2.1: Comparison of T-Distributions with Lenth's T

the computation of p-values for a simple orthogonal design. We then investigate how correlations among the estimates impact the null distribution of Lenth t-statistics. Analysis of two non-orthogonal designs follows. Finally, we consider an example with center point replication and suggest an improvement to Larntz and Whitcomb's [51] proposal for handling such cases. We conclude with a discussion of our results and possible extensions.

2.2 Simulations to Compute p-values

For a given set of least squares estimates $b_1, ..., b_m$ computed from the model matrix, $\mathbf{X}$, we may obtain p-values for the individual tests $H_0 : \beta_i = 0$ vs. $H_a : \beta_i \neq 0$ as follows:

- Step 1: Compute the standardized coefficients using (2.2), the PSE, and then the m Lenth t-statistics as in (2.5).
- Step 2: Compute the $m \times m$ matrix, $\mathbf{R}$, where the ij^{th} element of $\mathbf{R}$ is the correlation between b_i and b_j, given by $v_{ij}/(v_{ii}v_{jj})^{1/2}$.

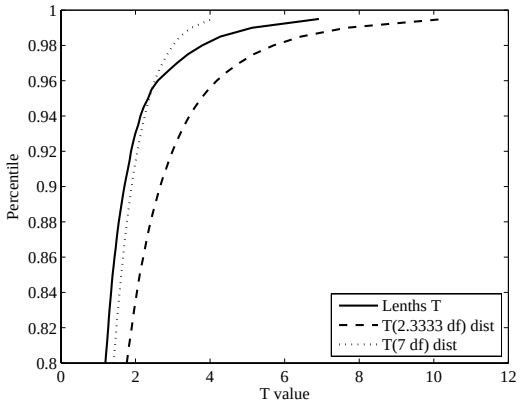

Figure 2.2: Upper Tail Comparison of Distributions

- Step 3: Generate N sets of m random variables, $\{z_1, z_2, ..., z_m\}$, from a multivariate normal distribution with mean zero and covariance, $\mathbf{R}$.

- Step 4: For each set in Step 3, compute Lenth's PSE and the statistics, $|z_i|$ /PSE $(i = 1, ..., m)$. Denote the Lenth t-statistics based on the kth simulated experiment as $t^{(k)}_{Lenth,i}$.

- Step 5: Compute the p-value for each b_i as the proportion of $\left|t^{(k)}_{Lenth,i}\right|$ that are greater than or equal to $|t_{Lenth,i}|$. That is, the (two-sided) p-value for testing b_i is

$$p_i = \frac{\#\left(\left|t^{(k)}_{Lenth,i}\right| \geq |t_{Lenth,i}|\right)}{N}, \tag{2.6}$$

  where '#()' means "number of".

Usually, only the first two digits of p-values are utilized to form a conclusion regarding a particular effect. The standard error of (2.6) is $se_p = [p(1-p)/N]^{1/2}$, where $p = \mathbb{E}[p_i]$, with maximum value $0.5/N^{1/2}$. Hence, N=40,000 ensures that $2se_p \leq 0.005$ for all p. We recommend a minimum of 40,000 $\{z_1, z_2, ..., z_m\}$ sets, as this gives a 95% confidence interval of $p_i \pm 0.005$ for p near 0.5 and $p_i \pm 0.002$ for p near 0.05.

Recall the discreteness of the distribution of $t_{Lenth,i}$ at 2/3. Since $P(t_{Lenth,i} = 2/3) > 0$, any rounding error will affect the p-value when $|t_{Lenth,i}| = 2/3$. However, the p-value will be close to 0.5, so the conclusion regarding effect significance is unaltered.

Note that when the c_i's are independent or equicorrelated, all m rows of $t^{(k)}_{Lenth,i}$'s have the same distribution. In such cases one can pool together all mN $t^{(k)}_{Lenth,i}$'s in order to estimate each p-value. However, in general, the m rows of $t^{(k)}_{Lenth,i}$ are not exchangeable, and each p_i is computed using only the N values of $t^{(k)}_{Lenth,i}$ corresponding to b_i.

JMP 7.0's "Fit Model" platform approximates p-values for Lenth t-statistics using the $t_{m/3}$ approximation, while p-values estimated by simulation are given by JMP 7.0's "Screening" platform. We recommend computing p-values directly for each estimate, even if correlated, whereas JMP computes p-values via simulation only for uncorrelated linear combinations of the original estimates. We suspect that practitioners have great difficulty correctly interpreting tests for these linear combinations, since they no longer correspond to the original effects of interest.

2.2.1 Example 1: An Unreplicated Orthogonal Design

Consider the 2^3 example from Hicks and Turner ([40], p. 249) shown in our Table 2.1. Per Hamada and Balakrishnan ([39], p. 25), the case m=7 often has insufficient power. However, it is useful for illustration. Table 2.2 contains the least squares estimates for the saturated model's coefficients. Since $\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1} = (1/8)\mathbf{I}$, each $c_i = 8^{1/2}b_i$, and the PSE=2.6517. Per Step 3 of the method outlined above, we simulated $N = 1,000,000$ sets of m=7 independent standard normal random variables. The resulting p_i are given in Table 2.2 to three decimal places since, with the large N, all standard errors are less than 0.0005. Based on these results, all three main effects have statistically significant estimates at the 0.05 level.

Table 2.1: Example from Hicks and Turner [40]

A	B	C	Y
-1	-1	-1	2
-1	-1	1	-12
-1	1	-1	15
-1	1	1	-2
1	-1	-1	-5
1	-1	1	-17
1	1	-1	13
1	1	1	-7

As seen earlier in Figure 2.1, neither Lenth's [53] suggestion of a t-distribution with $m/3 = 2.\bar{3}$ df nor a t_7 distribution, as was employed by JMP 6.0, is adequate here; refer to the last two columns of Table 2.2. As was apparent from Figure 2.1, use of $m/3 = 2.\bar{3}$ df produces approximate p-values that are too large for all $|t| \geq 2/3$, while 7 df produces approximate p-values that are close for $|t| \approx 2.5$, but are too large for $|t| \in [1, 2]$. Figure 2.2 highlights the upper tails of the three different distributions compared in Table 2.2. The difference in shape in the upper tail suggests that no t-distribution will fit well the upper 20% of the $|t_{Lenth,i}|$ null distribution.

2.3 The Impact of Correlated Parameter Estimates

For unreplicated 2^k full factorial, 2^{k-f} fractional factorial designs, and most other orthogonal arrays, the model matrix $\mathbf{X}$ for models of interest will have orthogonal columns. With the exception of these special cases, the columns of the model matrix are generally not orthogonal (i.e. $\mathbf{X}'\mathbf{X}$ is not a diagonal matrix). Correlated estimates also arise when intended orthogonal designs are missing one or more treatment combinations. For an n-run, unreplicated two-level orthogonal design missing one treatment combination, the factorial effect parameter estimates are equicorrelated with each other. For models with m factorial effects plus an intercept, the correlation

Table 2.2: Example 1: Lenth t Statistics and Comparison of P-value Approximations

Term	b_i	c_i	$t_{Lenth,i}$	Empirical P-values	P-values based on $t_{2.333}$	P-values based on t_7
A	-2.375	-6.7175	-2.533	0.040	0.109	0.039
B	6.375	18.0312	6.800	0.005	0.014	0.000
A*B	0.625	1.7678	0.667	0.578	0.565	0.526
C	-7.875	-22.2739	-8.400	0.003	0.009	0.000
A*C	-0.125	-0.3536	-0.133	0.910	0.905	0.898
B*C	-1.375	-3.8891	-1.467	0.138	0.263	0.186
A*B*C	-0.625	-1.7678	-0.667	0.578	0.565	0.526

matrix $\mathbf{R}$ is

$$\mathbf{R} = \begin{bmatrix} 1 & \pm\rho & \dots & \pm\rho \\ \pm\rho & 1 & \dots & \pm\rho \\ \vdots & \vdots & \ddots & \vdots \\ \pm\rho & \pm\rho & \dots & 1 \end{bmatrix} \tag{2.7}$$

with $\rho = 1/(n-m)$. For saturated models with one missing observation, $\rho = 0.5$ since $m + 1 = n - 1$. Before applying Lenth's method to two non-orthogonal designs, we present a small simulation study to illustrate the impact of ignoring such correlations among estimates.

Consider the case of $m = 6$ and 14 estimates, which correspond to fitting a saturated model to orthogonal designs of size 8 and 16 with one observation missing. We determine by simulation the correct Lenth t critical values for ρ from 0 to 0.9, to provide insight regarding the sensitivity of the Lenth null distribution to correlations. For each m and ρ combination, $N = 1,000,000$ sets of m contrasts were generated as in Step 3 above. By Step 4, we obtain 1,000,000m identically distributed Lenth t-statistics, and used these to estimate the upper $\alpha = 0.05$ percentile for $\left|t^{(k)}_{Lenth,i}\right|$, which we denote as C^{α}. In addition, we computed the proportion of $\left|t^{(k)}_{Lenth,i}\right|$ statistics that exceed those 0.05 IER critical values, $C^{0.05}$, obtained when $\rho = 0$. The simulation results appear in Table 2.3.

For both values of m in Table 2.3, the $\alpha = 0.05$ critical values decrease as ρ increases from 0 to 0.9. Thus, as the correlations increase, smaller sized $t_{Lenth,i}$ statistics can be deemed significant at a specified level for α. For instance, for m=14, $\rho = 0$, and $\alpha = 0.05$, an effect would be declared active if $|t_{Lenth,i}| > 2.153$. However, for $m = 14$, $\rho = 0.5$, and $\alpha = 0.05$, effects should be declared active if $|t_{Lenth,i}| > 1.918$. Obviously, accounting for the correlation is necessary to maintain the correct level of the test. If large correlations are ignored, Lenth's test becomes very conservative.

Table 2.3: $\alpha = 0.05$ Critical Values and IER for m Equicorrelated Contrasts

		Correct Critical Value	IER if Correlation is Ignored
m	ρ	$C^{0.05}$	Nominal $\alpha = 0.05$
6	0	2.172	0.050
6	0.2	2.120	0.047
6	0.5	1.938	0.037
6	0.7	1.749	0.028
6	0.9	1.430	0.016
14	0	2.153	0.050
14	0.2	2.098	0.046
14	0.5	1.918	0.034
14	0.7	1.753	0.025
14	0.9	1.460	0.014

For instance, Table 2.3 indicates that for $m = 14$ and $\rho = 0.5$, $|t_{Lenth,i}| > 2.153$ only 3.4% of the time, well below the nominal 0.05 level. However, ignoring correlations of $\rho = 0.2$ has little effect on the level of the test. Thus, for applications with small correlations among the estimates such as for our Example 3 later, use of Lenth t percentiles obtained by Ye and Hamada [99] may well be adequate. However, in general, the presence of correlations makes the use of case-specific tables of critical values impractical. We now consider two examples with correlated estimates.

2.3.1 Example 2: A 2^4 with One Observation Missing

Suppose we intended to conduct a full factorial for 4 two-level factors. This 16-run design would permit estimation of all four main effects and 11 interactions. However, while conducting the experiment, the run consisting of factors A, B, and C at their high level (+1) and factor D at its low level (-1) resulted in no useful data. Thus, the resulting design only has 15 runs as shown in Table 2.4. Due to this one missing observation, we must omit the four factor interaction from our model. We fit a saturated model (14 df) with four main effects, six two-factor interactions, and four three-factor interactions. All diagonal elements of $(\mathbf{X}'\mathbf{X})^{-1}$ are equal to $(\pm)$ 1/8. The off-diagonal elements of the correlation matrix $\mathbf{R}$ are equal to $(\pm)$ 1/2.

Based on the recommended N=40,000 simulated experiments, Table 2.5 gives the correct p-values for this example, obtained by simulations using the matrix $\mathbf{R}$. Based on these p-values, the B and D main effects and their interaction all have significant estimates at the 0.05 level. Furthermore, Table 2.5 compares the correct p-values with what one would obtain by simulation with uncorrelated z_i's, or with p-values obtained based on a t-distribution with $14/3 = 4.\bar{6}$ df. As expected, these alternatives produce conservative tests; both fail to indicate the significance of the D and B*D effects at the 0.05 level. Clearly, correlations this large must not be ignored.

Table 2.4: Example 2 - a 2^4 with one missing observation

A	B	C	D	Y
-1	-1	-1	-1	-3.451
-1	-1	-1	1	-3.156
-1	-1	1	-1	-2.235
-1	-1	1	1	-1.767
-1	1	-1	-1	0.854
-1	1	-1	1	4.068
-1	1	1	-1	2.677
-1	1	1	1	5.000
1	-1	-1	-1	-1.898
1	-1	-1	1	-1.857
1	-1	1	-1	-1.297
1	-1	1	1	-2.207
1	1	-1	-1	0.050
1	1	-1	1	4.740
1	1	1	1	1.455

Table 2.5: Example 2 Estimates and P-value Comparisons

Term	b_i	c_i	PSE	$t_{Lenth,i}$	p_i	p_i using $t_{4.\bar{6}}$	p_i ignoring correlation
A	-0.264	-2.1118	3.1677	-0.667	0.502	0.537	0.506
B	2.21846	17.7477	3.1677	5.603	0.001	0.003	0.002
A*B	-0.6826	-5.4608	3.1677	-1.724	0.069	0.150	0.097
C	0.06618	0.52946	3.1677	0.167	0.888	0.874	0.876
A*C	-0.6038	-4.8304	3.1677	-1.525	0.099	0.192	0.129
B*C	-0.2909	-2.3273	3.1677	-0.735	0.434	0.498	0.439
A*B*C	-0.3095	-2.4759	3.1677	-0.782	0.394	0.472	0.411
D	0.79963	6.39705	3.1677	2.019	0.043	0.104	0.060
A*D	0.01211	0.0969	3.1677	0.031	0.980	0.977	0.978
B*D	0.81296	6.5037	3.1677	2.053	0.038	0.099	0.056
A*B*D	0.21626	1.73007	3.1677	0.546	0.630	0.610	0.616
C*D	-0.2304	-1.8429	3.1677	-0.582	0.595	0.588	0.594
A*C*D	-0.1406	-1.1251	3.1677	-0.355	0.760	0.738	0.746
B*C*D	-0.1332	-1.0657	3.1677	-0.336	0.774	0.751	0.759

2.3.2 Example 3: Retchshaffner Design for Seven Factors

Correlated estimates are a natural consequence of nonorthogonal designs, since by name, they do not have diagonal $\mathbf{X}'\mathbf{X}$ matrices. In this example, we consider a design by Retchshaffner [85] who proposed a series of two-level designs for estimating all main effects and two-factor interactions with the minimal number of runs. For such models, these designs are saturated and thus have zero df for error.

LeThanh, Voilley, and Luu [55] investigated how seven factors influence the volatility of three different aromatic food additives. The response in each case was the measured vapor-liquid equilibrium coefficient. The design and response of interest are shown in Table 2.6. The model coefficients are correlated with small correlations of either ± 0.0862 or ± 0.0690 and are estimated with equal precision. We omit the correlation matrix to save space.

Based on the standardized estimates in Table 2.7, we calculate Lenth's PSE of 0.0806. For an orthogonal design with m=28, at the 0.05 level of significance, one would declare an effect active for $|t_{Lenth,i}| > 2.074$. Therefore, based on the t-statistics in Table 2.7, the following effects would be declared active: B, C, D, F, G, E*F, and C*G.

Table 2.7 provides the p-values based on accounting for and ignoring the correlations. In this example, there exists little difference between p-values based on independent estimates and p-values that incorporate the correlation structure of the model estimates. Such a result should be expected to occur since the correlations among the estimates are all small (less than 0.1) in magnitude.

Table 2.6: Example 3 - Retchshaffner's Nonorthogonal Design

A	B	C	D	E	F	G	Y
-1	-1	-1	-1	-1	-1	-1	-1.046
-1	-1	-1	-1	-1	1	1	-0.046
-1	-1	-1	-1	1	-1	1	-0.959
-1	-1	-1	-1	1	1	-1	-0.420
-1	-1	-1	1	-1	-1	1	-1.222
-1	-1	-1	1	-1	1	-1	-0.260
-1	-1	-1	1	1	-1	-1	-1.155
-1	-1	1	-1	-1	-1	1	-0.585
-1	-1	1	-1	-1	1	-1	0.097
-1	-1	1	-1	1	-1	-1	-0.796
-1	-1	1	1	-1	-1	-1	-1.046
-1	1	-1	-1	-1	-1	1	-0.638
-1	1	-1	-1	-1	1	-1	0.322
-1	1	-1	-1	1	-1	-1	-0.602
-1	1	-1	1	-1	-1	-1	-0.959
-1	1	1	-1	-1	-1	-1	-0.456
1	-1	-1	-1	-1	-1	1	-1.000
1	-1	-1	-1	-1	1	-1	-0.076
1	-1	-1	-1	1	-1	-1	-1.000
1	-1	-1	1	-1	-1	-1	-1.222
1	-1	1	-1	-1	-1	-1	-0.824
1	1	-1	-1	-1	-1	-1	-0.523
-1	1	1	1	1	1	1	0.322
1	-1	1	1	1	1	1	0.025
1	1	-1	1	1	1	1	-0.046
1	1	1	-1	1	1	1	0.386
1	1	1	1	-1	1	1	0.270
1	1	1	1	1	-1	1	-0.509
1	1	1	1	1	1	-1	-0.066

Table 2.7: Example 3 Estimates and Empirical P-values for Saturated Model

Term	b_i	c_i	PSE	$t_{Lenth,i}$	p_i	p_i (ignoring correlation)
A	-0.0127	-0.0565	0.0806	-0.701	0.474	0.472
B	0.1548	0.6901	0.0806	8.563	0.000	0.000
A*B	0.0022	0.0098	0.0806	0.122	0.903	0.909
C	0.121	0.5394	0.0806	6.693	0.000	0.000
A*C	-0.0215	-0.0958	0.0806	-1.188	0.230	0.230
B*C	-0.0227	-0.1013	0.0806	-1.257	0.207	0.203
D	-0.1007	-0.449	0.0806	-5.570	0.000	0.000
A*D	-0.0071	-0.0317	0.0806	-0.393	0.709	0.711
B*D	-0.0345	-0.154	0.0806	-1.911	0.066	0.068
C*D	-0.0048	-0.0212	0.0806	-0.263	0.802	0.803
E	-0.0208	-0.0925	0.0806	-1.148	0.243	0.243
A*E	-0.0026	-0.0115	0.0806	-0.143	0.891	0.891
B*E	0.0036	0.0163	0.0806	0.203	0.845	0.844
C*E	0.0068	0.0302	0.0806	0.375	0.722	0.724
D*E	0.0309	0.1377	0.0806	1.708	0.094	0.095
F	0.4095	1.8257	0.0806	22.651	0.000	0.000
A*F	-0.0072	-0.0321	0.0806	-0.399	0.705	0.703
B*F	-0.001	-0.0043	0.0806	-0.054	0.960	0.958
C*F	-0.0057	-0.0255	0.0806	-0.317	0.760	0.763
D*F	0.019	0.0847	0.0806	1.050	0.286	0.285
E*F	-0.0721	-0.3214	0.0806	-3.988	0.002	0.003
G	0.0722	0.3219	0.0806	3.994	0.002	0.003
A*G	-0.0118	-0.0525	0.0806	-0.651	0.529	0.530
B*G	-0.0148	-0.065	0.0806	-0.807	0.405	0.407
C*G	0.0503	0.2242	0.0806	2.782	0.017	0.017
D*G	0.005	0.022	0.0806	0.274	0.790	0.794
E*G	0.0198	0.0884	0.0806	1.097	0.262	0.262
F*G	0.0123	0.055	0.0806	0.682	0.485	0.485

2.4 Pure Error Degrees of Freedom

The addition of center point runs to a two-level factorial design serves several purposes. First, these runs allow one to determine if pure quadratic curvature is present, and thus provide an indicator whether the design should be augmented. Second, they provide pure error degrees of freedom. Montgomery ([72], pp. 247-251) offers further useful suggestions and comments regarding the use of center point runs.

Pure error df are obtained by replicating any design point(s). Given our assumption of homogeneous error variance, the mean square pure error (MSPE) provides a model independent estimate of the error variance, σ^2; we denote its degrees of freedom by df_{PE}. When the MSPE is available, ordinary t-statistics may be computed as $b_i/\sqrt{v_{ii}\text{MSPE}}$.

Consider the case where effect sparsity is reasonable and replication is minimal, e.g. $df_{PE} \leq 3$. Although one could estimate σ with the square root of MSPE, due to its few degrees of freedom such an estimator will be imprecise. Assuming effect sparsity, it seems that one should not ignore the information available in Lenth's PSE. Larntz and Whitcomb [51] proposed using a linear combination of MSPE and PSE^2, weighting each by its df (or pseudo df). We denote such combined standard error estimators as

$$\text{CSE}_\pi = \left[\pi(\text{PSE}^2) + (1-\pi)(\text{MSPE})\right]^{1/2}. \tag{2.8}$$

Larntz and Whitcomb's reasonable choice for π,

$$\pi^* = \frac{m/3}{(m/3) + df_{PE}}, \tag{2.9}$$

would produce the minimum variance unbiased estimator for σ^2 if PSE^2 were a multiple of a chi-square random variable (see Graybill and Deal [37]). Note that using

$\pi = 0$ in (2.8) ignores use of Lenth's PSE, while $\pi = 1$ ignores the MSPE.

In conjunction with the use of CSE_{π^*}, we propose to incorporate the model independent information provided by MSPE into the adaptive feature of Lenth's method by combining s_0^2 and MSPE before computing the PSE:

$$s_0^C(\omega) = \left[\omega s_0^2 + (1-\omega)\text{MSPE}\right]^{1/2} \tag{2.10}$$

for some weight ω. The use of $s_0^C(\omega)$ in the computation of

$$\text{PSE}_\omega = 1.5 * median_{|c_j|<2.5s_0^C(\omega)} |c_j| \tag{2.11}$$

makes the procedure more robust to violations of effect sparsity. PSE_ω for $\omega = 1$ corresponds to the Larntz and Whitcomb method. We have found via simulations that using $\omega = \pi^*$ in (2.10) generally maximizes power for conditions where Lenth's method is known to perform well. That is, if 20% or fewer active effects is assured, we recommend tha the same weight π^* be used for both (2.8) and (2.10). Choices of $\omega < \pi^*$ tend to decrease power slightly when there are very few active effects, while increasing power when the proportion of active effects exceeds 20%. After evaluating several choices, we recommend using

$$\omega^* = \frac{m/3}{m/3 + 5df_{PE}} \tag{2.12}$$

in (2.10) and (2.11) as a means of increasing the power when effect sparsity may not hold. We now describe a small power simulation that supports these choices.

For our first power simulation, active effects are each assigned a magnitude of $\beta_i/v_{ii}^{1/2} = 3\sigma$. Thus, in the simulation results, experiment size has minimal effect on power. In each simulation, we consider 1-3 degrees of freedom for pure error and up

to 40% of the effects active, in order to investigate cases where the effect sparsity assumption is violated. Now suppose, for example, that the first three effects are active, i.e. $\beta_i/v_{ii}^{1/2} = 3\sigma$ for $i = 1, 2, 3$. Then based on $N = 200,000$ simulated data sets, we calculate power by

$$power = \frac{\#\left(\left|t^{(k)}_{Lenth,1}\right| > C^{0.05}\right) + \#\left(\left|t^{(k)}_{Lenth,2}\right| > C^{0.05}\right) + \#\left(\left|t^{(k)}_{Lenth,3}\right| > C^{0.05}\right)}{3N}, \tag{2.13}$$

where $C^{0.05}$ is the appropriate 0.05 critical value obtained via simulation. We estimate power for tests of size $\alpha = 0.05$. See Tables 2.8-2.10 for a summary of simulations for $n = 8, 16$, and 32, respectively. For $\pi = 0$, we calculate power exactly using the non-central t distribution.

With $df_{PE} = 1$, use of the MSPE alone provides power inferior to Larntz and Whitcomb's recommended CSE_{π^*} for every case in Tables 2.8-2.10. For df_{PE}=2 and 3, CSE_{π^*} is preferred to the MSPE except for a few cases where more than 30% of the effects are active. Using the modified PSE_{π^*} (i.e., (2.10) and (2.11) with $\omega = \pi^*$) in the combined estimator increases power over that achieved by the Larntz and Whitcomb procedure for approximately half of the cases in Table 2.8-2.10 (based on

Table 2.8: Power Simulation [$m = 7$,$\pi^* = 2.\bar{3}/(2.\bar{3} + df_{PE})$, $\omega^* = 2.\bar{3}/(2.\bar{3} + 5df_{PE})$]

Number Active	df_{PE}	MSPE ($\pi = 0$)	CSE_{π^*} only	CSE_{π^*} and $s_0^C(\pi^*)$	CSE_{π^*} and $s_0^C(\omega^*)$
1	1	0.19	0.63	0.63	0.62
1	2	0.39	0.68	0.68	0.66
1	3	0.53	0.71	0.71	0.70
2	1	0.19	0.51	0.51	0.52
2	2	0.39	0.57	0.58	0.59
2	3	0.53	0.62	0.63	0.64
3	1	0.19	0.30	0.31	0.36
3	2	0.39	0.37	0.39	0.46
3	3	0.53	0.43	0.46	0.53

Table 2.9: Power Simulation [$m = 15$,$\pi^* = 5/(5 + df_{PE})$, $\omega^* = 5/(5 + 5df_{PE})$]

Number Active	df_{PE}	MSPE ($\pi = 0$)	CSE_{π^*} only	CSE_{π^*} and $s_0^C(\pi^*)$	CSE_{π^*} and $s_0^C(\omega^*)$
1	1	0.19	0.72	0.72	0.71
1	2	0.39	0.74	0.74	0.73
1	3	0.53	0.76	0.76	0.75
2	1	0.19	0.68	0.68	0.68
2	2	0.39	0.71	0.71	0.70
2	3	0.53	0.73	0.73	0.72
3	1	0.19	0.62	0.63	0.63
3	2	0.39	0.66	0.66	0.67
3	3	0.53	0.68	0.69	0.69
4	1	0.19	0.55	0.55	0.56
4	2	0.39	0.58	0.60	0.61
4	3	0.53	0.62	0.63	0.65
5	1	0.19	0.44	0.45	0.47
5	2	0.39	0.48	0.50	0.54
5	3	0.53	0.52	0.54	0.59
6	1	0.19	0.30	0.32	0.35
6	2	0.39	0.35	0.37	0.44
6	3	0.53	0.39	0.42	0.50

Table 2.10: Power Simulation [$m = 31$,$\pi^* = 10.\bar{3}/(10.\bar{3} + df_{PE})$, $\omega^* = 10.\bar{3}/(10.\bar{3} + 5df_{PE})$]

Number Active	df_{PE}	MSPE ($\pi = 0$)	CSE_{π^*} only	CSE_{π^*} and $s_0^C(\pi^*)$	CSE_{π^*} and $s_0^C(\omega^*)$
1	1	0.19	0.78	0.78	0.78
1	2	0.39	0.79	0.79	0.78
1	3	0.53	0.79	0.79	0.79
2	1	0.19	0.77	0.77	0.77
2	2	0.39	0.78	0.78	0.77
2	3	0.53	0.78	0.78	0.78
3	1	0.19	0.75	0.75	0.75
3	2	0.39	0.76	0.76	0.76
3	3	0.53	0.75	0.77	0.77
4	1	0.19	0.73	0.73	0.73
4	2	0.39	0.74	0.75	0.74
4	3	0.53	0.75	0.75	0.75
5	1	0.19	0.71	0.71	0.71
5	2	0.39	0.72	0.72	0.73
5	3	0.53	0.73	0.74	0.74
6	1	0.19	0.68	0.68	0.68
6	2	0.39	0.69	0.70	0.70
6	3	0.53	0.70	0.71	0.72
7	1	0.19	0.64	0.64	0.65
7	2	0.39	0.66	0.66	0.67
7	3	0.53	0.67	0.68	0.70
8	1	0.19	0.59	0.60	0.61
8	2	0.39	0.61	0.62	0.64
8	3	0.53	0.63	0.64	0.67
9	1	0.19	0.53	0.54	0.56
9	2	0.39	0.56	0.57	0.60
9	3	0.53	0.58	0.59	0.63
12	1	0.19	0.31	0.32	0.35
12	2	0.39	0.34	0.36	0.41
12	3	0.53	0.37	0.39	0.46

values rounded to the nearest 0.01), without lowering the power when the proportion of active effects is small. Still, when roughly 40% of the effects are active, MSPE is preferred for $df_{PE} \geq 2$.

To investigate the possibility of improving power when the proportion of active effects is above 20%, we evaluated various ω in the computation of the modified PSE (2.11). By using ω^* (see (2.12)), the power increased for every case with the proportion of active effects above 20%, with only the slightest decrease in power when the proportion of active effects is small. Thus, we recommend the use of the modified PSE_{ω^*} with CSE_{π^*} as a general rule if effect sparsity is not guaranteed.

To further validate the use of (2.12), we report a second series of simulations where the active effects were of multiple sizes. Again, we considered various ω and found that (2.12) was preferred to other choices. Table 2.11 summarizes these additional power simulation results for the recommended procedure that uses ω^* in the modified PSE, and π^* in the combined standard error (2.8). For each case in Table 2.11, an equal number of active effects $\beta_i/v_{ii}^{1/2}$ are of size 1σ, 2σ, and 3σ. As expected, the power is negligible for effects of size 1σ and moderately better for 2σ. For effects of size 3σ, the power is slightly higher in Table 2.11 than in the comparable cases in Tables 2.8-2.10. For example, with size of the 15 effects active (two each of size 1σ, 2σ, and 3σ), the power for detecting the 3σ effects ranges from 0.51-0.61 depending on df_{PE}. The power is notably less when all 40% of the active effects are of size 3σ.

The simulation required to estimate p-values for test statistics using CSE_π is very similar to the procedure outlined previously. As before, fit a saturated model and obtain the standardized regression coefficients, c_i. First, choose whether to use ω equal to π^* or ω^* depending on one's belief about effect sparsity. For the new Step 1, calculate MSPE, then $s_0^C(\omega)$ and the modified PSE_ω. Then compute CSE_{π^*} and use it to compute the relevant t-statistics. Step 2 is unchanged. For Step 3, in addition

Table 2.11: Power Simulation with Effects of Various Sizes [$\pi^* = (m/3)/((m/3) + df_{PE})$, $\omega^* = (m/3)/((m/3) + 5df_{PE})$]

m	Number Active of Size $1\sigma/2\sigma/3\sigma$	df_{PE}	Power (1σ)	Power (2σ)	Power (3σ)
7	1/1/1	1	0.06	0.22	0.48
7	1/1/1	2	0.07	0.27	0.55
7	1/1/1	3	0.08	0.30	0.61
15	1/1/1	1	0.1	0.34	0.66
15	1/1/1	2	0.11	0.36	0.68
15	1/1/1	3	0.12	0.38	0.71
15	2/2/2	1	0.06	0.23	0.51
15	2/2/2	2	0.07	0.27	0.56
15	2/2/2	3	0.08	0.29	0.61
31	1/1/1	1	0.13	0.42	0.76
31	1/1/1	2	0.13	0.43	0.76
31	1/1/1	3	0.14	0.43	0.77
31	2/2/2	1	0.11	0.37	0.71
31	2/2/2	2	0.11	0.38	0.72
31	2/2/2	3	0.12	0.39	0.73
31	3/3/3	1	0.08	0.31	0.64
31	3/3/3	2	0.09	0.33	0.66
31	3/3/3	3	0.09	0.34	0.68
31	4/4/4	1	0.05	0.24	0.55
31	4/4/4	2	0.06	0.26	0.58
31	4/4/4	3	0.07	0.28	0.61

to generating N sets of m standard normal variables, $\{z_1, z_2, ..., z_m\}$, generate a chi-square random variate (with df_{PE} degrees of freedom) divided by df_{PE}, which we shall denote as $s^{2(k)}$. For Step 4, from each simulated normal vector and $s^{2(k)}$, compute $s_0^C(\omega)$, (2.11), and $|z_i| \,/\, \left[\pi^* \text{PSE}_\omega^{2(k)} + (1-\pi^*)s^{2(k)}\right]^{1/2}$ for $i = 1, ..., m$. Step 5 proceeds the same as before.

2.4.1 Example 4: A Fractional Factorial with Center Point Replication

Consider the 2^{5-1} fractional factorial design with three center point runs and response given in Table 2.12. Suppose we fit a model with all main effects, two-factor interactions, plus one pure quadratic effect. We obtain the estimates shown in Table 2.13 and compute MSPE=1.48. The p-values computed using a t-distribution with 2 df for pure error are shown in Table 2.13. Note that this is what essentially all statistical software would provide. Based on this analysis, both the B and D main effects estimates are statistically significant, with p-values of 0.013 and 0.041, respectively. However, this traditional analysis ignores useful information if the sparsity of effects assumption is considered valid.

Table 2.13 provides two additional sets of p-values, one set based on Lenth t-statistics that ignore the pure error degrees of freedom, and a final set that utilizes our combined estimator. In particular, for m=16 and $df_{PE} = 2$,$\pi^* = 0.727$, and $\omega^* = 0.348$ (which should provide good power even if five or six effects are active). Thus, with median$|c_i| = 0.6938$ and MSPE=1.48, we obtain

$$s_0^C(\omega^*) = [0.348(1.5 * 0.694)^2 + 0.652(1.48)]^{1/2} = 1.158,$$

Lenth's PSE_{ω^*}=0.8525 (here using $s_0^C(\omega^*)$ instead of s_0 makes no difference in the

Table 2.12: Example 4- 2^{5-1} Design with 3 Center Point Runs

A	B	C	D	E	Y
-1	-1	-1	-1	1	-2.755
-1	-1	-1	1	-1	-2.687
-1	-1	1	-1	-1	-3.561
-1	-1	1	1	1	-1.125
-1	1	-1	-1	-1	1.229
-1	1	-1	1	1	3.759
-1	1	1	-1	1	0.96
-1	1	1	1	-1	4.185
1	-1	-1	-1	-1	-2.009
1	-1	-1	1	1	-0.642
1	-1	1	-1	1	-2.189
1	-1	1	1	-1	-1.375
1	1	-1	-1	1	-0.369
1	1	-1	1	-1	4.929
1	1	1	-1	-1	0.028
1	1	1	1	1	3.493
0	0	0	0	0	1.094
0	0	0	0	0	-1.309
0	0	0	0	0	0.223

Table 2.13: Example 4 Comparison of P-values based on MSPE, PSE, and CSE_{π}

Term	c_i	P-values for $c_i/\text{MSPE}^{1/2}$	P-values for c_i/PSE	P-values c_i/CSE_{π^*} using $s_0^C(\omega^*)$
A	0.465178	0.688	0.659	0.656
B	8.639344	0.013	0.000	0.000
C	-0.26012	0.819	0.804	0.803
D	4.80079	0.041	0.003	0.002
E	0.098104	0.931	0.924	0.924
A*B	-1.49181	0.274	0.128	0.136
A*C	-0.71611	0.548	0.439	0.467
B*C	-0.1812	0.873	0.862	0.864
A*D	0.671512	0.571	0.476	0.505
B*D	2.457857	0.133	0.028	0.031
C*D	0.168923	0.881	0.872	0.869
A*E	-0.73814	0.537	0.421	0.449
B*E	-1.36212	0.306	0.159	0.164
C*E	0.832536	0.493	0.365	0.385
D*E	0.118437	0.917	0.908	0.911
A*A	0.181308	0.873	0.861	0.864

PSE), and

$$\text{CSE}_{\pi^*} = [0.727(0.8525)^2 + 0.273(1.48)]^{1/2} = 0.9656.$$

We conclude that not only are the B and D main effects active, but also the B*D interaction. Using only the MSPE, its t-statistic $2.4579/1.480^{1/2}$=2.02 is not statistically significant because of the low df_{PE}. However, relative to the combined estimator CSE_{π^*} (or the PSE), its estimate stands out, with a p-value near 0.03. In practice, one might typically fit a reduced model, and using the reduced model, conclude that the B and D effects are not additive. However, such a conclusion is reached here by fitting the saturated model and using test statistics based on CSE_{π^*}, rather than with a conditional test that risks bias to the estimator for σ^2 from the selection of a reduced model.

