Title

Evaluation of the Design Effect for Optimizing the Model Discrimination Strength to Detect Non-Zero Interactions in Factorial Experiments

Author

Wales, Brandon Allen

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Evaluation of the Design Effect for Optimizing the Model Discrimination Strength to Detect Non-Zero Interactions in Factorial Experiments

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Brandon Allen Wales

December 2017

Dissertation Committee:
Dr. Subir Ghosh, Chairperson
Dr. Daniel Jeske
Dr. Christian Shelton

The dissertation of Brandon Allen Wales is approved:

____

____

____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

Before I started higher education, I had the incorrect assumption that higher education was some obstacle I needed to hurdle through to get a degree which hopefully increased my prospects at finding a suitable career. As I progressed through higher education, I realized that it was much more than attending lectures, reading books, and passing exams. Higher education offers experience, wisdom, intuition, character building, connections, and more that related to my success at the University of California, Riverside than I can ever list versus just learning from a textbook. There are many people who contributed to my success at UCR.

First, I would like to thank parents, my brother, and my three sisters for building the foundations of my personality which ultimately led me to higher education. There had been many difficult moments through higher education and my personality helped me to diligently conquer them and turn them into learning experiences. Without this inscribed dedication into my personality, I would not have the opportunity to write these words.

Next, I would like to thank Dr. Subir Ghosh, my wonderful advisor. I realized one of the biggest personal developments I have achieved through my graduate studies is my intuition between statistics and the world around me. Working with Subir as my advisor had not only changed how I think about statistics but developed key decision making skills that are useful for all applications in life. This skill has been essential and directly correlates to my success as a graduate student at UCR.

I also would like to thank Dr. Daniel Jeske. Working with Dr. Jeske had helped me understand the perspective of people outside the field of statistics and how to effectively maximize my communication ability when approaching statistics. I am thankful for the many opportunities Dr. Jeske has given me through my graduate education.

A special thank you to Dr. Linda Penas. Linda is the teaching assistant coordinator and faculty advisor for Mu Sigma Rho (statistics honor society) at UCR. Some of my most enjoyable moments have been a teaching assistant, associate instructor, and the president of Mu Sigma Rho and could not have been possible without Linda.

A special thank you to Dr. Christian Shelton who had not only agreed to be a committee member, but had enlightened me with machine learning and how it relates to statistics. This perspective has been invaluable at understanding a bigger picture.

A special thank you to Jill E. Smith and the Highlander Statistics Society where I had most of my social interactions between graduate and undergraduate students.

I understand that at some point the acknowledgement section must end and I give my final thanks to all the other professors, graduate students, undergraduate students, which made the entire higher education experience infinitely more rewarding than just reading a text book.

ABSTRACT OF THE DISSERTATION

Evaluation of the Design Effect for Optimizing the Model Discrimination Strength to Detect Non-Zero Interactions in Factorial Experiments

by

Brandon Allen Wales

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, December 2017
Dr. Subir Ghosh, Chairperson

Model discrimination is the capability of a design to separate two competing models within a class of models. Model identification and discrimination are the foundation for model selection (Srivastava, 1975). The rank conditions for design matrices are necessary and sufficient to identify and discriminate in the noiseless case (Srivastava,1975). In the noisy case, the conditions are necessary but not sufficient. The success in the correct model selection depends on the strength of noise which is measured by the error sum of squares. To characterize the design effect for model discrimination, we consider the pairwise differences for the error sum of squares between two competing models. It is desired for the true model to have the smallest error sum of squares, an incorrect model to have a larger error sum of squares, and a large distance between the two error sum of squares to maximize the success for model selection. The differenced error sum of squares depends on both the observed values and the difference of two projection matrices. By examining the eigenvalues and eigenvectors for each differenced projection matrix, we propose to determine the strength of discrimination between two

competing models. Using eigenvalue decomposition for a design, the expected differenced error sum of squares between two model comparisons is expressed as a weighted average between the model parameters. An example for comparing two designs obtained from the 12-run Plackett-Burman design (1946) the magnitude of the eigenvalues can be used a measure of model selection performance. Additionally, we propose a model selection technique using ridge regression estimation and have also used LASSO as a model selection procedure. While comparing the performances of four designs in model selection with respect to the search design methodology, ridge regression, and LASSO methods, we observed a strong interaction effect at model selection performance between the method and design used.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1

Introduction to Factorial Designs

1.1 Overview

Consider a study where the objective is to maximize the yield (amount) of corn using three different fertilizers (A, B, and C), four different variations in seeds (1, 2, 3, and 4), and two different temperature settings (low and high). This type of study is known as a $3 * 4 * 2$ factorial design since there are a total of $3 * 4 * 2 = 24$ different treatment combinations. One treatment combination could be fertilizer A, seed #1, at low temperature. For this factorial design, the experimental unit is corn, the response variable is the yield of corn, and there are three different factors of interest: fertilizers (3 levels), seeds (4 levels), and temperature (2 levels). A "run" in the factorial design is the resulting yield of corn for a specified treatment combination for a total of n runs. The 3 factors of interest are the effect of the fertilizers on the yield outcome, the effect of the seeds on the yield outcome, and the different temperature settings on the yield outcome. It may be possible that the temperature is more responsible for yield outcome than the type of fertilizer used. Interaction effects are also of interest, there may be some effect of using fertilizer B and seed #3 but also an additive effect for using that specific combination of fertilizer and seed known as a two-factor interaction. The three-factor interaction effects are also possible important effects. The problem at hand, which of the two-factor interactions (or higher order) are important? Chapter 1 provides the necessary framework to address which of the effects are important.

1.2.1 Complete Factorial Designs

For a general factorial design with m factors $(A_1, A_2, \dots, A_m)$ and levels $(a_1, a_2, \dots, a_m)$, there are a total of $a_1 * a_2 * \dots * a_m$ different treatment combinations. For the j^{th} factor A_j, denote the different levels by 0, 1, 2, … , $(a_j - 1)$ and the observation (the value of the response variable) for run i as y_i. A complete factorial design is a design where the n runs contain every treatment combination for a total of $n = a_1 * a_2 * \ldots * a_m$ different runs. A complete factorial design for the corn example is given (A_1 = fertilizer, A_2 = seed, A_3 = temperature),

Table 1.1 Complete factorial design for the corn yield example

Run	A_1	A_2	A_3	Observation
1	0	0	0	y_1
2	0	0	1	y_2
3	0	1	0	y_3
4	0	1	1	y_4
5	0	2	0	y_5
6	0	2	1	y_6
7	0	3	0	y_7
8	0	3	1	y_8
9	1	0	0	y_9
10	1	0	1	y_{10}
11	1	1	0	y_{11}
12	1	1	1	y_{12}
13	1	2	0	y_{13}
14	1	2	1	y_{14}
15	1	3	0	y_{15}
16	1	3	1	y_{16}
17	2	0	0	y_{17}
18	2	0	1	y_{18}
19	2	1	0	y_{19}
20	2	1	1	y_{20}
21	2	2	0	y_{21}
22	2	2	1	y_{22}
23	2	3	0	y_{23}
24	2	3	1	y_{24}

A statistical model is often used to interpret the effect of the factors $A_1, A_2, \dots, A_m$ have on the response variable Y. For each factor A_j, there are $(a_j - 1)$ main effects (often $A_j = 0$ is used as a reference variable and the remaining $(a_j - 1)$ main effects are relative effect to the reference variable). For each two-factor interaction between factors A_j and $A_{j'}$, there are a total of $(a_j - 1)(a_{j'} - 1)$ two-factor interaction effects. For each factor A_j, the design points x_j are chosen so that the estimated statistical model using x_j can be easily interpretable and relatable to the main effects. For example, if the factor A_j has two levels $(a_j = 2)$, then there is $(a_j - 1) = 1$ main effect for factor A_j, by defining the design points $x_j = -1$ if $A_j = 0$ or $x_j = 1$ if $A_j = 1$ the estimated effect using the design points x_j perfectly correlates to the estimated main effect for factor A_j. This is useful because if there is no main effect for factor A_j, then the same conclusion will be present when using the design points x_j.