2.5 Discussion

In this chapter, we have proposed and illustrated the usefulness of empirically determined p-values in analyzing unreplicated factorial experiments, as well as for experiments with limited replication. Estimation of p-values is performed easily using Monte Carlo simulation with minimal computation time. Copies of MATLAB programs that perform the necessary computations are available upon request. However, it is our hope that soon software packages will implement these methods, creating a more user-friendly approach to the analysis of unreplicated experiments using saturated models and an assumption of effect sparsity.

We have shown that no t-distribution approximation to the null distribution of Lenth's PSE is appropriate, even though a t-distribution with $m/3$ df is still the usual approximation. This important observation provides support for estimating p-values by simulation, since the use of any t_ν approximation will not provide accurate results

across the full range of possible values.

All tests considered were two-sided. For one-sided tests, one may simply divide the two-sided p-value in half if the estimate is of the sign anticipated by the alternative hypothesis, since the null distribution of Lenth's t is symmetric about zero. If there are many one-sided tests, then one could consider modifying the definition of PSE to only exclude estimates larger than $2.5s_0$ in the anticipated direction. However, this modification has a negligible effect on the null distribution, and hence on p-values, since very few estimates get excluded.

Two examples with correlated estimates were provided to illustrate how the correlation structure of contrast estimates can be easily incorporated into the computation of p-values. Current tables, such as Ye and Hamada [99], only provide critical values based on independent estimates. Given that correlated estimates can arise in many different situations (nonorthogonal designs, designs with missing observations, etc.), some general purpose approach such as we recommend is needed. JMP offers another general approach, using a Cholesky factorization to orthogonalize estimates before computing Lenth's PSE. However, for most modeling situations, the original columns of $\mathbf{X}$ correspond to more interpretable effects than the Cholesky factorization will provide. We suspect that practitioners mistakenly try to interpret such orthogonalized estimates as if they correspond to the original effects. For this reason, and because Cholesky factorization is not invariant to the order of columns in $\mathbf{X}$, we believe working with the correlated estimates is the preferred approach.

As mentioned earlier, JMP 7.0's Screening platform determines p-values by simulation of the null distribution of Lenth t-statistics. Analogous to Ye and Hamada's [99] experiment-wise critical values, JMP also provides simultaneous-test p-values, based on the null distribution of the $\max\{|z_1|, ..., |z_m|\}$ /PSE. These p-values account for the fact that we are conducting m tests. Such simultaneous p-values are informative

for the most extreme estimates, though not necessarily for other estimates. JMP's simultaneous p-values correspond to a single-step simultaneous test procedure. Westfall and Young ([92], ch.2) discuss adjusted p-values corresponding to both single-step and step-down simultaneous procedures, with the later having potentially greater power. In this chapter, we have opted instead to emphasize the simpler, individual test p-values, since these are typical in regression.

We evaluated a simple improvement to Larntz and Whitcomb's [51] method for combining Lenth's PSE with the usual estimate of error variance, MSPE, when limited replication is available. Our simulation results indicate that substantial power is gained by combining estimators instead of relying on MSPE alone, as software traditionally does, and that the use of $s_0^C(\omega^*)$ in computing a modified PSE improves the robustness of the method to moderate violations of the sparsity of effects assumption. If $df_{PE} > 3$, the benefit of using PSE_{ω^*} and CSE_{π^*} is diminished, especially if 40% or more of the effects are active.

Chapter 3

Supersaturated Designs: Are Our Results Significant?

3.1 Introduction and Motivation

It is common for the beginning stage of experimentation to consist of an initial experiment involving many factors. Such screening experiments often utilize two-level strength two orthogonal arrays (OA) of size n in order to potentially investigate up to n-1 factors of interest. Recall that strength two OA's allow for clear estimation of each factor's main effect, but confound two-factor interactions with main effects and other two-factor interactions. Thus, such designs are used for situations in which there are too many factors to study in great detail. Examples of such commonly used OA's include resolution III fractional factorial designs and Plackett-Burman (PB) designs.

If the number of factors is very large and/or experimental runs are very expensive, then even the use of the OA's mentioned above can become impractical. Supersaturated designs (SSDs) are designs that examine more than n-1 factors in n runs and were introduced to handle such situations. Naturally, SSDs have too few runs to allow for estimation of main effects of all the factors of interest, which is a source of ambiguous results in analysis. We now briefly review the history of SSDs, which

includes a discussion of construction and analysis methods.

The origin of SSDs is attributed to Satterthwaite [89] who proposed random balance experiments to identify a subset of important factors. Research in the area of SSDs remained rather stagnant for over 30 years until Lin [60] revived interest. In his paper, Lin constructed SSDs as half-fractions of Plackett-Burman (PB) designs with n runs and k factors by using one column of the PB design as a branching column in order to form two SSDs with $n/2$ runs and k-1 factors. That is, all runs in which the branching column is either at its high or low level are selected with the branching column deleted. This method of construction is illustrated in Table 3.1 for a 12-run PB design in which factor L is chosen as the branching column. Lin's method has become a popular strategy for SSD construction, due to its simplicity.

Wu [95] proposed constructing SSDs by augmenting Hadamard matrices with two-factor interaction columns. The interaction columns are constructed simply by multiplying signs of the corresponding contributing factors. A consequence of this approach is that nonorthogonality of the factors is mainly accumulated in the last factors. Lin [61] examines the maximum number of factors that can be accommodated when the number of runs is given and when the maximum degree of nonorthogonality between pairs of factors is specified. An example of a SSD with 24 runs and 138 factors is presented, which we shall revisit in a later section. Nguyen [79] developed an algorithmic approach to SSD construction based on cyclic balanced incomplete block designs that generalizes the method of Lin [60]. Li and Wu [59] build SSDs based on a D-optimal design search by applying columnwise-pairwise algorithms. This method contrasts with Wu [95] in that nonorthogonality does not mainly accumulate in the last factors.

More recently, Allen and Bernshteyn [2] develop a new class of SSDs that are constructed so that stepwise regression should be more effective in finding active

Table 3.1: Plackett-Burman design with n=12, k=11, and two SSDs with $n_{SS} = 6$ and $k_{SS} = 10$

A	B	C	D	E	F	G	H	J	K	L
Plackett-Burman ($n = 12$)										
1	1	1	1	1	1	1	1	1	1	1
-1	1	-1	1	1	1	-1	-1	-1	1	-1
-1	-1	1	-1	1	1	1	-1	-1	-1	1
1	-1	-1	1	-1	1	1	1	-1	-1	-1
-1	1	-1	-1	1	-1	1	1	1	-1	-1
-1	-1	1	-1	-1	1	-1	1	1	1	-1
-1	-1	-1	1	-1	-1	1	-1	1	1	1
1	-1	-1	-1	1	-1	-1	1	-1	1	1
1	1	-1	-1	-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
Supersaturated Design 1 ($n_{SS} = 6$)										
1	1	1	1	1	1	1	1	1	1	1
-1	-1	1	-1	1	1	1	-1	-1	-1	1
-1	-1	-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
Supersaturated Design 2 ($n_{SS} = 6$)										
-1	1	-1	1	1	1	-1	-1	-1	1	-1
1	-1	-1	1	-1	1	1	1	-1	-1	-1
-1	1	-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

factors. Based on optimization of their new design criteria, unbalanced designs are generated that may be considered preferable in situations where levels of certain factors are associated with high costs. Table 3.2 shows an example of one such design with 10 runs and 11 factors. Jones, Lin, and Nachtsheim [46] introduce a class of SSDs using Bayesian D-optimality. They found that those designs generated based on Bayesian D-optimality are equal to or better than other SSDs in terms of certain criteria, namely, $\mathbb{E}(s^2)$, which will be discussed next. An example of this type of design will be presented in a later section. At this point, there is still much more we could say regarding SSD construction. In particular, our discussion has and will continue to focus only on *two-level* SSDs and their role in factor screening. For work on SSDs with more than two levels, see Lu, Wu, and Zheng [65]. For a more extensive review of SSD construction in general, see Dejaegher and Vander Heyden [27], henceforth known as DVH.

One popular criterion for comparing and ranking SSDs is the $\mathbb{E}(s^2)$ criterion, originally proposed by Booth and Cox [11] and extended by Wu [95]. Let $\mathbf{D}$ represent the $n \times k$ supersaturated design matrix and $\mathbf{X} = [\mathbf{1}, \mathbf{D}]$ be the main effects model

Table 3.2: Unbalanced Supersaturated Design to Maximize Probability of Identifying Active Factors

A	B	C	D	E	F	G	H	J	K	L
1	1	-1	-1	1	-1	-1	-1	-1	-1	-1
1	1	1	-1	1	-1	1	1	-1	-1	1
1	1	-1	-1	1	1	1	1	-1	-1	1
1	-1	-1	-1	1	-1	-1	1	-1	-1	1
1	1	-1	1	1	-1	1	1	-1	-1	1
1	1	1	-1	1	-1	-1	-1	-1	-1	1
-1	1	-1	-1	1	-1	1	1	-1	-1	1
-1	1	-1	-1	-1	-1	-1	1	1	-1	1
1	1	-1	-1	1	-1	1	-1	1	-1	1
1	1	-1	-1	1	-1	1	1	-1	1	1

matrix where **1** is a vector of ones. Then, the main effects model can be written as

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \tag{3.1}$$

where $\mathbf{Y}$ is a response variable, β is a vector of unknown parameters, and ϵ is the error term with $\mathbb{E}(\epsilon) = 0$ and $\mathbb{V}(\epsilon) = \sigma^2$. For SSDs, $rank(\mathbf{X}) \leq k$. Thus, **X'X** is singular and no unique least squares estimate for β can be obtained.

Consider, for a moment, Supersaturated Design 1 given in Table 3.1. This 6-run design has

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6 & 2 & -2 & -2 & 2 & 2 & -2 & 2 & 2 \\ 0 & 2 & 6 & 2 & 2 & -2 & 2 & -2 & 2 & 2 \\ 0 & -2 & 2 & 6 & 2 & 2 & 2 & 2 & 2 & -2 \\ 0 & -2 & 2 & 2 & 6 & -2 & -2 & 2 & 2 & 2 \\ 0 & 2 & -2 & 2 & -2 & 6 & 2 & 2 & 2 & -2 \\ 0 & 2 & 2 & 2 & -2 & 2 & 6 & 2 & -2 & 2 \\ 0 & -2 & -2 & 2 & 2 & 2 & 2 & 6 & -2 & 2 \\ 0 & 2 & 2 & 2 & 2 & 2 & -2 & -2 & 6 & -2 \\ 0 & 2 & 2 & -2 & 2 & -2 & 2 & 2 & -2 & 6 \end{bmatrix}. \tag{3.2}$$

Gilmour [34] states that "if it were possible to reduce the off-diagonal elements in absolute value, the matrix could be made nonsingular. Best of all would be if the off-diagonal elements were all zero, in which case all main effects would be estimated independently." Although, we certainly cannot obtain a rank larger than 6 in the above $\mathbf{X}'\mathbf{X}$ matrix based on a design with n=6 and k=10, a good SSD may be one in which the off-diagonal elements are made as small as possible. Thus, Booth and

Cox [11] suggested two criterion based on the off-diagonal elements of $\mathbf{X}'\mathbf{X}$, which will be denoted as s_{ij}. The first criterion is to choose a design that minimizes $max_{i<j}\,|s_{ij}|$. The second criterion is the $\mathbb{E}(s^2)$ criterion which chooses a design that minimizes

$$\mathbb{E}(s^2) = \frac{2}{k(k-1)} \sum_{i<j} s_{ij}^2. \tag{3.3}$$

For designs with equal occurrence of -1 and +1, the $\mathbb{E}(s^2)$ criterion is equivalent to minimizing the variance of the s_{ij}'s. This criterion has become the most commonly used criterion in the literature on supersaturated designs and it should come as no surprise that methods for constructing $\mathbb{E}(s^2)$-optimal supersaturated designs are available. See, for example, Cheng [21], Liu, Ruan, and Dean [62], and Nguyen and Cheng [78].

Deng, Lin, and Wang [28] present a criterion for measuring goodness of a supersaturated design based on the projection properties of the design known as resolution rank (r-rank for short). As before, let $\mathbf{D} = (d_1, d_2, ..., d_k)$ represent an $n \times k$ supersaturated design matrix. The *r-rank* is defined as

$$r = max\left\{c|\text{for any } (d_1, ..., d_c) \text{ of } \mathbf{D}, d_1, ..., d_c \text{ are linearly independent}\right\}. \tag{3.4}$$

Basically, if a supersaturated design has an r-rank of r, then when $\mathbf{D}$ is projected to any submatrix of r (or fewer) factors, the main effects of the projected design are all estimable. One downside in the use of this criterion is its computation. That is, r-rank must be computed by conducting an exhaustive search of all models until one of a given size is discovered in which the columns of $\mathbf{D}$ are linearly dependent. Generally r-rank $\leq n/2$. The r-rank criterion will be seen later in a model selection context.

DVH make an important observation that "although many construction methods

can be found in the literature, much less is published concerning the analysis of supersaturated design results and the identification of significant factors." We now turn to the analysis aspect of SSDs. DVH comment that "one of the requirements of a proper method of analysis is that the important factors should indeed be found. In other words, the method should be able to distinguish between the estimated effects representing noise, i.e. the unimportant effects, and those have a real effect on the response." An important assumption and justification underlying analysis methods of SSDs is *effect sparsity* which states that the number of important effects is relatively small.

One of the most commonly used analysis methods for SSDs is stepwise regression (forward and/or mixed) advocated in Lin [61]. Westfall, Young, and Lin [93] point out that an appeal of forward stepwise selection is that nominal Type I error significance levels are utilized. However, they comment that if effect sparsity holds, Type I error rates can be quite high and hence propose the use of adjusted p-values, based on Monte Carlo simulation, to better control Type I errors. A simulation study showed that while the adjusted p-values do well in controlling the familywise Type I errors, significant power is lost due to premature stopping of the forward selection procedure. They conclude their paper with the following comments:

- Identification of significant variables in SSDs is very tricky,
- Many Type I and Type II errors are expected using forward variable selection.

Abraham, Chipman, and Vijayan [1] note problems when using forward selection. Their simulations indicate that there is a high chance of missing the real active factors and selecting inactive ones instead. In particular, assignment of factors to columns is critical due to the correlation structure among the factors. One simulation study utilized supersaturated subsets of the data of Williams [94], a 28-run PB design in

23 factors. The Williams data has become a benchmark data set for measuring the success of SSDs and is displayed in Table 3.3. In Williams [94], factors 15, 19, and 16 are identified as having large effects while 4, 21, 14, and 8 have moderate effects. Although we will reinvestigate the full Williams data later, it is worthy to note that forward selection was unable to find all of the large and moderate effects in any of the supersaturated subsets. In light of this, Abraham, Chipman, and Vijayan [1] indicate that all-subsets variable selection should be used instead of stepwise regression. However, they do point out that even all-subsets can be problematic as it does not provide a means for deciding how many factors are active.

Kelly and Voelkel [48] state the all-subsets procedure offers better promise than forward selection. They further suggest, in situations where all-subsets would prove to be too computationally burdensome, to examine all-subsets of effects up to size m, where m is chosen to be at least as large as the maximum number of effects expected. Assuming effects sparsity holds, a potential rule of thumb when $k >> n$ would be to set $m = \lfloor min\{k/5, n/2\} \rfloor$. This rule of thumb is not reasonable, however, for only marginally oversaturated designs. For example, if k=11 and n=10, then m=2, which is too small. Thus, another worthy choice of m is the r-rank, which, although computationally intensive to compute, could be utilized for any SSD. Although an m too large is preferable to one that is too small, computational feasibility must always be taken into account when dealing with all-subsets regression.

Holcomb, Montgomery, and Carlyle [42] evaluate several methods for analyzing SSDs including forward and all-subsets regression. They also propose a *many-models* method that examines many possible models (but not all). This method proceeds as follows:

1. Create all possible two-factor models.

2. Select a fraction of the best models based on some criterion such as R^2.

Table 3.3: Data from Williams [94]

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Y
1	1	1	-1	-1	-1	1	1	1	1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	-1	1	133
-1	1	-1	-1	-1	-1	1	1	1	1	-1	1	-1	-1	1	1	1	1	1	-1	1	1	-1	49
1	-1	-1	-1	-1	-1	1	1	1	-1	-1	-1	1	1	1	-1	1	-1	-1	1	1	-1	-1	62
1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	1	1	1	1	-1	-1	-1	-1	1	1	-1	45
1	1	-1	-1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	-1	1	1	1	-1	1	88
1	1	-1	1	-1	1	-1	-1	-1	1	1	-1	1	-1	1	-1	1	1	1	-1	-1	-1	-1	52
-1	-1	1	1	1	1	1	1	-1	1	-1	1	1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	300
-1	-1	1	1	1	1	-1	1	1	-1	-1	-1	1	-1	1	1	-1	-1	1	-1	1	1	1	56
-1	-1	1	1	1	1	1	-1	1	1	1	-1	-1	1	1	1	1	1	1	1	1	-1	-1	47
-1	-1	-1	-1	1	-1	-1	1	-1	1	-1	1	1	1	-1	1	1	1	1	1	-1	-1	1	88
1	-1	1	-1	-1	1	-1	-1	1	-1	-1	-1	1	1	-1	1	1	1	-1	-1	-1	1	-1	116
-1	1	1	1	-1	-1	1	-1	-1	1	1	-1	1	1	-1	1	1	-1	-1	1	1	1	1	83
-1	1	1	-1	-1	1	-1	1	-1	1	-1	-1	-1	-1	-1	-1	-1	1	-1	1	1	1	-1	193
-1	-1	-1	1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	230
1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	51
-1	1	1	-1	1	-1	-1	-1	1	-1	1	1	1	-1	1	1	-1	1	-1	1	-1	-1	-1	82
-1	-1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	32
1	-1	1	1	-1	-1	-1	1	-1	-1	1	1	1	-1	1	-1	1	1	1	1	1	1	1	58
1	-1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	1	-1	-1	1	-1	-1	201
1	1	1	-1	1	1	-1	1	1	1	1	1	-1	1	1	-1	1	-1	-1	-1	1	-1	1	56
-1	1	-1	1	-1	1	1	-1	1	-1	-1	1	1	1	-1	-1	-1	1	1	-1	1	-1	1	97
1	1	1	1	-1	1	1	1	-1	-1	-1	1	-1	1	1	1	-1	1	-1	1	-1	-1	1	53
-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	-1	-1	1	1	-1	-1	-1	1	1	276
1	-1	-1	-1	1	1	1	-1	1	1	1	1	1	-1	-1	-1	-1	1	-1	1	1	1	1	145
1	1	1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1	-1	1	-1	1	1	-1	1	-1	130
-1	1	-1	-1	1	1	1	1	-1	-1	1	-1	1	1	1	-1	-1	-1	1	1	-1	1	-1	55
1	-1	-1	1	-1	1	-1	1	1	1	1	1	-1	1	-1	1	-1	-1	1	1	-1	1	-1	160
-1	-1	1	-1	-1	-1	-1	-1	-1	-1	1	1	-1	1	-1	-1	-1	-1	1	-1	1	-1	-1	127

3. Order the factors by frequency that they appear in the retained models.

4. Retain those factors that appear most often (in at least 75% of the best models).

5. Create all possible three-factor models based on the factors identified in 4).

6. Repeat the method until a specified number of terms are included. The most frequently occurring factors at the final stage are then counted as active.

Although the *many-models* approach has potential merit, we found via simulation that this method often misses at least one active factor when selecting a fraction of best models in the two-factor all-subsets stage. Unlike all-subsets, this leads to an inability to consider dropped factors in subsequent stages. However, this approach would be preferable to stepwise regression when m is large. Holcomb, et.al. [42] recommend that the philosophy SSD use should be to eliminate as many inactive factors as possible from further experimentation and not to clearly identify all of the active factors. They further assert that SSDs may find practical use in conjunction with well-planned follow-up experiments.

Chipman, Hamada, and Wu [23] propose a Bayesian variable selection strategy that uses prior information on the parameters to compute posterior probabilities of models. In particular, they use independent prior distributions for each main effect being active. The prior for $\beta_j,\ j = 1, 2, ..., k$ was a mixture of normals, namely, $N(0, \tau_j^2)$ with prior probability $1 - \pi_j$ and $N(0, c_j\tau_j^2)$ with prior probability π_j, where c_j is much larger than 1. The prior for σ^2 is the usual inverse gamma. The method then proceeds with the the stochastic search variable selection (SSVS) method of George and McCullough [33] to obtain posterior probabilities that a particular β_j is active.

A two-stage extension to this Bayesian approach was proposed in Beattie, Fong, and Lin [8]. The first stage is identical to the strategy of Chipman, Hamada, and

Wu [23] where candidate models are ranked and factors chosen accordingly. All models and factors selected in the first stage are combined into one "encompassing" model. The second stage utilizes a Bayes Factor approach on the "encompassing" model to select the best model.

We briefly mention several other appraoches. Genetic algorithms were applied to the analysis of SSDs in Cela, Martinez, and Carro [20] with favorable performance. Li and Lin [56] introduce a variable selection procedure based on nonconvex penalized least squares for SSDs constructed from Hadamard matrices. Lu and Wu [66] investigate a staged dimensionality reduction strategy based on stepwise regression.

More recently, Zhang, Zhang, and Liu [101] introduce the Partial Least Squares Variable Selection (PLSVS) method based on the variable importance in projection (VIP). Partial Least Squares (PLS) is a technique that generalizes and combines features from principal component analysis, canonical correlation analysis, and regression. The first PLS component is that which contains as much information about the explanatory variables as possible and in which the correlation between the response and the first component is maximized. It can be shown that when only one component is selected, the PLSVS procedure is equivalent to forward stepwise regression. Phoa, Pan, and Xu [80] consider variable selection for SSDs based on the Dantzig selector. The Dantzig selector chooses active factors by solving a very simple convex program, which can be recast as a linear program. They propose both a graphical and automated procedure to accompany the method.

Although the latter analysis methods mentioned above usually outperform stepwise (forward) regression, they are more difficult to apply and the results are not always clearly interpretable. In some instances, prior knowledge of the parameters is necessary (i.e., the Bayesian approaches). Thus, the use of these methods becomes less feasible for real-life situations. Other analysis methods not mentioned here as

well as further details on many of those included in this chapter can be found in DVH.

With regards to any analysis strategy, Gilmour [34] states that although methods for analyzing data from SSDs have been proposed, none of them seem very convincing. Therefore, although methods exist for constructing SSDs with good statistical properties, current methods of analysis are not as convincing. In this chapter, we propose a non-Bayesian approach for the analysis of SSDs.

Whether using forward selection or all-subsets regression, it is common to select models from SSDs that explain a very large percentage of the total variation in a response. The naïve p-values one sees for the selected model can persuade the user that the included factors are clearly active. This has contributed to the poor performance of the stepwise and even the all-subsets procedure in terms of Type I error and power as demonstrated in the aforementioned papers. The forward selection procedure also suffers from an inability to entertain and compare multiple models. That is, since the aliasing structure inherent in the SSD can hide real effects or encourage identifying nonactive effects as active, it is common for a stepwise procedure to easily be led astray by the entry of a nonactive effect. Often it is difficult or impossible for the stepwise procedure to recover from such errors. Thus, we recommend, as others have done, that all-subsets regression be utilized instead of forward selection.

This chapter proceeds as follows. In section 2, we begin by assessing the null model performance of supersaturated designs when using all-subsets and forward selection regression. The propensity for model selection criteria to overfit is highlighted. We subsequently propose a strategy for analyzing supersaturated designs that combines all-subsets regression and permutation tests. That is, in conjunction with all-subsets regression, we show how permutation procedures may be used to more appropriately select candidate models of interest and ascertain statistical significance for individual coefficients. Also, we will show how the power for detecting active effects decreases

as the number of factors in the SSD increases. Section 3 illustrates the method for several real and simulated data sets. Section 4 concludes the chapter with a discussion and comments on future research.

3.2 A Strategy for Analyzing Supersaturated Designs

3.2.1 More on All-Subsets vs. Forward Selection

The selection of a subset of predictor variables in regression is often done by forward selection. However, as we have alluded to, this procedure does not necessarily find the model with the minimum residual sum of squares for each subset size. All-subsets regression, which evaluates every possible subset of factors, has been recommended as an alternative. Berk [9] shows that the difference in favor of all-subsets can be "arbitrarily large in examples where there are predictors which do poorly alone but do very well together". Furthermore, Berk [9] states that "there is a tendency for forward selection to agree with all-subsets for small subset sizes, but not for large subset sizes" and that "it is doubtful that all-subsets can ever do very badly".

One of the criticisms of all-subsets regression is the computational burden. Beattie, Fong, and Lin [8] consider all-subsets to be impractical even when a moderate number of factors are active. For instance, a SSD with 23 factors and at most six active factors leads to consideration of 145,498 models, which is deemed to be a formidable comparison by those authors. However, our own simulations show that conducting 500 all-subsets regressions for a SSD with 23 factors and at most seven active factors ($\sum_{i=1}^{7} \binom{23}{i} = 390{,}655$ model comparisons) took less than 5 minutes (i.e. less than 1 second for each all-subsets regression) using SAS or R software packages

on a 1.1GHz Pentium with 256MB of RAM. Therefore, the ever increasing rise in modern computing abilities makes all-subsets regression more practical for everyday use.

We now turn our attention to the null performance of SSDs when using all-subsets and forward selection in conjunction with model selection criterion such as Akaike's Information Criterion (AIC). For details, Burham and Anderson [19] describe the foundations and underlying philosophy of AIC as well as its extensions. AIC is given by

$$AIC = -2log(L(\hat{\theta}|data)) + 2p \tag{3.5}$$

where $L(\hat{\theta}|data))$ is the likelihood function and p is the number of parameters in the fitted model. Suppose we compute AIC for two competing models: Model 1 with AIC_1 and Model 2 with AIC_2. Then, if $\text{AIC}_1 < \text{AIC}_2$, Model 1 is considered a better fit than Model 2.

For the special case of linear models with normally distributed errors, we express AIC as

$$AIC = nlog(\hat{\sigma}^2) + 2p \tag{3.6}$$

where

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{\epsilon}_i^2}{n}, \tag{3.7}$$

and the $\hat{\epsilon}_i$ are the estimated residuals from the fitted model. When p is large relative to the sample size n (as is the case for SSDs), there is a small sample (second order bias correction) version of AIC denoted as AIC_c,

$$AIC_c = -2log(L(\hat{\theta}|data)) + 2p + \frac{2p(p+1)}{n-p-1}. \tag{3.8}$$

Burham and Anderson [19] state that a "pervasive mistake in the model selection

literature is the use of AIC when AIC_c really should be used...People often conclude that AIC overfits because they failed to use the second order criterion." Unfortunately, the use of AIC_c does not always prevent overfitting, as we now demonstrate.

Consider again a SSD with n=14 runs and k=23 factors. In order to better understand the behavior of AIC_c under the null model, for each of 500 simulated response vectors, with each $y_1, ..., y_{14}$ from a $N(0, 1)$ distribution, we fit and score models of size $p = 1, 2, ..., 7$ chosen by randomly selecting columns from the 14×23 design matrix. Figure 3.1 shows the average AIC_c versus the number of terms in the fitted model. From this plot, we clearly see that the null model is chosen as best among the competing models. This is the result we wanted to see.

Now suppose that instead of choosing seven models at random for each response, we simulate 500 response vectors with each $y_1, ..., y_{14}$ from $N(0, 1)$, run both forward selection and all-subsets for $p = 1, 2, ..., 7$, and score AIC_c for the seven models chosen by these procedures. Figure 3.2 indicates the average AIC_c versus number of terms for both forward selection and all-subsets. Based on this, we see the propensity for even AIC_c to overfit when using forward and all-subsets selection procedures. In particular, when considering models up to 7 terms, AIC_c is smallest for a model with 7 terms on average. Thus, although the null model is true, the largest (7 variable) model is generally identified as best. We conjecture that this result occurs due to the small run size and the large number of candidate models explored via forward selection and all-subsets. Such overfitting is not beneficial, however, and some modification will be necessary in order to successfully utilize AIC_c for model selection in the context of SSDs.

Although the null model is true, Figure 3.2 indicates that all-subsets either outperforms or performs similarly to forward selection for every model size considered. For $p = 6, 7$, all-subsets drastically outperforms forward selection. Note that our use

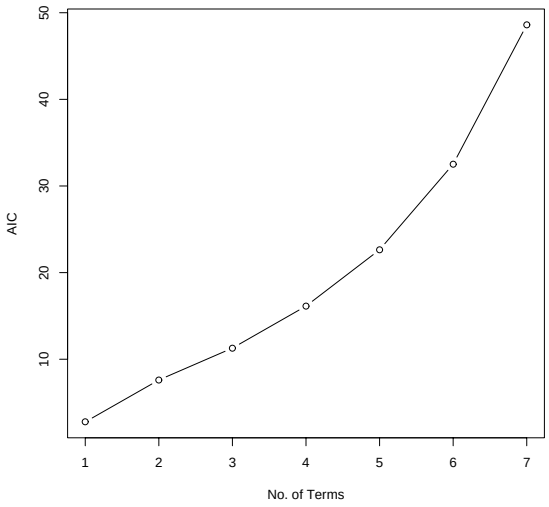

Figure 3.1: Mean AIC_c vs. Number of Factors- Null Model (Random Models)

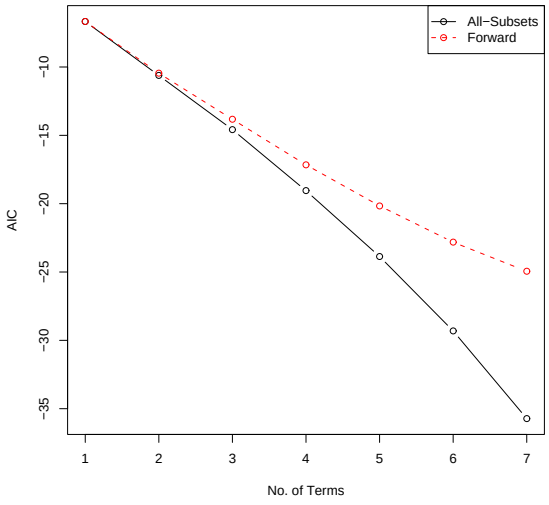

Figure 3.2: Mean AIC_c vs. Number of Factors -Null Model (All-Subsets and Forward Selection)

of the word 'outperform' is meant to convey the ability of all-subsets to locate models with lower AIC_cs than forward selection. This result provides further evidence that one should utilize all-subsets rather than forward selection. Also, we see that forward selection provides an upper bound on AIC_c (or likewise, a lower bound on R^2) for all-subsets. This result will especially prove useful when dealing with a large number of factors as will be seen later.

We now briefly and loosely explore how adding factors to a SSD affects its ability to detect active effects. To illustrate, consider augmenting the 28-run Plackett-Burman design in 27 factors with two-factor interaction columns. In particular, one could construct a 28-run SSD with up to 378 factors using only the main effect and two-factor interaction columns. Following this construction method, we investigate five designs: the 27-factor Plackett-Burman design and four SSDs with k=50, 100, 150, 200. For each design, we simulate 100 response vectors with each $y_1, ..., y_{28}$ from a $N(0, 1)$ distribution and fit models of size p=1,2,...,10 found via forward selection. For each fitted model, R^2 is computed as a measure of model adequacy. Note that forward selection is utilized here for ease in simulation as it provides a lower bound for all-subsets. Figure 3.3 displays the average R^2 for each design and model size.

Assuming the null model is true, Figure 3.3 clearly indicates that as the number of factors under consideration increases, the perceived systematic variation explained by the fitted model also increases. One should expect to see similar or larger R^2 statistics when using all-subsets. Therefore, true effects may potentially be masked in SSDs with a large number of factors vs. run size. That is, if on average the best s-variable model explains >80% of the variation when there are no true effects, our ability to identify true effects has decreased. This is meant to provide a healthy warning to practitioners considering the use of SSDs. That is, although these designs may prove useful in cases where there are many factors to study and runs are expensive, one

should not be over zealous in their use. Having provided some details regarding the use of forward selection and all-subsets regression with SSDs, we now move on to describe another component of our proposed analysis strategy: permutation tests.

3.2.2 Permutation Tests

Permutation (also known as randomization) tests calculate the probability of getting a value equal to or more extreme than the observed value of a test statistic under a specified null hypothesis by recalculating the test statistic after random shuffling of the data. The earliest descriptions of permutation tests for linear models can be traced back to the first half of the 20th century in the works of Fisher [31] and Pitman ([81], [82], [83]). Due to the computational intensity of these tests, however, their use did not receive much attention until the emergence of widely accessible computing power. More recent work in this area can be found in Kennedy and Cade [49], Anderson and Legendre [5], and Anderson and Robinson [6]. Anderson [4] provides a thorough review of permutation test procedures, consolidates recent findings, and provides practical recommendations for practitioners. For book length treatments, see Edgington [30], Manly [67], and Good [36].

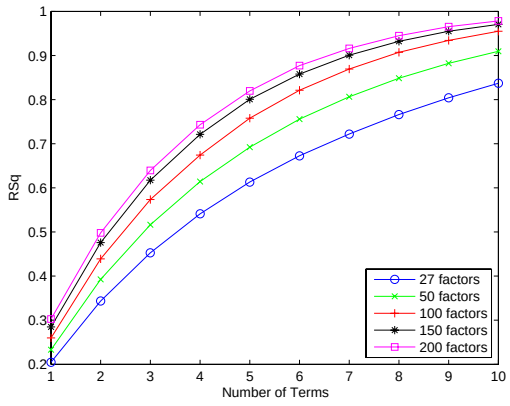

Figure 3.3: R^2 vs. Number of Factors - Null Model

Following Anderson [4], we briefly review permutation procedures in the context of linear regression. Consider the model,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \epsilon \tag{3.9}$$

and suppose we want to test the global null hypothesis $H_0 : \beta_1 = \beta_2 = ... = \beta_p = 0$. An appropriate test statistic, among others, is the usual coefficient of multiple determination (R^2). In order to conduct a permutation test, one needs to think about what is exchangeable under a true null hypothesis (i.e. $Y = \beta_0 + \epsilon$). Under the assumption that the errors, ϵ, are i.i.d., the observations are exchangeable, which means that if Y has no relationship with any of the explanatory variables, $X_1, ..., X_p$, then the values obtained for Y could have been observed in any order. Thus, an exact p-value for the above hypothesis test, conditional on the observed y's, is obtained by randomly permuting Y, leaving $X_1, ..., X_p$ fixed, and recalculating R^2 for each of B permutations (denoted by $R^{2(b)}$). That is, we calculate

$$p - value = \frac{\#(R^{2(b)} \geq R^2)}{B} \tag{3.10}$$

where # means 'number of'.