Consider the data of size n, $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_n, y_n)$ where $\boldsymbol{x}_i$ is one treatment combination for the m factors denoted by $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})$ and with the mean of observation y_i depending on $\boldsymbol{x}_i$ denoted by $E[y_i|\boldsymbol{x}_i]$. Denote the observation vector $\underline{y} = (y_1, y_2, \dots, y_n)'$, the j^{th} main effect $\underline{x_j} = (x_{1j}, x_{2j} \dots x_{nj})'$, two-factor interactions between j and j': $\underline{x_j x_{j'}} = \left(x_{1j}x_{1j'}, x_{2j}x_{2j'}, \dots, x_{nj}x_{nj'}\right)'$, or higher order terms in the same structure. The columns of the design matrix X are constructed using a column of ones $\underline{1}$ for the general mean and selected columns of main effects, interaction terms, and higher order terms for a total of p columns and n rows. For a constructed design matrix X, let $\underline{\beta} = \left(\beta_1, \beta_2, \dots, \beta_p\right)'$ be the set of unknown p parameters of interest, one unknown parameter for each column of X, and the linear model is expressed by,

$$M: E\left[\underline{y}|X\right] = X\underline{\beta},$$

$$Var\left(\underline{y}\right) = \sigma^2 I,$$

$$dim(X) = n \text{ x } p \quad , \quad \dim\left(\underline{\beta}\right) = p \text{ x } 1.$$

A complete factorial design can estimate all main effects, all two-factor interactions, and all higher order interactions. If the number of main effects m increases, then the number of runs required for a complete factorial design increases as the number of treatment combinations increase. If it is assumed that all three-factor and higher order interaction effects are zero, then a fractional design can be used which limits some capability of estimating three-factor or higher order interactions but uses a much smaller number of runs.

1.2.2 Fractional Factorial Designs

Consider a factorial experiment with m main effects $A_1, A_2, \ldots, A_m$ each with two levels $(a_1 = 2, a_2 = 2, \ldots, a_m = 2)$, this is known as a $2 * 2 * \ldots * 2 = 2^m$ factorial experiment. There is only one main effect for each factor j since $\left(a_j - 1\right) = 1$ for a total of m main effects and only one two-factor interaction effect for each interaction between factors j and j' since $\left(a_j - 1\right)\left(a_{j'} - 1\right) = 1$ for a total of $\binom{m}{2}$ two-factor interaction effects (the notation $\binom{m}{k}$ is used to denote the number of possible subsets of size k for the total of m objects). A complete factorial design is one with $n = 2^m$ runs can estimate all main effects and two-factor or higher order interactions. Observe in this simple case, as m increases, the number of runs exponentially increases so it may not be feasible to collect data using all 2^m treatment combinations for many practical limitations. If it is assumed that three-factor interaction effects are negligible, then it is not needed to collect all 2^m treatment combinations. A fractional factorial design uses a carefully chosen subset of the complete factorial design so that all effects of interest (main effects and two-factor interactions) can still be estimated while only using a fraction of the runs. Let $m = 4$, table 1.2 lists all possible 2^4 treatment combinations,

Table 1.2 Level combinations of A_1, A_2, A_3, and A_4 for a complete 2^4 factorial design

Run	A_1	A_2	A_3	A_4	Observation
1	0	0	0	0	y_1
2	0	0	0	1	y_2
3	0	0	1	0	y_3
4	0	0	1	1	y_4
5	0	1	0	0	y_5
6	0	1	0	1	y_6
7	0	1	1	0	y_7
8	0	1	1	1	y_8
9	1	0	0	0	y_9
10	1	0	0	1	y_{10}
11	1	0	1	0	y_{11}
12	1	0	1	1	y_{12}
13	1	1	0	0	y_{13}
14	1	1	0	1	y_{14}
15	1	1	1	0	y_{15}
16	1	1	1	1	y_{16}

Observe that 8 treatment combinations satisfy the condition $mod(A_1 + A_2 + A_3 + A_4, 2) = 0$ and the other 8 treatment combinations satisfy the condition $mod(A_1 + A_2 + A_3 + A_4, 2) = 1$. Consider the set of treatment combinations that satisfy the equation $mod(A_1 + A_2 + A_3 + A_4, 2) = 0$ as presented in table 1.3

Table 1.3 Level combinations of A_1, A_2, A_3, and A_4 for a 2^{4-1} factorial design

Run	A_1	A_2	A_3	A_4
1	0	0	0	0
2	0	0	1	1
3	0	1	0	1
4	0	1	1	0
5	1	0	0	1
6	1	0	1	0
7	1	1	0	0
8	1	1	1	1

The equation, $mod(A_1 + A_2 + A_3 + A_4, 2) = 0$, that separates out the 8 treatment combinations from the total of 16 treatment combinations is known as a "word". From here on forward, words will be shorthanded from $mod(A_1 + A_2 + A_3 + A_4, 2) = 0$ to $A_1 + A_2 + A_3 + A_4 = 0$. A collection of words, multiple equations to further subset the treatment combinations, is known as the defining relation.

The fractional factorial design with $n = 2^{4-1}$ runs can estimate all main effects assuming two-factor interactions, three-factor interactions, and the four-factor interaction are all zero. This fractional factorial design is a resolution IV plan, resolution plans are summarized below,

Resolution III – The general mean and main effects can be estimated assuming two-factor and higher order interactions to be zero.

Resolution IV – The general mean and main effects can be estimated assuming three-factor and higher order interactions to be zero and not all two-factor interactions can be estimated.

Resolution V – The general mean, main effects, and all two-factor interactions can be estimated assuming three-factor and higher order interactions to be zero.

For a fractional factorial experiment where none of the two-factor interaction effects zero, a resolution V or higher order plan must be used otherwise valid inferences will not be possible.

1.3 Problem: Selecting Important Two-Factor Interactions

For a 2^m factorial experiment, there is one general mean effect, m main effects, $\binom{m}{2}$ two-factor interaction effects, $\binom{m}{3}$ three-factor interaction effects, …, $\binom{m}{m} = 1$; m-factor interaction effect. If we assume all three-factor and higher order interaction effects are zero, we focus only on the $p = 1 + m + \binom{m}{2}$ effects that are of interest. A fractional factorial design with n runs must be at least a resolution V plan otherwise some two-factor interaction effects will be aliased with each other since no two-factor interaction effects are assumed to be zero. Let X be the design matrix where the columns correspond to the general mean, m main effects, $\binom{m}{2}$ two-factor interactions for a total of $p = 1 + m + \binom{m}{2}$ columns; the rows of the design matrix X correspond to the n runs. Define the true model to be the model that correctly describes the relationship between the m factors and the respond variable Y and is not known except there are k two-factor interactions which are non-zero. Consider the class of models where each model has the general mean, m main effects, and k two-factor interactions for a total of $v_k = \binom{\binom{m}{2}}{k}$ different models. Let model M_u be in the class of models where $u = 1,2, \dots, v_k$, then there are v_k different ways of defining the design matrix denoted by $X^{(u)}$ by taking all k subsets of X with their corresponding parameters of interest $\underline{\beta}^{(u)}$ each estimating model M_u is expressed by,

$$M_u: E\left[\underline{y}|X^{(u)}\right] = X^{(u)}\underline{\beta}^{(u)},$$

$$Var\left(\underline{y}\right) = \sigma^2 I.$$

The problem is selecting the correct k two-factor interactions within the different v_k models. The ability to pick the correct model within the class of models depends on the design chosen, the method at estimating the models, and the model selection procedure used.

1.4.1 Balanced Designs

Section 1.4.1 – section 1.4.3 discusses minor topics regarding factorial experiments that are used throughout the chapters. For a given design d, consider the design matrix X_d that consists of the general mean, m main effects, and $\binom{m}{2}$ two-factor interaction effects for a total of $p = 1 + m + \binom{m}{2}$ interactions. We focus on a class of models that include subsets of k two-factor interactions for $v_k = \binom{\binom{m}{2}}{k}$ different models expressed by,

$$M_u: E\left[\underline{y}|X^{(u)}\right] = X^{(u)}\underline{\beta}^{(u)},$$

$$Var\left(\underline{y}\right) = \sigma^2 I.$$

Denote the variance-covariance matrix for model M_u as $V^{(u)} = \left(X^{(u)'}X^{(u)}\right)^{-1}$, then design is balanced with respect to model M_u if the variance covariance matrix $V^{(u)}$ is invariant under a permutation of the factor symbols (Srivastava and Chopra, 1971). Figure 1.1 presents the structure of a balanced variance-covariance matrix for $m = 4$ (Srivastava and Chopra, 1971),

Figure 1.1 Structure for a balanced design

	$\hat{\mu}$	$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$	$\hat{A}_4$	$\hat{A}_{12}$	$\hat{A}_{13}$	$\hat{A}_{14}$	$\hat{A}_{23}$	$\hat{A}_{24}$	$\hat{A}_{34}$
$\hat{\mu}$	v1	v2	v2	v2	v2	v3	v3	v3	v3	v3	v3
$\hat{A}_1$	v2	v4	v5	v5	v5	v6	v6	v6	v7	v7	v7
$\hat{A}_2$	v2	v5	v4	v5	v5	v6	v7	v7	v6	v6	v7
$\hat{A}_3$	v2	v5	v5	v4	v5	v7	v6	v7	v6	v7	v6
$\hat{A}_4$	v2	v5	v5	v5	v4	v7	v7	v6	v7	v6	v6
$\hat{A}_{12}$	v3	v6	v6	v7	v7	v8	v9	v9	v9	v9	v10
$\hat{A}_{13}$	v3	v6	v7	v6	v7	v9	v8	v9	v9	v10	v9
$\hat{A}_{14}$	v3	v6	v7	v7	v6	v9	v9	v8	v10	v9	v9
$\hat{A}_{23}$	v3	v7	v6	v6	v7	v9	v9	v10	v8	v9	v9
$\hat{A}_{24}$	v3	v7	v6	v7	v6	v9	v10	v9	v9	v8	v9
$\hat{A}_{34}$	v3	v7	v7	v6	v6	v10	v9	v9	v9	v9	v8

To illustrate the difference between a balanced design and an unbalanced design with respect to model M_u, consider these two variance-covariance matrices of the estimates for the general mean, main effects, and two-factor interactions,

Table 1.4.1 Variance-covariance matrix for the estimates of the general mean, main effects, and two-factor interaction effects example 1 (Balanced)

	$\hat{\mu}$	$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$	$\hat{A}_4$	$\hat{A}_5$	$\hat{A}_{12}$
$\hat{\mu}$	0.0938	0	0	0	0	0	-0.0312
$\hat{A}_1$	0	0.1071	-0.0179	-0.0179	-0.0179	-0.0179	0
$\hat{A}_2$	0	-0.0179	0.1071	-0.0179	-0.0179	-0.0179	0
$\hat{A}_3$	0	-0.0179	-0.0179	0.1071	-0.0179	-0.0179	0
$\hat{A}_4$	0	-0.0179	-0.0179	-0.0179	0.1071	-0.0179	0
$\hat{A}_5$	0	-0.0179	-0.0179	-0.0179	-0.0179	0.1071	0
A_{12}	-0.0312	0	0	0	0	0	0.0938

Table 1.4.2 Variance-covariance matrix for the estimates of the general mean, main effects, and two-factor interaction effects example 2 (Unbalanced)

	$\hat{\mu}$	$\hat{A}_1$	$\hat{A}_2$	$\hat{A}_3$	$\hat{A}_4$	$\hat{A}_5$	$\hat{A}_{12}$
$\hat{\mu}$	0.0833	0	0	0	0	0	0
$\hat{A}_1$	0	0.0833	0	0	0	0	0
$\hat{A}_2$	0	0	0.0833	0	0	0	0
$\hat{A}_3$	0	0	0	0.0972	0.0139	-0.0139	0.0417
$\hat{A}_4$	0	0	0	0.0139	0.0972	-0.0139	0.0417
$\hat{A}_5$	0	0	0	-0.0139	-0.0139	0.0972	-0.0417
A_{12}	0	0	0	0.0417	0.0417	-0.0417	0.125

Observe for example 1, the value of v1 is 0.0938, the value of v2 is 0, the value of v3 is -0.0312, the value of v4 is 0.1071, the value of v5 is -0.0179, the value of v6 and v7 are both 0, and the value of v8 is 0.0938. Using these values, example 1 perfectly fits the structure of a balanced design. Example 2 is close to a balance design, but the value of v2 is both 0.0833 and 0.0972. The 4 x 4 submatrix almost fits a balanced design but is off by a few values which are negative.

1.4.2 Orthogonal Designs

An orthogonal design is a design in which the estimates are uncorrelated with each other. Let d^* be an orthogonal design with n runs, m main effects, with full estimation capacity for all two-factor interactions for a total of $p = 1 + m + \binom{m}{2}$ parameters. The design matrix X_{d^*} which has columns for the general mean, m main effects, and all $\binom{m}{2}$ two-factor interactions has dimensions n x p and satisfies the equation,

$$X'_{d^*} X_{d^*} = nI_p.$$

Observe for any model M_u and the corresponding submatrix $X_{d^*}^{(u)}$ of dimensions n x p_u will also satisfy the equation,

$$X_{d^*}^{(u)'} X_{d^*}^{(u)} = nI_{p_u}.$$

For any orthogonal design, the variance-covariance matrix $V^{(u)} = \left(X_{d^*}^{(u)'} X_{d^*}^{(u)}\right)^{-1} = \frac{1}{n} I_{p_u}$ hence an orthogonal design is also a balanced design with diagonal elements $\frac{1}{n}$ and off diagonal elements 0.

1.4.3 Optimality Criterion Functions

One way to compare the efficiency of a design is using the six optimality functions. For a design d and model M_u calculate the variance-covariance matrix $V_d^{(u)} = \sigma^2 \left(X_d^{(u)'} X_d^{(u)}\right)^{-1}$. Using $V_d^{(u)}$ for the class of models $u = 1,2,\dots,v_k$ we can calculate these six optimality criterion functions,

$$AD(d) = \frac{1}{v_k} \sum_{u=1}^{v_k} \left|V_d^{(u)}\right| \qquad GD(d) = \left(\prod_{u=1}^{v_k} \left|V_d^{(u)}\right|\right)^{\frac{1}{v_k}}$$

$$AT(d) = \frac{1}{v_k} \sum_{u=1}^{v_k} trace\left(V_d^{(u)}\right) \qquad GT(d) = \left(\prod_{u=1}^{v_k} trace\left(V_d^{(u)}\right)\right)^{\frac{1}{v_k}}$$

$$AMCR(d) = \frac{1}{v_k} \sum_{u=1}^{v_k} \max\left(ev\left(V_d^{(u)}\right)\right) \qquad GMCR(d) = \left(\prod_{u=1}^{v_k} \max\left(ev\left(V_d^{(u)}\right)\right)\right)^{\frac{1}{v_k}}$$

"A" represents the arithmetic mean and "G" represents the geometric mean for the calculated v_k values. The "D" criterion functions is based on the volume of the variance-covariance matrices and known as the D-optimality criterion function for a design. The "T" criterion function is based on minimizing the average variance for the variance-covariance matrix and known as the A-optimality criterion function for a design. The "MCR" criterion function is based on the max amount of variance when accounting for the covariance and known as the E-optimality criterion function for a design.

1.5 Dissertation Summary

There are eight chapters in this dissertation about the model selection problem. Chapter 2 describes the necessary conditions for a design to have full estimation capacity and the ability to discriminate between all pairwise models within a class of models. For a design that satisfies the necessary conditions as described in chapter 2, chapter 3 dissects the search design methodology to understand how the choice of design can affect the model selection performance results. Chapter 4 compares two very similar designs where each design is balanced with the main effects and two-factor interactions. Chapter 5 compares two designs that have the same values of the six optimality criterion functions and are balanced designs with respect to main effects but not with the two-factor interactions. Chapter 6 understands the ridge regression estimation method and proposes a model selection procedure using ridge regression. Chapter 7 analyzes the estimation technique for LASSO and its use as a model selection procedure. Chapter 8 is a simulation study that compares the model selection performance using multiple designs and multiple model selection procedures as described in chapters 2 through 7.