It is often the case that one is interested in something more specific than the global null hypothesis given above. Rather, we frequently want to test $H_0 : \beta_i = 0$ for individual $i = 1, 2, ..., p$. One statistic for testing the relationship between Y and X_i given $X_1, X_2, ..., X_{i-1}, X_{i+1}, ..., X_p$ is

$$F = \frac{SSR(X_i | X_1, X_2, ..., X_{i-1}, X_{i+1}, ..., X_p)/1}{SSE(X_1, X_2, ..., X_p)/(n - p - 1)} \tag{3.11}$$

where $SSR(X_i | X_1, X_2, ..., X_{i-1}, Xi + 1, ..., Xp)$ represents the reduction in the error

sums of squares associated with X_i given that $X_1, X_2, ..., X_{i-1}, X_{i+1}, ..., X_p$ are already in the model and $SSE(X_1, X_2, ..., X_p)$ is the error sums of squares for the full model. Other appropriate test statistics include the squared partial correlation coefficient, AIC_c, etc. For now, let us denote the chosen test statistic for the partial permutation test as TS.

Under a true null hypothesis, we have $Y = \beta_0 + \beta_1 X_1 + ... + \beta_{i-1} X_{i-1} + \beta_{i+1} X_{i+1} + ... + \beta_p X_p + \epsilon$. In this case, the errors ϵ after removing the effect of X are exchangeable and not the observations Y. That is, the units that should be shuffled for an exact permutation test are $\epsilon = Y - (\beta_0 + \beta_1 X_1 + ... + \beta_{i-1} X_{i-1} + \beta_{i+1} X_{i+1} + ... + \beta_p X_p)$. Unfortunately, the parameters are not known. Therefore, no exact permutation test for individual coefficients is possible. However, several approximations are available.

Anderson [4] states that the best method for partial regression permutation tests is that of Freedman and Lane [32]. Their idea is to simply replace the unknown parameters with their estimates. That is, compute $e = Y - (b_0 + b_1 X_1 + ... + b_{i-1} X_{i-1} + b_{i+1} X_{i+1} + ... + b_p X_p)$ which approximate the errors (ϵ) that are exchangeable under the null hypothesis. Then, for each of B reorderings of the residuals, e, we compute TS (denoted by $TS^{(b)}$). The p-value is computed as the proportion of $TS^{(b)}$s that exceed TS for the original data. That is,

$$p - value = \frac{\#(TS^{(b)} \geq TS)}{B}. \tag{3.12}$$

It is worthy to point out that that this permutation test is not exact, but is asymptotically exact. Such a test has a Type I error that asymptotically approaches the significance level chosen for the test, with increases in n.

Also, we should comment that when the assumptions of the normal theory tests are satisfied, the permutation test and normal theory test converge. Instead of using

permutation tests, Westfall, Young, and Lin [93] developed a Monte Carlo procedure for computing adjusted p-values based on simulating a response (Y) from a $N(0,1)$ distribution. Reconsider Figure 3.3 in which the average R^2 values for five different designs were all computed based on simulating a $N(0,1)$ response. Instead, suppose that we compute the average R^2 after permuting a response Y simulated from some non-null model. In most instances, we found that the average R^2 computed from a permuted response vector vs. $N(0,1)$ responses were nearly identical. However, if Y resembled a very skewed distribution with large outliers, the permutation test actually performed mildly better giving slightly smaller R^2 values than the normal test as the number of factors of interest increased. Although the differences were not large, this result as well as the intuition gained provides some justification for the use of permutation procedures.

3.2.3 Proposed Analysis Strategy

1. *All-Subsets Regression.* Our procedure begins by performing all-subsets regression and retaining the best, say 5, models of each size under consideration. We have discussed the advantages and disadvantages of all-subsets previously and believe it is well warranted for use in the context of SSDs. The user has considerable freedom in this step with regards to the maximum model size, m, as well as the number of candidate models retained for further exploration.

2. *Global Model Test.* For each model under consideration, it is useful to first perform a test of the global null hypothesis ($H_0 : \beta_1 = \beta_2 = ... = \beta_p = 0$). Any model failing this test need not be examined any further. In fact, this test provides a reality check with regards to the propensity for model selection techniques to select models from SSDs that explain a very large percentage of

the total variation in a response. That is, a candidate model with an R^2=0.95 or higher may easily fail the global model test. A permutation test for the global null hypothesis is conducted as follows:

(a) Compute R^2 for a model of size p of interest.

(b) For each of B permutations (shuffling) of the response, Y, perform all-subsets regression for models of size p and select the model with largest R^2. Denote this R^2 as $R^{2(b)}$.

(c) Compute $p = \#(R^{2(b)} \geq R^2)/B$. Thus, we compute the probability of finding an R^2 greater than that observed for the model fit to the observed data.

In general, we recommend using B=1000. However, performing 1000 all-subsets in (b) may become quite computationally burdensome for large k. That is, for a large number of factors, say $k \geq 100$, greater consideration of computational time must be utilized with all-subsets regression. There are several possible options to remedy this problem. For large k, one simple solution is to use forward selection in place of all-subsets. Recalling that forward selection provides a lower bound on R^2 for all-subsets, a p-value that fails to reject the global null hypothesis using forward selection will also fail to reject the null hypothesis using all-subsets. Therefore, only if one obtains a significant p-value (i.e. p-value less than some specified significance level α) when using forward selection would the need arise to perform all-subsets repeatedly. If all-subsets must be conducted for large k, one possibility to reduce computation is to choose a smaller B. We will utilize this option when necessary; although it comes with the unfortunate cost of a loss of precision in computing the global p-value.

3. *Select candidate models based on AIC_c.* A rejected global null hypothesis reveals

that at least one factor in the model affects the response. Thus, the permutation test described in 2) may reveal many models of interest. Our use of permutation tests in conjunction with AIC_c will assist in advising the user on candidate models. Furthermore, permutation tests will provide us with a necessary adjustment to AIC_c to help alleviate the problem of overfitting alluded to above. Therefore, we suggest selecting candidate models as follows:

(a) Select a model (M1) of interest of size p for $p = 1, 2, ..., m$. Compute AIC_c for M1.

(b) Find the best model (M2) of size $p-1$ that is nested within M1 by removing the term in M1 with the smallest t-statistic.

(c) Compute the residuals of M2.

(d) For each of B permutations of the residuals calculated in c), run forward selection up to size p assuming that the $p-1$ variables of M2 are forced to enter. Compute AIC_c for each model found via forward selection and denote this by $\text{AIC}_c^{(b)}$.

(e) Compute $\Delta\text{AIC}_c = \overline{\text{AIC}_c} - \text{AIC}_c$ where

$$\overline{\text{AIC}_c} = \frac{\sum_{b=1}^{B} \text{AIC}^{(b)}}{B}. \tag{3.13}$$

A negative ΔAIC_c indicates that the AIC_c for M1 is worse than the average AIC_c based on models found using random permutations of the residuals of M2. Therefore, M1 need not be considered as a candidates for further exploration. On the other hand, a positive ΔAIC_c indicates that the AIC_c for M1 is better than the average AIC_c based on random permutations of the residuals of M2. Thus, M1 is worthy of further consideration. In particular, the larger the ΔAIC_c, the more evidence in favor (support)

for the candidacy of M1. Some simple rules of thumb are often useful in assessing the relative merits of models: models having $\Delta\text{AIC}_c \leq 1$ show little to no support, those where $3 \leq \Delta\text{AIC}_c \leq 7$ would show considerably more support, while models with $\Delta\text{AIC}_c \geq 10$ would have substantial support. The ΔAIC_c are easy to interpret and allow for an easy "strength" comparison or ranking of candidate models.

4. *Partial Tests.* Having narrowed ourselves down to one or more candidate models, we would like to determine which terms are worthy of inclusion in further experimentation. Thus, we now conduct a test of the partial null hypothesis $H_0 : \beta_i = 0$ for $i = 1, 2, ..., p$ using permutation tests as follows:

   (a) For a candidate model of interest (again, denote by M1), compute the t-statistic for each term in the model (denoted by t_i).

   (b) For each $i = 1, 2, ..., p$, let $\text{M2}_{(-i)}$ represent the model M1 without the ith term. Compute the residuals of $\text{M2}_{(-i)}$.

   - For each of B permutations of the residuals, run forward selection up to size p assuming that the $p - 1$ variables of $\text{M2}_{(-i)}$ are forced to enter. Calculate the t-statistic for the term chosen via forward selection (denoted by $t_i^{(b)}$).

   (c) For each $i = 1, 2, ..., p$, compute

$$p_i = \frac{\#(t_i^{(b)} \geq t_i)}{B}. \tag{3.14}$$

We recommend using a significance level of 10% to make decisions regarding each individual term. Those terms selected as having a significant impact on the response should now be investigated further in follow-up experiments.

3.3 Examples

3.3.1 Williams Data

Full Williams Data

In this section, we explore the Williams data as given in Table 3.3. Recall that supersaturated subsets of this 28-run, 23-factor design has become a benchmark for studying SSDs. Although the full design is an orthogonal array, we analyze the data using both Lenth's method as well as the proposed strategy for comparison. Table 3.4 displays Lenth's empirical p-values computed using JMP 7.0's Screening Platform. C24-C27 represent the remaining four columns of a Hadamard matrix of order 28. These four columns serve as branching columns for constructing SSDs using the method of Lin [60]. Based on the results, we see factors 15 and 19 significant at a 5% level and factors 4 and 16 significant at a 10% level.

In order to apply the proposed analysis strategy, we first begin with an all-subsets regression. For the sake of simplicity, we choose m=6. Table 3.5 displays the best five models of each size based on AIC_c. For brevity, we now choose the best model of each size for further consideration. The next step is to perform a test of the global null hypothesis as described above. The results are shown in Table 3.6 and clearly indicate that the global null hypothesis is rejected for each model considered. Therefore, we move on to pick candidate models based on computing ΔAIC_c and are given in Table 3.7. Based on this, the model with only factor 15 is worthy of further consideration based on a ΔAIC_c of 6.89. All other ΔAIC_cs are negative. This result is somewhat consistent with the traditional analysis as factor 15 was the most significant with a p-value of 0.0023.

Table 3.4: Analysis of Full Williams Data

Term	Estimate	t Ratio	Empirical p-value
1	-13.03571	-1.25	0.2096
2	-10.03571	-0.96	0.3307
3	-3.392857	-0.32	0.7673
4	18.25	1.75	0.0939
5	6.25	0.60	0.5836
6	-5.892857	-0.56	0.6062
7	-6.75	-0.65	0.5503
8	14.821429	1.42	0.1609
9	7.6071429	0.73	0.4635
10	7.1071429	0.68	0.4958
11	-1.821429	-0.17	0.8780
12	-7.892857	-0.76	0.4480
13	-11.46429	-1.10	0.2652
14	-15.46429	-1.48	0.1452
15	-43.17857	-4.13	0.0023
16	-21.39286	-2.05	0.0547
17	-6.821429	-0.65	0.5457
18	-1.75	-0.17	0.8833
19	-24.39286	-2.34	0.0307
20	-4.178571	-0.40	0.7161
21	-16.10714	-1.54	0.1326
22	-5.964286	-0.57	0.6011
23	-6.178571	-0.59	0.5868
C24	13.5357	1.30	0.1948
C25	6.5357	0.63	0.5644
C26	-6.6786	-0.64	0.5550
C27	-3.6786	-0.35	0.7451

Table 3.5: Full Williams Data - All-Subsets

Subset						AIC_c
15						230.78
19						240.04
16						240.88
4						241.63
21						242.06
15	19					227.61
15	16					229.03
4	15					230.26
15	21					230.96
14	15					231.15
15	16	19				225.02
4	15	19				226.57
15	19	21				227.45
14	15	19				227.69
8	15	19				227.91
4	15	16	19			223.37
15	16	19	21			224.46
14	15	16	19			224.76
8	15	16	19			225.04
1	15	16	19			225.74
4	15	16	19	21		222.41
4	14	15	16	19		222.77
4	8	15	16	19		223.1
1	4	15	16	19		223.95
14	15	16	19	21		224.02
4	14	15	16	19	21	221.45
4	8	15	16	19	21	221.85
4	8	14	15	16	19	222.26
1	4	15	16	19	21	222.85
1	4	14	15	16	19	223.25

Table 3.6: Full Williams Data - Global Model Test

Model						$P(R^{2(b)} \geq R^2)$
15						0.008
15	19					0.020
15	16	19				0.027
4	15	16	19			0.044
4	15	16	19	21		0.053
4	14	15	16	19	21	0.066

Table 3.7: Full Williams Data - ΔAIC_c

Model						ΔAIC_c
15						6.89
15	19					-1.13
15	16	19				-2.98
4	15	16	19			-5.42
4	15	16	19	21		-7.07
4	14	15	16	19	21	-8.41

Supersaturated Subset of Williams Data 1

One possible supersaturated subset of the Williams data, among others, consists of rows 3, 6, 8, 10, 12, 13, 15, 16, 17, 19, 20, 21, 25, and 27. The r-rank of this SSD was determined to be 6 using an exhaustive search. Thus, we perform an all-subsets regression for models up to size 6. Retaining the best five models of each size, the results of all-subsets and the global model test, and ΔAIC_c are given in Table 3.8. Based on Table 3.8, only a handful of the models appear "remarkable" and worthy of further study. Of those models that do reject the global null hypothesis, we compute ΔAIC_c and display the results in Table 3.9. Clearly, one can see the benefit of using ΔAIC_c for protection against overfitting.

As with the full Williams data, the model with only factor 15 (Model 1) stands out above the others with a large positive ΔAIC_c of 5.59. However, the model with 15 and 23 (Model 2) could also be selected for consideration with ΔAIC_c of 1.7. As can be seen, these two models are the only two with positive ΔAIC_cs. In order to determine which terms in the selected models most significantly affect the response, we continue with the partial permutation tests as described in the previous section. Clearly, the partial p-value for the model with only factor 15 is the same as the global p-value of 0.012. Thus, factor 15 should be retained for follow-up experiments. The partial p-values are shown in Table 3.10 for Model 2. These p-values indicate the significance of factor 15 with a p-value of 0.008. Factor 23 is not significant. Thus, for this supersaturated subset, only factor 15 appears active.

Supersaturated Subset of Williams Data 2

We now consider another supersaturated subset of the Williams data for comparison purposes. This SSD consists of rows 1, 3, 4, 5, 7, 9, 13, 15, 16, 17, 18, 21, 23, and 27 of Table 3.2. As with the first subset, we determine the r-rank of this SSD to be 6

Table 3.8: Analysis of Supersaturated Subset of Williams Data 1 - Global Test

Subset						AIC_c	$P(R^{2(b)} \geq R^2)$
15						103.71	0.01
23						110.52	0.67
17						112.66	0.98
13						113.36	0.996
8						113.62	0.997
15	23					96.87	0.005
15	17					102.03	0.107
14	15					103.21	0.183
13	15					103.44	0.222
8	15					103.93	0.251
15	17	23				92.82	0.031
13	15	23				95.7	0.086
8	15	23				96.61	0.121
15	20	23				98.22	0.197
15	21	23				98.47	0.186
13	15	17	23			90.16	0.067
8	15	17	23			91.81	0.105
15	17	19	23			92.86	0.136
13	15	21	23			94.77	0.169
15	17	21	23			95.33	0.215
8	13	15	17	23		87.29	0.085
13	15	17	21	23		87.56	0.108
13	15	17	19	23		90.75	0.182
8	15	17	19	23		92.92	0.295
1	14	15	17	19		92.96	0.249
8	13	15	17	19	23	87.38	0.21
1	8	14	15	17	19	88.31	0.26
1	7	14	15	17	19	89.49	0.263
8	13	15	17	21	23	90.24	0.305
13	15	16	18	21	23	90.37	0.336

Table 3.9: Analysis of Supersaturated Subset of Williams Data 1 - ΔAIC$_c$s

Subset					ΔAIC$_c$
15					5.59
15	23				1.7
15	17	23			-2.72
13	15	23			-5.73
13	15	17	23		-5.03
8	13	15	17	23	-6.25

Table 3.10: Analysis of Supersaturated Subset of Williams Data 1 - Partial Tests

Factor	$\|t - ratio\|$	P-value
15	5.09	0.008
23	3.42	0.150

and thus perform all-subsets regression for models up to this size. The results of the analysis are given in Tables 3.11, 3.12, and 3.13. In particular, we again see the largest ΔAIC_c for the single variable model with factor 15. However, the four variable model with factors 1, 5, 8, and 15 also appears worthy of investigation with ΔAIC_c=4.173. Interestingly, all of the partial p-values shown in Table 3.13 are significant at the α=0.05 level. This is perhaps a surprising result since none of factors 1, 5, or 8 were deemed active by the full Williams data.

We suspect that further experimentation will indeed reveal that only factor 15 is active among the four factors explored here. Furthermore, one should take note of the differences between the two Williams subsets in terms of the all-subsets procedure. As pointed out in Abraham, Chipman, and Vijayan [1], "whether a single model or many models are considered, the important point is that different designs lead to identification of different factors as active. Even all-subsets regression, which identifies more models than stepwise, can still mislead in some cases."

Supersaturated Subset of Williams Data 3

Let us consider one more supersaturated subset of the Williams data consisting of rows 2, 6, 8, 10, 11, 12, 14, 19, 20, 22, 24, 25, 26, and 28. All-subsets for models up to size $m = 6$ was performed and the results of the global model test are given in Table 3.14. The best model of each size $i = 1, 2, ..., 6$ has an R^2 of 0.48, 0.67, 0.79, 0.87, 0.94, and 0.95, respectively. Clearly, however, none of the models reject the global null hypothesis. That is, the permutation test indicates that the models found using all-subsets are not as "remarkable" as they appear based on R^2 . Therefore, we need not proceed any further with the proposed analysis strategy for this particular subset. Thus, the past three examples have illustrated that the choice of subset does indeed impact what factors are chosen as worthy of further consideration.

Table 3.11: Analysis of Supersaturated Subset of Williams Data 2 - Global Test

Subset						AIC_c	$P(R^{2(b)} \geq R^2)$
15						117.58	0.009
8						121.74	0.273
16						126.68	1.000
21						126.75	1.000
1						126.89	1.000
15	19					113.98	0.012
8	15					116.42	0.132
15	16					118.13	0.329
15	21					118.24	0.297
1	15					118.48	0.393
5	8	15				109.31	0.090
3	15	19				113.62	0.304
1	8	15				113.62	0.319
2	15	19				113.95	0.357
8	15	19				114.01	0.341
1	5	8	15			97.36	0.026
5	8	15	19			106.57	0.092
5	8	12	15			111.15	0.333
3	4	15	19			111.22	0.292
5	8	15	21			111.83	0.392
1	5	8	15	20		96.67	0.039
1	5	8	12	15		98.28	0.061
1	5	8	15	21		98.97	0.053
1	5	8	15	19		100.8	0.078
1	5	8	15	22		100.82	0.108
1	5	8	11	15	21	97.6	0.128
1	5	8	12	15	20	98.69	0.145
5	6	8	10	17	20	99.44	0.159
1	5	8	12	15	21	102.51	0.277
1	5	8	15	19	20	102.8	0.318

Table 3.12: Analysis of Supersaturated Subset of Williams Data 2 - ΔAIC_cs

Subset					ΔAIC_c
15					5.27
15	19				-0.973
5	8	15			1.05
1	5	8	15		4.173
5	8	15	19		-4.73
1	5	8	15	20	-7.78
1	5	8	12	15	-9.71
1	5	8	15	21	-10.03
1	5	8	15	19	-12.1

Table 3.13: Analysis of Supersaturated Subset of Williams Data 2 - Partial Tests

Factor	$\lvert t - ratio \rvert$	P-value
1	4.6181	0.030
5	5.6786	0.007
8	7.7871	0.000
15	6.0302	0.005

Table 3.14: Analysis of Supersaturated Subset of Williams Data 3 - Global Test

Subset						AIC_c	$P(R^{2(b)} \geq R^2)$
2						111.268	0.369
6						114.341	0.954
15						115.071	0.995
19						115.291	0.997
8						115.326	0.999
2	13					110.139	0.743
2	6					110.596	0.765
2	19					111.869	0.9
2	8					111.915	0.891
2	4					112.296	0.933
2	6	19				108.719	0.829
2	8	13				108.72	0.822
2	3	6				109.814	0.995
2	13	20				110.468	0.934
2	6	16				110.886	0.951
2	6	16	19			103.737	0.596
2	12	13	20			108.273	0.887
2	3	8	13			109.082	0.909
2	6	14	19			109.184	0.912
2	7	8	13			109.815	0.944
2	6	14	16	19		104.135	0.817
2	4	6	16	19		104.933	0.857
2	6	16	19	23		106.364	0.917
2	6	12	16	19		106.778	0.924
2	6	10	16	19		106.805	0.936
2	6	10	14	16	19	103.138	0.862
2	6	4	16	19	21	103.936	0.833
2	4	6	16	19	23	104.811	0.91
2	3	6	10	16	19	107.112	0.973
2	6	14	16	19	23	107.729	0.968

3.3.2 A Bayesian D-Optimal SSD

Bayesian D-Optimal SSDs were introduced recently in Jones, Lin, and Nachtsheim [46]. For these designs and taking $\sigma^2 = 1$, the prior distribution is specified as $\beta \sim N(\beta_0, \mathbf{R}^{-1})$, the conditional distribution of Y given β is $\mathrm{Y}|\beta \sim N(\mathbf{X}\beta, \mathbf{I})$, and the posterior distribution for β given Y is $\beta|\mathrm{Y} \sim N(\beta^*, \mathbf{D})$ where $\beta^* = (\mathbf{X}'\mathbf{X} + \mathbf{R})^{-1}(\mathbf{X}'\mathrm{Y}+\mathbf{R}\beta_0)$ and $\mathbf{D} = (\mathbf{X}'\mathbf{X} + \mathbf{R})^{-1}$.

For a design, ξ, in the design space Ξ, the Bayesian D-optimal design, ξ_B satisfies:

$$\xi_B = argmax_{\xi \in \Xi} \left| \mathbf{X}'\mathbf{X} + \mathbf{R} \right|. \tag{3.15}$$

For the prior distribution specified above, the information matrix of the parameters is $\mathbf{R} = \mathbf{K}/\tau^2$ where

$$\mathbf{K} = \begin{bmatrix} \mathbf{0}_{p_1 \times p_1} & \mathbf{0}_{p_1 \times p_2} \\ \mathbf{0}_{p_2 \times p_1} & \mathbf{I}_{p_2 \times p_2} \end{bmatrix}, \tag{3.16}$$

τ^2 is the prior variance, and p_1 and p_2 represent the number of primary and potential terms, respectively. In the SSD case, all terms except the intercept are potential terms unless the experimenter has knowledge of certain factors being active.

In this example, we examine a Bayesian D-Optimal SSD with 25 runs and 52 factors. In order to construct this design, we make use of JMP 7.0's Custom Design platform. By specifying the desired number of runs, factors, and designating the estimability of main effects as "If Possible", we obtain the design in Table 3.15. The response, Y, is simulated by first generating a 52×1 vector from a $N(0, 0.2)$ distribution, which we denote as β_{init}. Next, five factors are chosen at random and given a value of ± 2. This ± 2 is then added to the corresponding rows of β_{init} to

Table 3.15: Bayesian D-Optimal Design

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	-1	1	1	1	1	-1	1	-1	1	1	1	-1	1	-1	1	-1	1	1
-1	1	1	-1	-1	-1	-1	-1	1	1	-1	1	1	1	1	-1	-1	1	-1	1
1	-1	-1	-1	1	-1	-1	1	1	1	1	1	-1	1	-1	1	-1	-1	1	-1
1	-1	1	1	1	-1	1	-1	-1	1	-1	1	1	1	-1	-1	1	1	1	-1
-1	-1	1	1	-1	1	-1	-1	-1	1	1	1	-1	1	1	-1	1	-1	-1	-1
-1	1	-1	1	1	-1	-1	1	1	1	-1	-1	1	-1	1	-1	1	-1	-1	-1
1	1	1	1	1	1	-1	-1	-1	1	1	-1	1	1	1	1	1	1	-1	-1
1	1	-1	-1	1	1	1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1
1	-1	-1	1	1	-1	-1	1	-1	-1	-1	-1	-1	1	1	-1	1	1	1	1
-1	-1	-1	1	1	1	1	-1	1	-1	-1	1	-1	1	-1	1	1	-1	1	1
1	-1	-1	-1	-1	-1	1	1	1	-1	-1	-1	1	-1	-1	-1	-1	1	-1	-1
1	1	1	1	1	-1	1	-1	1	1	1	-1	1	-1	-1	1	-1	1	1	1
1	1	-1	-1	-1	1	1	1	-1	1	1	1	-1	-1	-1	-1	-1	-1	-1	1
-1	-1	-1	-1	-1	-1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	-1	1	-1
1	-1	1	1	-1	1	-1	1	1	1	-1	-1	-1	-1	1	1	-1	1	1	-1
1	1	1	1	-1	1	-1	1	-1	-1	-1	-1	1	1	-1	1	-1	-1	1	-1
-1	1	-1	-1	-1	1	-1	1	1	-1	1	1	1	-1	-1	-1	1	1	-1	-1
-1	-1	1	1	-1	-1	1	1	-1	-1	-1	1	-1	-1	-1	1	-1	1	-1	1
1	1	1	-1	-1	1	1	1	1	-1	1	-1	-1	1	-1	1	1	1	1	1
1	1	-1	1	1	1	-1	1	-1	-1	1	1	-1	-1	-1	-1	-1	1	-1	1
1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	-1	-1	1	1	1	-1
-1	1	-1	1	-1	-1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1	-1	1
-1	-1	1	-1	1	-1	-1	1	-1	-1	1	1	1	-1	1	1	1	1	1	1
-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
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
-1	1	-1	1	-1	1	1	1	1	1	-1	1	-1	-1	1	1	-1	-1	1	-1
1	-1	-1	1	1	1	1	-1	1	1	1	1	-1	1	1	1	1	1	-1	-1
1	-1	-1	-1	1	1	1	-1	1	-1	1	-1	1	1	-1	-1	-1	-1	1	-1
1	-1	-1	-1	-1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	-1	-1	1
-1	1	1	1	-1	1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	-1	1	-1	-1
1	1	1	1	-1	-1	1	1	1	-1	1	-1	1	-1	-1	-1	-1	-1	-1	1
-1	-1	1	-1	1	1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	-1	1	1
-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1
1	-1	-1	-1	1	1	1	-1	1	-1	1	-1	1	1	-1	-1	-1	-1	1	-1
1	1	-1	1	1	-1	-1	1	1	1	1	1	1	1	1	1	1	1	1	1
-1	-1	1	-1	-1	-1	1	1	-1	1	-1	-1	1	1	1	-1	-1	1	1	-1
-1	-1	-1	1	1	-1	-1	-1	1	-1	-1	1	1	-1	-1	-1	1	1	1	-1
1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	-1	-1	1	1	-1	1	1
-1	-1	1	1	1	-1	1	1	-1	1	-1	-1	1	1	1	1	-1	1	1	-1
1	-1	-1	1	-1	1	-1	1	-1	-1	1	1	1	1	1	1	-1	-1	1	-1
-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	-1	-1	-1	1	1	1	1	-1	1
-1	-1	-1	-1	1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1
1	1	-1	1	-1	-1	-1	-1	1	-1	1	1	-1	1	-1	-1	-1	-1	1	1
41	42	43	44	45	46	47	48	49	50	51	52	Y							
1	1	1	1	-1	-1	-1	-1	-1	1	-1	1	-0.97							
1	1	1	1	1	-1	1	1	1	1	1	1	-5.8							
-1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1.83							
-1	1	1	1	1	1	-1	1	-1	-1	-1	-1	-4.44							
-1	1	-1	-1	-1	1	-1	1	-1	1	1	1	-4.99							
1	-1	1	1	1	1	1	1	1	1	-1	-1	-3.58							
-1	1	-1	-1	-1	-1	1	1	1	1	1	-1	-4.63							
1	-1	-1	1	1	1	1	-1	1	-1	1	-1	6.39							
-1	-1	1	-1	-1	-1	1	-1	1	-1	1	1	0.81							
-1	-1	-1	-1	1	-1	-1	1	1	1	-1	-1	4.36							
-1	-1	-1	-1	1	-1	1	1	-1	1	1	1	6.72							
-1	-1	1	-1	1	1	1	1	1	-1	1	1	5.54							
1	1	-1	-1	1	-1	1	1	-1	-1	1	-1	13.52							
1	1	1	-1	-1	1	-1	-1	1	-1	-1	-1	0.37							
1	-1	-1	-1	-1	1	1	-1	1	-1	-1	1	-3.02							
1	-1	-1	1	-1	-1	-1	1	1	1	-1	-1	-6.24							
-1	-1	1	-1	-1	1	-1	-1	-1	-1	-1	-1	0.4							
1	-1	1	1	-1	1	-1	1	-1	1	1	1	-2.17							
-1	1	1	-1	-1	1	1	-1	-1	1	1	1	-4.04							
-1	1	1	1	-1	1	1	1	1	1	-1	-1	4.81							
1	-1	1	-1	1	-1	-1	-1	-1	-1	1	-1	-6.4							
-1	1	-1	-1	-1	-1	-1	-1	1	-1	-1	1	6.7							
1	1	-1	1	1	1	-1	-1	1	1	1	1	-9.46							
-1	-1	-1	1	1	1	1	1	-1	-1	-1	-1	-2.92							
1	-1	1	1	-1	-1	1	-1	-1	-1	1	-1	0.06							

create β. Thus, we obtain

$$\begin{aligned}\beta = [&0.3970 \quad -0.2924 \quad -2.2015 \quad -0.5662 \quad -0.0577 \quad 0.1377 \quad 2.0890\\ &-0.0362 \quad 0.1663 \quad 0.2762 \quad 0.0431 \quad -0.3071 \quad -0.0441 \quad -0.2518\\ &-0.2885 \quad -0.0973 \quad 0.1158 \quad -0.0904 \quad -0.1256 \quad 0.1626 \quad -0.0355\\ &0.0035 \quad 0.1481 \quad -0.0539 \quad 0.0659 \quad 0.1483 \quad -1.7600 \quad -1.5800\\ &-0.1589 \quad -0.0368 \quad -0.0503 \quad 0.4359 \quad 0.1233 \quad 0.1933 \quad -0.0575\\ &0.0897 \quad 0.5396 \quad -0.2903 \quad 0.1549 \quad 0.1424 \quad -0.0567 \quad -0.3019\\ &0.0031 \quad -0.1919 \quad -0.1669 \quad -0.0839 \quad 1.7600 \quad 0.0388 \quad -0.0429\\ &-0.0313 \quad 0.3189 \quad -0.0912]'.\end{aligned}$$

Finally, we let $\mathrm{Y} = \mathbf{D}\beta + \epsilon$ where $\mathbf{D}$ is the design matrix and $\epsilon \sim N(0, 1)$. For this example, factors 3, 27, and 28 are assigned a value of -2, while factors 7 and 47 are assigned a value of 2.

Table 3.16 displays the results of all-subsets for up to m=7 factors, the corresponding AIC_cs, and p-values for a test of the global null hypothesis. Although our rule of thumb would suggest using $m = \lfloor min\{k/5, n/2\} \rfloor = \lfloor min\{52/5, 25/2\} \rfloor = \lfloor 12.5 \rfloor = 12$, we choose m=7 since we know the true model and to save time and space. Note that all models with 5 or more terms pass the global test. However, as with previous examples, this simulated data set shows how the use of $\Delta\mathrm{AIC}_c$ helps prevent selection of models that overfit. Table 3.17 shows the $\Delta\mathrm{AIC}_c$s for those models in Table 3.16 that have significant p-values at $\alpha = 0.1$. All $\Delta\mathrm{AIC}_c$s for models larger than 5 are negative, which indicates that they need not be considered further. Since all of the models with positive $\Delta\mathrm{AIC}_c$ are nested within the five variable model with factors 3, 7, 27, 28, and 47, we need only consider partial tests for these term (shown in Table 3.18). Clearly, all of the true factors are deemed significant and should be investigated more closely in subsequent experiments.

Table 3.16: Analysis of Bayesian D-Optimal SSD - Global Test

Subset							AIC_c	$P(R^{2(b)} \geq R^2)$
7							70.86	0.176
3							74.75	0.882
50							75.28	0.950
28							76.05	0.994
17							76.30	0.996
7	28						64.55	0.120
7	47						67.30	0.342
7	17						68.52	0.482
3	7						69.07	0.530
7	21						69.47	0.585
7	28	47					56.84	0.050
7	21	28					60.16	0.153
3	7	28					62.09	0.253
3	7	47					63.90	0.402
7	28	39					63.96	0.404
3	7	28	47				50.50	0.039
3	7	27	47				52.94	0.075
7	21	28	47				53.16	0.077
7	22	28	47				54.40	0.130
7	28	39	47				55.21	0.145
3	7	27	28	47			30.39	0.000
3	7	21	28	47			47.59	0.087
3	7	28	39	47			48.92	0.110
3	7	28	44	47			49.76	0.129
3	7	28	29	47			49.92	0.141
3	7	27	28	44	47		27.22	0.002
3	5	7	27	28	47		29.17	0.001
3	7	27	28	31	47		29.19	0.002
3	7	27	28	47	49		30.46	0.003
3	4	7	27	28	47		30.58	0.003
3	7	27	28	44	47	49	26.10	0.007
3	4	7	27	28	44	47	26.35	0.003
3	7	27	28	40	44	47	26.48	0.009
3	7	27	28	43	44	47	26.95	0.011
3	5	7	25	27	28	47	28.01	0.007

Table 3.17: Analysis of Bayesian D-Optimal SSD - ΔAIC_cs

Subset							ΔAIC_c
7	28	47					0.0165
3	7	28	47				-2.796
3	7	27	47				1.800
3	7	27	28	47			9.638
3	7	21	28	47			-7.538
3	7	27	28	44	47		-8.687
3	5	7	27	28	47		-10.272
3	7	27	28	31	47		-10.483
3	7	27	28	47	49		-11.819
3	4	7	27	28	47		-11.868
3	7	27	28	44	47	49	-11.470
3	4	7	27	28	44	47	-11.405
3	7	27	28	40	44	47	-12.046
3	7	27	28	43	44	47	-12.428
3	5	7	25	27	28	47	-11.957

Table 3.18: Analysis of Bayesian D-Optimal Design - Partial Tests

Factor	$\lvert t - ratio \rvert$	P-value
3	6.52	0.000
7	7.92	0.000
27	5.46	0.002
28	5.91	0.001
47	7.67	0.000

3.3.3 AIDS Data

Lin [61] illustrates the use of forward selection on a SSD with 24 runs and 138 factors based on a case study for testing and validating an acquired immune deficiency syndrome (AIDS) model. The response is the AIDS incidence rate of a population of size 100,000 and the large number of factors were generated by categorizing ages so that variables might have an instantiation for ages 15-29, 30-45, 45+, etc. The design and response are based on details frm Lin [61]. Figure 3.4 displays a histogram of the response and clearly indicates the presence of a severe outlier. Since this outlying observation can easily distort ones attempt to identify models that explain the systematic variation in the data, we believe that a logarithmic transformation of the response is appropriate.

Employing forward selection, Lin [61] identified 11 factors as active with $R^2 =$ 99%. The first eight of these (118, 25, 129, 13, 91, 93, 86, and 76) were selected for further study. Using the log transformed data, we now compare these results with those obtained via the proposed analysis strategy. All-subsets regression up to $m = \lfloor min\{k/5, n/2\} \rfloor = \lfloor min\{138/5, 24/2\} \rfloor = 12$ with 138 factors is formidable even amidst modern computing power. Therefore, for the sake of time and simplicity, the results of an all-subsets regression for $m = 5$ is shown in Table 3.19. This choice of m will allow us to make a reasonable enough comparison with Lin's results.

Given the large number of factors under consideration, the global model test

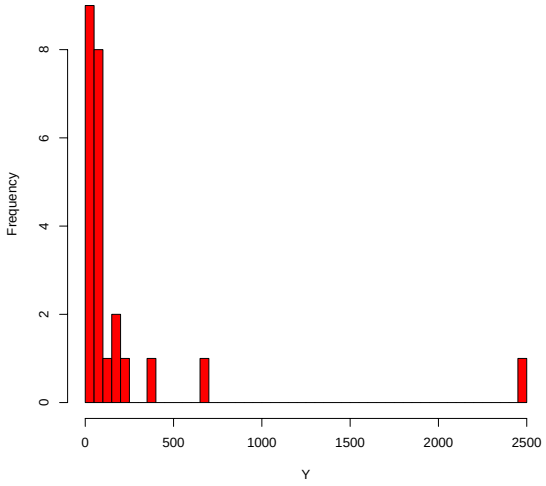

Figure 3.4: Histogram of Untransformed AIDS Data

Table 3.19: Analysis of AIDS Data - Global Test

Subset					AIC_c	$P(R^{2(b)} \geq R^2)$
118					6.408	0.194
63					8.962	0.531
39					9.656	0.688
87					11.364	0.964
3					12.823	1
87	118				-1.249	0.138
66	118				2.235	0.444
74	118				2.41	0.453
40	118				2.822	0.522
18	63				3.432	0.565
55	87	118			-6.752	0.211
6	87	118			-6.665	0.238
85	87	118			-5.298	0.344
25	87	118			-5.194	0.35
87	118	121			-4.446	0.405
18	58	63	105		-13.89	0.256
3	75	85	87		-13.561	0.247
6	87	118	121		-12.607	0.293
55	87	103	118		-12.118	0.367
87	103	118	130		-12.109	0.346
3	48	75	85	87	-21.241	0.252
6	63	66	118	130	-20.712	0.262
18	58	63	105	120	-20.392	0.287
6	87	94	98	118	-20.228	0.281
87	103	114	118	130	-19.717	0.307

is performed using forward selection instead of all-subsets as suggested in section 2.3 since forward selection provides a lower bound for the global p-value using all-subsets. All-subsets may then be performed on those models that reject the global null hypothesis if desired. From Table 3.19, we see that none of the models found using all-subsets reject the global null hypothesis. Furthermore, note that only factor 118 is common with Lin's analysis.

Although we are unable to say that no model is appropriate for the AIDS data, the candidate models in Table 3.19 do not appear worthy of further investigation. It is unlikely, then, that we can learn anything useful from the AIDS data using this particular SSD, especially if many factors are active or interactions are present. Based on this, the results of Lin [61] are questionable.

3.4 Discussion

In this chapter, we have proposed an analysis strategy for SSDs based on all-subsets regression and permutation tests. By assessing the null model performance of all-subsets vs. forward selection, we see that all-subsets regression outperforms forward selection and thus, we recommend its use even while considering the added computational burden. That is, with the ever increasing amount of computing power available, performing all-subsets regression is becoming more feasible for a moderate to large number of factors (say, $k \leq 100$). However, with the use of permutation tests, it potentially becomes necessary to perform all-subsets regression many times. We presented several suggestions for such use of all-subsets. For instance, one may consider running all-subsets up to some predetermined number of factors. Another solution, at the cost of precision, is to reduce the number of simulations performed. We combine both of the above suggestions in this chapter to ease computational requirements and

believe this to be a reasonable choice.

If the loss of precision due to the reduction in the number of all-subsets performed is too great, one may consider a different approach based on sampling theory. Cochrane [25] and Lohr [63] discuss the use of regression estimation in sampling to estimate the mean of some population, call it Y, based on the mean of another population, X. Suppose that y_i and x_i are each obtained for every unit in a sample and that $\mathbb{E}(\mathrm{X})$, the population mean for X is known. Then the regression estimator of $\mathbb{E}(\mathrm{Y})$ is the predicted value of Y from the fitted regression model when X=$\mathbb{E}$(X):

$$\widehat{\mathbb{E}(\mathrm{Y})} = \bar{y} + b(\mathbb{E}(\mathrm{X}) - \bar{x}), \tag{3.17}$$

where b is the ordinary least squares slope. The general situation in which such estimation is helpful is when one can make a rapid estimate, x_i, of some characteristic for every unit and can also, by some more costly method, determine the correct value, y_i, of the characteristic for a simple random sample of the units. For our case, suppose that given its computational ease, we run forward selection on 1000 permuted data sets and compute R^2 for each selected model. Now, suppose computational feasibility allows us to run all-subsets on 50 randomly selected data sets from the 1000 and compute R^2 for each model chosen. From the sample of 50, we can compute $\bar{R}^2{}_{all-subsets}$ and $\bar{R}^2{}_{forward}$, the mean R^2 from all-subsets and forward selection, respectively. Then, using the mean of all 1000 R^2s as $\mathbb{E}(R^2_{forward})$, we obtain an estimate of $\mathbb{E}(R^2_{all-subsets})$ using the above equation.

Of course, one is not often interested in the average R^2 produced via all-subsets. Rather, we would like to obtain an estimate of the proportion of R^2s that exceed that of the model based on the original data. This idea does not appear to have been treated in the literature and is worthy of future research. Rueda, Arcos, Martinez-

Miranda, and Roman [87] discuss estimation of quantiles using regression estimates. This may provide a stepping stone for such research.

Permutation tests provide an intuitive approach to model selection in the context of SSDs and allow for fewer assumptions than the usual normal theory tests. We have shown how a global model test can be performed using permutation tests to indicate how "remarkable" a particular model appears to be. The use of such a test can serve as an initial screening of candidate models. That is, those models that fail to reject the global null hypothesis need not be considered further.

The use of standard model selection criterion such as AIC tend to overfit when using forward selection or all-subsets regression. However, we have seen that even the bias-corrected AIC_c also suffers from this problem. Thus, we have introduced a new criterion, ΔAIC_c, that combines AIC_c with permutation tests in order to help prevent such overfitting. In particular, ΔAIC_c measures the deviation of AIC_c for a particular model from the average AIC_c based on the permutation of residuals of a nested model. Our examples indicate that this criterion is indeed successful in its endeavor to prevent overfitting and thereby helps to further reduce the number of candidate models.

Recall that Westfall, Young, and Lin [93] (WYL) introduce the use of adjusted p-values for better control of Type I error rates in forward selection. Such control comes with the cost of a substantial reduction in power due to premature stopping of the forward selection procedure. The adjusted p-values are computed by generating residuals from a $N(0, 1)$ distribution instead of permuting residuals from a reduced model. We have seen that the p-values computed for tests of partial regression by permutation procedures are actually very close in value to WYLs p-values. WYL also interestingly point out that the "ordinary Bonferroni method provides a very accurate approximation to the true critical level of the multiple significance tests". We verify

this to be true and note that the Bonferroni method also does well to approximate the permutation p-values for partial tests. As WYL mention, this fact may prove to be very useful for the practitioner, since it is easy to implement. Our proposed analysis strategy, however, is less concerned with Type I error control since the use of SSDs is mainly exploratory in nature. That is, we believe that follow-up experimentation is expected in order to draw appropriate conclusions regarding those factors that most impact the response of interest.

It is also worthy of mention that our proposed analysis strategy could be combined with any method for analysis of SSDs. That is, instead of all-subsets regression, the permutation procedures presented could easily be applied to forward selection, Bayesian variable selection, the Dantzig Selector, genetic algorithms, etc. Therefore, even though we suggest the use of all-subsets regression, we believe the proposed permutation procedures will assist in clarifying the significance of factors chosen by any SSD analysis procedure. Although we provide a strategy for ascertaining significance of factors for SSDs, the analysis of SSDs continues to be an open and viable area for further research.

Chapter 4

A New Class of Designs for Factor Screening and Response Surface Exploration

4.1 Introduction and Motivation

Cheng and Wu [22] (henceforth CW) introduce a novel method for exploring a response surface using only one design. This is in contrast to the common practice in response surface optimization, which, as discussed in Chapter 1, is to use a sequential experimentation process that consists of beginning with a screening experiment in order to identify important factors, moving to a new region using steepest ascent if the initial experimental region proved to be inadequate, and finally fitting a response surface model using a second-order design. Instead, CW proposes to use 3-level regular or nonregular designs (as opposed to the usual 2-level designs) in order to first screen a large number of factors and then project from the larger factor space onto a smaller factor space to perform response surface exploration. Designs with more than two-levels are necessary in order to fit a second-order model with quadratic curvature. That is, assuming that all factors are quantitative and are denoted by $x_1, x_2, ..., x_t$,

the second-order model is given by

$$y = \beta_0 + \sum_{i=1}^{t} \beta_i x_i + \sum_{i=1}^{t} \beta_{ii} x_i^2 + \sum_{i<j}^{t} \beta_{ij} x_i x_j + \epsilon, \tag{4.1}$$

where β_i are the linear effects, β_{ii} are the quadratic effects, β_{ij} are the linear-by-linear interactions, and ϵ is the error term. The one-step approach to RSM is also described in Lawson [52].

While the sequential nature of RSM is usually viewed as advantageous, as it gives the experimenter an opportunity to learn from each experiment, CW mention the case in which experimental preparation is time-consuming or its duration long as a disadvantage to the sequential framework. For instance, running experiments on a production line may require a change of work schedule, training of operators, and/or trial runs.

In the first stage of CW's method, factor screening is performed to identify which of a potentially large number of factors are active. This step usually involves a main effects only analysis, which can have the unfortunate consequence of missing important interactions and can lead to a misspecification of the response surface. In CW, interactions are considered only in the projected design space. Bingham [10] comments that because the first step considers main effects only, an underlying assumption is that every factor affecting the response has a significant main effect and thus, one is assuming that all factors with significant interaction effects also have significant main effects (this is known as *strong effect heredity*). Bingham [10] further comments that "while it is convenient to perform the first stage of the analysis ignoring interactions, it seems unrealistic to expect that the strong heredity assumption will apply in most applications". However, strong effect heredity does not seem unreasonable with a main effects only analysis that considers both linear and quadratic main effects. In

this case, strong effect heredity requires only that the linear or quadratic main effect for a factor involved in a two-factor interaction not be zero. However, CW ignore the bias caused by interactions in the first step. If this bias results in missing an active effect, the loss is substantial.

A key step in CW's method is the projection step (stage 2). This provides the link between the screening stage and the response surface exploration. In particular, in order to sufficiently explore the response surface in the projected factors, the projected design needs to be a second-order design. CW terms such projected designs to be *eligible*. Otherwise, the projection is said to be *ineligible*. A useful result from CW is that if a p-factor projected design of a design with t factors is ineligible, then the t-factor design is also ineligible. This result assists in concluding that any regular three-level fractional factorial design with resolution III is ineligible since it contains one nine-run three-factor projected design, which is ineligible since a second-order model in three factors has ten parameters. The possibility that a projection for a desired set of factors may be ineligible is another drawback of CW's method. Consider the following example from CW.

Example 1

The experiment is a 27-run experiment to study the PVC insulation for electric wire with nine continuous factors denoted by A-H and J. The design matrix and response (denoted as Response1) are given in Table 4.1. This design is a regular 3^{9-6} with $C = AB$, $D = A^2B$, $F = AE$, $G = A^2E$, $H = B^2E$, and $J = AB^2E$.

In the screening step of CW's analysis of Table 4.1's Response1 data, factors *A, B, C, D* and *G* are identified as significant. The authors add that factors *A* and *B* are more significant than the other three identified factors. This is easily verified by their respective p-values of less than 0.0001. The second stage of the analysis would

Table 4.1: Example from CW

Run	A	B	C	D	E	F	G	H	J	Response1	Response2
1	0	0	0	0	0	0	0	0	0	5	-13
2	0	0	0	0	1	1	1	1	1	2	-16.9
3	0	0	0	0	2	2	2	2	2	8	-23.2
4	0	1	1	1	0	0	0	2	2	-15	0.054
5	0	1	1	1	1	1	1	0	0	-6	-0.84
6	0	1	1	1	2	2	2	1	1	-10	0.216
7	0	2	2	2	0	0	0	1	1	-28	12.42
8	0	2	2	2	1	1	1	2	2	-19	16.5
9	0	2	2	2	2	2	2	0	0	-23	23.2
10	1	0	1	2	0	1	2	0	1	-13	-8.11
11	1	0	1	2	1	2	0	1	2	-17	-9.58
12	1	0	1	2	2	0	1	2	0	-7	-8.86
13	1	1	2	0	0	1	2	2	0	-23	13.58
14	1	1	2	0	1	2	0	0	1	-31	1.305
15	1	1	2	0	2	0	1	1	2	-23	6.835
16	1	2	0	1	0	1	2	1	2	-34	-2.49
17	1	2	0	1	1	2	0	2	0	-37	7.361
18	1	2	0	1	2	0	1	0	1	-29	2.414
19	2	0	2	1	0	2	1	0	2	-27	-2
20	2	0	2	1	1	0	2	1	0	-27	3.69
21	2	0	2	1	2	1	0	2	1	-30	-7.41
22	2	1	0	2	0	2	1	2	1	-35	-7.52
23	2	1	0	2	1	0	2	0	2	-35	-12.5
24	2	1	0	2	2	1	0	1	0	-38	-0.84
25	2	2	1	0	0	2	1	1	0	-39	9.75
26	2	2	1	0	1	0	2	2	1	-40	11.09
27	2	2	1	0	2	1	0	0	2	-41	9.663

then be to fit a second-order model in the five active factors. However, there is no eligible projected design of five factors in the 3^{9-6} design. Recall that $C = AB$ and $D = A^2B$. Thus, C and D cannot be considered for projection once A and B are chosen. CW then proceeds to find an eligible projection in the three most significant factors (A,B, and G). By not considering C and D in subsequent analyses, it is possible that important interactions involving these factors will be missed, and that some estimates for effects in our model will be biased, as the following example illustrates. □

Example 2

Reconsider Example 1. Since all of the factors are quantitative, we consider only the linear-by-linear interaction effects. Suppose that we simulate a new response variable from the following simple model:

$$y = 10x_B + 8x_C + 6x_{CG} + \epsilon. \tag{4.2}$$

where $\epsilon \sim N(0, 1)$. Note that the model does not satisfy strong effect heredity. The simulated data is shown as Response2 in Table 4.1. If we fit a model with all nine main effects and pure quadratic terms, we obtain the ANOVA table shown in Table 4.2. Clearly, factors B, C, and J stand out as active. Nothing else can be identified as significant. We, therefore, proceed to project the 3^{9-6} design onto B, C, and J (which is eligible) and fit a second-order model. The results of this fit are shown in Table 4.3. From this, we see no new effects declared as active. Thus, CW's analysis strategy identified an insignificant main effect as active and failed to identify the important interaction. □

Examples 1 and 2 have helped to illustrate two problems with CW's method.

Table 4.2: Simulated Example using CW's Design - Pure Quadratic Model

Term	Estimate	Std Error	t Ratio	P-value
Intercept	-1.387519	4.076913	-0.34	0.7424
A	0.3040556	1.145514	0.27	0.7974
B	9.7376667	1.145514	8.5	< 0.0001
C	7.4886111	1.145514	6.54	0.0002
D	0.3103889	1.145514	0.27	0.7933
E	-0.037	1.145514	-0.03	0.9750
F	-0.145056	1.145514	-0.13	0.9024
G	0.3057222	1.145514	0.27	0.7963
H	0.0812778	1.145514	0.07	0.9452
J	-2.819944	1.145514	-2.46	0.0392
A*A	-0.140944	1.984088	-0.07	0.9451
B*B	0.2198889	1.984088	0.11	0.9145
C*C	-0.295611	1.984088	-0.15	0.8852
D*D	0.1023889	1.984088	0.05	0.9601
E*E	0.2472222	1.984088	0.12	0.9039
F*F	-0.257278	1.984088	-0.13	0.9000
G*G	0.3717222	1.984088	0.19	0.8560
H*H	-0.248611	1.984088	-0.13	0.9034
J*J	2.3507222	1.984088	1.18	0.2701

Table 4.3: Simulated Example using CW's Design - Second-Order Model

Term	Estimate	Std Error	t Ratio	P-value
Intercept	-1.339906	1.670014	-0.8	0.4334
B	9.7328584	0.773066	12.59	$< .0001$
C	7.4841753	0.773066	9.68	$< .0001$
J	-2.816217	0.773066	-3.64	0.0020
B*B	0.224221	1.33899	0.17	0.8690
B*C	-0.206054	0.946809	-0.22	0.8303
C*C	-0.292266	1.33899	-0.22	0.8298
B*J	-0.007785	0.946809	-0.01	0.9935
C*J	1.0432398	0.946809	1.1	0.2859
J*J	2.3486659	1.33899	1.75	0.0974

In particular, with the usual screening design scenario with many factors in a small number of runs, the initial screening stage may not be able to entertain the possibility of two-factor interactions forcing one to potentially miss important effects that will not be considered in the second-order model. Furthermore, and as seen in Example 1, the projection onto the factors of interest does not always yield a second-order design. That is, the projected design is ineligible. These issues would certainly be problematic to the experimenter seeking to implement CW's method.

CW propose a *projection efficiency* criterion in order to help find suitable designs for the dual purposes of factor screening and interaction detection. They investigate regular 27-run orthogonal designs, projections of a nonregular orthogonal array for 13 factors in 27 runs, denoted as OA(27,3^{13}), an OA(18,3^7), and an OA(36,3^{12}). Their criterion is given as follows:

1. The number of eligible projected designs should be large and lower-dimensional projections are more important than higher-dimensional projections.

2. Among the eligible projected designs, the estimation efficiency should be high. One such measure of efficiency is D-efficiency given by

$$D_{eff} = \left(|\mathbf{M}(\mathbf{d})| \,/\, |\mathbf{M}(\mathbf{d}^*)|\right)^{1/p} \tag{4.3}$$

   where $\mathbf{M}$ is the moment matrix defined as $\mathbf{X}'\mathbf{X}/n$, $|\mathbf{M}(\mathbf{d}^*)| = max_d\,|\mathbf{M}(\mathbf{d})|$, and $\mathbf{X}$ denotes the $n \times p$ second-order model matrix for a t-factor design. That is, $\mathbf{d}^*$ is the D-optimum continuous design for the second-order model given above. Note that $p = (t+1)(t+2)/2$.

CW's justification for the projection efficiency criterion is based on the *factor sparsity* principle (i.e., the number of relatively important factors is small, say 20% of the factors are active) and thus deem an eligible projection for a lower dimension to be

more important than that for a higher dimension since a lower-dimensional projected design causes any higher-dimensional projected designs that contain it to be ineligible. Furthermore, they state that if a design has fewer ineligible projected designs on lower dimensions, it should have a better chance of getting fewer ineligible projected designs on higher dimensions. While the factor sparsity assumption provides a reasonable rationalization for the projection efficiency criterion, it is likely that the *effect sparsity* assumption, which states that the number of important effects (rather than factors) is relatively small, is more appropriate when many factors are under consideration. That is, the assumption of effect sparsity may still hold even when factor sparsity does not. Therefore, it makes sense to search for designs that can screen and project onto more than just a few factors.

Xu, Cheng, and Wu [97] (henceforth XCW) further propose an optimality criterion, known as *projection aberration*, to assess the performance of projections of three combinatorially non-isomorphic OA(18,3^7)s and three combinatorially non-isomorphic OA(27,3^{13})s. The 18 and 27-run OAs can screen up to 7 and 13 factors, respectively. Their criterion is based on the generalized word-length pattern of Xu and Wu [98],$(A_1, A_2, ..., A_t)$, where A_i serves to measure the overall aliasing between all i-factor effects and the general mean and is defined to be

$$A_i = n^{-2} \sum_{j=1}^{t_i} \left| \sum_{h=1}^{n} x_{hj}^{(i)} \right|^2 \tag{4.4}$$

where t_i is the number of all i-factor effect contrasts and $x_{hj}^{(i)}$ is the hth component of the jth-factor effect contrast.

When a design with t factors is projected onto any three factors, it produces $\binom{t}{3}$ three-factor projected designs. Each of these designs has an A_3 value, known as the *projected* A_3 value. XCW comment that for an OA with smaller A_3, its main

effects suffer less contamination when a main effects model is fitted, and thus factor screening is more effective as long as strong effect heredity is present. As alluded to above, there is a close connection between projected A_3 values and eligibility. That is, the presence of projections with three-letter words causes low projection efficiency. The frequency of the projected A_3 values is called the *projection frequency*.

The projection aberration criterion sequentially minimizes the projection frequency starting from the largest projected A_3 value. XCW utilize this criterion to screen out poor OA and then considered designs obtained by level permutations from remaining OA's using the projection efficiency criterion of CW. They recommend an 18 and 27-run OA, which we will examine in more detail later. Unfortunately, these designs still suffer from the shortcomings illustrated in Examples 1 and 2 above. For instance, consider the following example of an 18-run 7-factor OA recommended by XCW.

Example 3

The 18-run 7-factor OA mentioned above is shown in Table 4.4. The response, y, is simulated from the following model (which we suppose is unknown to the experimenter):

$$y = 10x_D + 9x_E + 8x_G - 6x_{AD} - 6x_{AG} + 6x_{D^2} + \epsilon. \tag{4.5}$$

The following polynomial model,

$$y = \beta_0 + \sum_{i=1}^{k} (\beta_i x_i + \beta_{ii} x_i^2) \tag{4.6}$$

is fit to the data in Table 4.4 and the results are shown in Table 4.5. Based on Table 4.5, the linear main effects, B, D, E, F, and G are identified as significant using $\alpha = 0.1$. This is certainly problematic since the only linear-by-linear interactions that

Table 4.4: 18-run OA from XCW

A	B	C	D	E	F	G	y
-1	-1	-1	-1	-1	-1	-1	-29.830996
-1	0	0	0	0	0	0	-0.8688412
-1	1	1	1	1	1	1	50.0542329
0	-1	-1	0	0	1	1	7.89254832
0	0	0	1	1	-1	-1	11.24119
0	1	1	-1	-1	0	0	-14.512017
1	-1	0	-1	1	0	1	7.27908994
1	0	1	0	-1	1	-1	-13.269653
1	1	-1	1	0	-1	0	9.9487719
-1	-1	1	1	0	0	-1	3.96702798
-1	0	-1	-1	1	1	0	-0.2854042
-1	1	0	0	-1	-1	1	4.96974603
0	-1	0	1	-1	1	0	7.77942863
0	0	1	-1	0	-1	1	-0.4972899
0	1	-1	0	1	0	-1	0.94619691
1	-1	1	0	1	-1	0	9.19619749
1	0	-1	1	-1	0	1	9.33056596
1	1	0	-1	0	1	-1	5.32400436

Table 4.5: Example 3 Results

Term	Estimate	Std Error	t Ratio	P-value
Intercept	-5.17863	4.921967	-1.05	0.3700
A	-0.016399	1.556463	-0.01	0.9923
B	4.2039698	1.556463	2.7	0.0737
C	3.0780679	1.556463	1.98	0.1424
D	10.403653	1.556463	6.68	0.0068
E	9.4970357	1.556463	6.1	0.0088
F	4.3722948	1.556463	2.81	0.0673
G	8.3875935	1.556463	5.39	0.0125
A*A	2.5095525	2.695872	0.93	0.4206
B*B	4.3094247	2.695872	1.6	0.2082
C*C	-3.209088	2.695872	-1.19	0.3195
D*D	3.5055179	2.695872	1.3	0.2844
E*E	-0.719489	2.695872	-0.27	0.8069
F*F	4.1865611	2.695872	1.55	0.2183
G*G	2.9075328	2.695872	1.08	0.3598

are truly active involve factor A. Thus, if one follows the analysis plan of CW, we lose important information that could have been identified if detecting the presence of two-factor interactions was plausible using the OA's of XCW.

Furthermore, there is no clear-cut strategy to handle the need to project the 18-run OA onto 5-factors as deemed necessary by Table 4.5. Since no such projection is eligible, the experimenter would be forced to choose the most significant factors to project onto and again, no such choice involves factor A. Thus, if a second-order model is fit in the "most significant" factors, then both spurious main effects and two-factor interactions may be identified. Note that factors D, E, and G are significant at $\alpha = 0.05$ and the projection of the 18-run OA onto these three factors is eligible. A second-order model in these three factors only identifies the D, E, and G main effects as active and fails to recognize the pure quadratic term involving factor D. □

Ye, Tsai, and Li [100] also select 18-run and 27-run designs for factor screening and response surface exploration. As with other authors, their purpose is to choose designs that are useful for projection. Their choices of designs are based on both model estimation and model discrimination criteria. In particular, they make use of two of six non-Bayesian criterion proposed by Jones, Li, Nachtsheim, and Ye [47], which are *Average Expected Prediction Difference*

$$AEPD = \frac{1}{r} \sum_{1 \leq i < j \leq r} \mathbb{E}(\|\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_j\| \,|\, ||\mathbf{y} = 1\|) \tag{4.7}$$

and *Minimum Maximum Prediction Difference*

$$MMPD = min_{1 \leq i < j \leq r} max_{\|\mathbf{y}=1\|} \|\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_j\| \tag{4.8}$$

where r is the number of models, $\mathbf{y}$ is the response vector, and $\hat{\mathbf{y}}_i$ is the fitted value of the ith model.

It should be made clear that we do not dispute the ability of XCW's or Ye, Tsai, and Li [100]'s criteria to find good designs for projection. However, as the above examples illustrate, unless one can correctly identify the important factors to explore further, the projection properties of a design become of little importance. XCW briefly mentions this drawback and suggests the "more elaborate" Bayesian procedure for factor screening in the first stage. See Box and Meyer [17] and Chipman, Hamada, and Wu [23] for more on this approach. Although this approach is advocated in Ye, Tsai, and Li [100], it does have the drawback of being more computationally intensive and requires specification of prior information for possible models.

Thus, in this chapter, we propose a new class of three-level designs and an analysis strategy that provides the experimenter with the opportunity to explore the presence of interactions in stage 1 and therefore be more appropriately advised to the choice of active factors. Once the appropriate factors are identified, these t-factor designs will allow for projections of up to $t-1$ factors. In section 2, we introduce the new designs, discuss their construction method, and compares them with existing designs based on established design criterion as well as a simulation study. Section 3 proposes an analysis strategy based on the new designs and provides several illustrative examples. Section 4 concludes the chapter with a discussion and suggestions for future research.

4.2 New Three-Level Designs for Screening and Response Surface Exploration

4.2.1 Motivation

In this section, we assume a spherical design region, which will allow for response surface exploration with wider ranges for each factor. In other words, we intend to

suggest initial designs that are spread out in order to fill a large region of interest with the intent of detecting big effects. While narrow spacing is appropriate when the initial design is followed by steepest ascent, here we need a larger range for each factor. We assume that the response surface on the experimental region is appropriate for studying second-order curvature.

Mee [69] explains how to utilize optimal design algorithms to construct three-level second order designs for spherical regions. In particular, D-optimal designs of this type are constructed for various run sizes using an appropriately chosen candidate set of points from a specific orbit of points the same distance from the center. For instance, with a second-order design in two factors, the candidate set consists of two orbits with four points each. The inner orbit contains all face centered axial points while the outer orbit contains the factorial corner points. In general, for t factors, there are t orbits with $\binom{t}{k}2^k$ points in the kth orbit $(k = 1, 2, ..., t)$.

For spherical regions, D-efficiencies are calculated using the formula

$$D - eff = \left[\frac{|\mathbf{X}'\mathbf{X}/n|}{D_\infty}\right]^{1/p}, \tag{4.9}$$

where

$$D_\infty = 2^t(t+1)^{-p}(t+2)^{-t(t+2)}(t+3)^{p-1} \tag{4.10}$$

and, for calculation purposes, we require the t-factor design matrix $\mathbf{D}$ to be constrained (scaled) to the unit hypersphere (i.e. all diagonal elements of $\mathbf{DD}' \leq 1$). This scaling is also necessary to compare designs with points from different orbits.

The designs presented in this section are motivated from an initial optimal design search from a specified orbit of points using SAS's PROC OPTEX. The intent of such a search was to seek out structure in D-optimal designs of an economical run size for the purposes of screening and interaction detection. Since the run sizes

chosen were less than the number of parameters in a second-order model (in order to be economical), it was necessary to utilize the Bayesian formulation of DuMouchel and Jones [29] in the search algorithm. To be specific, we partition the terms in the linear model, $y = \mathbf{X}\beta$ into two groups,

$$y = \mathbf{X}_{pri}\beta_{pri} + \mathbf{X}_{pot}\beta_{pot} \tag{4.11}$$

where β_{pri} denotes the q primary parameters, which are believed to be required in the model and β_{pot} denotes the s potential parameters. The Bayesian D-optimal design maximizes the determinant of

$$\mathbf{M}_\alpha(\mathbf{d}) = (1-\alpha)K + \alpha\mathbf{M}(\mathbf{d}) \tag{4.12}$$

where

$$\alpha = \frac{n}{n_0 + n} = \frac{n\tau^2}{\sigma^2 + n\tau^2} \tag{4.13}$$

and $n_0 = \sigma^2/\tau^2 = n(1-\alpha)/\alpha$, σ^2 is the error variance, τ^2 is a small positive value, and

$$K = \begin{bmatrix} \mathbf{0}_{q\times q} & \mathbf{0}_{q\times s} \\ \mathbf{0}_{q\times s} & I_s \end{bmatrix}. \tag{4.14}$$

DuMouchel and Jones [29] state that σ^2=1 and τ^2=1 are reasonable choices, although they illustrate the sensitivity of the Bayesian D-optimal designs to different choices of τ^2. As $\alpha \to 1$, the design tends to the D-optimum design when the model with all $r + s$ parameters is of interest (assuming $rank(\mathbf{X}) \leq n$). On the other hand, as $\alpha \to 0$, the design tends to the D-optimum designs for the model with just the primary terms.

The OPTEX procedure allows specification of prior information for the potential

terms by asking the user to input the value of n_0, which serves to represent the number of runs of prior information we have concerning the potential terms. In our search, the linear and quadratic main effects are denoted as the primary terms while the linear-by-linear interactions are the potential terms. For the search, we specify varying values for n_0 ranging from 1 to $\frac{1}{2}n$ to represent almost negligible to much larger amounts of prior information available. By default, OPTEX prints

$$D - eff = \frac{|\mathbf{X}'\mathbf{X}|^{1/p}}{n} \tag{4.15}$$

or in the case of a Bayesian D-optimal search,

$$D - eff = \frac{\left|\mathbf{X}'\mathbf{X} + \frac{1}{\tau^2}K\right|^{1/p}}{n}. \tag{4.16}$$

Although we typically compute efficiencies relative to the optimum continuous design as specified in (4.9) above, maximizing (4.15) or (4.16) is equivalent to optimizing (4.9) since these measures are monotonically related to $|\mathbf{X}'\mathbf{X}|$. See Atkinson, Donev, and Tobias [7] for more. In what follows, when referring to OPTEX results, we make use of the (4.16) efficiency. The orbit of candidate points utilized was largely based on the recommendations of Mee [69]. The run sizes chosen for each search were based on the orbit chosen, economy, and "hope" for structure. For instance, n=21 was chosen for t=6 factors involving the full 3rd orbit since there are $\binom{6}{3}$=21 triples of factors. Other run size choices were simply chosen based on economy alone.

Table 4.6 lists a selection of Bayesian D-optimal searches that were conducted using OPTEX and $n_0 = \frac{1}{2}n$. We omit details of searches using other choices of n_0. Unfortunately, none of the designs (with the exception of one for t=7 factors and $n_0 = \frac{1}{2}n$) yielded any recognizable and/or useful structure. This motivated a new search that was conducted based on using subsets of an orbit as candidate runs

rather than the entire orbit.

Mee [69] discusses Box-Behnken (BB) designs (Box and Behnken [14]) for estimating second-order models. For 3-5 factors, the BB designs involve all points from the second orbit whereas for more factors, these designs contain just a subset of points from the tth orbit, where $t > 2$. BB designs have a nice construction method and are based on incomplete block designs (IBD). Before proceeding, the following IBD notation will be useful:

- t: number of factors,
- k: block size,
- b: number of blocks,
- r: number of replicates of each of the k factors,
- $\lambda = r(k-1)/(t-1)$.

If possible, the BB designs are based on balanced IBDs (BIBD) where pairs of factors occur together within a block λ times. However, for cases where a BIBD does not exist or where run sizes become too large with a BIBD, regular graph IBDs (RG) are used in which all pairs of treatment levels occur together within a block either $\lambda_1 = \lceil \lambda \rceil$ or $\lambda_2 = \lfloor \lambda \rfloor$ blocks. Table 4.7 illustrates a 6-factor BB design, which is taken

Table 4.6: First Bayesian D-optimal Search using OPTEX

Number of Factors	Run Size	Orbit	D-efficiency
6	21	3	83.71
7	29	3	86.91
8	41	3	90.13
8	41	4	91.24
9	25	4	62.79
9	49	3	88.62
10	61	3	88.40

as a subset of points from the 3rd orbit and is based on a RG design with $t = 3$, $\lambda_1 = 2$, and λ_2=1. In the table, the symbol $(\pm 1, \pm 1, ..., \pm 1)$ means that all combinations of plus and minus levels for a full factorial in t factors are to be run (e.g., for t=3, a full 2^3 factorial). Thus, with the inclusion of center-point runs, the 6-factor BB design consists of 49 distinct points. Myers and Montgomery [75] discuss BB designs and state that they are "an efficient option and indeed an important alternative to the central composite design". However, as the number of factors increases, both the central composite design and BB design have too many runs to be practical for factor screening. For instance, the 7-factor BB design has 56 runs (plus center point runs) while Montgomery [72] recommends a minimum of 80 runs (including center point runs) for a central composite design.

The second Bayesian D-optimal search mentioned above was based on using BB-type designs as the candidate set of points. Again, the search identified linear and quadratic main effects as primary terms and linear-by-linear interactions as potential terms. Consider again the 6 factor, 49 run design shown in Table 4.7, consisting of eight full 2^3 factorials in different sets of three factors. Therefore, it makes sense to search for a D-optimal subset of this design in 25 runs so as to allow for the possibility of six blocks of size four (plus a center-point run) with some easily identifiable structure. For brevity, Table 4.8 summarizes only a subset of this second search with

Table 4.7: Box-Behnken Design for 6 factors

A	B	C	D	E	F
± 1	± 1	0	± 1	0	0
0	± 1	± 1	0	± 1	0
0	0	± 1	± 1	0	± 1
± 1	0	0	± 1	± 1	0
0	± 1	0	0	± 1	± 1
± 1	0	± 1	0	0	± 1
0	0	0	0	0	0

$n_0 = \frac{1}{2}n$ so as to compare them with a couple of the Table 4.6 designs. It is worthy to note that the choice of n_0 had little to no effect on the D-optimal designs generated using a more restrictive candidate set of points.

Only two designs from Table 4.6 can be compared to those from Table 4.8 due to their runs sizes being the same: 7 factors in 29 runs and 9 factors in 49 runs. In both cases, only a moderate loss of efficiency occurs by using the more restrictive candidate set of points. For instance, the Bayesian D-efficiency of the 7 factor, 29 run design using the entire third orbit is 86.91 while that of the 7 factor, 29 run design using only a subset of the third orbit is 81.97. What is most revealing is that this second search did produce designs with structure and is the motivation for the new designs that we now present.