Chapter 2

Model Identification and Discrimination

2.1 Introduction

Model identification and discrimination are two different objectives for a design. For any design and a class of models, it is desirable to uniquely estimate all models or have capability of discriminating between all pairwise model comparison within the class of models. This chapter describes the different rank conditions required for a design to successfully identify and discriminate between all pairwise model comparisons in the class of models. The chapter concludes with an example design and how its chosen treatment combinations affect model identification and discrimination.

2.2 Definitions

For fractional factorial designs with n runs based on a 2^m fractional experiment, let the columns of the design matrix X_1 represent the general mean, the m main effects, and the columns of the design matrix X_2 represent the two-factor interaction effects. The 3-factor and higher order interactions are assumed to be zero, the n x 1 column vector of observations for the n runs are denoted as $\underline{y}$. Define a run using the notation $(x_1, x_2, \dots, x_m)$, a row of X_1 is $(1, x_1, \dots, x_m)$, and a row for X_2 is $(x_1x_2, x_1x_3, \dots, x_{m-1}x_m)$ thus having a total of n rows for both X_1 and X_2.

The parameters representing the general mean and m main effects are β_0 and $(\beta_1, \beta_2, \dots, \beta_m)$. Denote $\underline{\beta_1} = (\beta_0, \beta_1, \dots, \beta_m)'$ and $\underline{\beta_2}$ as a $\binom{m}{2}$ x 1 column vector of parameters representing all the two-factor interaction effects with a total of $p = 1 + m + \binom{m}{2}$ parameters.

For a choice of k, we can populate the $v_k = \binom{\binom{m}{2}}{k}$ subsets of $\underline{\beta_2}$ of size k are denoted by $\underline{\beta_2}^{(u)}$ and the corresponding submatrices of X_2 are denoted by $X_2^{(u)}$, $u = 1,2, \dots, v_k$. Let model M_u be expressed as,

$$M_u\colon\ E\left[\underline{y}\right] = X_1\underline{\beta_1} + X_2^{(u)}\underline{\beta_2}^{(u)}\ ,\ \ Var\left(\underline{y}\right) = \sigma^2 I_n,\ \ u = 1,2, \dots, v_k \qquad (2.1)$$

When considering two models M_u and $M_{u'}$, denote the pairwise model comparison as $\{M_u, M_{u'}\}$, there are $\binom{v_k}{2}$ such comparisons. For the model comparison $\{M_u, M_{u'}\}$, the number of common parameters in $\underline{\beta_2}^{(u)}$ and $\underline{\beta_2}^{(u')}$ is $c_{uu'}$ where $0 \leq c_{uu'} \leq (k-1)$ and $M_u \neq M_{u'}$. The total number of common parameters in $\{M_u, M_{u'}\}$ is $1 + m + c_{uu'}$ and the total number of uncommon parameters is $2(k - c_{uu'})$. The total number of parameters that need to be estimated $\{M_u, M_{u'}\}$ is $1 + m + (2k - c_{uu'})$. For model M_u, subset the vector of all two-factor interactions $\underline{\beta_2}$ into $\underline{\beta_2}^{(u)}$ which contains the k two-factor interactions in model M_u and $\underline{\beta_2}^{(\bar{u})}$ which contains the $\binom{m}{2} - k$ two-factor interactions not in model M_u. $\underline{\beta_2}^{(u')}$ and $\underline{\beta_2}^{(\bar{u}')}$ is constructed identically for a model $M_{u'}$.

When considering the comparison $\{M_u, M_{u'}\}$, there may be some elements of $\underline{\beta_2}^{(u)}$ and $\underline{\beta_2}^{(u')}$ in common and some elements not in common. Let $\underline{\beta_2}^{(uu')}$ be the elements in common of $\underline{\beta_2}^{(u)}$ and $\underline{\beta_2}^{(u')}$, $\underline{\beta_2}^{(\bar{u}u')}$ be the elements in $\underline{\beta_2}^{(u')}$ but not in $\underline{\beta_2}^{(u)}$, $\underline{\beta_2}^{(u\bar{u}')}$ be the elements in $\underline{\beta_2}^{(u)}$ but not in $\underline{\beta}_2^{(u')}$, and $\underline{\beta}_2^{(\bar{u}\bar{u}')}$ be the elements not in $\underline{\beta}_2^{(u)}$ or $\underline{\beta}_2^{(u')}$. Let $X_2^{(uu')}, X_2^{(\bar{u}u')}, X_2^{(u\bar{u}')}$, and $X_2^{(\bar{u}\bar{u}')}$ be the corresponding design matrices for $\underline{\beta}_2^{(uu')}$, $\underline{\beta}_2^{(\bar{u}u')}, \underline{\beta}_2^{(u\bar{u}')}$, and $\underline{\beta}_2^{(\bar{u}\bar{u}')}$ respectively.

Table 2.1 Summary of notations with vector lengths underneath

$M_u \backslash M_{u'}$	$M_{u'}$	$M_{\bar{u}'}$	Total
M_u	$\underline{\beta}_2^{(uu')}$ $c_{uu'}$	$\underline{\beta}_2^{(u\bar{u}')}$ $k - c_{uu'}$	$\underline{\beta}_2^{(u)}$ k
$M_{\bar{u}}$	$\underline{\beta}_2^{(\bar{u}u')}$ $k - c_{uu'}$	$\underline{\beta}_2^{(\bar{u}\bar{u}')}$ $\binom{m}{2} - 2k + c_{uu'}$	$\underline{\beta}_2^{(\bar{u})}$ $\binom{m}{2} - k$
Total	$\underline{\beta}_2^{(u')}$ k	$\underline{\beta}_2^{(\bar{u}')}$ $\binom{m}{2} - k$	$\underline{\beta}_2$ $\binom{m}{2}$

In M_u, the $(1 + m + k)$ β parameters are denoted by $\underline{\beta}^{(u)}$ and the design matrix by $X^{(u)}$ where $X^{(u)} = \left(X_1 : X_2^{(u)}\right)$ and $\underline{\beta}^{(u)} = \left(\underline{\beta}_1, \underline{\beta}_2^{(u)}\right)'$, $u = 1,2, \dots, v_k$. The model from equation (2.1) becomes,

$$M_u: \ E\left[\underline{y}\right] = X^{(u)}\underline{\beta}^{(u)}, Var\left(\underline{y}\right) = \sigma^2 I_n, \ \ u = 1,2, \dots, v_k \tag{2.2}$$

2.3 Rank Conditions

For a specified class of models, it is desirable for a design to be able to identify all models M_u and discriminate between all pairwise model comparisons M_u and $M_{u'}$. There are two objectives to consider,

(1) Model identification,

(2) Model Discrimination

A model M_u is identifiable if the least squares estimation (LSE) of $\underline{\beta}^{(u)}$ is unique. The least squares estimation (LSE) of $\underline{\beta}^{(u)}$ is possible when,

$$Rank\left(X^{(u)}\right) = 1 + m + k. \tag{2.3}$$

This is the case since the LSE of $\underline{\beta}^{(u)}$ is $\underline{\hat{\beta}}^{(u)} = \left(X^{(u)'}X^{(u)}\right)^{-1} X^{(u)'}\underline{y}$. If the model identification rank condition is not satisfied, it is known that $\left(X^{(u)'}X^{(u)}\right)^{-1}$ is not unique. For discriminating between two models M_u and $M_{u'}$ it is necessary to be able to estimate all the unique parameters, $\underline{\beta}_1$, $\underline{\beta}_2^{(uu')}$, $\underline{\beta}_2^{(\bar{u}u')}$, and $\underline{\beta}_2^{(u\bar{u}')}$. The condition for having the least squares estimations of $\underline{\beta}_1$, $\underline{\beta}_2^{(uu')}$, $\underline{\beta}_2^{(\bar{u}u')}$, and $\underline{\beta}_2^{(u\bar{u}')}$ are possible when,

$$\begin{aligned} Rank\left(X_1 : X_2^{(uu')} : X_2^{(\bar{u}u')} : X_2^{(u\bar{u}')}\right) &= 1 + m + c_{uu'} + 2(k - c_{uu'}) \\ &= 1 + m + 2k - c_{uu'}. \end{aligned} \tag{2.4}$$

Theorem 2.1 (Srivastava, 1975)

Model identification is necessary and sufficient if (2.3) holds for every model M_u in the class of models. Model discrimination is necessary and sufficient in the noiseless case, necessary in the noisy case if (2.4) is satisfied for every pairwise model comparison $\{M_u, M_{u'}\}$.