4.2.2 Resolution III Subsets of Four Runs (Fractional Box-Behnken Designs)

The Bayesian D-optimal designs that resulted from the restricted candidate set for t=6 factors in 25 runs, 7 factors in 29 runs, and 9 factors in 37 runs consisted of t Resolution III blocks with run size of 4. Although other resulting structures were found for different cases and will be mentioned briefly later, the Resolution III structure is straightforward to construct and will be the main focus of this chapter. We will see later that these designs are useful for both factor screening and interaction detection

Table 4.8: Subset of Second Bayesian D-optimal Search using OPTEX

Number of Factors	Run Size	Candidate Set	D-efficiency
6	25	RG: 6×2^3	84.86
7	29	BIBD: 7×2^3	81.97
9	37	BIBD: 12×2^3	72.07
9	49	BIBD: 12×2^3	84.93

since they possess a much simpler aliasing structure than the more commonly used OA's. Therefore, we propose to utilize this structure in order to construct a class of three-level designs of various run sizes. Let us begin by examining the 7-factor, 29 run design as produced by OPTEX and shown in Table 4.9. In this case, the symbol $(\pm 1, \pm 1, ..., \pm 1)$ represents a 2^{3-1}_{III}. We take a moment now to compare this design with a 18-run OA in 7 factors and a 27-run OA in 7 factors.

In order to compare these designs in terms of precision for estimating linear and quadratic main effects, we rescale the levels so that the points lie inside or on the boundary of the assumed spherical design region (i.e., the unit sphere). To do this, simply divide each run by the largest distance from the design center. For the 18-run OA, after rescaling, we see that $\mathbb{V}(b_i) = 0.58\bar{3}\sigma^2$ and $\mathbb{V}(b_{ii}) = 12.25\sigma^2$ whereas for the 29-run design in Table 4.9, we have that $\mathbb{V}(b_i) = 0.25\sigma^2$ and $\mathbb{V}(b_{ii}) = 2\sigma^2$. Finally, for the 27-run, 7 factor OA (taken as a projection from the 27-run, 13 factor OA), $\mathbb{V}(b_i) = 0.38\bar{\sigma}^2$ and $\mathbb{V}(b_{ii}) = 8.1\bar{6}\sigma^2$. Then, we see that, assuming no interactions, the new 29-run design outperforms the 18 and 27-run OAs in spherical regions in terms of precision for estimating both linear and quadratic main effects. Figures 4.1, 4.2 and Table 4.10 compares the aliasing of the three 7-factor designs with factors denoted by the letters A-G. It is easily seen that both of the OA's have much more complex aliasing structures than does the new 29-run design.

Table 4.9: Bayesian D-optimal Design for 7 factors in 29 runs

A	B	C	D	E	F	G	Design Generator
±1	0	0	0	0	±1	±1	G=AF
±1	0	±1	0	±1	0	0	E=AC
±1	±1	0	±1	0	0	0	D=AB
0	0	0	±1	±1	±1	0	F=DE
0	0	±1	±1	0	0	±1	G=CD
0	±1	0	0	±1	0	±1	G=BE
0	±1	±1	0	0	±1	0	F=BC

Intercept = - 0.5*A - 0.5*C + 0.5*D + 0.5*A*A + A*C + 0.5*C*C - A*D - C*D + 0.5*D*D
= - 0.5*A - 0.5*B - 0.5*C - F + 0.5*A*A + 1.5*B*B + A*C - 0.5*C*C - A*D - C*D - 2*A*E
= 0.16667*A - 0.16667*B - 0.16667*C + 0.33333*D + 0.33333*E - 0.66667*F + 0.33333*G + 1.16667*A*A
+ 0.16667*B*B + 0.33333*A*C + 0.66667*B*C + 0.16667*C*C - 0.33333*A*D - 0.66667*B*D
- 0.33333*C*D + 0.66667*B*E - 0.66667*C*E
= - 0.5*A + 0.3*B - 0.1*C + 0.8*D + 0.4*E - 0.6*F - 0.4*G + 0.5*A*A + 0.4*A*B
+ 0.3*B*B + A*C + 1.2*B*C + 0.7*C*C - A*D - C*D - 1.2*D*E
= - 0.5*A - 0.25*C + 0.5*D + 0.25*E + 0.5*G + 0.5*A*A + A*B + A*C + 0.25*C*C - A*D - C*D + 0.75*E*E
= - 0.5*A - 0.5*B + 1.5*C + 2*E + F + 0.5*A*A + 1.5*B*B - A*C - 0.5*C*C + 3*A*D - C*D + 2*A*F
= 0.16667*A - 0.11111*B - 0.22222*D + 0.44444*E - 0.44444*F - 0.33333*G + 1.16667*A*A - 0.44444*A*B
- 0.11111*A*C + 0.66667*B*C + 0.33333*C*C - 0.11111*A*D - 0.66667*B*D - 0.33333*C*D + 0.66667*B*E
- 0.66667*B*F
= - 0.5*A - 0.5*B - 2.5*C + D + 3*E - 2*F + 5*G + 0.5*A*A + 2*A*B - 1.5*B*B + 5*A*C + 2.5*C*C
- 3*A*D - 6*B*D - C*D + 6*C*F
= - 0.5*A - 0.83333*B + 0.5*C - 0.66667*D - 0.66667*E - 0.33333*F + 0.5*A*A - 1.33333*A*B + 0.5*B*B
- 0.33333*A*C + 0.5*C*C - 0.33333*A*D - C*D + 2*B*E - 2*D*F
= - 0.5*A + 2*B - C - 3*D - E - F + 2*G + 0.5*A*A - 3*A*C + C*C + A*D + 5*C*D - 6*E*F
= - 0.5*A - 0.5*B - 0.25*C - 0.5*D + 0.25*F - G + 0.5*A*A - A*B - A*C + 0.25*C*C - C*D + 0.75*F*F
= - 0.5*A - B - 4*C + 2*E + 2*F - G + 0.5*A*A + 3*A*C + C*C - 5*A*D - 7*C*D + 6*B*G
= 0.08333*B - 0.25*C + 0.16667*D + 0.16667*E - 0.16667*F + A*A - 0.16667*A*B + 0.25*B*B + 0.33333*A*C
+ 0.5*B*C + 0.25*C*C - 0.16667*A*D - 0.5*B*D - 0.5*C*D + 0.5*B*E - 0.5*D*G
= - 0.5*A - 2*B + 5*C - 2*D + 4*F - G + 0.5*A*A + 2*A*B + 3*B*B - A*C - 2*C*C + 3*A*D
+ 6*B*D - C*D - 6*E*G
= - 0.5*A - 0.75*B - 0.25*C + 0.5*D + E - G + 0.5*A*A - 0.5*A*B + 0.75*B*B - 0.5*A*C + 1.5*B*C
+ 0.25*C*C + 0.5*A*D - C*D - 1.5*F*G
= - 0.5*A - 0.25*B + 0.5*C + 0.5*E + 0.5*F - 0.25*G + 0.5*A*A + 0.75*B*B - 0.5*C*C + A*D - C*D + 0.75*G*G

B = - D - G - A*B - 2*A*C + A*D - A*G
= - 3*C + 4*D - 2*E + 2*F + 2*A*B - 3*B*B + 2*A*C + 3*C*C - 4*A*D - 6*B*E + 6*C*G

Figure 4.1: Aliasing Structure of XCW 18-run OA in 7 factors

Intercept = - 0.83333*A - 0.16667*C + 0.33333*D - 0.5*E - 0.33333*F + G + 1.16667*A*A - 1.33333*A*B
+ 0.33333*B*B - A*C - 1.33333*B*C + 0.16667*C*C - 0.66667*A*D - 0.66667*B*D - 0.66667*C*D - 0.33333*A*E
+ 0.33333*B*E + C*E + 1.33333*D*E - 0.16667*E*E - 0.66667*A*F + 1.33333*B*F + 0.33333*C*F + 1.33333*D*F
= - 0.16667*A - 0.66667*B - 0.16667*C - 0.33333*D + 0.16667*E - 0.33333*F + G + 0.5*A*A - 0.33333*B*B - A*C
- 1.33333*B*C + 0.16667*C*C - 0.66667*A*D - 0.66667*B*D - 0.66667*C*D + 0.66667*D*D - 0.33333*A*E
+ 0.33333*B*E + C*E + 0.5*E*E + 0.66667*A*F + 0.33333*C*F + 1.33333*E*F
= - 0.16667*A - 0.66667*B - 0.16667*C + 0.33333*D + 0.16667*E - 0.33333*F + 0.33333*G + 0.5*A*A - 0.33333*B*B
- A*C + 0.16667*C*C - 0.66667*B*D - 0.66667*C*D + 0.66667*D*D - 0.33333*A*E + 0.33333*B*E + C*E
+ 0.5*E*E + 0.66667*A*F + 0.33333*C*F - 0.66667*A*G
= - 0.83333*A - 0.66667*B - 0.16667*C + 0.33333*D + 0.16667*E - 0.33333*F + G + 1.16667*A*A - 1.33333*A*B
- 0.33333*B*B - A*C - 1.33333*B*C + 0.16667*C*C - 0.66667*A*D - 0.66667*B*D - 0.66667*C*D - 0.33333*A*E
+ 0.33333*B*E + C*E + 1.33333*D*E + 0.5*E*E - 0.66667*A*F + 1.33333*B*F + 0.33333*C*F + 1.33333*C*G
= - 0.16667*A - 0.66667*B - 0.16667*C + 0.33333*D + 0.16667*E - 0.33333*F + 0.33333*G + 1.16667*A*A
- 0.33333*B*B - A*C + 0.16667*C*C - 0.66667*B*D - 0.66667*C*D - 0.33333*A*E + 0.33333*B*E + C*E
+ 0.5*E*E + 0.66667*A*F + 0.33333*C*F - 0.66667*D*G
= - 0.16667*A - 0.66667*B - 0.16667*C + 0.33333*D + 0.16667*E - 0.33333*F + 0.33333*G + 0.5*A*A - 0.33333*B*B
- A*C + 0.16667*C*C - 0.66667*A*D - 0.66667*B*D - 0.66667*C*D - 0.33333*A*E + 0.33333*B*E
+ C*E + 0.5*E*E + 0.66667*A*F + 0.33333*C*F + 0.66667*G*G

B = C + E - F + 2*A*B - B*B - 2*A*C + C*C - 2*A*E + 2*C*E + E*E + 2*A*F - 2*B*F - F*F
= E + 2*A*B - B*B - 2*A*C - 2*A*E + 2*C*E - 2*D*E + E*E + 2*A*F - 2*B*F + 2*B*G
= E - B*B - 2*A*C - 2*B*D + 2*C*E + E*E + 2*A*F - 2*B*F + 2*E*G
= E - B*B - 2*A*C - 2*C*D + E*E + 2*A*F + 2*F*G

Figure 4.2: Aliasing Structure of XCW 27-run OA in 7 factors

The designs introduced in this section can be easily constructed based on BIBDs and regular graph (RG) IBDs, of which a few are cyclic IBDs. In particular, cyclic IBDs are constructed as follows:

- The first (or initial) block consists of a selection of k distinct factor labels.
- the second block is obtained from the first block by cycling the factor labels - replace factor 1 with 2, 2 with 3,...$t-1$ with t, and t with 1. The third block is obtained from the second block by cycling the treatment labels again, and so on until the tth block is created.

John [44] and John [43] list efficient cyclic designs which can be readily utilized. In particular, when possible, we choose cyclic designs with k=3 (note that r=k when t=b) since they are the least costly in terms of run size economy. If more runs can be afforded, one could choose a cyclic design with larger r, or double r and k, and begin with two distinct initial blocks.

Following Mee [68], we desire incomplete block designs that minimize the diagonal elements of $(\mathbf{N}\mathbf{N}')^{-1}$, where, for our new designs, $\mathbf{N}$ is the $b\times t$ incidence matrix for the incomplete block designs. That is, the ith row of $\mathbf{N}$ contains three ones corresponding to the treatments contained in the ith block and $t-3$ zeros. For example, for the

Table 4.10: Aliasing Structure of new 29-run design in 7 factors

Linear Main Effect	Aliasing
G	+ C*D + B*E + A*F
F	+ B*C + D*E + A*G
E	+ A*C + D*F + B*G
D	+ A*B + E*F + C*G
C	+ A*E + B*F + D*G
B	+ A*D + C*F + E*G
A	+ B*D + C*E + F*G

29-run design illustrated in Table 4.9, which is a balanced IBD with $\lambda = 1$, we have

$$\mathbf{N} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{4.17}$$

We denote the diagonal elements of $(\mathbf{NN}')^{-1}$ as the vector d_N. For this example, d_N is a 7×1 vector with entries 0.4444. Suppose, instead that the design was generated as a cyclic IBD with the initial block (1, 3, 5). Then, d_N is a 7×1 vector with entries 1.4444. Thus, our use of a balanced IBD is a better choice. Mee [68] states that this criterion is contrary to maximizing the efficiency of pairwise comparisons for incomplete block designs. In particular, minimizing the diagonal elements of $(\mathbf{NN}')^{-1}$ is equivalent to finding a balanced IBD, when it exists. Otherwise, the optimum is a regular graph design. If more than one regular graph design exists, we choose the one that minimizes d_N. Finding either a balanced or regular graph IBD is important since it helps to minimize the number of pairs of factors that do not appear together at all. To illustrate, let us consider the construction of an 8-factor design in 32 runs (plus center point runs). Using John [44]'s initial block recommendation, we begin with the block (1,2,4) and obtain

$$\begin{bmatrix} 1 & 2 & 4 \\ 2 & 3 & 5 \\ 3 & 4 & 6 \\ 4 & 5 & 7 \\ 5 & 6 & 8 \\ 6 & 7 & 1 \\ 7 & 8 & 2 \\ 8 & 1 & 3 \end{bmatrix} \Rightarrow \begin{bmatrix} \pm1 & \pm1 & 0 & \pm1 & 0 & 0 & 0 & 0 \\ 0 & \pm1 & \pm1 & 0 & \pm1 & 0 & 0 & 0 \\ 0 & 0 & \pm1 & \pm1 & 0 & \pm1 & 0 & 0 \\ 0 & 0 & 0 & \pm1 & \pm1 & 0 & \pm1 & 0 \\ 0 & 0 & 0 & 0 & \pm1 & \pm1 & 0 & \pm1 \\ \pm1 & 0 & 0 & 0 & 0 & \pm1 & \pm1 & 0 \\ 0 & \pm1 & 0 & 0 & 0 & 0 & \pm1 & \pm1 \\ \pm1 & 0 & \pm1 & 0 & 0 & 0 & 0 & \pm1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

This 32-run design has an aliasing structure as shown in Table 4.11. The structure remains simple, yet it is clear that not every pair of factors appear together in this design. The interactions A*E, B*F, C*G, and D*H are inestimable, regardless what is assumed about other terms. In this case, follow-up runs will be necessary in order to estimate certain interactions should their corresponding factors be deemed worthy of further investigation. Thus, although we wish to keep run sizes as small as possible, it will be clearly advantageous to construct designs based on balanced or regular graph IBDs with $\lambda_2 = 1$ in order to avoid this issue. Table 4.12 presents a summary of our proposed designs. The column labeled "Number of Runs" assumes one center point run. Additional center point runs can be added as desired. The number after the decimal for each design name represents the number of center point runs. Note that we do not provide a proposed design for t=8 as no BIBD or RG design exists of a reasonable and/or useful run size. See John and Mitchell [45] and Clatworthy [24] for extensive tables of IBDs.

Table 4.11: Aliasing Structure of 8-factor, 33-run design (Cyclic IBD Construction)

Linear Main Effect	Aliasing
A	+ B*D + F*G + C*H
B	+ A*D + C*E + G*H
C	+ B*E + D*F + A*H
D	+ A*B + C*F + E*G
E	+ B*C + D*G + F*H
F	+ C*D + A*G + E*H
G	+ D*E + A*F + B*H
H	+ A*C + E*F + B*G

Table 4.12: Summary of New Three-Level Designs

Design	t	Design Structure	Number of Parameters (Second-Order Model)	Number of Runs
4.1	4	BIBD: $4 \times 2^{3-1}$	15	17
5.1	5	RG: $5 \times 2^{3-1}$	21	21
6.1	6	RG: $6 \times 2^{3-1}$ (Cyclic)	28	25
7.1	7	BIBD: $7 \times 2^{3-1}$ (Cyclic)	36	29
9.1	9	BIBD: $12 \times 2^{3-1}$	55	49
10.1	10	RG: $20 \times 2^{3-1}$ (Cyclic)	66	81
11.1	11	RG: $22 \times 2^{3-1}$ (Cyclic)	78	89
12.1	12	RG: $24 \times 2^{3-1}$	91	97
13.1	13	BIBD: $26 \times 2^{3-1}$	105	105

Design Comparisons

In this section, we compare the recommended designs of XCW with the new three-level designs (which we shall denote as FBBD for Fractional Box-Behnken Designs) in terms of the number of eligible projections and average D-efficiency (denote as $\overline{D}$) for eligible projections. Also, we compare the FBBDs with D-optimal designs of the same size generated using OPTEX. In cases where the D-optimal design has fewer runs than the number of parameters in the second-order model, the Bayesian D-optimal framework was utilized. In such instances, we take the average number of eligible projections as well as the average $\overline{D}$ for Bayesian D-optimal designs constructed with $n_0 = 1, \frac{1}{4}n, \frac{1}{2}n$. Such is the case for t=6,7,9. The results are displayed in Tables 4.13, 4.14, 4.15, and 4.16.

When compared with the 18-run OA from XCW for $t \leq 7$, the FBBDs (with run sizes of 17, 21, 25, and 29) clearly outperform in terms of $\overline{D}$ and are comparable in terms of the number of eligible projections. While our 25-run 6-factor and 29-run 7-factor designs can project onto up to 5 and 6 factors, respectively, it is an unfair comparison to say that they *outperform* the 18-run OA in terms of number of eligible projections given that this design does not have enough runs to estimate a second-order model in 5 factors (21 runs are needed).

The recommended 27-run OA from XCW can screen up to 13-factors and all projections are eligible for up to 5 factors. However, this design's inability to project beyond 5-factors and its complex aliasing scheme make it difficult to successfully use in practical situations. For $t \leq 7$, where run sizes are most comparable, the FBBDs have better average D-efficiency for two-factor projections for $t = 4$, three and four-factor projections for t=5, 6, and 7, and five-factor projections for t=6 and 7. When $t = 9$, the FBBD has higher D-efficiency for five-factor projections and for $t \geq 10$, the 27-run OA has higher D-efficiency for all three, four, and five-factor

Table 4.13: Eligible Projections and D-Efficiencies of Table 4.12 Designs

t	Run Size	Projection	Number of Projections	Number of Eligible Projections	$\overline{D}$
4	17	2	6	6	0.728
		3	4	4	0.588
		4	1	0	*
5	21	3	10	10	0.664
		4	5	5	0.582
		5	1	0	*
6	25	3	20	20	0.668
		4	15	15	0.521
		5	6	6	0.616
7	29	3	35	35	0.632
		4	35	35	0.558
		5	21	21	0.563
		6	7	7	0.639
9	49	3	84	84	0.507
		4	126	126	0.503
		5	126	126	0.436
		6	84	84	0.519
		7	36	36	0.599
		8	9	9	0.655
10	81	3	120	120	0.44
		4	210	210	0.427
		5	252	252	0.395
		6	210	210	0.452
		7	120	120	0.533
		8	45	45	0.616
		9	10	10	0.693
11	89	3	165	165	0.406
		4	330	330	0.407
		5	462	462	0.374
		6	462	462	0.401
		7	330	330	0.472
		8	165	165	0.567
		9	55	55	0.621
		10	11	11	0.684
12	97	3	220	220	0.376
		4	495	495	0.388
		5	792	792	0.353
		6	924	924	0.363
		7	792	792	0.423
		8	495	495	0.49
		9	220	220	0.559
		10	66	66	0.623
		11	12	12	0.673
13	105	3	286	286	0.35
		4	715	715	0.368
		5	1287	1287	0.341
		6	1716	1716	0.333
		7	1716	1716	0.383
		8	1287	1287	0.445
		9	715	715	0.509
		10	286	286	0.572
		11	78	78	0.63
		12	13	13	0.667

Table 4.14: Eligible Projections and D-Efficiencies of XCW's Recommended 18-run OA

t	Projection	Number of Projections	Number of Eligible Projections	$\overline{D}$
4	2	6	6	0.725
	3	4	4	0.589
	4	1	1	0.465
5	3	10	10	0.589
	4	5	5	0.465
	5	1	0	*
6	3	20	20	0.589
	4	15	15	0.465
	5	6	0	*
7	3	35	34	0.579
	4	35	31	0.445
	5	21	0	*

Table 4.15: Eligible Projections and D-Efficiencies of XCW's Recommended 27-run OA

t	Projection	Number of Projections	Number of Eligible Projections	$\overline{D}$
4	2	6	6	0.725
	3	4	4	0.616
	4	1	1	0.555
5	3	10	10	0.607
	4	5	5	0.527
	5	1	1	0.445
6	3	20	20	0.601
	4	15	15	0.512
	5	6	6	0.408
7	3	35	35	0.598
	4	35	35	0.505
	5	21	21	0.407
8	3	56	56	0.597
	4	70	70	0.502
	5	56	56	0.402
9	3	84	84	0.596
	4	126	126	0.501
	5	126	126	0.401
10	3	120	120	0.595
	4	210	210	0.498
	5	252	252	0.396
11	3	165	165	0.594
	4	330	330	0.496
	5	462	462	0.393
12	3	220	220	0.593
	4	495	495	0.495
	5	792	792	0.391
13	3	286	286	0.593
	4	715	715	0.494
	5	1287	1287	0.390

Table 4.16: Eligible Projections and D-Efficiencies of D-optimal Designs

t	Run Size	Projection	Number of Projections	Number of Eligible Projections	$\overline{D}$
4	17	2	6	6	0.719
		3	4	4	0.575
		4	1	1	0.742
5	21	3	10	10	0.560
		4	5	5	0.643
		5	1	1	0.828
6	25	3	20	20	0.578
		4	15	15	0.525
		5	6	6	0.602
7	29	3	35	31	0.465
		4	35	28.33	0.468
		5	21	16	0.543
		6	7	4.67	0.551
9	49	3	84	68	0.463
		4	126	82.33	0.420
		5	126	63.67	0.430
		6	84	33	0.513
		7	36	10.33	0.523
		8	9	1.67	0.573
10	81	3	120	120	0.427
		4	210	210	0.403
		5	252	252	0.392
		6	210	210	0.466
		7	120	120	0.560
		8	45	45	0.664
		9	10	10	0.778
11	89	3	165	165	0.393
		4	330	330	0.385
		5	462	462	0.356
		6	462	462	0.400
		7	330	330	0.477
		8	165	165	0.562
		9	55	55	0.655
		10	11	11	0.755
12	97	3	220	220	0.356
		4	495	495	0.347
		5	792	792	0.318
		6	924	924	0.345
		7	792	792	0.405
		8	495	495	0.473
		9	220	220	0.546
		10	66	66	0.622
		11	12	12	0.702
13	105	3	286	286	0.315
		4	715	715	0.290
		5	1287	1287	0.266
		6	1716	1716	0.296
		7	1716	1716	0.344
		8	1287	1287	0.394
		9	715	715	0.446
		10	286	286	0.500
		11	78	78	0.550
		12	13	13	0.595

projections. Although we cannot compare the designs beyond 5-factor projections, it is worthy to note that $\overline{D}$ is monotonically decreasing as the projection size increases for the 27-run OA. In contrast, the FBBDs show no such monotonicity for the lower-order projections, but do appear to have a monotonically increasing $\overline{D}$ for projections beyond 5-factors. We will show in the next section via a simple simulation study how the 27-run OA performs when the number of active factors is known vs. unknown. In general, we safely recommend the FBBDs for $t \leq 7$ and for cases when factor sparsity is not expected to hold for $t \geq 9$.

Finally, we take a moment to compare the FBBDs with D-optimal designs of the same run size. Overall, and as expected, the D-optimal designs have similar or superior efficiency for higher factor projections except for t=13. For instance, when t=10, the D-optimal 81-run design has higher average efficiency for the 6-9 factor projections. On the other hand, for almost every t, the FBBDs have higher average efficiency for lower order projections. Thus, the usefulness of this comparison is to illustrate the the FBBDs can indeed compete with their "optimal" counterparts in terms of projection efficiency. Furthermore, the FBBDs are more intuitive than D-optimal designs as they have a recognizable structure and do not require the use of software.

Simulation Study

This small-scale simulation study is intended to illustrate the performance of XCW's OAs and the FBBDs with regards to their ability to identify the correct factors for projection and response surface exploration. Suppose that we simulate a response from the following model:

$$\begin{aligned} Y &= 57.3 + 1.5X_1 - 2.1X_2 + 1.8X_3 - 4.7X_1^2 - 6.3X_2^2 \\ &- 5.2X_3^2 - 7.1X_1X_2 - 3.3X_1X_3 - 2.7X_2X_3 + \epsilon \end{aligned} \tag{4.18}$$

where X_1, X_2, X_3 are three randomly chosen factors from $\binom{t}{3}$ possibilities and $\epsilon \sim N(0, \sigma^2)$. The coefficients in the above model are the least squares estimates for an example in Box and Draper [13]. We choose $\sigma = 1, 2, 3$ for this study. The above model is a reasonable choice since all of the linear main effects are small relative to the quadratic main effects and two-factor interactions. For brevity in what follows, we show results for $t = 5, 6, 7, 9, 12$.

First, consider the following two scenarios:

1. There are t factors under consideration. However, it is known that there are only three factors that have an effect on the shape of the response surface.

2. There are t factors under consideration. It is unknown which or how many factors are active.

In scenario 1, given that the ultimate objective is to locate the three active factors, fit a second-order model in those factors, and determine optimum settings, it makes sense to fit all three-factor second-order models to the data and pick the best model based on some criterion such as R^2. Thus, we simulate scenario 1 1000 times, each time randomly choosing a new set of three active factors, and then computing the percentage of the time 0, 1, 2, and 3 correct factors appear in the best 3-factor, second-order model. Table 4.17 displays the results with NCF standing for "Number of Correct Factors" and PCF represents the "Percent of Correct Factors" found out of 1000. It is clear from the results that XCW's OAs as well as the FBBDs perform

Table 4.17: Simulation Results (Scenario 1)

		18-run OA	27-run OA	FBBD
		PCF	PCF	PCF
t	NCF	$\sigma = 1/2/3$	$\sigma = 1/2/3$	$\sigma = 1/2/3$
4	3	1.000/1.000/0.996	1.000/1.000/1.000	1.000/0.999/0.973
	2	0.000/0.000/0.004	0.000/0.000/0.000	0.000/0.001/0.027
	1	0.000/0.000/0.000	0.000/0.000/0.000	0.000/0.000/0.000
	0	0.000/0.000/0.000	0.000/0.000/0.000	0.000/0.000/0.000
5	3	1.000/1.000/0.982	1.000/1.000/1.000	1.000/1.000/0.976
	2	0.000/0.000/0.017	0.000/0.000/0.000	0.000/0.000/0.024
	1	0.000/0.000/0.001	0.000/0.000/0.000	0.000/0.000/0.000
	0	0.000/0.000/0.000	0.000/0.000/0.000	0.000/0.000/0.000
6	3	1.000/1.000/0.946	1.000/1.000/1.000	1.000/1.000/0.990
	2	0.000/0.000/0.033	0.000/0.000/0.000	0.000/0.000/0.010
	1	0.000/0.000/0.011	0.000/0.000/0.000	0.000/0.000/0.000
	0	0.000/0.000/0.010	0.000/0.000/0.000	0.000/0.000/0.000
7	3	1.000/1.000/0.932	1.000/1.000/0.997	1.000/1.000/0.999
	2	0.000/0.000/0.050	0.000/0.000/0.003	0.000/0.000/0.001
	1	0.000/0.000/0.012	0.000/.000/0.000	0.000/0.000/0.000
	0	0.000/0.000/0.006	0.000/0.000/0.000	0.000/0.000/0.000
9	3	-	1.000/1.000/0.997	1.000/1.000/1.000
	2	-	0.000/0.000/0.003	0.000/0.000/0.000
	1	-	0.000/0.000/0.000	0.000/0.000/0.000
	0	-	0.000/0.000/0.000	0.000/0.000/0.000
12	3	-	1.000/1.000/0.990	1.000/1.000/1.000
	2	-	0.000/0.000/0.008	0.000/0.000/0.000
	1	-	0.000/0.000/0.001	0.000/0.000/0.000
	0	-	0.000/0.000/0.001	0.000/0.000/0.000

well under scenario 1. That is, given that the true number of active factors is known (and assuming the run size is large enough to fit a second-order model), the designs are able to discriminate among competing models of the same size. This is certainly the result that one would hope to see despite some of the shortcomings of the OAs mentioned above.

Scenario 2 is a bit different. In this case, since we do not know how many factors are active, it is necessary to employ the strategy of CW and fit a pure quadratic model in all the factors under consideration before proceeding to the projection stage. Although we have seen in the examples presented in section 1 that this strategy often fails to identify important factors, we take a moment to compare designs using this approach. In particular, in simulating scenario 2, we use a significance level of $\alpha = 0.1$, which seems a reasonable choice for screening purposes. As before, we simulate scenario 2 1000 times, each time randomly choosing a new set of three active factors, and then computing the percentage of the time 0, 1, 2, and 3 correct factors are found. In addition, we also report the mean number of effects declared active (denoted by MEA). The results are displayed in Table 4.18.

The 18-run OA clearly has a difficult time finding the correct factors using the method of CW. Only for t=7 does the MEA exceed (or even come close to) 3. Therefore, relatively few effects are being declared active. However, even when MEA$>$ 3, the percentage of time all three active factors are found is less than 38%. As expected, the 27-run OA and the FBBDs perform better. The MEA for the 27-run OA hovers around 2 for all cases considered except for $t = 12$, in which case MEA$>$ 4. Even so, for this case, PCF is less than 35%. For $t = 4, 5$, the 27-run OA outperforms the smaller FBBDs in terms of PCF in spite of MEA being smaller. However, in comparison, for $t \geq 6$, we see much larger PCFs and MEAs for the FBBDs.

Therefore, even though we do not advocate CWs analysis strategy alone, the

Table 4.18: Simulation Results (Scenario 2)

		18-run OA		27-run OA		FBBD	
		PCF	MEA	PCF	MEA	PCF	MEA
t	NCF	$\sigma = 1/2/3$	$\sigma = 1/2/3$	$\sigma = 1/2/3$	$\sigma = 1/2/3$	$\sigma = 1/2/3$	$\sigma = 1/2/3$
4	3	0.003/0.029/0.047	1.113/1.146/1.145	0.284/0.244/0.177	2.124/1.997/1.829	0.053/0.095/0.105	3.076/2.652/2.414
	2	0.341/0.288/0.247		0.555/0.461/0.411		0.296/0.267/0.258	
	1	0.405/0.409/0.401		0.161/0.273/0.346		0.602/0.462/0.410	
	0	0.251/0.274/0.305		0.000/0.022/0.066		0049/0.176/0.227	
5	3	0.000/0.015/0.026	0.398/0.665/0.785	0.255/0.204/0.197	2.182/2.040/1.952	0.227/0.203/0.123	2.900/2.676/2.126
	2	0.066/0.128/0.158		0.578/0.514/0.423		0.191/0.220/0.211	
	1	0.263/0.337/0.324		0.162/0.240/0.293		0.452/0.377/0.384	
	0	0.671/0.520/0.492		0.005/0.042/0.087		0.130/0.200/0.282	
6	3	0.000/0.001/0.009	0.010/0.171/0.397	0.231/0.227/0.169	2.399/2.402/2.159	0.867/0.607/0.431	4.390/3.857/3.273
	2	0.000/0.024/0.056		0.599/0.495/0.435		0.101/0.279/0.318	
	1	0.010/0.118/0.223		0.166/0.253/0.311		0.032/0.100/0.191	
	0	0.990/0.857/0.712		0.004/0.025/0.085		0.000/0.014/0.060	
7	3	0.379/0.318/0.318	3.764/3.421/3.390	0.272/0.214/0.193	2.622/2.416/2.355	0.997/0.881/0.662	4.910/4.444/3.866
	2	0.122/0.151/0.121		0.464/0.440/0.378		0.003/0.103/0.235	
	1	0.015/0.056/0.122		0.235/0.273/0.332		0.000/0.016/0.083	
	0	0.484/0.475/0.439		0.029/0.073/0.097		0.000/0.000/0.020	
9	3	-	-	0.156/0.144/0.134	2.240/2.233/2.262	1.000/0.996/0.914	6.504/6.020/5.440
	2	-		0.426/0.388/0.358		0.000/0.004/0.074	
	1	-		0.351/0.363/0.346		0.000/0.000/0.012	
	0	-		0.067/0.105/0.162		0.000/0.000/0.000	
12	3	-	-	0.348/0.332/0.284	4.642/4.307/4.050	1.000/1.000/0.988	7.220/7.160/6.930
	2	-		0.154/0.146/0.160		0.000/0.000/0.012	
	1	-		0.149/0.132/0.156		0.000/0.000/0.000	
	0	-		0.349/0.390/0.400		0.000/0.000/0.000	

FBBDs show better performance over the OAs for $t \geq 6$ given its ability to declare more effects active and thus have a higher likelihood of finding all of the important factors. It is worth a brief mention that if one were to force the four effects with the smallest p-values to be active in the scenario 2 simulation, we would obtain better results for the OAs. For instance, for the 27-run OA with $t = 6$ and $\sigma = 1$, we see that the the percentage of time all three active factors are found increases from 23.1% to 81.5%. Likewise, for $t = 7$ and $\sigma = 1$, we see an increase from 27.2% to 75.9%. This is meant to illustrate that if the OAs were able to detect more active effects, it would presumably have a greater propensity for finding all three factors.

4.2.3 Other Designs

We now briefly mention two other types of designs that appeared during the Bayesian D-optimal search using BB-type designs as candidate sets of points. Although their general construction and use will not be developed here, we suspect it is a worthy topic for future research.

Resolution II Subsets with Four Runs

Table 4.19 illustrates a Bayesian D-optimal design for 10 factors in 41 runs using the balanced IBD with 10 blocks of size 4 as the candidate set of points. That is, the candidate set consists of 10 $\times 2^4$=160 points from the 4th orbit. Note that each resolution II subset is paired with another "opposite" subset. To achieve such structure, an even number of factors is required. For an odd number of factors, one needs only to pair together as many subsets as possible. Although not as complicated as the OA's examined earlier, the aliasing structure of this design is more complex than when using resolution III subsets.

Table 4.19: 41-run 10 factor design with Resolution II Blocks

A	B	C	D	E	F	G	H	J	K	
-1	-1	0	0	1	0	0	0	0	-1	A = K
-1	1	0	0	-1	0	0	0	0	-1	
1	-1	0	0	-1	0	0	0	0	1	
1	1	0	0	1	0	0	0	0	1	
-1	0	-1	0	0	-1	0	0	1	0	C = -J
-1	0	1	0	0	1	0	0	-1	0	
1	0	-1	0	0	1	0	0	1	0	
1	0	1	0	0	-1	0	0	-1	0	
-1	0	0	-1	0	0	1	1	0	0	G = H
-1	0	0	1	0	0	-1	-1	0	0	
1	0	0	-1	0	0	-1	-1	0	0	
1	0	0	1	0	0	1	1	0	0	
-1	0	0	0	0	0	0	-1	1	1	A = -K
-1	0	0	0	0	0	0	1	-1	1	
1	0	0	0	0	0	0	-1	-1	-1	
1	0	0	0	0	0	0	1	1	-1	
0	-1	-1	0	0	0	1	-1	0	0	G = -H
0	-1	1	0	0	0	-1	1	0	0	
0	1	-1	0	0	0	-1	1	0	0	
0	1	1	0	0	0	1	-1	0	0	
0	-1	0	-1	0	1	0	0	-1	0	B = -F
0	-1	0	1	0	1	0	0	1	0	
0	1	0	-1	0	-1	0	0	1	0	
0	1	0	1	0	-1	0	0	-1	0	
0	-1	0	0	0	-1	-1	0	0	1	B = F
0	-1	0	0	0	-1	1	0	0	-1	
0	1	0	0	0	1	-1	0	0	-1	
0	1	0	0	0	1	1	0	0	1	
0	0	-1	-1	1	0	0	0	0	-1	D = -E
0	0	-1	1	-1	0	0	0	0	1	
0	0	1	-1	1	0	0	0	0	1	
0	0	1	1	-1	0	0	0	0	-1	
0	0	-1	0	-1	0	1	0	-1	0	C = J
0	0	-1	0	1	0	-1	0	-1	0	
0	0	1	0	-1	0	-1	0	1	0	
0	0	1	0	1	0	1	0	1	0	
0	0	0	-1	-1	-1	0	1	0	0	D = E
0	0	0	-1	-1	1	0	-1	0	0	
0	0	0	1	1	-1	0	-1	0	0	
0	0	0	1	1	1	0	1	0	0	
0	0	0	0	0	0	0	0	0	0	

Singletons

During the first design search discussed above, we mentioned how none of the resulting designs yielded any recognizable structure with the exception of the 7-factor design in 28 runs (plus a center-point run). In this case, each of the 28 runs consisted of a different triple of factors (i.e. no triple appearing more than once). Note that with 7 factors, there are $\binom{7}{3}$=35 possible triples of factors that could appear together. Since the number of runs was specified to be n=29, all triples could not appear in this design. A subsequent *D*-optimal search, however, with a specified run size of n=36, did not yield a design with all possible triples. Although not as simple as with resolution III blocks, the aliasing is not as complex as the OA's shown earlier.

4.3 Analysis Strategy

In this section, we introduce a simple analysis strategy for the FBBDs constructed in the previous section. Again, our purpose is to present a method that will allow for interaction detection and thus, aid the experimenter in choosing an appropriate set of factors for projection. In what follows, we make the following assumptions:

- 1. *Effect Hierarchy:* Main effects are more likely to be important than interaction effects and effects of the same order are equally important.
- 2a. *Effect Sparsity:* The number of important effects is relatively small.
- 2b. *Interaction Sparsity:* The number of important two-factor interactions is relatively small.
- 3. *Weak Effect Heredity:* An interaction is more likely to be important if one or more of its parent factors is also important.

Recall that the method of CW assumes strong effect heredity which states that an interaction is active only if both parent factors are active. Here, we relax this stronger assumption.

We now propose a new analysis strategy that will aid in determining which main effects and two-factor interactions are most likely to be important. By constructing the designs in the manner discussed in section 4.2.2 (i.e. resolution III blocks), we are able to obtain r independent estimates of each factor's linear main effect. Note that each of the estimates is aliased with one two-factor interaction. By obtaining these estimates, we will show how one can utilize them to aid in factor screening and interaction detection. In what follows, we outline and illustrate two different approaches to analysis based on the degree of interaction sparsity assumed.

In general, by assuming interaction sparsity, stage 1 of either approach of our proposed analysis strategy may utilize a Lenth-type procedure in order to obtain a robust scale estimate. This scale estimate can then be utilized to aid in identifying active effects. Let $\gamma_{ij},\ i = 1, 2, ..., r;\ j = 1, 2, ..., t$ be the ith estimate of the jth factor's linear main effect and Γ be the $r \times t$ matrix with ij^{th} element γ_{ij}. In order to obtain an initial estimate of each factor's linear main effect, we compute the median of each column of Γ, denoted as m_j. Now, let δ_{ij} denote the deviation of γ_{ij} from m_j. If interaction sparsity holds, each of the δ_{ij}'s should be close to zero and large deviations from m_j are likely to represent the presence of a two-factor interaction. Thus, if we can estimate the scale of the null δ_{ij}'s, then, presumably, the δ_{ij}'s that are large compared to this scale estimate would correspond to non-null interactions.

Haaland and O'Connell [38] describe the procedure for computing robust scale estimators as two steps:

1. Obtain an initial estimate of scale using all estimated effects. This estimate

should be of the form

$$s_0(q) = a_0(q) * quantile(q; \left|\hat{\theta}_i\right|) \tag{4.19}$$

where $a_0(q)$ is a "consistency constant" and $quantile(q; \left|\hat{\theta}_i\right|)$ is the qth quantile of the absolute values of the estimates $\hat{\theta}_i$. For simplicity of notation, hereafter, we drop the dependence of s_0 on q.

2. Obtain a final scale estimate (or pseudo standard error (PSE)) using all estimated effects that are smaller than some multiple of the initial scale estimate. That is, compute

$$\hat{\sigma}(q, c) = a(q, c) * quantile_{|\hat{\theta}_i|<c*s_0}(q; \left|\hat{\theta}_i\right|) \tag{4.20}$$

where $a(q, c)$ is another "consistency constant" and c is a tuning constant for computing the pruning threshold of $c * s_0$.

For example, Lenth's [53] method uses $q = 0.5$ (i.e., the median), $a_0(q) = 1.5$, $a(q, c) = 1.5$, and $c = 2.5$.

The consistency constants are defined to be dependent on the choice of q. In particular, $a_0(q)$ "can be determined directly from the sampling distribution of absolute values of a standard normal random variable" as $a_0(q) = 1/\Phi^{-1}((q + 1)/2)$ where Φ is the cumulative standard normal distribution and $a(q, c)$ can be determined empirically, according to Haaland and O'Connell [38]. Although termed "consistency constants", these values are used to adjust for bias in finite samples.

As stated, our intention is to construct a robust scale estimator based on Lenth's method using the δ_{ij}'s introduced above. Thus, our initial choices of q, $a_0(q)$, $a(q, c)$, and c are identical to Lenth. We investigate these choices for each design in Table

4.12 (i.e. for $r = 3, 4, 6$) and find that it will be required to adjust s_0 and/or the PSE in order to correct for downward bias. That is, for each choice of r, $\mathbb{E}(PSE) < \sigma$. The trouble with bias is most pronounced with $r = 3$ (i.e. $\mathbb{E}(PSE) << \sigma$) since, necessarily, at least t of the δ_{ij}'s are equal to zero.

Rather than require $a_0(q)$ or $a(q, c)$ to be larger than the usual 1.5 to correct for bias when $r = 3$, we restrict ourselves to the $(t + 1)^{st}$ to $(rt)^{th}$ order statistics of the δ_{ij}'s in order to eliminate the t guaranteed zeros. After doing so, use of the median ($q = 0.5$) is not the best choice for computing s_0 (i.e. we obtain a PSE that is slightly upward biased). Instead, we utilize the 45th percentile, which was suggested to be a reasonable alternative in Schoen and Kaul [90]. Thus, for $r = 3$, we make use of the following: $q_1 = 0.45$, $a_0(q_1) = 1.5$, $q_2 = 0.5$, $a(q_2, c)$=1.5, and c=2.5. That is, for $r = 3$, the procedure is identical to Lenth's method with the exception of the use of the 45th percentile to obtain s_0.

On the other hand, the bias for the $r = 4$ or 6 designs is not as severe. Thus, a simple adjustment to $a(q, c)$ is all that is required to obtain an unbiased estimator of σ. For instance, for $r = 4$, it was empirically determined (based on simulating $2\gamma_{ij}$'s from a $N(0, 1)$ distribution) that $\mathbb{E}(PSE) \approx 0.689$. (Note that since $\mathbb{V}(\gamma_{ij}) = \sigma^2/4$, then $\mathbb{V}(2\gamma_{ij}) = \sigma^2$.) Thus, $1.45 * \mathbb{E}(PSE) \approx 1$ and thus, we choose $a(q, c) = 1.45 * 1.5 = 2.175$ in order to construct an approximately unbiased estimate of σ. Then, for $r = 4$, we have $q = 0.5$, $a_0(q) = 1.5$, $a(q, c) = 2.175$, and $c = 2.5$. In the same manner, for $r = 6$, we determine that $a(q, c) = 2.025$ is best (q, $a_0(q)$, and c remain the same as from $r = 4$). Thus, the only modification from Lenth's method for $r = 4$ and 6 is in the multiplier $a(q, c)$.

It is worth mentioning that although these corrections are able to adequately adjust for bias, there is a loss of efficiency as r decreases, as may be expected. However, since the purpose of stage 1 is to simply gauge the importance of factors, we believe

that the proposed strategy provides a sufficiently reasonable estimator for σ in order to conduct tests of significance.

4.3.1 Analysis Approach 1

In the first approach to analysis, we suppose that the experimenter is willing to assume *strong* interaction sparsity (i.e. only very few interactions are likely to be active). In particular, we assume that at most 10% of the interactions are important and no more than one interaction is active within each group of r interactions aliased with a linear main effect. We now outline approach 1:

1. Compute $m_j = median(\Gamma^{(j)})$ where $\Gamma^{(j)}$ is the jth column of Γ.
2. Compute $\delta_{ij} = \gamma_{ij} - m_j$.
3. Compute s_0.
   - For Designs 4.1, 5.1, 6.1, and 7.1, $s_0 = 1.5 \times quantile(0.45; \{\left|\delta^{(t+1)}\right|, ..., \left|\delta^{(rt)}\right|\})$ where $\delta^{(t+1)}$ and $\delta^{(rt)}$ represent the $(t+1)^{st}$ and $(rt)^{th}$ order statistics of the δ_{ij}'s, respectively.
   - For Designs 9.1, 10.1, 11.1, 12.1, and 13.1, $s_0 = 1.5 \times median\,|\delta_{ij}|$
4. Compute an adjusted Lenth's PSE.
   - For Designs 4.1, 5.1, 6.1, and 7.1:
     $PSE = 1.5 \times median_{|\delta_{ij}|<2.5s_0}\,|\delta_{ij}|$.
   - For Design 9.1:
     $PSE = 2.175 \times median_{|\delta_{ij}|<2.5s_0}\,|\delta_{ij}|$.
   - For Designs 10.1, 11.1, 12.1, and 13.1:
     $PSE = 2.025 \times median_{|\delta_{ij}|<2.5s_0}\,|\delta_{ij}|$.

5. Compute pseudo estimates of the linear main effects as $\tilde{b}_j = median_{|\delta_{ij}|<2.5s_0}(\Gamma^{(j)})$ and the t-like statistics, $t_j^{(l)} = \frac{\tilde{b}_j}{PSE/\sqrt{4r}}$. Note that $median_{|\delta_{ij}|<2.5s_0}(\Gamma^{(j)})$ represents the median of the γ_{ij}'s computed after pruning away those γ_{ij}'s that correspond to $|\delta_{ij}| < 2.5s_0$. If $t_j^{(l)} > t_\alpha^{(l)}$, then factor j is worthy of further investigation.

6. Compute the usual least squares estimates, $b = (\mathbf{X'X})^{-1}\mathbf{X'Y}$ where $\mathbf{X}$ is the model matrix of a pure quadratic model (i.e. all linear and quadratic main effects). Then, the unbiased quadratic main effects, denoted as b_{jj}, are the $(t+1)^{st}$ to $(2t)^{th}$ elements of b. We further compute the t-like statistics, $t_j^{(q)} = \frac{b_{jj}}{\sqrt{v_{jj}}PSE}$ where v_{jj} is the diagonal element of $(\mathbf{X'X})^{-1}$ corresponding to b_{jj}. Then, if $t_j^{(q)} > t_\alpha^{(q)}$, we also consider factor j worthy of further consideration.

7. Compute pseudo estimates of the two-factor interactions as $\tilde{b}_{ij} = \gamma_{ij} - \tilde{b}_j$ and the statistics, $t_{ij}^{(int)} = \frac{\tilde{b}_{ij}}{PSE/2}$. If $t_{ij}^{(int)} > t_\alpha^{(int)}$, then the factors involved in the two-factor interaction are worthy of further consideration (assuming that the two-factor interaction satisfies weak effect heredity). For this calculation, it is possible for $\tilde{b}_{ij}$ to have a value of zero if the number of δ_{ij}'s pruned in each group of r estimates is even and $r = 3$ or if the number of δ_{ij}'s pruned is odd and r=4 or 6. A value of zero simply indicates that the two-factor interaction associated with the estimate is unlikely.

As in Chapter 2, we standardize all the estimates so that they are estimated with equal precision before utilizing them to compute Lenth t-statistics. In doing so, the PSE is an estimate of σ. In what follows, we will always refer to standardized estimates. It is worthwhile to note that under the null distribution, $\mathbb{V}(\tilde{b}_j) > \mathbb{V}(b_j) = \sigma^2/4r$ and $\mathbb{V}(\tilde{b}_{ij}) < \mathbb{V}(\gamma_{ij}) = \sigma^2/4$. Rather than estimate $\mathbb{V}(\tilde{b}_j)$ and $\mathbb{V}(\tilde{b}_{ij})$ under the null, we simply absorb the differences among the variances into our critical values.

The critical values, $t_\alpha^{(l)}$, $t_\alpha^{(q)}$, and $t_\alpha^{(int)}$ are provided in Table 4.20 for each proposed design. They were calculated via simulation for $\alpha = 0.01, 0.05, 0.1$ in a manner similar to as was done in Chapter 2, section 2. In particular, 20,000 sets of rt contrasts are generated from a standard normal distribution. For each set, we follow the steps outlined above to obtain three sets of t-like statistics. Then, for $\alpha = 0.01, 0.05, 0.1$, the critical values are approximated as follows:

- $t_\alpha^{(l)}$ is the $(20,000 * \alpha * t)^{th}$ largest of the $(20,000 * t)$ $t^{(l)}$-statistics.
- $t_\alpha^{(q)}$ is the $(20,000 * \alpha * t)^{th}$ largest of the $(20,000 * t)$ $t^{(q)}$-statistics.
- $t_\alpha^{(int)}$ is the $(20,000 * \alpha * rt)^{th}$ largest of the $(20,000 * rt)$ $t^{(int)}$-statistics.

Note that in step 6 the pure quadratic effects are likewise generated from their respective null distribution. It is recommended that a significance level of 10% be utilized given that stage 1 is exploratory in nature. Furthermore, since the proposed designs can project up to $t - 1$ factors, being somewhat liberal in determining factor importance is acceptable.

In addition to the above steps, which allow for a more objective approach to determining factor importance, we also recommend a simple graphical method that displays the m_j's and δ_{ij}'s. In particular, one plots the m_j's on the x-axis and the δ_{ij}'s on the y-axis labeled with its corresponding interaction. Factors with m_j's that stand away from zero would tend to indicate the importance of a linear main effect, whereas a δ_{ij} that deviates from zero indicates the importance of a two-factor interaction. This first approach is further developed best through example. Our examples are intended to illustrate the success of this approach as well as provide warnings regarding its use when strong interaction sparsity is violated.

Table 4.20: Critical Values for Analysis Approach 1 for Table 4.12 Designs

t	n	α	$t_\alpha^{(l)}$	$t_\alpha^{(q)}$	$t_\alpha^{(int)}$
4	17	0.01	8.174	6.498	5.766
		0.05	4.008	3.203	2.342
		0.1	2.782	2.251	1.554
5	21	0.01	7.314	5.658	5.372
		0.05	3.738	2.98	2.332
		0.1	2.686	2.189	1.596
6	25	0.01	6.341	5.087	4.896
		0.05	3.495	2.77	2.241
		0.1	2.555	2.072	1.585
7	29	0.01	5.711	4.511	4.632
		0.05	3.284	2.642	2.189
		0.1	2.465	2.011	1.58
9	49	0.01	4.898	3.903	4.043
		0.05	3.116	2.476	2.493
		0.1	2.416	1.932	1.632
10	81	0.01	4.196	3.181	3.446
		0.05	2.871	2.195	2.262
		0.1	2.283	1.776	1.663
11	89	0.01	4.152	3.152	3.402
		0.05	2.824	2.183	2.238
		0.1	2.263	1.773	1.653
12	97	0.01	4.118	3.076	3.356
		0.05	2.827	2.163	2.229
		0.1	2.254	1.764	1.654
13	105	0.01	4.037	3.027	3.321
		0.05	2.785	2.146	2.206
		0.1	2.232	1.754	1.648

Example 4 - Simulated Data 1

Instead of utilizing the 18 or 27-run OA as suggested by XCW, suppose the experimenter constructs a 29-run 7-factor design as in section 4.2.2 using an initial block of (1 2 4). Furthermore, suppose some response, Y, is simulated from the following model:

$$Y = 5x_D + 4x_G + 2.5x_{DG} - 3x_{AD} - 3x_{AG} + 2x_{A^2} + 3x_{D^2} + \epsilon \qquad (4.21)$$

where $\epsilon \sim N(0, 1)$. The design and response are shown in Table 4.21 while Table 4.22 provides the aliasing of this design. Table 4.23 displays preliminary estimates of the linear main effects and two-factor interactions as well as least squares estimates of the quadratic main effects. The t-ratios provided are based on $PSE = 1.13$, which is only slightly larger than the true σ.

Based on the critical values provided in Table 4.20 and using the suggested significance level of $\alpha = 0.1$, the following effects are identified as significant: D, G, A^2, D^2, B*D, A*D, A*G, and D*G. Therefore, using the terms identified above, the second stage of the analysis strategy is to project onto factors A, B, D, and G. It should be noted, at this point, that the main effects only analysis of CW would also point to the significance of the linear main effects D and G. However, CW's strategy is to then project only onto factors D and G. Figure 4.3 provides a plot of the δ_{ij}'s vs. the m_j's. This graphic provides a visual confirmation to the significance test results.

Recall, again, that the purpose of this analysis strategy using the newly proposed designs is to better identify factors that are involved in interaction effects but that do not necessarily show up in a main effects analysis. Using the terminology of CW, the identified projection is eligible. Table 4.24 displays the results of this analysis. We obtain R^2=0.986 for the second-order model in all the projected factors and it is clear that we have been able to correctly detect the active interaction effects that

Table 4.21: Design and Response for Example 4

A	B	C	D	E	F	G	Y
-1	-1	0	1	0	0	0	13.571
-1	1	0	-1	0	0	0	-3.3999
1	-1	0	-1	0	0	0	3.69
1	1	0	1	0	0	0	7.8156
0	-1	-1	0	1	0	0	0.71191
0	-1	1	0	-1	0	0	1.2902
0	1	-1	0	-1	0	0	0.6686
0	1	1	0	1	0	0	1.1908
0	0	-1	-1	0	1	0	-3.2025
0	0	-1	1	0	-1	0	7.9802
0	0	1	-1	0	-1	0	-2.1567
0	0	1	1	0	1	0	6.3959
0	0	0	-1	-1	0	1	-0.2427
0	0	0	-1	1	0	-1	-4.5565
0	0	0	1	-1	0	-1	2.9151
0	0	0	1	1	0	1	13.695
-1	0	0	0	-1	1	0	2.5287
-1	0	0	0	1	-1	0	2.2193
1	0	0	0	-1	-1	0	1.0781
1	0	0	0	1	1	0	-0.17067
0	-1	0	0	0	-1	1	3.9408
0	-1	0	0	0	1	-1	-5.0106
0	1	0	0	0	-1	-1	-3.3855
0	1	0	0	0	1	1	4.5077
-1	0	-1	0	0	0	1	10.692
-1	0	1	0	0	0	-1	-4.4087
1	0	-1	0	0	0	-1	0.3564
1	0	1	0	0	0	1	3.3803
0	0	0	0	0	0	0	-1.0091

Table 4.22: Aliasing of Example 4 Design

Linear Main Effect	Aliasing
G	+ A*C + D*E + B*F
C	+ B*E + D*F + A*G
F	+ C*D + A*E + B*G
A	+ B*D + E*F + C*G
E	+ B*C + A*F + D*G
D	+ A*B + C*F + E*G
B	+ A*D + C*E + F*G

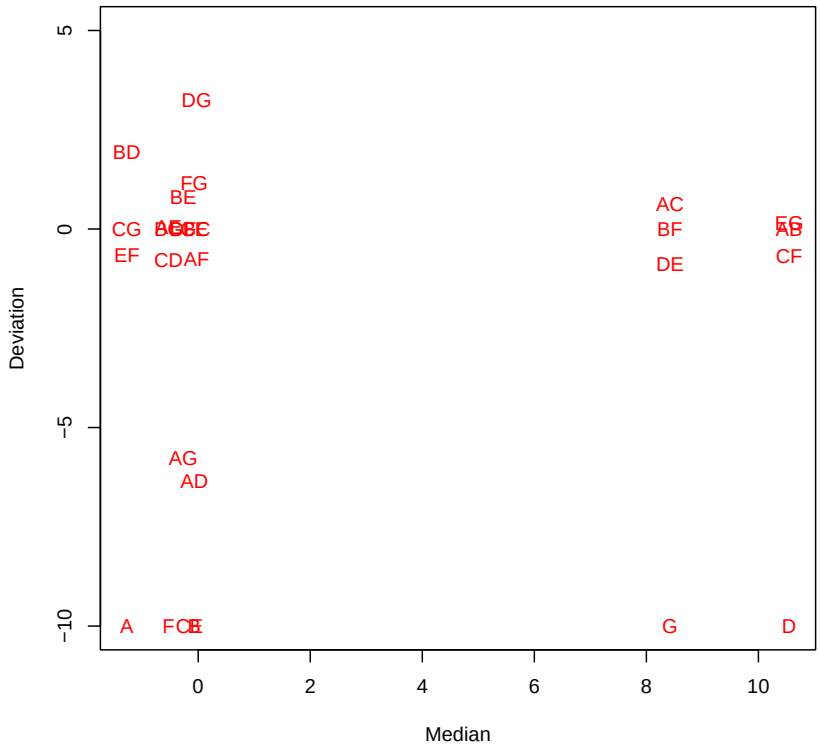

Figure 4.3: Example 4 - Analysis Stage 1

Table 4.23: Example 4 - Analysis Stage 1

Term	Standardized Estimate	t Ratio
A	-2.206	-1.957
B	0.887	0.788
C	0.243	0.217
D	18.270	16.217
E	-0.699	-0.620
F	-0.916	-0.814
G	14.588	12.949
A^2	5.130	4.553
B^2	2.011	1.785
C^2	1.297	1.151
D^2	6.496	5.766
E^2	0.881	0.782
F^2	-0.870	-0.773
G^2	1.028	0.912
BD	1.947	1.723
EF	-0.649	-0.574
CG	0.000	0.000
AD	-6.956	-6.156
CE	-0.585	-0.518
FG	0.585	0.518
BE	0.411	0.364
DF	-0.411	-0.364
AG	-6.198	-5.485
AB	0.000	0.000
CF	-0.683	-0.604
EG	0.157	0.139
BC	0.376	0.333
DG	3.648	3.228
AF	-0.376	-0.333
CD	-0.789	-0.698
AE	0.060	0.053
BG	0.000	0.000
DE	-0.878	-0.777
BF	0.000	0.000
AC	0.642	0.568

Table 4.24: Second-Order Analysis Results for Example 4

Term	Estimate	Std Error	t Ratio	P-value
Intercept	-0.516106	0.384308	-1.34	0.2007
A	-0.798396	0.303822	-2.63	0.0199
B	0.2561613	0.303822	0.84	0.4133
D	5.143075	0.303822	16.93	$< .0001$
G	4.1719083	0.248069	16.82	$< .0001$
A*A	2.2743497	0.337411	6.74	$< .0001$
A*B	0.13105	0.526235	0.25	0.807
B*B	0.8041597	0.337411	2.38	0.0319
A*D	-3.467486	0.526235	-6.59	$< .0001$
B*D	1.1320212	0.526235	2.15	0.0494
D*D	2.9183959	0.337411	8.65	$< .0001$
A*G	-3.0192	0.429669	-7.03	$< .0001$
B*G	-0.26455	0.429669	-0.62	0.548
D*G	1.616525	0.429669	3.76	0.0021
G*G	0.3407459	0.337411	1.01	0.3297

would have been missed had we followed CW's analysis strategy.

Before concluding this example, we show how violations of strong interaction sparsity affect the efficiency of the pseudo estimates of the linear main effects and two-factor interactions (i.e. $\tilde{b}_j$ and $\tilde{b}_{ij}$). Table 4.25 provides $\mathbb{V}(\tilde{b}_j)$ and $\mathbb{V}(\tilde{b}_{ij})$ based on 10,000 simulated responses of variations of the above model (4.21). From this, we see that the addition of each two-factor interaction to the true model increases $\mathbb{V}(\tilde{b}_j)$ from approximately $\sigma^2/8$ to $\sigma^2/5$ for those $\tilde{b}_j$ aliased with active interactions. (Note that this example illustrates the case where no more than one interaction is active within each group of r estimates.) Furthermore, $\mathbb{V}(\tilde{b}_{ij})$ for the active interactions increases from roughly $\sigma^2/5$ to $\sigma^2/2$. Since the critical values in Table 4.20 are based on the null distribution, the loss of efficiency that results from active interactions can lead to more Type I errors than expected when conducting significance tests. Furthermore, this loss calls into question the validity of the critical values. Thus, although we were met with apparent success in this example using approach 1, one must be careful when interpreting the results if many interactions can be expected. □

Example 5 - Simulated Data 2

In this example, we consider a 9-factor, 49-run FBBD constructed using a BIBD with $r = 4$. Note that this design has 22 more runs than the recommended 27-run OA. However, we anticipate much better performance with the new design. Now, suppose we simulate from the following model:

$$Y = 2x_A - 1.5x_E + 2x_G + 2.5x_{E^2} - 3x_{H^2} + 4x_{AC} - 5x_{CG} + 3.5x_{EH} - 4x_{GH} + \epsilon \quad (4.22)$$

where $\epsilon \sim N(0, 1)$. The two-factor interactions have been purposely made larger than the linear and quadratic main effects in order to investigate what issues, if any,

Table 4.25: Variance of Pseudo-Estimates - Example 4

	Null Case	Number of Interactions		
Term	(Main Effects Only)	1 (D*G)	2 (Add A*D)	3 (Add A*G)
A	0.1233	0.1171	0.1181	0.1132
B	0.1203	0.1173	0.1904	0.1901
C	0.1186	0.1194	0.1162	0.1886
D	0.1200	0.1154	0.1155	0.1151
E	0.1197	0.2109	0.2145	0.2187
F	0.1221	0.1162	0.1162	0.1134
G	0.1230	0.1200	0.1144	0.1128
BD	0.2030	0.2064	0.1929	0.1974
EF	0.2042	0.1983	0.1958	0.1992
CG	0.2065	0.1936	0.1992	0.1889
AD	0.2066	0.1983	0.4724	0.4791
CE	0.1979	0.1919	0.1722	0.1747
FG	0.2047	0.2052	0.1803	0.1612
BE	0.2062	0.2039	0.1977	0.1797
DF	0.2034	0.2020	0.1976	0.1635
AG	0.2012	0.1955	0.2040	0.4824
AB	0.2053	0.1971	0.2019	0.1994
CF	0.2087	0.1993	0.1998	0.1983
EG	0.2008	0.2005	0.1995	0.1993
BC	0.1932	0.1911	0.1841	0.1842
DG	0.2054	0.5055	0.5115	0.5199
AF	0.2044	0.1906	0.1981	0.1910
CD	0.2084	0.2083	0.1979	0.1997
AE	0.2003	0.2020	0.2001	0.2006
BG	0.2057	0.2003	0.2021	0.2011
DE	0.2050	0.2006	0.1992	0.1999
BF	0.2100	0.1966	0.1962	0.1977
AC	0.2038	0.2043	0.1963	0.2017

may arise in such a situation. The design and response are given in Table 4.26. The aliasing of the linear main effects with two-factor interactions is shown in Table 4.27.

The stage 1 analysis is shown in Table 4.28 and further supported by Figure 4.4. Following similarly to Example 4, based on $PSE = 0.968$ and using the critical values in Table 4.20 for $\alpha = 0.1$, the following effects are deemed active: A, E, G, E^2, H^2, A*C, E*H, C*G, C*E, and G*H. All of the identified interactions satisfy weak effect heredity and will therefore be investigated further. Based on Figure 4.4, the interactions F*G and D*J aliased with factor B also appear to stand out aside from those identified using the Lenth test . The D*J interaction is most unlikely since neither the D or J main effects are identified as important. However, the F*G interaction does satisfy the weak effect heredity assumption. If this interaction is indeed active, then three of the four interactions aliased with factor B would be important, which is furthermore unlikely due to the interaction sparsity assumption. Thus, although we advocate plots like Figure 4.4 to gauge factor importance, the objectiveness of the Lenth-type test helps to better solidify the results.

It is worthwhile to note that upon simulating data from the same model using the 27-run OA of XCW, no linear or quadratic main effects were detected at $\alpha = 0.1$. In fact, the smallest p-value is 0.143 for the quadratic main effect of factor A. Thus, it is clear that their recommended design would be unsuccessful for this model.

In order to consider all of the possible interactions, we must project the 9-factor design onto factors A, C, E, G, and H. Table 4.29 displays the second-order model involving the 5 factors of interest ($R^2 = 0.981$). The following effects are identified as active at $\alpha = 0.05$: A, C, E, G, E^2, H^2, A*C, A*E, C*E, C*G, C*H, E*H, and G*H. Although C, A*E, C*E, and C*H are spurious effects identified as significant, it is clear that we have been successful in identifying both important main effects and interactions in this analysis using a FBBD.

Table 4.26: Design and Response for Example 5

A	B	C	D	E	F	G	H	J	Y
-1	-1	1	0	0	0	0	0	0	-5.068
-1	1	-1	0	0	0	0	0	0	1.430
1	-1	-1	0	0	0	0	0	0	-3.499
1	1	1	0	0	0	0	0	0	5.950
-1	0	0	-1	0	0	1	0	0	0.553
-1	0	0	1	0	0	-1	0	0	-3.917
1	0	0	-1	0	0	-1	0	0	1.578
1	0	0	1	0	0	1	0	0	3.669
-1	0	0	0	-1	0	0	0	1	2.795
-1	0	0	0	1	0	0	0	-1	-1.785
1	0	0	0	-1	0	0	0	-1	4.737
1	0	0	0	1	0	0	0	1	3.667
-1	0	0	0	0	-1	0	1	0	-6.393
-1	0	0	0	0	1	0	-1	0	-6.301
1	0	0	0	0	-1	0	-1	0	-1.605
1	0	0	0	0	1	0	1	0	-2.489
0	-1	0	-1	0	0	0	0	1	0.559
0	-1	0	1	0	0	0	0	-1	-0.277
0	1	0	-1	0	0	0	0	-1	-1.294
0	1	0	1	0	0	0	0	1	-0.888
0	-1	0	0	-1	0	0	1	0	-3.487
0	-1	0	0	1	0	0	-1	0	-5.572
0	1	0	0	-1	0	0	-1	0	2.085
0	1	0	0	1	0	0	1	0	0.806
0	-1	0	0	0	-1	1	0	0	0.609
0	-1	0	0	0	1	-1	0	0	-1.670
0	1	0	0	0	-1	-1	0	0	-1.402
0	1	0	0	0	1	1	0	0	2.147
0	0	-1	-1	0	0	0	1	0	-3.101
0	0	-1	1	0	0	0	-1	0	-5.635
0	0	1	-1	0	0	0	-1	0	-2.972
0	0	1	1	0	0	0	1	0	-3.876
0	0	-1	0	-1	0	1	0	0	10.735
0	0	-1	0	1	0	-1	0	0	-6.328
0	0	1	0	-1	0	-1	0	0	5.842
0	0	1	0	1	0	1	0	0	-1.420
0	0	-1	0	0	-1	0	0	1	0.240
0	0	-1	0	0	1	0	0	-1	-0.351
0	0	1	0	0	-1	0	0	-1	0.892
0	0	1	0	0	1	0	0	1	1.578
0	0	0	-1	-1	1	0	0	0	2.892
0	0	0	-1	1	-1	0	0	0	0.974
0	0	0	1	-1	-1	0	0	0	2.889
0	0	0	1	1	1	0	0	0	1.751
0	0	0	0	0	0	-1	-1	1	-8.500
0	0	0	0	0	0	-1	1	-1	-1.517
0	0	0	0	0	0	1	-1	-1	2.441
0	0	0	0	0	0	1	1	1	-5.753
0	0	0	0	0	0	0	0	0	0.926

Table 4.27: Aliasing of Example 5 Design

Linear Main Effect	Aliasing
J	+ B*D + A*E + C*F + G*H
E	+ D*F + C*G + B*H + A*J
D	+ E*F + A*G + C*H + B*J
F	+ D*E + B*G + A*H + C*J
B	+ A*C + F*G + E*H + D*J
A	+ B*C + D*G + F*H + E*J
C	+ A*B + E*G + D*H + F*J
H	+ C*D + B*E + A*F + G*J
G	+ A*D + C*E + B*F + H*J

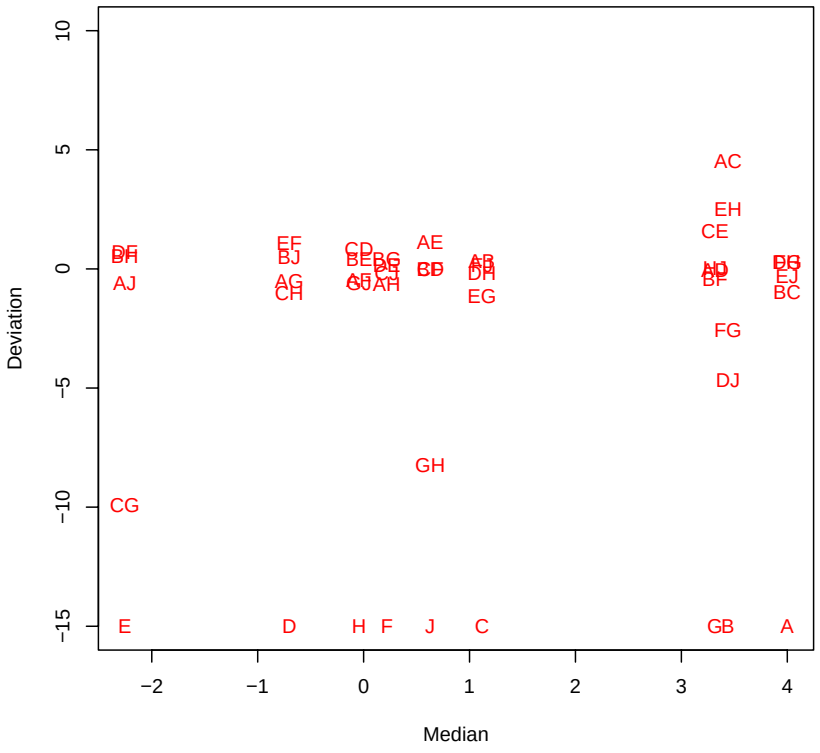

Figure 4.4: Example 5 - Analysis Stage 1

Table 4.28: Example 5 - Analysis Strategy Step 1

Term	Standardized Estimate	t-Ratio	Term	Standardized Estimate	t-Ratio
A	7.996	8.257	DH	-0.174	-0.179
B	0.000	0.000	EG	-1.110	-1.146
C	2.235	2.308	FJ	0.174	0.179
D	-1.404	-1.450	AG	-0.487	-0.503
E	-3.365	-3.475	BJ	0.487	0.503
F	0.437	0.452	CH	-1.017	-1.050
G	6.633	6.849	EF	1.090	1.125
H	-0.085	-0.088	AJ	-1.143	-1.180
J	1.277	1.319	BH	0.000	0.000
A^2	-0.829	-0.856	CG	-10.479	-10.822
B^2	-1.386	-1.432	DF	0.155	0.160
C^2	-0.619	-0.639	AH	-0.615	-0.635
D^2	-0.910	-0.940	BG	0.416	0.430
E^2	4.417	4.561	CJ	-0.171	-0.177
F^2	-0.745	-0.769	DE	0.171	0.177
G^2	-0.109	-0.113	AD	-0.036	-0.037
H^2	-9.527	-9.838	BF	-0.403	-0.416
J^2	-0.210	-0.217	CE	1.584	1.636
BC	-0.954	-0.985	HJ	0.036	0.037
DG	0.307	0.317	AF	-0.445	-0.460
EJ	-0.302	-0.312	BE	0.445	0.460
FH	0.302	0.312	CD	0.857	0.885
AC	7.973	8.233	GJ	-0.563	-0.582
DJ	-1.232	-1.272	AE	1.117	1.153
EH	5.975	6.170	BD	-0.018	-0.019
FG	0.904	0.933	CF	0.000	0.000
AB	0.358	0.370	GH	-8.227	-8.496