If a design satisfies the rank conditions for model discrimination for all model comparisons $\{M_u, M_{u'}\}$ given k, then that design satisfies the rank conditions for model identification since for any positive integer k we have $1 + m + k \leq 1 + m + 2k - c_{uu'}$ implying that (2.4) is a stronger property than (2.3). Replicating any design that satisfies the desired rank condition results, the resulting design is a design with $(r_1 + \cdots + r_n)$ runs that satisfies the desired rank condition.

2.4 Example

To illustrate the two rank conditions, consider the balanced fractional factorial design d_1 presented from Ghosh, Tian (2004) that contains $n = 12$ runs, with the main effect A_5 modified by taking the negative of that column. The structure of d_1 is each main effect A_i is positive 7 of the 12 runs in such a way that there is all $\binom{5}{3} = 10$ three treatment combinations runs, a run where we have no treatments present, and a run where all treatments are present.

d_1

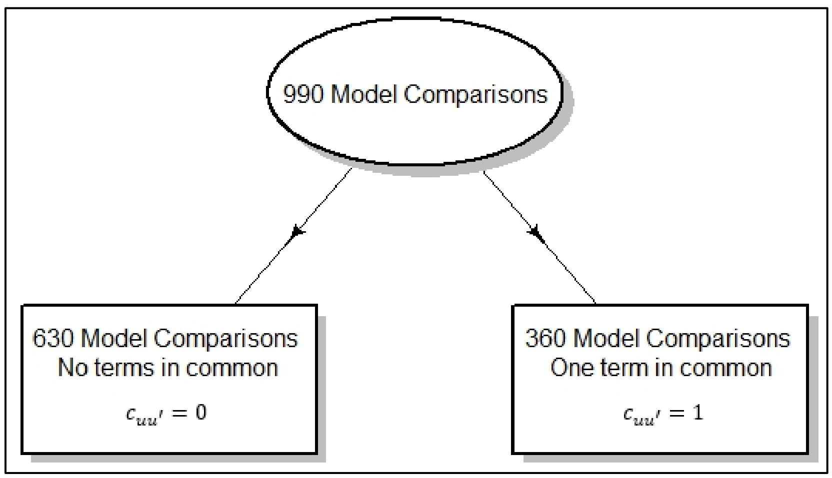
Run	Treatment	$\underline{\boldsymbol{x}}_1$	$\underline{\boldsymbol{x}}_2$	$\underline{\boldsymbol{x}}_3$	$\underline{\boldsymbol{x}}_4$	$\underline{\boldsymbol{x}}_5$
1	(1)	-1	-1	-1	-1	-1
2	125	1	1	-1	-1	1
3	135	1	-1	1	-1	1
4	145	1	-1	-1	1	1
5	235	-1	1	1	-1	1
6	245	-1	1	-1	1	1
7	345	-1	-1	1	1	1
8	12345	1	1	1	1	1
9	123	1	1	1	-1	-1
10	124	1	1	-1	1	-1
11	134	1	-1	1	1	-1
12	234	-1	1	1	1	-1

There are $m = 5$ main effects and $\binom{5}{2} = 10$ two-factor interactions considered. The design d_1 satisfies the rank conditions for model identification up to $k = 5$ and model discrimination up to $k = 2$. For $k = 2$, the class of models consists of $v_k = \binom{10}{2} = 45$ models that contain 2 of the 10 two-factor interactions and a total of $\binom{45}{2} = 990$ different pairwise model comparisons. For any model comparison $\{M_u, M_{u'}\}$, there is either one two-factor interaction in common $c_{uu'} = 1$ or no two-factor interactions in common $c_{uu'} = 0$. Since models M_u and $M_{u'}$ are chosen such that $M_u \neq M_{u'}$ there cannot be 2 two-factor interactions in common.

The number of different model comparisons that do not have any terms in common $c_{uu'} = 0$ will be $\frac{\binom{10}{2} * \binom{8}{2}}{2} = 630$. That is for M_u selects 2 out of the 10 two-factor interactions, $M_{u'}$ we select 2 out of the remaining eight two-factor interactions, and the assignment of M_u and $M_{u'}$ is not important therefore the number of combinations is divided by 2.

The number of different model comparisons that have one term in common $c_{uu'} = 1$ is $\frac{\binom{10}{1}*\binom{9}{1}\binom{8}{1}}{2} = 360$. That is select 1 out of the 10 two-factor interactions that is in common, give M_u one of the 9 remaining two-factor interactions, and then $M_{u'}$ one of the remaining 8 two-factor interactions. The assignment of M_u and $M_{u'}$ is not important and the total number of choices is divided by two. The following information is summarized in the flowchart below,

Figure 2.1 Partitioning the 990 pairwise model comparisons for d_1



For the 630 model comparisons with no terms in common $c_{uu'} = 0$, d_1 for all M_u and $M_{u'}$ satisfies,

$$Rank\left(X_1 : X_2^{(u)} : X_2^{(u')}\right) = 1 + m + 2k - c_{uu'} = 10,$$

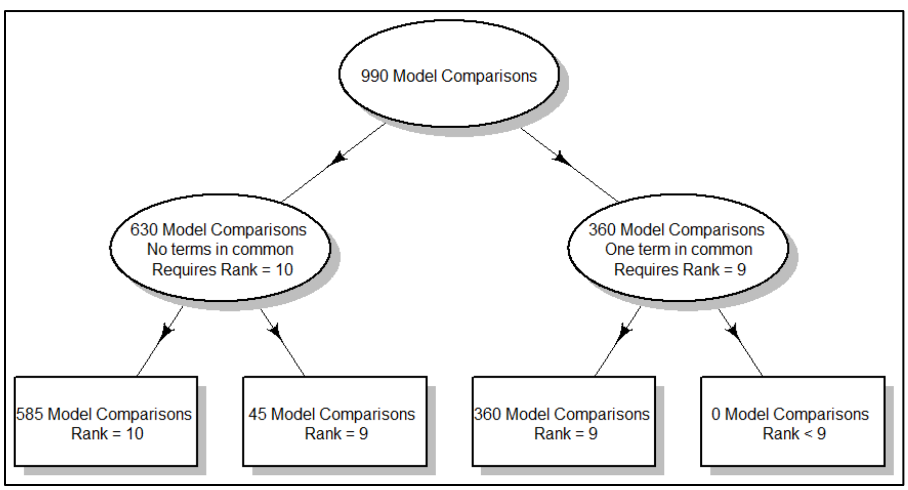
and for the 360 model comparisons with one term in common $c_{uu'} = 1$, d_1 for all M_u and $M_{u'}$ satisfies,

$$Rank\left(X_1 : X_2^{(u)} : X_2^{(u')}\right) = 1 + m + 2k - c_{uu'} = 9.$$

The different treatment combinations affect the rank conditions for the $v_k = 45$ models and the 990 pairwise model comparisons. To illustrate this, create a $n = 11$ run design by removing run i from d_1 denoted as $d_1^{(-i)}$ for $i = 1,2, \dots ,12$. All 12 of these new designs $d_1^{(-1)}, \dots , d_1^{(-12)}$ will still satisfy the rank condition for model identification of the 45 different models as specified in (2.3) but no longer satisfy the rank conditions for model discrimination for all 990 model comparisons as specified in (2.4), the model comparison $\{M_u, M_{u'}\}$ depend which of the 12 runs we remove.

By removing run 1 or run 8 then the resulting design $d_1^{(-1)}$ or $d_1^{(-8)}$ will not satisfy the rank conditions for model discrimination for 45 of 990 different model comparisons. For each pair of M_u there will be two $M_{u'}$ that do not satisfy the specified rank condition.