Table 4.29: Second-Order Analysis Results for Example 5

Term	Estimate	Std Error	t Ratio	P-value
Intercept	-0.012667	0.22602	-0.06	0.9557
A	1.918375	0.169515	11.32	$< .0001$
C	0.6183333	0.195739	3.16	0.0038
E	-1.005833	0.195739	-5.14	$< .0001$
G	1.5911667	0.195739	8.13	$< .0001$
H	0.0155625	0.169515	0.09	0.9275
A*A	-0.065979	0.215165	-0.31	0.7614
A*C	3.98675	0.33903	11.76	$< .0001$
C*C	0.0252708	0.215165	0.12	0.9073
A*E	0.8775	0.33903	2.59	0.0151
C*E	0.8590833	0.391478	2.19	0.0367
E*E	2.2056042	0.215165	10.25	$< .0001$
A*G	-0.59475	0.33903	-1.75	0.0903
C*G	-5.075417	0.391478	-12.96	$< .0001$
E*G	-0.614583	0.391478	-1.57	0.1277
G*G	0.2461042	0.215165	1.14	0.2624
A*H	-0.198	0.33903	-0.58	0.5639
C*H	-0.8595	0.33903	-2.54	0.0171
E*H	2.9875	0.33903	8.81	$< .0001$
G*H	-3.79425	0.33903	-11.19	$< .0001$
H*H	-3.831896	0.215165	-17.81	$< .0001$

As with Example 4, let us consider the loss of efficiency of the pseudo estimates of the linear main effects and two-factor interactions when simulating analysis approach 1 (see Table 4.30). The addition of the A*C interaction to the above model (4.22) with only active main effects increases the variance of factor B's linear main effect estimate from $0.092\sigma^2$ to $0.127\sigma^2$ and the variance of A*C's estimate from $0.213\sigma^2$ to $0.375\sigma^2$. We see a similar increase in the variance of the estimates of E and C*G upon adding C*G to the model. However, there is a more severe loss of efficiency when a second interaction aliased with B (E*H) is added. In particular, the variance of B's estimate increases to $0.481\sigma^2$. This loss of efficiency is due to our estimating the linear main effects based on the median of the γ_{ij}'s after pruning away those associated with large δ_{ij}'s. If, in a simulation of (4.22), the same δ_{ij}'s are always pruned, one would obtain more precise $\tilde{b}_j$'s than when the pruning of δ_{ij}'s is more erratic. For instance, for the simulated response shown in Table 4.26, all of the δ_{ij}'s associated with factor B are pruned and we were able to correctly identify the important interactions. However, a different simulated response may have led to a different estimate of factor B's linear main effect and thus, a different conclusion. □

Example 6 - Simulated Data 3

Here, we consider one more simulated data set based on the 105-run 13-factor FBBD, which has r=6. In particular, we generate a response from the following model:

$$\begin{aligned} Y &= 5x_A + 2.5x_B + 3x_D + 3x_F + 4x_J - 6x_M + 3_{A^2} - 3x_{D^2} + 3x_{J^2} - 2.5x_{BC} - 5x_{FG} \\ &- 5x_{DF} + 5x_{GJ} - 4x_{AG} + 6x_{DG} + 2.5x_{AM} - 3.5x_{FJ} + \epsilon \end{aligned} \quad (4.23)$$

where $\epsilon \sim N(0, \sigma = 2)$. This example not only considers the situation of a larger number of effects, but also investigates a larger error variance. Thus, we provide an

Table 4.30: Variance of Pseudo-Estimates - Example 5

	Null Case	Number of Interactions			
Term	(Main Effects Only)	1 (A*C)	2 (Add C*G)	3 (Add E*H)	4 (Add G*H)
A	0.0946	0.0906	0.0893	0.0855	0.0828
B	0.0919	0.1274	0.1239	0.4809	0.6180
C	0.0948	0.0909	0.0892	0.0834	0.0840
D	0.0899	0.0904	0.0880	0.0838	0.0818
E	0.0936	0.0901	0.1316	0.1280	0.1228
F	0.0904	0.0897	0.0864	0.0834	0.0831
G	0.0925	0.0932	0.0892	0.0849	0.0825
H	0.0904	0.0908	0.0879	0.0845	0.0845
J	0.0923	0.0903	0.0865	0.0838	0.1199
BC	0.2250	0.2135	0.2194	0.2051	0.2059
DG	0.2214	0.2054	0.2112	0.2089	0.2075
EJ	0.2072	0.2129	0.2206	0.2054	0.2004
FH	0.2146	0.2173	0.2061	0.2085	0.2133
AC	0.2127	0.3751	0.3686	0.7901	0.9500
DJ	0.2196	0.2023	0.2073	0.5999	0.6981
EH	0.2189	0.2134	0.2104	0.8796	1.0191
FG	0.2169	0.2114	0.2115	0.5921	0.6954
AB	0.2164	0.2116	0.2171	0.2059	0.2125
DH	0.2182	0.2158	0.2125	0.2145	0.2062
EG	0.2190	0.2181	0.2105	0.2144	0.2119
FJ	0.2243	0.2223	0.2102	0.2151	0.2070
AG	0.2144	0.2193	0.2204	0.2078	0.2075
BJ	0.2110	0.2177	0.2100	0.2059	0.2065
CH	0.2124	0.2106	0.2080	0.2135	0.2105
EF	0.2204	0.2159	0.2099	0.2090	0.2151
AJ	0.2218	0.2167	0.2206	0.2138	0.2090
BH	0.2218	0.2003	0.2059	0.2141	0.2064
CG	0.2170	0.2194	0.3857	0.3755	0.3764
DF	0.2129	0.2146	0.2116	0.2036	0.2072
AH	0.2203	0.2176	0.2103	0.2066	0.2073
BG	0.2139	0.2184	0.2117	0.2081	0.2067
CJ	0.2141	0.2139	0.2147	0.2062	0.2014
DE	0.2097	0.2142	0.2067	0.2085	0.2091
AD	0.2187	0.2161	0.2142	0.2082	0.2092
BF	0.2242	0.2178	0.2194	0.2128	0.2113
CE	0.2185	0.2283	0.2170	0.2055	0.2096
HJ	0.2146	0.2167	0.2124	0.2133	0.2030
AF	0.2162	0.2163	0.2159	0.2092	0.2081
BE	0.2167	0.2155	0.2084	0.2085	0.2096
CD	0.2159	0.2212	0.2143	0.2029	0.2102
GJ	0.2209	0.2172	0.2176	0.2092	0.2010
AE	0.2273	0.2197	0.2071	0.2120	0.2058
BD	0.2209	0.2155	0.2124	0.2020	0.2025
CF	0.2200	0.2185	0.2106	0.2081	0.2096
GH	0.2128	0.2141	0.2187	0.2070	0.3691

example in which effect sparsity holds, but not factor sparsity. Given that six linear main effects are active, we are unable to use the 27-run OA of XCW for such a model. Due to the run size, we omit the design and response in order to save space. It is available upon request from the author. Table 4.31 displays the aliasing while Table 4.32 and Figure 4.5 provide the stage 1 analysis.

The following effects are deemed important in stage 1 by the Lenth test with $PSE = 1.8895$ using the $\alpha = 0.1$ critical values: A, B, D, F, J, K, M, A^2, D^2, J^2, F*G, M*N, D*F, G*J, B*H, C*F, A*G, B*D, H*K, A*J, B*N, F*J, D*G, A*M. A smaller α-level critical value would have obviously trimmed this list down. Such trimming may have the unfortunate consequence, however, of missing out on a potentially important factor(s). We stress again that the use of a larger α allows one to be more liberal in determining factor importance. Note that all of the two-factor interactions satisfy the weak effect heredity assumption. From these results, one would proceed to project the 13-factor design onto factors A, B, C, D, F, G, H, J, K, M, and N (a feat which is clearly impossible with the designs of XCW) and fit a second-order model. The results are shown in Table 4.33. Although a few spurious interactions are declared active, we are able to identify the true model with the exception of the BC interaction (recall that this interaction is one of the smallest in size). □

Example 7 - Real Data Set

This last example illustrating approach 1 involves a real data set taken from Alvarez, et.al. [3]. In particular, they seek to optimize a very-large-scale-integrated (VLSI) process, device, and circuit design through computer simulation using a six factor, 50-run Box-Behnken design. According to Rubin [86], "such circuits are becoming increasingly common due to their ease of manufacture, low cost, and simplified design methodologies. No longer must the designer study electronics and physics to build

Table 4.31: Aliasing of Example 6 Design

Linear Main Effect	Aliasing
N	+ E*G + C*H + D*J + F*K + B*L + A*M
M	+ D*G + F*H + C*J + B*K + E*L + A*N
L	+ C*G + D*H + F*J + A*K + E*M + B*N
H	+ B*E + A*J + G*K + D*L + F*M + C*N
J	+ B*G + A*H + E*K + F*L + C*M + D*N
G	+ A*F + B*J + H*K + C*L + D*M + E*N
K	+ C*D + G*H + E*J + A*L + B*M + F*N
E	+ A*D + C*F + B*H + J*K + L*M + G*N
C	+ A*B + E*F + D*K + G*L + J*M + H*N
D	+ A*E + B*F + C*K + H*L + G*M + J*N
F	+ B*D + C*E + A*G + J*L + H*M + K*N
B	+ A*C + D*F + E*H + G*J + K*M + L*N
A	+ B*C + D*E + F*G + H*J + K*L + M*N

Table 4.32: Example 6 - Analysis Strategy Step 1

Term	Standardized Estimate	t-Ratio	Term	Standardized Estimate	t-Ratio	Term	Standardized Estimate	t-Ratio
A	19.998	10.584	GJ	9.934	5.257	CN	0.000	0.000
B	13.466	7.127	KM	-0.663	-0.351	DL	1.230	0.651
C	0.165	0.087	LN	1.208	0.639	FM	-0.098	-0.051
D	14.184	7.507	AB	1.328	0.703	GK	-1.101	-0.583
E	-3.609	-1.910	DK	-0.494	-0.261	AH	-0.984	-0.521
F	13.506	7.148	EF	1.766	0.935	BG	1.019	0.540
G	0.324	0.171	GL	-2.320	-1.228	CM	-1.749	-0.926
H	-3.474	-1.838	HN	0.494	0.261	DN	2.860	1.514
J	15.830	8.378	JM	-0.529	-0.280	EK	-0.220	-0.116
K	-4.719	-2.497	AE	1.320	0.699	FL	0.220	0.116
L	-1.458	-0.772	BF	-1.088	-0.576	AL	0.182	0.096
M	-28.806	-15.246	CK	0.402	0.213	BM	-0.604	-0.320
N	1.882	0.996	GM	0.195	0.103	CD	-0.549	-0.290
A^2	7.822	4.140	HL	-0.195	-0.103	EJ	-0.182	-0.096
B^2	-0.659	-0.349	JN	-0.826	-0.437	FN	2.270	1.201
C^2	-1.867	-0.988	AD	-2.091	-1.106	GH	1.436	0.760
D^2	-7.833	-4.146	BH	4.576	2.422	AK	0.866	0.458
E^2	-0.731	-0.387	CF	5.210	2.757	BN	4.928	2.608
F^2	-1.071	-0.567	GN	2.133	1.129	CG	-1.824	-0.966
G^2	1.361	0.721	JK	0.000	0.000	DH	-0.866	-0.458
H^2	-0.669	-0.354	LM	-0.726	-0.384	EM	0.887	0.469
J^2	7.227	3.825	AG	-6.754	-3.574	FJ	-7.874	-4.167
K^2	-1.194	-0.632	BD	-3.911	-2.070	AN	-0.961	-0.509
L^2	-0.781	-0.413	CE	0.000	0.000	BK	2.016	1.067
M^2	0.070	0.037	HM	0.013	0.007	CJ	-2.891	-1.530
N^2	-0.874	-0.462	JL	1.175	0.622	DG	11.140	5.896
BC	-0.653	-0.346	KN	-1.951	-1.032	EL	0.000	0.000
DE	2.943	1.557	AF	0.933	0.494	FH	0.672	0.356
FG	-8.127	-4.301	BJ	-0.857	-0.454	AM	6.772	3.584
HJ	0.218	0.116	CL	-0.182	-0.096	BL	-0.158	-0.084
KL	-0.218	-0.116	DM	0.182	0.096	CH	0.592	0.314
MN	5.215	2.760	EN	-1.730	-0.916	DJ	0.391	0.207
AC	-1.568	-0.830	HK	3.726	1.972	EG	-2.759	-1.460
DF	-11.003	-5.824	AJ	4.440	2.350	FK	0.000	0.000
EH	0.663	0.351	BE	0.496	0.263			

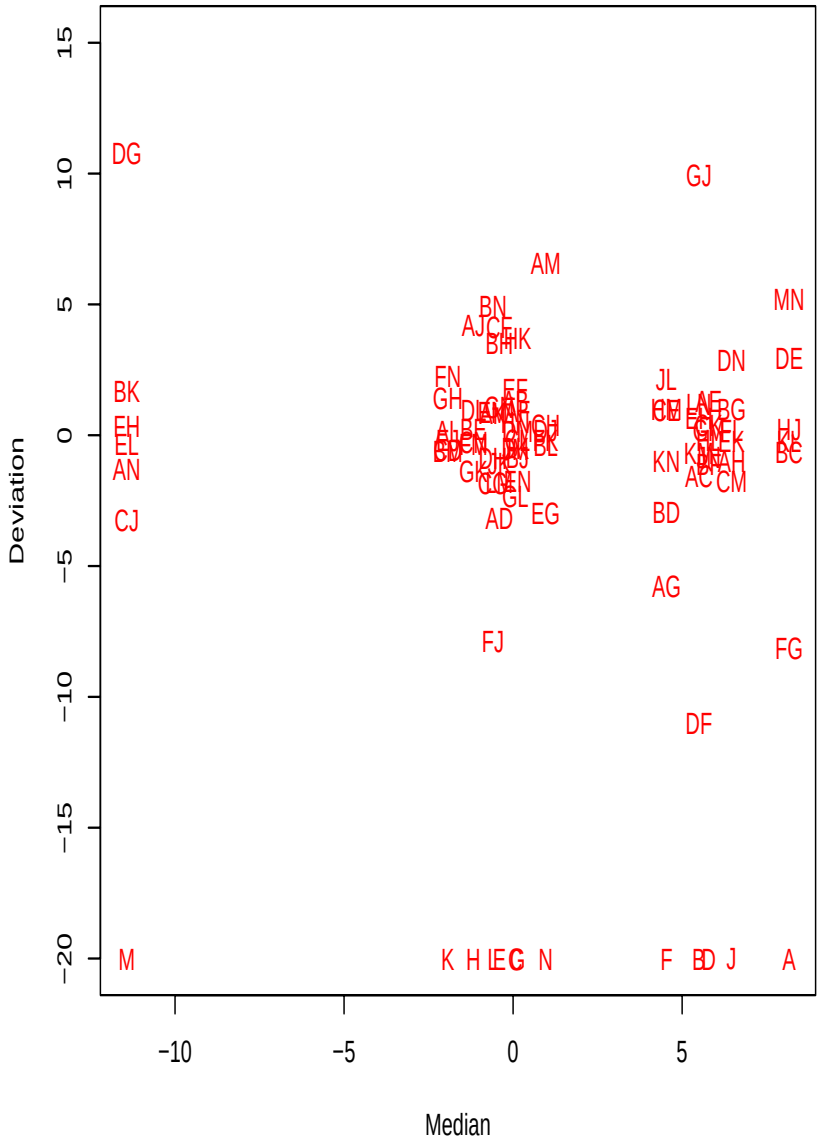

Figure 4.5: Example 6 - Analysis Stage 1

Table 4.33: Second-Order Analysis Results for Example 6

Term	Estimate	Std Error	t Ratio	P-value	Term	Estimate	Std Error	t Ratio	P-value
Intercept	-0.097008	0.668896	-0.15	0.8858	H*H	-0.021784	0.43446	-0.05	0.9604
A	4.7632537	0.579281	8.22	< .0001	A*J	1.7884363	1.003343	1.78	0.0859
B	3.2163437	0.579281	5.55	< .0001	B*J	0.0493525	1.003343	0.05	0.9611
C	-0.104712	0.579281	-0.18	0.8579	C*J	-1.4454	1.158561	-1.25	0.2229
D	3.1766175	0.579281	5.48	< .0001	D*J	0.9249463	1.003343	0.92	0.3648
F	3.0506	0.579281	5.27	< .0001	F*J	-4.234725	0.819226	-5.17	< .0001
G	-0.412078	0.579281	-0.71	0.483	G*J	4.4992813	1.003343	4.48	0.0001
H	-0.277486	0.579281	-0.48	0.6358	H*J	-0.571904	1.003343	-0.57	0.5734
J	3.2312937	0.579281	5.58	< .0001	J*J	3.1198044	0.43446	7.18	< .0001
K	-0.963244	0.579281	-1.66	0.1079	A*K	0.135385	0.819226	0.17	0.87
M	-5.880075	0.819226	-7.18	< .0001	B*K	1.008275	1.158561	0.87	0.3918
N	-0.345154	0.579281	-0.6	0.5562	C*K	-0.080292	1.003343	-0.08	0.9368
A*A	3.3567804	0.43446	7.73	< .0001	D*K	-0.108413	1.003343	-0.11	0.9148
A*B	0.802405	1.003343	0.8	0.4308	F*K	0.7293288	1.003343	0.73	0.4735
B*B	-0.017941	0.43446	-0.04	0.9674	G*K	-0.982264	1.003343	-0.98	0.3363
A*C	-1.251686	1.003343	-1.25	0.2229	H*K	2.3409775	1.003343	2.33	0.0273
B*C	-1.007661	1.003343	-1	0.3241	J*K	-0.736747	0.819226	-0.9	0.3764
C*C	-0.498673	0.43446	-1.15	0.2611	K*K	-0.230855	0.43446	-0.53	0.5995
A*D	-1.782107	0.819226	-2.18	0.0385	A*M	4.1148788	1.003343	4.1	0.0003
B*D	-2.249375	1.003343	-2.24	0.0334	B*M	-0.301856	1.003343	-0.3	0.7658
C*D	-0.274281	1.003343	-0.27	0.7867	C*M	-0.874819	1.003343	-0.87	0.3909
D*D	-2.872686	0.43446	-6.61	< .0001	D*M	0.5692525	1.003343	0.57	0.5752
A*F	0.9446525	1.003343	0.94	0.3548	F*M	-0.480364	1.003343	-0.48	0.636
B*F	-0.825592	1.003343	-0.82	0.4178	G*M	-0.184142	1.003343	-0.18	0.8558
C*F	1.868275	0.819226	2.28	0.0307	H*M	-0.28705	1.003343	-0.29	0.777
D*F	-5.969319	1.003343	-5.95	< .0001	J*M	-0.125812	1.003343	-0.13	0.9011
F*F	-0.181935	0.43446	-0.42	0.6787	K*M	-0.798794	1.003343	-0.8	0.4329
A*G	-3.670675	1.003343	-3.66	0.0011	M*M	0.2721804	0.43446	0.63	0.5363
B*G	0.5096813	1.003343	0.51	0.6156	A*N	-0.4806	1.158561	-0.41	0.6815
C*G	-1.2098	0.819226	-1.48	0.1513	B*N	2.1665125	0.819226	2.64	0.0135
D*G	5.5698	1.158561	4.81	< .0001	C*N	-0.431586	1.003343	-0.43	0.6705
F*G	-4.744679	1.003343	-4.73	< .0001	D*N	1.4300638	1.003343	1.43	0.1655
G*G	0.7860134	0.43446	1.81	0.0816	F*N	1.1351688	1.003343	1.13	0.2678
A*H	-0.491844	1.003343	-0.49	0.6279	G*N	0.329635	0.819226	0.4	0.6906
B*H	1.551	0.819226	1.89	0.0691	H*N	0.3852292	1.003343	0.38	0.704
C*H	1.025516	1.003343	1.02	0.3158	J*N	-0.694525	1.003343	-0.69	0.4947
D*H	-0.730672	0.819226	-0.89	0.3803	K*N	-1.269325	1.003343	-1.27	0.2166
F*H	0.336025	1.158561	0.29	0.774	M*N	1.9265713	1.003343	1.92	0.0655
G*H	0.7179938	1.003343	0.72	0.4804	N*N	-0.103312	0.43446	-0.24	0.8138

an integrated circuit. Digital electronic design is taught widely and is accessible to people with any scientific background."

Unfortunately, the authors are somewhat vague about the factors under study. Therefore, we shall simply refer to them as A-F. In summary, the authors identify the main effects (both linear and quadratic) of factors D and E and the D*E interaction as active. They report the coefficients of this second-order model and little else. Table 2 of their paper displays only half of the Box-Behnken design used and response for their experiment (the remaining half is omitted from the article). We suspect they do so for brevity as no other indication as to why the rest of the data is missing is provided. Regardless, we intend to use only the available Box-Behnken half to illustrate our proposed analysis strategy.

The 25-run FBBD and response is given in Table 4.34 and the aliasing is provided in Table 4.35. Note that the 6-factor FBBD ($r = 3$) is based on a regular graph design with $\lambda_1 = 2$ and $\lambda_2 = 1$ instead of a BIBD. Thus, we see three interactions that appear twice in the aliasing structure. This structure, however, does not deter us from analyzing the design in the exact same manner as done with the previous examples, which were based on use of BIBDs. Table 4.36 and Figure 4.6 displays the results of a stage 1 analysis based on $PSE = 2.52$. Again, utilizing the critical values in Table 4.20 for $\alpha = 0.1$, we find the following effects to be important: D, E, D^2, E^2, D*E, B*E, A*E, A*D. The same linear main effects and interactions also stand out in Figure 4.6. All of the identified interactions satisfy weak effect heredity.

We now proceed to project the 6-factor FBBD onto factors A, B, D, and E and fit a second-order model in these four factors. The results are given in Table 4.37. This fitted model has an R^2=0.9977 and identifies D, E, D^2, E^2, A*E, and D*E at a α=0.01 level of significance. A reduced model with only these five terms produces an R^2 of 99.2%. Therefore, we are able to identify the same effects as Alvarez, et.al [3]

Table 4.34: Design and Response for Example 7

A	B	C	D	E	F	Y
1	1	0	-1	0	0	107.8
1	-1	0	1	0	0	44.46
-1	1	0	1	0	0	45.01
-1	-1	0	-1	0	0	109.32
0	1	1	0	-1	0	84.08
0	1	-1	0	1	0	65.36
0	-1	1	0	1	0	66.25
0	-1	-1	0	-1	0	83.58
0	0	1	1	0	-1	45.3
0	0	1	-1	0	1	112.4
0	0	-1	1	0	1	45.51
0	0	-1	-1	0	-1	107.15
1	0	0	1	-1	0	58.84
1	0	0	-1	1	0	117.76
-1	0	0	1	1	0	45.31
-1	0	0	-1	-1	0	153.08
0	1	0	0	1	-1	65.05
0	1	0	0	-1	1	89.93
0	-1	0	0	1	1	65.25
0	-1	0	0	-1	-1	79.98
1	0	1	0	0	-1	64.12
1	0	-1	0	0	1	64.03
-1	0	1	0	0	1	65.19
-1	0	-1	0	0	-1	63.4
0	0	0	0	0	0	63.07

Table 4.35: Aliasing of Example 7 Design

Linear Main Effect	Aliasing
A	-B*D -D*E -C*F
C	-B*E -A*F -D*F
B	-A*D -C*E -E*F
D	-A*B -C*F -A*E
E	-B*C -A*D -B*F
F	-C*D -B*E -A*C

Table 4.36: Example 7 - Analysis Stage 1

Term	Standardized Estimate	t Ratio
A	-1.087	-0.431
B	-0.338	-0.134
C	1.628	0.646
D	-111.019	-44.056
E	-34.303	-13.612
F	4.728	1.876
A^2	2.632	1.045
B^2	-4.931	-1.957
C^2	-0.467	-0.185
D^2	28.501	11.310
E^2	28.068	11.138
F^2	-0.013	-0.005
BD	-0.408	-0.162
DE	-10.267	-4.074
CF	0.408	0.162
AD	-0.290	-0.115
CE	0.000	0.000
BE	5.070	2.012
BE	-0.245	-0.097
DF	1.580	0.627
AF	0.000	0.000
AB	0.273	0.108
CF	-0.273	-0.108
AE	-19.248	-7.638
BC	1.780	0.706
AD	-4.620	-1.833
BF	0.000	0.000
CD	0.000	0.000
BE	2.345	0.931
AC	-1.880	-0.746

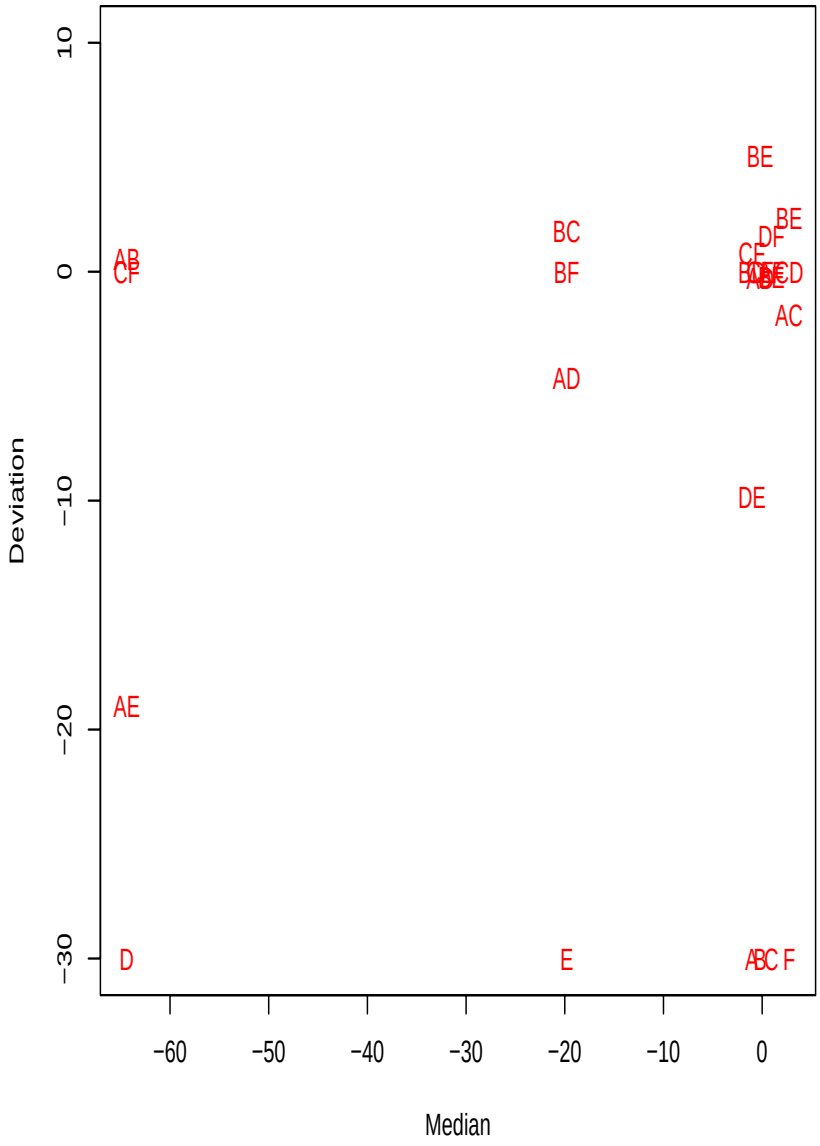

Figure 4.6: Example 7 - Analysis Stage 1

Table 4.37: Second-Order Analysis Results for Example 7

Term	Estimate	Std Error	t Ratio	P-value
Intercept	62.780968	0.974543	64.42	$< .0001$
A	-0.11	1.025423	-0.11	0.9167
B	1.39375	0.661908	2.11	0.0615
D	-32.185	1.025423	-31.39	$< .0001$
E	-9.68125	0.661908	-14.63	$< .0001$
A*A	1.4522043	0.908496	1.6	0.141
A*B	-0.2725	1.450167	-0.19	0.8547
B*B	-2.466962	0.908496	-2.72	0.0217
A*D	2.08375	0.888042	2.35	0.0409
B*D	0.4075	1.450167	0.28	0.7844
D*D	14.857204	0.908496	16.35	$< .0001$
A*E	9.4875	1.450167	6.54	$< .0001$
B*E	-1.4425	0.725083	-1.99	0.0747
D*E	5.3375	1.450167	3.68	0.0042
E*E	14.633038	0.908496	16.11	$< .0001$

in half as many runs. □

4.3.2 Analysis Approach 2

We have seen through example that the proposed approach 1 to analysis can be successful in identifying important effects. However, we have also seen how a loss of efficiency of the pseudo estimates occurs due to a lack of interaction sparsity and can lead to potentially ambiguous results. Furthermore, more Type I errors than expected are likely to occur. Therefore, we now outline our second approach to analysis for those less comfortable with assuming strong interaction sparsity. This approach, however, does not allow for individual testing of each effect. Rather, it is developed in the spirit of group factor screening.

The essential characteristic of group screening is the intentional confounding of effects with the aim of reducing the number of runs required to identify those that are most important. Upon identifying important groups of effects, additional runs may be added to estimate those of interest. See Morris [73] for more details on group screening. By constructing FBBDs as described, we do intentionally (partially) alias each linear main effect with r two-factor interactions. Approach 2 proceeds as follows:

1. Compute PSE as described in Steps 1-4 of approach 1. (Note that we compute PSE from $2\gamma_{ij}$ so that PSE is an estimate of σ).

2. Compute the standard deviation of each column of Γ. Denote this as s_j.

3. Compute the statistics, $U_j = s_j/PSE$. If $U_j > U_\alpha$, then it is likely that one or more interactions in the jth group are important. The critical values, U_α, which were determined via simulation are given in Table 4.38. A large standard deviation likely indicates the presence of one or more interactions.

4. Same as step 6 of approach 1. That is, compute least squares estimates of the quadratic main effects and construct t-like statistics using the PSE. One can use the critical values, $t_{\alpha}^{(q)}$, in Table 4.20 to determine if factor j's quadratic main effect is worthy of further consideration.

5. Construct a plot of the δ_{ij}'s vs. the m_j's. Use this plot as a gauge to determine likely linear main effects and interactions.

6. Using expert opinion and/or effect heredity, if possible, remove unlikely interactions from the groups deemed important.

7. Based on the remaining interactions:

   - If desired, augment the FBBD with carefully chosen 2^{3-1} fraction(s) in order to estimate and test the interactions in those groups under consideration.
   - Project the FBBD onto those factors likely to be important as determined in the above steps.

We now briefly revisit several of the previous examples to illustrate the second approach.

Table 4.38: Critical Values for Approach 2

t	$U_{0.01}$	$U_{0.05}$	$U_{0.1}$
4	5.41	2.80	2.02
5	4.95	2.60	1.92
6	4.29	2.44	1.85
7	3.89	2.34	1.81
9	3.18	2.13	1.74
10	2.34	1.77	1.53
11	2.28	1.74	1.51
12	2.25	1.72	1.50
13	2.19	1.70	1.49

Example 8 - Example 5 Revisited

Recall from Example 5, that we obtain PSE=0.968 for the simulated data. Table 4.39 displays s_j and $U_j = s_j/PSE$. Then, based on the critical values in Table 4.38, we see that interactions aliased with factors B, E, and J are most likely at $\alpha = 0.05$. For instance, the B linear main effect is partially aliased with A*C, F*G, E*H, and D*J. Refer to Table 4.27 to recall the full aliasing of this design. Furthermore, as in approach 1, using the critical values of Table 4.20 for quadratic main effects, we see E^2 and H^2 stand out. Therefore, factors E and H are worthy of further consideration. Finally, using Figure 4.4, we see the most likely linear main effects and interactions to be A, B, E, G, C*G, G*H, A*C, E*H, F*G, and D*J.

By relying on the main effects identified above and weak effect heredity, one may be able to discount some of the interactions aliased with B, E, and J. In particular, none of D*J, D*F, or C*F satisfy weak effect heredity. Thus, if desired, one may now proceed to project the FBBD onto factors A, B, C, E, F, G, and H and fit a second-order model in these seven factors (we omit these results). On the other hand, one may not be comfortable with relying mainly on the results of Figure 4.4 to make a judgment regarding important interactions. Therefore, augmenting the FBBD with

Table 4.39: Example 8 - Example 5 Approach 2

Factor Group	s_j	U_j
A	0.600	0.620
B	4.291	4.433
C	0.653	0.674
D	0.948	0.979
E	5.108	5.277
F	0.447	0.462
G	0.880	0.909
H	0.690	0.713
J	4.329	4.472

additional 2^{3-1} fractions would provide a means to estimate the desired interactions. For example, suppose that we still continue to discount D*J, D*F, and C*F. There are now three interactions left to estimate within each group. To do so requires seven additional half fractions (28 runs)(i.e. J=-BD, J=-AE, J=-GH, E=-CG, E=-BH, G=-AB, G=-BF). Although the new design now has (49+28)=77 runs, it is still fewer than the 96+ runs required for a Box-Behnken design. Upon estimating the desired interactions and determining their importance, one may then proceed to project the (now 77-run) FBBD onto the factors of interest. □

Example 9 - Example 6 Revisited

Example 6 explored approach 1 with a 13-factor 105-run FBBD. Refer back to Table 4.31 to view the aliasing structure of this design. For the simulated data, we computed PSE=1.89 (slightly less than the true $\sigma = 2$). Table 4.40 displays the s_j's and U_j's. At $\alpha = 0.05$, we see that one or more interactions partially aliased with factors A, B, L, and M are likely to be active. If one also considers $\alpha = 0.1$ critical values, we add factors E, F, and N to the list. Thus, one might conjecture that those interactions partially aliased with factors E, F, and N are not as likely to be active. Recall from Example 6 that the quadratic main effects A^2, D^2, and J^2 are declared active using the critical values of Table 4.20. Finally, Figure 4.5 indicates the likely presence of the following linear main effects and two-factor interactions: A, B, D, F, J, M, D*G, D*F, F*J, F*G, G*J, and perhaps A*G and A*M.