Figure 2.2 Rank conditions of the 990 pairwise model comparisons for $\boldsymbol{d_1^{(-1)}}$ or $\boldsymbol{d_1^{(-8)}}$



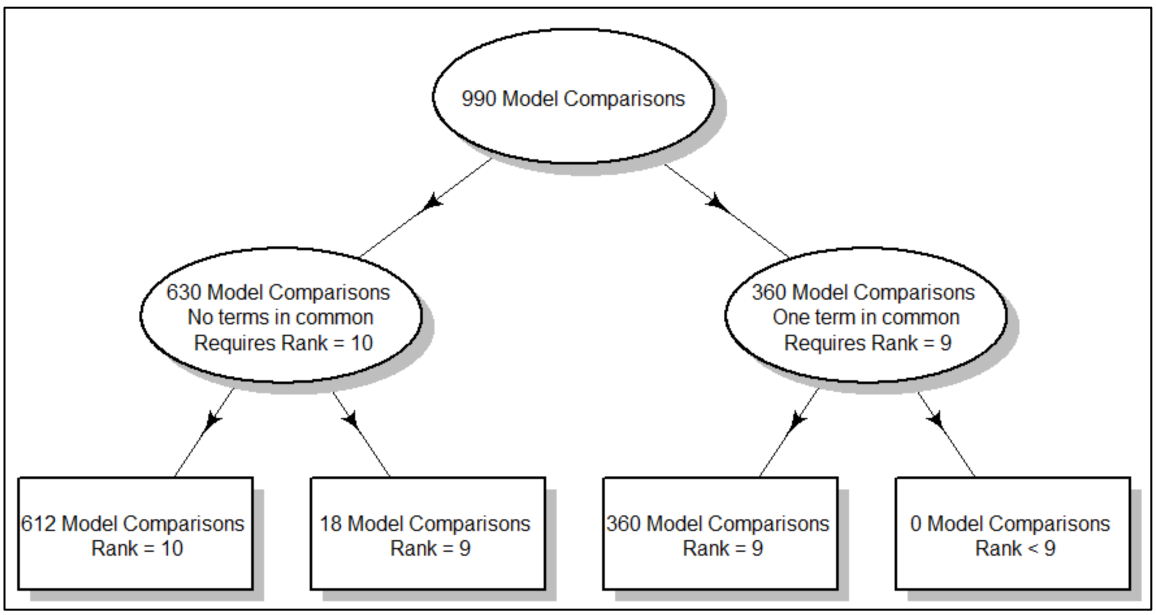
For any model M_u, table 2.2 presents the corresponding $M_{u'}$ such that the rank condition for model discrimination is not met. In the table below, there are a total of 90 $M_{u'}$ that correspond to M_u since duplicates are presented, E.g. the pair $M_u = x_1x_2\ x_1x_5$ and $M_{u'} = x_2x_5\ x_3x_4$ do not satisfy the rank condition for model discrimination, also presented is the reverse $M_{u'} = x_1x_2\ x_1x_5$ and $M_u = x_2x_5\ x_3x_4$. Table 2.2 demonstrates if the true model were any M_u, then we would have trouble discriminating between the two corresponding $M_{u'}$. The importance is that by deleting run 1 or 8, every model M_u will have trouble discriminating between 2 of the other 44 $M_{u'}$. It will not be desirable to use either $d_1^{(-1)}$ or $d_1^{(-8)}$ for any true model M_u for searching 2 two-factor interactions as we would be able to discriminate with other models.

Table 2.2 Models that do not satisfy the model discrimination properties for $d_1^{(-1)}$

u		u'				u		u'			
x1x2	x1x3	x1x4	x1x5	x2x3	x4x5	x1x4	x3x4	x2x4	x4x5	x1x3	x2x5
x1x2	x1x4	x1x3	x1x5	x2x4	x3x5	x1x5	x3x4	x2x3	x2x4	x1x2	x2x5
x1x3	x1x4	x1x2	x1x5	x2x5	x3x4	x2x3	x3x4	x1x3	x3x5	x1x5	x2x4
x1x2	x1x5	x2x5	x3x4	x1x3	x1x4	x2x4	x3x4	x1x4	x4x5	x1x5	x2x3
x1x3	x1x5	x2x4	x3x5	x1x2	x1x4	x2x5	x3x4	x1x3	x1x4	x1x2	x1x5
x1x4	x1x5	x2x3	x4x5	x1x2	x1x3	x1x2	x3x5	x3x4	x4x5	x1x4	x2x4
x1x2	x2x3	x2x4	x2x5	x1x3	x4x5	x1x3	x3x5	x1x5	x2x4	x2x3	x3x4
x1x3	x2x3	x3x4	x3x5	x1x2	x4x5	x1x4	x3x5	x1x2	x2x4	x2x3	x2x5
x1x4	x2x3	x2x5	x3x5	x1x5	x4x5	x1x5	x3x5	x2x5	x4x5	x1x3	x2x4
x1x5	x2x3	x2x4	x3x4	x1x4	x4x5	x2x3	x3x5	x1x4	x2x5	x1x3	x3x4
x1x2	x2x4	x2x3	x2x5	x1x4	x3x5	x2x4	x3x5	x1x2	x1x4	x1x3	x1x5
x1x3	x2x4	x1x5	x3x5	x2x5	x4x5	x2x5	x3x5	x1x5	x4x5	x1x4	x2x3
x1x4	x2x4	x1x2	x3x5	x3x4	x4x5	x3x4	x3x5	x1x2	x4x5	x1x3	x2x3
x1x5	x2x4	x2x3	x3x4	x1x3	x3x5	x1x2	x4x5	x1x3	x2x3	x3x4	x3x5
x2x3	x2x4	x1x2	x2x5	x1x5	x3x4	x1x3	x4x5	x1x2	x2x3	x2x4	x2x5
x1x2	x2x5	x1x5	x3x4	x2x3	x2x4	x1x4	x4x5	x1x5	x2x3	x2x4	x3x4
x1x3	x2x5	x1x4	x3x4	x2x4	x4x5	x1x5	x4x5	x1x4	x2x3	x2x5	x3x5
x1x4	x2x5	x1x3	x3x4	x2x3	x3x5	x2x3	x4x5	x1x2	x1x3	x1x4	x1x5
x1x5	x2x5	x1x2	x3x4	x3x5	x4x5	x2x4	x4x5	x1x3	x2x5	x1x4	x3x4
x2x3	x2x5	x1x4	x3x5	x1x2	x2x4	x2x5	x4x5	x1x3	x2x4	x1x5	x3x5
x2x4	x2x5	x1x3	x4x5	x1x2	x2x3	x3x4	x4x5	x1x4	x2x4	x1x2	x3x5
x1x2	x3x4	x3x5	x4x5	x1x5	x2x5	x3x5	x4x5	x1x5	x2x5	x1x2	x3x4
x1x3	x3x4	x2x3	x3x5	x1x4	x2x5						

Consider removing any other run than 1 or 8 then the resulting design will not satisfy the rank conditions for model discrimination for 18 of 990 different model comparisons. The 18 model comparisons that do not satisfy the rank condition depends on which of the 10 runs is removed. Here is an example of using $d_1^{(-2)}$ which is found by removing run 2 or treatment 125,

Figure 2.3 Rank conditions for $\boldsymbol{d_1^{(-2)}}$



Observe that comparing $d_1^{(-1)}$ and $d_1^{(-2)}$ that $d_1^{(-1)}$ has 45 of the 990 different model comparisons not satisfy the rank condition for model discrimination. For any model M_u, there are two corresponding models for $M_{u'}$ that do not satisfy the model discrimination rank conditions. The second observation is that $d_1^{(-2)}$ has a smaller subset of model comparisons that do not satisfy the rank conditions for model discrimination. Since there are only 18 model comparisons that do not satisfy the rank conditions for model discrimination, it is expected that there exist some models M_u that satisfy the rank conditions for model discrimination for all other $M_{u'}$. This is observed by the following table,

Table 2.3 Models that do not satisfy the model discrimination properties for $d_1^{(-1)}$

u	u'		u	u'	
x1x2 x1x3			x1x4 x3x4	x2x4 x4x5	x1x3 x2x5
x1x2 x1x4			x1x5 x3x4	x2x3 x2x4	x1x2 x2x5
x1x3 x1x4	x2x5 x3x4		x2x3 x3x4	x1x3 x3x5	x1x5 x2x4
x1x2 x1x5	x2x5 x3x4		x2x4 x3x4	x1x4 x4x5	x1x5 x2x3
x1x3 x1x5			x2x5 x3x4	x1x3 x1x4	x1x2 x1x5
x1x4 x1x5			x1x2 x3x5	x3x4 x4x5	
x1x2 x2x3			x1x3 x3x5	x2x3 x3x4	
x1x3 x2x3	x3x4 x3x5		x1x4 x3x5		
x1x4 x2x3			x1x5 x3x5		
x1x5 x2x3	x2x4 x3x4		x2x3 x3x5	x1x3 x3x4	
x1x2 x2x4			x2x4 x3x5		
x1x3 x2x4			x2x5 x3x5		
x1x4 x2x4	x3x4 x4x5		x3x4 x3x5	x1x2 x4x5	x1x3 x2x3
x1x5 x2x4	x2x3 x3x4		x1x2 x4x5	x3x4 x3x5	
x2x3 x2x4	x1x5 x3x4		x1x3 x4x5		
x1x2 x2x5	x1x5 x3x4		x1x4 x4x5	x2x4 x3x4	
x1x3 x2x5	x1x4 x3x4		x1x5 x4x5		
x1x4 x2x5	x1x3 x3x4		x2x3 x4x5		
x1x5 x2x5	x1x2 x3x4		x2x4 x4x5	x1x4 x3x4	
x2x3 x2x5			x2x5 x4x5		
x2x4 x2x5			x3x4 x4x5	x1x4 x2x4	x1x2 x3x5
x1x2 x3x4	x3x5 x4x5	x1x5 x2x5	x3x5 x4x5	x1x2 x3x4	
x1x3 x3x4	x2x3 x3x5	x1x4 x2x5			

There are three different cases of model discrimination for $d_1^{(-2)}$, models that satisfy the rank conditions for all other 44 models, models that satisfy the rank conditions for all but 1 other models, and models that satisfy the rank conditions for all but 2 other models. If the true model contains the two-factor interactions x_1x_2 and x_1x_3 then the rank conditions will be satisfied using any of the other 44 models. If the true model contains the two-factor interactions x_1x_3 and x_1x_4 then the rank conditions will not be satisfied for the model that contains x_2x_5 and x_3x_4. If the true model contains the two-factor interactions x_1x_2 and x_3x_4 then the rank conditions will not be satisfied for the two models that contain x_3x_5 and x_4x_5 along with x_1x_5 and x_2x_5.

The other 9 designs we can make by removing a run that involves a three-treatment combination will have similar results to $d_1^{(-2)}$ but not satisfy the rank conditions for a different set of 18 models. If there were some information about the true model it may be possible to strategically implement a design $d_1^{(-i)}$ for $i \neq 1,8$ with 11 runs that would not have trouble discriminating with the other 44 models.

Chapter 3

Model Comparisons

3.1 Introduction

In planning an experiment to study the relationship between a response variable Y and factors $A_1, A_2, \ldots, A_m$ for a total of m factors, the parameters of interest are a general mean β_0, m main effects β_j, and k $\left(0 < k < \binom{m}{2}\right)$ two-factor interactions $\beta_{j,j'}$ $(j \neq j', j, j' = 1,2, \ldots m)$. The additive model between the mean of the response variable $E[Y]$ and the parameters of interest are considered for a total of $v_k = \binom{\binom{m}{2}}{k}$ different models. A design d must be proposed consisting of n treatment combinations of the m factors, each treatment combination is defined as a run. For a chosen k, the design d must have at least n runs where $n \geq 1 + m + k$ to satisfy the model identification rank conditions for all v_k models. For model discrimination, the design d must satisfy the rank condition for all $\binom{v_k}{2}$ pairwise model comparisons of $\{M_u, M_{u'}\}$ and have at least n runs where $n \geq 1 + m + 2k$.

In section 2.4, the design d_1 with $n = 12$ runs, $m = 5$ main effects, and $\binom{5}{2} = 10$ two-factor interactions. For $k = 2$, the design d_1 can identify all $v_k = \binom{10}{2} = 45$ models and discriminating between all $\binom{45}{2} = 990$ pairwise comparisons. Assume the true model is within the class of models, in the noiseless case, only the true model will have perfect fit and can be identified as the model with zero error sum of square errors. In the noisy case, instead the model that minimizes the error sum of squares is chosen and there is a non-zero probability of picking the true model.

To understand how each treatment combination affects the two rank conditions, removing run i of design d_1 was considered was considered. We observe that $d_1^{(-i)}$ satisfies the rank conditions for model identification for all 45 models but no longer satisfies the rank conditions for model discrimination for all 990 models for any $i = 1,2, \ldots ,12$. We also observe that removing the treatment combination (1) or 12345 results in a design that any model M_u cannot be discriminated with two other models $M_{u_1'}$ and $M_{u_2'}$. Removing any of the other 10 runs results in a design that is possible to have a model M_u that can discriminate between all other models $M_{u'}$ the exact models depending on which of the 10 runs.

These rank conditions for model identification and discrimination will only provide information to whether a design can successfully identify or discriminate between all $\binom{v_k}{2}$ pairwise model comparisons. The rank conditions for model discrimination does not provide evaluation on how well a design can discriminate between all $\binom{v_k}{2}$ pairs of models.

The error sum of squares is a performance metric to find the model that best describes the data. All $\binom{v_k}{2}$ pairwise error sum of squares is considered between the comparisons $\{M_u, M_{u'}\}$. If a design results in the zero difference for error sum of squares between the comparison $\{M_u, M_{u'}\}$ and will not have any discrimination strength. If the design is capable of larger values of the difference, then the design has more capability of discriminating between models M_u and $M_{u'}$.

3.2 Model Selection for a Fixed Design

For a chosen k, we have a class of models as described in the introduction to chapter 3. Consider a set of designs $\mathcal{D}$ with n runs each in which every design $d \in \mathcal{D}$ is capable of model discrimination for all $\binom{v}{2}$ model comparisons. For a design $d \in \mathcal{D}$ and chosen model M_u, the design matrix $X_d^{(u)}$ consists of n rows as the runs and $1 + m + k = p_u$ columns to represent the overall mean, m main effects, and k two-factor interactions. Define the projection matrix be a function of the design matrix $X_d^{(u)}$ and expressed as $H_d^{(u)} = X_d^{(u)} \left(X_d^{(u)'} X_d^{(u)} \right)^{-1} X_d^{(u)'}$. Once the observation vector $\underline{y}_d$ is obtained, the least squares fitted values $\underline{\hat{y}}_d$ will satisfy,

$$\underline{\hat{y}}_d^{(u)} = H_d^{(u)} \underline{y}_d. \tag{3.1}$$

The error sum of squares for model M_u is used as a performance metric and denoted as $SSE_u(d)$ and can be expressed as the observation vector $\underline{y}_d$ and the projection matrix $H_d^{(u)}$,

$$\begin{aligned} SSE_u(d) &= \left(\underline{y}_d - \underline{\hat{y}}_d^{(u)} \right)' \left(\underline{y}_d - \underline{\hat{y}}_d^{(u)} \right) \\ &= \underline{y}_d' \left[I_n - H_d^{(u)} \right] \underline{y}_d. \end{aligned} \tag{3.2}$$

A small value of error sum of squares implies the fitted values are close to the observed values. For a fixed design d, we select the model M_{u^*} within our class of models that obtains the smallest value of the error sum of squares from all other models M_u,

$$SSE_{u^*}(d) \leq SSE_u(d), \qquad u = 1,2, \dots, v_k \tag{3.3}$$

To find the model M_{u^*} that satisfies equation (3.3), there are a total $\binom{v_k}{2}$ pairwise model comparisons $\{M_u, M_{u'}\}$ that are evaluated. We are interested in the $\binom{v_k}{2}$ differenced error sum of squares,

$$SSE_u(d) - SSE_{u'}(d) = \underline{y}_d' \left[H_d^{(u')} - H_d^{(u)}\right] \underline{y}_d.$$

Let the differenced projection matrix of $H_d^{(u')}$ and $H_d^{(u)}$ be $H_d^{(u,u')}$ for convenience, then express the differenced sum of squares errors for model comparison $\{M_u, M_{u'}\}$ to be,

$$SSE_u(d) - SSE_{u'}(d) = \underline{y}_d' \left[H_d^{(u,u')}\right] \underline{y}_d. \qquad (3.4)$$

For a design d, there is only one observation vector $\underline{y}_d$ and $\binom{v_k}{2}$ $H_d^{(u,u')}$ matrices. If the differenced error sum of squares for the model comparison $\{M_u, M_{u'}\}$ is zero, we understand this implies the error sum of squares between these two models are equivalent.