Using the same approach as in Example 8, we may consider using the above information to disregard interactions that do not satisfy the weak effect heredity assumption or expert opinion. For instance, one may find the interaction K*L aliased with factor A unlikely. Likewise, E*H and L*N aliased with factor B would be unlikely. Augmenting this design with additional 2^{3-1} fractions in order to estimate

Table 4.40: Example 9 - Example 6 Approach 2

Factor Group	s_j	U_j
A	4.525	2.395
B	6.696	3.544
C	1.485	0.786
D	0.877	0.464
E	2.957	1.565
F	2.978	1.576
G	1.888	0.999
H	1.928	1.020
J	1.620	0.857
K	1.171	0.620
L	4.228	2.238
M	4.925	2.607
N	3.165	1.675

those interactions aliased in likely groups would require an additional 76 runs (19 additional fractions) assuming that we only wish to consider interactions aliased with factors A, B, L, and M. The now 181-run FBBD still has fewer runs than a 210+ run Box-Behnken design in 13 factors. With so many additional runs required, it may be advantageous in such situations to be contently advised regarding interactions by plots such as Figure 4.5.

Before concluding this example, it is worthwhile to point out that in the first approach, the Lenth-type significance test for interactions identified the spurious A*J, H*K, C*F, and B*H as active. However, by using the second analysis approach, the factor groups containing these interactions were not identified as likely. These interactions also did not stand out in Figure 4.5. □

Example 10 - Example 7 Revisited

Our last example will reconsider the real data of Example 7. Table 4.41 displays the standard deviations and their corresponding t-like statistics based on PSE=2.52. Based on these results, we see that interactions partially aliased with factors A and D are most likely at $\alpha = 0.1$. Figure 4.6 indicates that the following effects are likely active: D, E, A*E, A*D, B*E, and D*E. Note, however, that B*E and A*E do not fall within either group of three interactions aliased with A and D. Thus, we remove these from consideration.

Abiding by weak effect heredity (i.e. based on D and E main effects as active), one may conclude that C*F and A*B are also unlikely to be active. Therefore, one may decide to simply project the FBBD onto factors A, B, D, and E and fit a second-order model. On the other hand, if we augment the 25-run design with two 2^{3-1} fractions (i.e. E=AD and D=AB) we may estimate the interactions (B*D, A*E, and D*E) of interest and subsequently determine their significance. Based on the results

Table 4.41: Example 10 - Example 7 Approach 2

Factor Group	s_j	U_j
A	5.942	2.358
B	3.014	1.196
C	0.991	0.393
D	11.116	4.411
E	3.303	1.311
F	2.117	0.840

of Example 7, we know that B*D is not active. Assuming we reach that conclusion through augmentation, one would project the now (25+8=)33-run FBBD onto factors A, D, and E. □

4.4 Discussion

In this chapter, we have proposed new three-level designs for the purpose of both screening and response surface exploration. These designs are intended to compete with the existing designs proposed by CW and XCW, who recommend the use of 18 and 27-run nonregular orthogonal arrays. Although of good run size economy, CW and XCW's designs suffer from two drawbacks:

- Inability to project onto a large number of factors (i.e. > 5).
- Inability to identify important factors using their recommended main-effects only analysis.

We have attempted to address both of these issues with the new designs.

The Fractional Box-Behnken Designs (FBBDs) are three-level designs based on BIBDs or RG designs with $\lambda_i \geq 1$, $k = 3$, and $b \geq t$. In particular, we construct these designs with subsets of resolution III 2^{3-1} designs. This construction method allows for r independent estimates of each factor's linear main effect. Since each

of the rt estimates is aliased with one two-factor interaction, it is clear that each linear main effect is partially aliased with r two-factor interactions. Although the restriction to such RG and BIBDs requires more runs (especially for $t \geq 10$) than the CW/XCW designs, their simple construction and aliasing structure allows for an improved analysis strategy that cannot be applied to the CW/XCW designs.

One suggestion for constructing FBBDs with smaller run sizes would be to utilize RG designs that allows for some pairs of factors to not appear together in any subset (i.e. $\lambda_2 = 0$). For instance, rather than requiring the 10-factor FBBD to have 81 runs, it would be possible to construct one with 41 distinct runs using a RG design with $r = 3$, $k = 3$, and $b = 10$. The analysis of these smaller designs could be performed similarly to that proposed in this chapter. However, if two factors of interest do not appear together in any subset of 4-runs, their respective two-factor interaction will not be estimable. Thus, augmentation with additional runs to allow such estimation will be required. On the other hand, if the practitioner has some prior knowledge regarding which interactions may be most important, then those factors could be assigned to columns in such a way to guarantee that they appear together in at least one subset of runs. This may prove to be a worthy topic of future research.

When compared to existing designs in terms of number of eligible projections and average D-efficiency of eligible projections, we see that the FBBDs compete well with its counterparts. For instance, whereas the 27-run OA can project onto up to 5 factors of interest, the FBBDs can project up to $t - 1$ factors. This increased eligibility is one advantage of FBBDs, especially when factor sparsity is not expected to be valid. Comparisons in terms of D-efficiency show favorable results for both the OAs and the FBBDs depending on the number of factors and projection size. These results provide us with an indication that the CW/XCW designs do have good model discrimination ability provided that the the number of important factors is known.

Simulation results indicate, however, that the FBBDs are superior to the CW/XCW designs in terms of their ability to identify important factors when performing a main-effects only analysis (the strategy suggested by those authors). For instance, for $t = 7$ factors (three of which are important) and a error standard deviation of $\sigma = 2$, the 27-run OA can only identify the three important factors 21.4% of the time whereas the FBBD can do so 88.1% of the time (see Table 4.18).

Regardless of the ability of the FBBDs to identify important factors using a main-effects only analysis, we propose a new analysis strategy that takes advantage of the useful structure of these designs. Our first approach to analysis provides a method to obtain pseudo-estimates of each linear main effect and two-factor interaction (quadratic main effects are unbiased and can be obtained using ordinary least squares) and conduct tests of significance for each effect based on a modified Lenth test, which is based on the interaction sparsity assumption. However, a loss of efficiency of the pseudo estimates occurs due to a lack of interaction sparsity and can lead to potentially ambiguous results as well as more Type I errors than expected. Thus, approach 1 is most appropriate when strong interaction sparsity holds.

A useful graphical tool is also proposed that displays the median of the r estimates of each linear main effect on the x-axis and each estimates deviation from this median on the y-axis. Deviations far from zero tend to indicate the presence of large two-factor interactions. As an alternative to conducting Lenth-type tests using pseudo-estimates, a second (simpler) approach is also proposed in which groups of interactions are declared likely or unlikely based on the standard deviation of the r estimates of each factor's linear main effect. This approach is recommended to handle cases when strong interaction sparsity is not expected to hold.

The pseudo-estimate of a linear main effect computed in analysis approach 1 hinges on the assumption of interaction sparsity within each group of r estimates.

That is, it is unlikely that we will have more than $r/2$ active interactions within each group. For instance, with $r = 3$, one would expect no more than one active two-factor interaction within each group. Likewise, with $r = 6$, no more than 3 two-factor interactions are expected. Once those estimates associated with large deviations from the original median are pruned away, the median of the remaining estimates should provide a reasonable enough estimate of the linear main effect for determining factor importance.

If the practitioner has available any prior information regarding the size or sign of certain interactions, one could utilize this information to better aid in identifying important effects. For instance, one could be certain to assign large interactions to different groups of aliased interactions. Doing so could possibly help alleviate any ambiguity caused by having two or more large interactions in the same group. Such ambiguity could lead to the unfortunate consequence of improperly identifying important main effects or other interactions. Furthermore, there is some expectation that the practitioner will consider the use of confirmation runs or follow-up experimentation to verify results. Since the aliasing of the FBBDs is simple, additional runs would likely help untangle any previous confusion regarding choice of factors.

The designs in Table 4.12 and the critical values in Table 4.20 are based on only one center point run. However, as we noted in Chapter 2, the addition of center point runs provides a model independent estimate of the error variance, σ^2. For a small number of center point runs, say 2-3, this estimator may be too imprecise as discussed in Chapter 2. Thus, one could combine the mean square pure error with the pseudo standard error (computed as in the previous section) in the same manner as done in Chapter 2 by weighting each estimate by its df (or pseudo df). Additional tables of critical values could then easily be developed based on this combined standard error and the number of center point runs.

On the other hand, four or more center point runs should allow for a precise enough estimate of σ^2 to warrant its use alone over any combined estimator. In this case, one need only skip steps 3 and 4 in the analysis strategy of section 3 and replace the pseudo estimate of the linear main effects by $\tilde{b}_j = median_{|\delta_{ij}|<2.5\hat{\sigma}}(\Gamma^{(j)})$ where $\hat{\sigma}$ is the pure error estimate of σ. As with the combined case, it would be a straightforward task to construct tables of critical values for these designs with four or more center point runs. We leave this task for a later time.

As with any analysis strategy, more case studies should be investigated to see how results can be affected. It may also be interesting to see if the structure of the FBBDs could be utilized in a Bayesian context for determining factor importance. As mentioned earlier, XCW suggests that a Bayesian analysis could provide an alternative to the main effects only analysis. However, there are no examples of such a strategy being conducted for their designs in the literature. It would be worthwhile to investigate the Bayesian approach for the CW/XCW designs as well as the FBBDs. This may also prove to be very interesting topic for research.

By proposing designs that can be easily constructed, possess a simple aliasing structure, and have a straightforward analysis method, we hope that practitioners will find the FBBDs a useful alternative to the existing designs recommended for screening and response surface exploration. Further study regarding FBBDs is planned.

Chapter 5

Concluding Remarks

It has been the purpose of this dissertation to propose and illustrate new design and analysis techniques for screening experiments in which one of the underlying assumptions is effect sparsity. The discussions in the previous chapters have been as concise as possible for the purpose of creating a setting conducive for the practitioner and to keep the dissertation from being too long. However, if interested, the reader is encouraged to refer to the appropriate cited sources to obtain extra details. We now provide some concluding remarks with suggestions for future research.

Lenth's method for analyzing unreplicated experiments has become a popular method since its introduction due to its simplicity and favorable properties. In particular, a recent search on Web of Science revealed that Lenth [53] has been cited approximately 140 times in both statistical and non-statistical journals ranging from chemistry to animal science. In Chapter 2, we proposed the use of empirically determined p-values for Lenth t-statistics to aid in making the analysis of unreplicated experiments using Lenth's method more straightforward and user-friendly. Since no t-distribution provides a reasonable approximation to Lenth's t-statistics, the use of Monte Carlo p-values is an appropriate alternative choice. Although tables of critical values exist (e.g. Ye and Hamada [99]), they have only been developed to handle certain cases with uncorrelated parameter estimates. We have shown through example

how correlated estimates can adversely affect the results of an analysis. Therefore, we recommend the use of the Monte Carlo p-values for ordinary least squares estimates and any saturated design scenario.

Although Lenth's method has been proposed for unreplicated experiments, we have investigated the case of limited replication in which an estimate of the error variance could be obtained, but will likely be too imprecise to be useful. In such cases, we suggest computing Monte Carlo p-values using an estimator that combines the pure error mean square with a modified Lenth's pseudo standard error (PSE). Two mixing proportions, π^* and ω^* are suggested for use in computing the combined standard error and modified PSE, respectively. Our results indicate that the use of these combined estimates increases power substantially over the use of the pure error mean square alone. Furthermore, power simulations indicate that the use of the combined estimators perform well in situations where effect sparsity is mildly violated. If the PSE^2 were indeed a multiple of a chi-square random variable, then π^* would be optimal. On the other hand, our choice of ω^* is based on simulation and observational evaluation. Thus, we make no claim as to the optimality of ω^* and further investigation is required in order to make this claim. Unless other choices of ω^* show a reasonable increase in power over our current choice, it is unlikely that such an investigation is worthy of future research.

With more factors than runs, the analysis of supersaturated designs (SSDs) has only more recently become a topic of continued interest. When many factors are under consideration and experimental runs are expensive, they have the appearance of a worthwhile design choice. The analysis of these designs requires the assumptions of effect sparsity and additivity. While many published articles discuss the construction of SSDs, the development of useful methods of analysis is still a "work in progress". This is likely due to the difficulty involved in analyzing these designs due to the $\mathbf{X}'\mathbf{X}$

matrix being singular, where $\mathbf{X}$ denotes a main-effects-only model matrix. Therefore, even the main effects cannot be estimated independently of each other.

In our initial investigation of SSDs, we attempted to develop a Lenth-type procedure based on using the simple regression estimate of each factor. However, these simple estimates were too biased to prove useful and thus the Lenth approach was abandoned. Instead, in Chapter 3, we proposed an analysis strategy based on all-subsets regression and permutation tests to better ascertain factor importance. For instance, while Lin's [61] forward selection analysis of a SSD involving an AIDS study identified 11 active factors out of 138, our analysis strategy suggests that it is unlikely that anything useful can be learned from this SSD.

In general, we make the following conclusions/recommendations regarding the analysis of SSDs:

- As the number of factors under study increases, the perceived systematic variation explained by the fitted model also increases. True effects may potentially be masked in SSDs with a large number of factors vs. run size.

- For a moderate to large number of factors (say, less than 100), all-subsets regression should be used instead of forward-selection. The ability to entertain many more possible models provides one clear advantage.

- The use of permutation tests to conduct a global test of model significance provides a sensible way to determine if a SSD is capable of finding active factors or if nothing useful can be found. When feasible, all-subsets regression should be used to conduct the global test. However, forward selection provides a lower bound on p. Thus, if a global p-value using forward selection is large (say larger than 10%), there is no need to proceed with all-subsets permutation calculations.

- Even bias-corrected information criteria such as AIC_c are not resistant to model overfitting when using forward selection or all-subsets. One should be wary about the choice of models chosen by such criterion. We proposed a permutation-based modification to help prevent overfitting.

- Partial p-values for individual factors can be computed using permutation procedures. However, the Bonferronni approximation is surprisingly accurate and can be utilized to obtain similar results.

- There is much literature available pertaining to the optimality properties of SSDs. Given that it is clear that the choice of SSD impacts which effects are chosen to be active, it would be advantageous to determine if there is a useful link between optimal SSDs and power of analysis.

- Although we have proposed a strategy to aid in SSD analysis, one should continue to be careful regarding their use. It is still possible to be misled.

As a final note regarding SSDs, in Steinberg [91], a recent article regarding the future of industrial statistics, the following comment was made: "The numerous articles in recent years presenting even more elaborate, complex multivariate statistical process control methods and supersaturated design and analysis methodology for dozens of experimental factor are, in my opinion, good examples of the disconnect between the outside view of what researchers think might be useful and the reality of what actually is useful." Thus, although continued research in SSD analysis is necessary for an increased understanding of how these designs may be useful, more real data case studies should be investigated rather than the simpler simulated data sets with small error variance currently utilized as "benchmark data" in the literature for gauging the ability of a SSD to identify important factors.

Cheng and Wu [22] (CW) and Xu, Cheng, and Wu [97] (XCW) present orthogonal arrays (OAs) that serve the dual purposes of factor screening and response surface exploration. Their main focus lies with 18-run and 27-run OAs for screening up to 7 and 13 factors, respectively. Although a novel idea, in Chapter 4, we have illustrated how their designs suffer from two shortcomings:

- Inability to identify important factors during the screening stage.
- Projection onto the factors of interest does not always yield a second-order design.

To address these shortcomings, we propose a new class of three-level designs known as Fractional Box-Behnken Designs (FBBDs) using resolution III subsets of four runs. As their name suggests, they are based on subsets of Box-Behnken designs and are constructed using incomplete block designs. One main advantage of these designs over the CW/XCW designs is that they, for t factors under consideration, have the ability to estimate a 2nd order model in any $t-1$ factors. Thus, it is not necessary to require the assumption of factor sparsity. Furthermore, we propose and illustrate an analysis strategy that enables one to better identify the important factors before the projection step. Based on this strategy, we construct a table of critical values for each design with only one center point run. Our simulated and real examples indicate that this strategy should perform well in practice.

We make the following recommendations/conclusions with regard to FBBDs:

- FBBDs provide a viable alternative to the OAs of XCW for the purpose of implementing the two-stage strategy of CW.
- More real case studies involving Box-Behnken designs would help to better examine the proposed analysis strategy and consider what situations most likely arise in practice.

- Further study is required to see what properties FBBDs for $t = 9-13$ with fewer runs possess and if they may prove to be as useful as their larger counterparts.

- Critical value tables should be developed for cases of limited replication of center point runs. Combined estimators as proposed in Chapter 2 should be utilized in this situation.

- Although mentioned as a possible analysis strategy by XCW, there are very few examples in the literature that utilize a Bayesian approach to analyze designs with complex aliasing. A Bayesian analysis should be investigated further for both the FBBDs and the OAs of XCW in order to see what benefit this approach may bring. In fact, the final comments we offer involve the Bayesian approach to model selection.

Throughout this dissertation, we have either utilized or attempted to utilize Lenth's method for obtaining an estimate of scale in order to conduct tests of hypotheses for determining the importance of factorial effects when no model independent estimate of error variance is available. However, unless we are able to obtain unbiased estimates for the model coefficients of interest and we can assume that the error variance is relatively small in the experimental region, it is unlikely that Lenth's method will perform well. As mentioned earlier, a Lenth test was attempted for SSDs, but did not produce useful results.

The Bayesian approach to model selection has been criticized as being too computationally intensive and for requiring the specification of prior information. On the other hand, it has been recommended for screening experiments with complex aliasing as a way to deal with model uncertainty. It was termed an "elaborate procedure" by XCW, but suggested anyway to circumvent the problem of large bias in the main effects only, stage 1 analysis (XCW provides no example, however).

Box and Meyer [17] (BM) and Chipman, Hamada, and Wu [23] (CHW) have proposed Bayesian strategies to analyze designs with complex aliasing. BM's strategy is factor focused and is summarized briefly as follows. Suppose there are t factors of interest. The main idea is to evaluate the posterior probability of 2^t models, each of which corresponds to a subset of the t candidate factors and includes all main effects and interactions of these factors. The posterior model probabilities are computed directly, which can be computationally intensive. Active factors are identified using marginal posterior probabilities (i.e. the sum of posterior probabilities for all the models containing a particular factor).

CHW describe BM's approach as factor based in that effects involving a certain factor are forced into the model if that factor is included in the model. CHW's procedure is termed effect focused in that individual effects have freedom to be included in the model or not. Using a Gibbs algorithm, the CHW method is able to compute marginal posterior probabilities as well as posterior probabilities of models. Their method requires less computation than an all-subsets or exhaustive search because the search through the model space is done stochastically. Also, the search is focused on a class of more reasonable models through the specification of hierarchical priors. XCW suggests CHW's approach while Ye, Tsai, and Li [100] suggest the approach of BM. It is clear, then, that both of these methods require some comparison and illustration in order to determine their advantages and disadvantages. More research is highly recommended for this area, especially as it pertains to the one design approach to screening and response surface exploration.

Bibliography

Bibliography

[1] B. Abraham, H. Chipman, and K. Vijayan. Some risks in the construction and analysis of supersaturated designs. *Technometrics*, 41(2):135–141, 1999.

[2] T.T. Allen and M. Bernshteyn. Supersaturated designs that maximize the probability of identifying active factors. *Technometrics*, 45(1):90–97, 2003.

[3] A.R. Alvarez, B.L. Abdi, D.L. Young, H.D. Weed, J. Teplik, and E.R. Herald. Application of statistical design and response surface methods to computer-aided vlsi device design. *IEEE Transactions on Computer-Aided Design*, 7:272–288, 1988.

[4] M.J. Anderson. Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Science*, 58:626–639, 2001.

[5] M.J. Anderson and P. Legendre. An empirical comparison of permutation test of partial regression coefficients in a linear model. *Journal of Statistical Computation and Simulation*, 62:271–303, 1999.

[6] M.J. Anderson and J. Robinson. Permutation tests for linear models. *Australian and New Zealand Journal of Statistics*, 43:75–88, 2001.

[7] A.C. Atkinson, A.N. Donev, and R.D. Tobias. *Optimum Experimental Designs, with SAS*. Oxford University Press, Oxford, U.K., 2007.

[8] S.D. Beattie, D.K.H. Gong, and D.K.J. Lin. A two-stage Bayesian model selection strategy for supersaturated designs. *Technometrics*, 44:55–63, 2002.

[9] K.N. Berk. Comparing subset regression procedures. *Technometrics*, 20:1–6, 1978.

[10] D. Bingham. Comment on factor screening and response surface exploration. *Statistica Sinica*, 11:580–583, 2001.

[11] K.H. Booth and D.R. Cox. Some systematic supersaturated designs. *Technometrics*, 4:489–495, 1962.

[12] G.E.P. Box. Sequential experimentation and sequential assembly of designs. *Quality Engineering*, 5(2):321–330, 1992.

[13] G.E.P. Box and N.R. Draper. *Response Surfaces, Mixtures, and Ridge Analyses*. John Wiley and Sons, New York, NY, 2 edition, 2007.

[14] G.E.P. Box and Behnken D.W. Some new three level designs for the study of quantitative factors. *Technometrics*, 2(4):455–475, 1960.

[15] G.E.P. Box, J.S. Hunter, and W.G. Hunter. *Statistics for Experimenters: Design, Innovation, and Discovery*. John Wiley and Sons, New York, NY, 2005.

[16] G.E.P. Box and R.D. Meyer. An anaylsis of unreplicated fractional factorials. *Technometrics*, 28:11–18, 1986.

[17] G.E.P. Box and R.D. Meyer. Finding the active factors in fractionated screening experiments. *Journal of Quality Technology*, 25:94–105, 1993.

[18] G.E.P. Box and K.B. Wilson. On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society Series B*, 13:1–45, 1951.

[19] K.P. Burham and D.R. Anderson. Multimodel inference: Understanding aic and bic in model selection. *Sociological Methods and Research*, 33(2):261–304, 2004.

[20] R. Cela, E. Martinez, and A.M. Carro. Supersaturated experimental designs: New approaches to building and using it, part ii: Solving supersaturated designs by genetic algorithms. *Chemometrics and Intelligent Laboratory Systems*, 57:75–92, 2001.

[21] C.S. Cheng. $\mathbb{E}(s^2)$-optimal supersaturated designs. *Statistica Sinica*, 7:929–939, 1997.

[22] S.W. Cheng and C.F.J. Wu. Factor screening and response surface exploration (with comment). *Statistica Sinica*, 11:553–604, 2001.

[23] H. Chipman, M. Hamada, and C.F.J. Wu. A bayesian variable selection approach for analyzing designed experiments with complex aliasing. *Technometrics*, 39:372–381, 1997.

[24] W.H. Clatworthy. *Tables of Two-Associate Class Partially Balanced Designs*. National Bureau of Standards, Applied Math Series 63, 1973.

[25] W.G. Cochrane. *Sampling Techniques*. John Wiley and Sons, New York, 1963.

[26] A. Dean and D. Voss. *Design and Analysis of Experiments*. Springer, New York, NY, 1999.

[27] B. Dejaegher and Y. Vander Heyden. Supersaturated designs: Set-ups, data interpretation, and analytical applications. *Annals of Bioanalytical Chemistry*, 390:1227–1240, 2008.

[28] L.Y. Deng, D.K.J. Lin, and J. Wang. A resolution rank criterion for supersaturated designs. *Statistica Sinica*, 9:605–610, 1999.

[29] W. DuMouchel and B. Jones. A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model. *Technometrics*, 36:37–47, 1994.

[30] E.S. Edgington. *Randomization Tests*. Marcel-Dekker, New York, 1995.

[31] R.A. Fisher. *Design of Experiments*. Oliver and Boyd, Edinburgh, 1935.

[32] D. Freedman and D. Lane. A nonstochastic interpretation of reported significance levels. *Journal of Business and Economic Statistics*, 1:292–298, 1983.

[33] E.I. George and R.E. McCullough. Variable selection with gibbs sampling. *Journal of the American Statistical Association*, 88:881–889, 1993.

[34] S.G. Gilmour. *Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*, chapter Factor Screening via Supersaturated Designs, pages 169–190. Springer, New York, NY, 2006.

[35] S.G. Gilmour and R. Mead. Fixing a factor in the sequential design of two-level fractional factorial experiments. *Journal of Applied Statistics*, 23(1):21–29, 1996.

[36] P.I. Good. *Permutation Tests: A practical guide to resampling methods for testing hypotheses*. Springer-Verlag, Berlin, 2000.

[37] F. Graybill and R. Deal. Combining unbiased estimators. *Annals of Mathematical Statistics*, 15(4):543–550, 1959.

[38] P.D. Haaland and M.A. O'Connell. Inference for effect-saturated fractional factorials. *Technometrics*, 37:82–93, 1995.

[39] M. Hamada and N. Balakrishnan. Analyzing unreplicated factorial experiments: A review with some new proposals (with discussion). *Statistica Sinica*, 8:1–41, 1998.

[40] C. Hicks and K. Turner. *Fundamental Concepts in the Design of Experiments*. Oxford University Press, New York, NY, 1999.

[41] W.J. Hill and W.G. Hunter. A review of response surface methodology: A literature survey. *Technometrics*, 8(4):571–590, 1966.

[42] D.R. Holcomb, D.C. Montgomery, and W.M. Carlyle. Analysis of supersaturated designs. *Journal of Quality Technology*, 35:13–27, 2003.

[43] J.A. John. Cyclic incomplete block designs. *Journal of the Royal Statistical Society, Series B*, 28(2):345–360, 1966.

[44] J.A. John. Efficient cyclic designs. *Journal of the Royal Statistical Society, Series B*, 43(1):76–80, 1981.

[45] J.A. John and T.J. Mitchell. Optimal incomplete block designs. *Journal of the Royal Statistical Society, Series B*, 39:39–43, 1977.

[46] B. Jones, D.K.J. Lin, and C.J. Nachtsheim. Bayesian d-optimal supersaturated designs. *Journal of Statistical Planning and Inference*, 138:86–92, 2008.

[47] B.A. Jones, W. Li, C.J. Nachtsheim, and K.Q. Ye. Model discrimination - another perspective on model-robust designs. *Journal of Statistical Planning and Inference*, 137:1576–1583, 2007.

[48] H.W. Kelly and J.O. Voelkel. Asymptotic-power problems in the analysis of supersaturated designs. *Statistics and Probability Letters*, 47:317–324, 2000.

[49] P.E. Kennedy and B.S. Cade. Randomization tests for multiple regression. *Communications in Statistics, Simulation and Computation*, 25:923–936, 1996.

[50] A.I. Khuri and J.A. Cornell. *Response Surfaces: Design and Analyses*. Marcel Dekker, New York, NY, 2 edition, 1996.

[51] K. Larntz and P. Whitcomb. Use of replication in almost unreplicated factorials. In *Proceedings of the 1998 Fall Technical Conference*, Minneapolis, MN, 1998. Stat-Ease.

[52] J. Lawson. One-step screening and process optimization. *The American Statistician*, 57(1):15–20, 2003.

[53] R.V. Lenth. Quick and easy analysis of unreplicated factorials. *Technometrics*, 31(4):469–473, 1989.

[54] M. Lentner and T. Bishop. *Experimental Design and Analysis*. Valley Book Company, Blacksburg, VA, 1993.

[55] M. LeThanh, A. Voilley, and R.P.T. Luu. The influence of the composition of model liquid culture medium on vapor liquid partition coefficient of aroma substances. *Sciences Del Aliments*, 13:699–710, 1993.

[56] R. Li and D.K.J. Lin. Analysis methods for supersaturated designs: Some comparisons. *Journal of Data Science*, 1:249–260, 2003.

[57] W. Li and D.K.J. Lin. Optimal foldover plans for two-level fractional factorial designs. *Technometrics*, 45(2):142–149, 2003.

[58] W. Li, D.K.J. Lin, and K.Q. Ye. Optimal foldover plans for two-level nonregular orthogonal designs. *Technometrics*, 45(4):347–351, 2003.

[59] W. Li and C.F.J. Wu. Columnwise-pairwise algorithms with applications to the construction of supersaturated designs. *Technometrics*, 39:171–179, 1997.

[60] D.K.J. Lin. A new class of supersaturated designs. *Technometrics*, 35:28–31, 1993.

[61] D.K.J. Lin. Generating systematic supersaturated designs. *Technometrics*, 37(2):213–225, 1996.

[62] Y. Liu, S. Ruan, and A. Dean. Construction and analysis of es^2 efficient supersaturated designs. *Journal of Statistical Planning and Inference*, 137:1516–1529, 2007.

[63] S.L. Lohr. *Sampling: Design and Analysis*. Duxbury, Pacific Grove, 1999.

[64] T.M. Loughin. Calibration of the Lenth test for unreplicated factorial designs. *Journal of Quality Technology*, 30:171–175, 1998.

[65] X. Lu, W. Wu, and Y. Zheng. A systematical procedure in the construction of multi-level supersaturated designs. *Journal of Statistical Planning and Inference*, 115:287–310, 2003.

[66] X. Lu and X. Wu. A strategy of searching active factors in supersaturated screening experiments. *Journal of Quality Technology*, 36:392–399, 2004.

[67] B.F.J. Manly. *Randomization, Bootstrap, and Monte Carlo Methods in Biology*. Chapman and Hall, London, 1997.

[68] R.W. Mee. New Box-Behnken designs. *University of Tennessee Department of Statistics Technical Report Series*, 2000.

[69] R.W. Mee. Optimal three-level designs for response surfaces in spherical experimental regions. *Journal of Quality Technology*, 39(4):340–354, 2007.

[70] R.W. Mee and M. Peralta. Semifolding 2^{k-p} designs. *Technometrics*, 42(2):122–134, 2000.

[71] R.D. Meyer, D.M. Steinberg, and G.E.P. Box. Follow-up designs to resolve confounding in multifactor experiments. *Technometrics*, 38(4):303–313, 1996.

[72] D.C. Montgomery. *Design and Analysis of Experiments*. John Wiley and Sons, New York, NY, 6 edition, 2005.

[73] M. Morris. *Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*, chapter An Overview of Group Factor Screening, pages 191–206. Springer, New York, NY, 2006.

[74] R.H. Myers. Response surface methodology-current status and future directions. *Journal of Quality Technology*, 31(1):30–44, 1999.

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

[76] R.H. Myers, D.C. Montgomery, G.G. Vining, C.M Borror, and S.M. Kowalski. Response surface methodology: A retrospective and literature survey. *Journal of Quality Technology*, 36(1):53–77, 2004.

[77] B.J. Nelson, D.C. Montgomery, R.J. Elias, and E. Maass. A comparison of several design augmentation strategies. *Quality and Reliability Engineering International*, 16:435–449, 2000.

[78] N. Nguyen and C.S. Cheng. New $\mathbb{E}(s^2)$-optimal supersaturated designs constructed from incomplete block designs. *Technometrics*, 50:26–31, 2008.

[79] N.K. Nguyen. An algorithmic approach to constructing supersaturated designs. *Technometrics*, 38:69–73, 1996.

[80] F.K.H. Phoa, Y.H. Pan, and H. Xu. Analysis of supersaturated designs via Dantzig selector. *Department of Statistics Technical Report, UCLA*, 2007.

[81] E.J.G. Pitman. Significance tests which may be applied to samples from any population. *Journal of the Royal Statistical Society, Ser. B*, 4:119–130, 1937.

[82] E.J.G. Pitman. Significance tests which may be applied to samples from any population ii: The correlation coefficient test. *Journal of the Royal Statistical Society, Ser. B*, 4:225–232, 1937.

[83] E.J.G. Pitman. Significance tests which may be applied to samples from any population iii: The analysis of variance test. *Journal of the Royal Statistical Society, Ser. B*, 4:322–335, 1937.

[84] R.L. Plackett and Burman J.P. The design of optimum multifactorial experiments. *Biometrika*, 33:305–325, 1946.

[85] R.L. Retchshaffner. Saturated fractions of 2^n and 3^n factorial designs. *Technometrics*, 9:569–575, 1967.

[86] S.M. Rubin. *Computer Aids for VLSI Design*. Addison-Wesley, 1987.

[87] M.M. Rueda, A. Arcos, M.D. Martinez-Miranda, and Y. Roman. Some improved estimators of finite population quantile using auxiliary information in sample surveys. *Computational Statistics and Data Analysis*, 45:825–848, 2003.

[88] T.P. Ryan. *Modern Experimental Design*. John Wiley and Sons, New York, NY, 2007.

[89] F.E. Satterthwaite. Random balance experimentation. *Technometrics*, 1:111–137, 1959.

[90] E.D. Schoen and E.A.A. Kaul. Three robust scale estimators to judge unreplicated experiments. *Journal of Quality Technology*, 32:276–283, 2000.

[91] D.M. Steinberg. The future of industrial statistics. *Technometrics*, 50:103–127, 2008.

[92] P.H. Westfall and S.S. Young. *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment*. John Wiley and Sons, New York, NY, 1993.

[93] P.H. Westfall, S.S. Young, and D.K.J. Lin. Forward selection error control in the analysis of supersaturated designs. *Statistica Sinica*, 8:101–117, 1997.

[94] K.R. Williams. Designed experiments. *Rubber Age*, 100:65–71, 1968.

[95] C.F.J. Wu. Construction of supersaturated designs through partially aliased interactions. *Biometrika*, 80:661–669, 1993.

[96] C.F.J. Wu and M. Hamada. *Experiments: Planning, Analysis, and Parameter Design Optimization*. John Wiley and Sons, New York, NY, 2000.

[97] H. Xu, S.W. Cheng, and C.F.J. Wu. Optimal projective three-level designs for factor screening and interaction detection. *Technometrics*, 46(3):280–292, 2004.

[98] H. Xu and C.F.J. Wu. Generalized minimum aberration for asymmetrical fractional factorial designs. *Annals of Statistics*, 29:1066–1077, 2001.

[99] K. Ye and M. Hamada. Critical values of the Lenth method for unreplicated factorial designs. *Journal of Quality Technology*, 32:57–66, 2000.

[100] K.Q. Ye, K.J. Tsai, and W. Li. *Optimal Orthogonal Three-Level Factorial Designs for Factor Screening and Response Surface Exploration*, pages 221–228. Contributions in Statistics. Physica-Verlag HD, 2007.

[101] Q.Z. Zhang, R.C. Zhang, and M.Q. Liu. A method for screening active factors in supersaturated designs. *Journal of Statistical Planning and Inference*, 137:2068–2079, 2006.

Vita

David Joseph Edwards is a Ph.D. candidate and graduate teaching associate in the Department of Statistics, Operations, and Management Science at the University of Tennessee. He received a B.S. in Mathematics with a minor in Statistics from Virginia Polytechnic Institute and State University in 2002 and an M.S. in Statistics from the University of Tennessee in 2004. As an M.S. student, David was awarded Chancellor's Citations for Extraordinary Academic Achievement and Extraordinary Professional Promise and the Graduate Student Excellence Award. He is currently pursuing a Ph.D. in Statistics with plans to graduate in August 2008. David has presented research at the Massachusetts Institute of Technology's International Conference on Information Quality in 2003, the Fall Technical Conference in 2007, and Design and Analysis of Experiments (DAE) in 2007. He served as a Statistics Instructor for the Ronald McNair Post-Baccalaureate Achievement Program during summers 2003-2005. He has taught Introduction to Statistics for eight semesters while at the University of Tennessee and was a finalist for the Graduate Teaching Award in 2007. After obtaining a Ph.D., David will move to Richmond, VA with his wife, Elizabeth, where he will serve as Assistant Professor of Statistics at Virginia Commonwealth University.