3.3 Min/Max of the Standardized Differenced Error Sum of Squares

Results from Bellman (1960) can be used to bound the differenced error sum of squares. Consider the standardized differenced error sum of squares for the model comparison $\{M_u, M_{u'}\}$ by taking the differenced error sum of squares and dividing the sum of squares of the observation vector $\underline{y}_d$,

$$\frac{SSE_u(d) - SSE_{u'}(d)}{\underline{y}_d'\underline{y}_d} = \frac{\underline{y}_d' \left[H_d^{(u,u')}\right] \underline{y}_d}{\underline{y}_d'\underline{y}_d}.$$

Bellman's results bound the standardized difference error sum of squares,

$$\text{Min}\left\{EV\left(H_d^{(u,u')}\right)\right\} \le \frac{\underline{y}_d'\left[H_d^{(u,u')}\right]\underline{y}_d}{\underline{y}_d'\underline{y}_d} \le \text{Max}\left\{EV\left(H_d^{(u,u')}\right)\right\}.$$

Let λ^* be the maximum eigenvalue for the differenced projection matrix between models M_u and $M_{u'}$. Chapter 3 section 4 demonstrates that for a design $d \in D$, if λ is an eigenvalue then so is $-\lambda$ and it follows $-\lambda$ is the calculated minimum eigenvalue for the differenced projection matrix for the model comparison $\{M_u, M_{u'}\}$ and the standardized difference error sum of squares bounds are,

$$\left|\frac{SSE_u(d) - SSE_{u'}(d)}{\underline{y}_d'\underline{y}_d}\right| \le \lambda^*. \qquad (3.5)$$

The eigenvalues of the differenced projection matrix can be calculated without our observation vector $\underline{y}_d$. It is clearly seen that if $\lambda^* = 0$, then $SSE_u(d) = SSE_{u'}(d)$ for any observation vector $\underline{y}_d'$ and we will not be able to discriminate between models M_u and $M_{u'}$ as would be the case if the rank conditions for model discrimination are not satisfied.

3.4.1 Properties of the Differenced Projection Matrix

Consider two different designs d and d' in our class of designs $\mathcal{D}$. There are some instances where design d and d' will have the same exact eigenvalues for all $\binom{v_k}{2}$ pairwise model comparisons. For a design d, let X_d be the matrix with n rows and $p = 1 + m + \binom{m}{2}$ columns with the intercept, main effects, all two-factor interactions. A similar matrix can be constructed for d'.

Define an isomorphism between designs d and d' if there exists a p x p non-singular square matrix T that satisfies $X_d = X_{d'}T$. For any model M_u we can construct the submatrix $X_d^{(u)}$ and $T^{(u)}$ by taking the corresponding columns of X_d and T and which will also satisfy $X_d^{(u)} = X_{d'}^{(u)}T^{(u)}$. Since T is a non-singular square matrix, the corresponding submatrix $T^{(u)}$ is non-singular and invertible. This allows us to understand when the eigenvalues of d and d' will be the same for $\binom{v_k}{2}$ model comparisons.

Theorem 3.1 If design d and d' are isomorphic, then the eigenvalues of the differenced projection matrix for any model comparison $\{M_u, M_{u'}\}$ will be equivalent for both d and d'. The proof follows from the definition of an isomorphic design. For any model M_u the projection matrix for design d is expressed as,

$$H_d^{(u)} = X_d^{(u)}\left(X_d^{(u)'}X_d^{(u)}\right)X_d^{(u)'}.$$

Since design d is isomorphic to d' by the definition, there exists a $T^{(u)}$ that satisfies $X_d^{(u)} = X_{d'}^{(u)}T^{(u)}$ and is invertible,

$$H_d^{(u)} = X_{d'}^{(u)}T^{(u)}\left(T^{(u)'}X_{d'}^{(u)'}X_{d'}^{(u)}T^{(u)}\right)^{-1}T^{(u)'}X_{d'}^{(u)'}$$

$$= X_{d'}^{(u)}\left(X_{d'}^{(u)'}X_{d'}^{(u)}\right)X_{d'}^{(u)'} = H_{d'}^{(u)}.$$

It must follow for any pairwise model comparison the differenced projection matrix is equivalent between design d and d' and therefore the eigenvalues will be the same. It is not necessary to compare the eigenvalues for isomorphic designs.

Theorem 3.2 For any design d that satisfies the rank conditions of model identification for both models M_u and $M_{u'}$, the sum of the eigenvalues for $H_d^{(u,u')}$ is 0. This is observed by looking at the trace of the differenced projection matrix,

$$Trace\left(H_d^{(u,u')}\right) = Trace\left(H_d^{(u')}\right) - Trace\left(H_d^{(u)}\right).$$

The trace for the projection matrix of either M_u or $M_{u'}$ can be calculated to be the number of columns of the design matrix $X_d^{(u)}$,

$$Trace\left(H_d^{(u)}\right) = Trace\left(X_d^{(u)}\left(X_d^{(u)\prime} X_d^{(u)}\right) X_d^{(u)\prime}\right) = Trace\left(\left(X_d^{(u)\prime} X_d^{(u)}\right) X_d^{(u)\prime} X_d^{(u)}\right)$$

$$= Trace\left(I_{p_u}\right) = p_u.$$

For any of the v_k models, we have $p_1 = p_2 = \cdots = p_u = \cdots = p_{v_k}$. Then the difference of the traces is equivalently 0 for any pairwise model comparison of M_u and $M_{u'}$

$$Trace\left(H_{d_i}^{(u')}\right) - Trace\left(H_{d_i}^{(u)}\right) = p_{u'} - p_u = 0.$$

A side note, if a matrix has trace 0 it is also a “Commutator”. The differenced projection matrix can be expressed as n x n square matrices A and B that satisfy the equation,

$$H_d^{(u,u')} = AB - BA.$$

To understand additional properties of the differenced projection matrix, a short discussion about permutation matrices is needed.

3.4.2 Permutation Matrices

A permutation matrix P_m can be constructed by starting out with an identity matrix then permuting the rows or columns. When pre-multiplying by the permutation P_m with a matrix A, the resulting matrix P_mA is the **row** permuted form of the matrix A. If the permutation matrix P_m was constructed by taking the identity matrix and permuting the first two rows then $P_m A$ will be the matrix A with the first two rows permuted. By post-multiplying A with P_m, AP_m the resulting matrix is the first two columns permuted. Let A to be a square matrix, if we define P_m to be a set of row permutations from an identity matrix, P_m' will be the corresponding column permutations from an identity matrix then $P_m A P_m'$ will be the square matrix A with the corresponding row and column permutations. Permutation matrices have the following properties,

1) Any permutation matrix P_m will satisfy the equations,

$$P_m' P_m = P_m P_m' = I_n$$

$$(P_m')^{-1} = P_m$$

$$(P_m)^{-1} = P_m'$$

2) If a permutation matrix P_m is composed of disjoint permutations in the form $i \leftrightarrow i'$, the permutation matrix P_m is also symmetric and $P_m = P_m' = P_m^{-1}$. An example is a 4 x 4 square permutation matrix where the only permutation is 1 ↔ 2.

3) $Trace(P_m)$ is the number of permutations where $i \leftrightarrow i$ or that the row/column was not permuted. In the example for 2), the trace of that permutation matrix would be equal to two as only row/column 1 and 2 were permuted out of the 4 rows/columns.

3.4.3 Properties of the Differenced Projection Matrix (Continued)

It is sometimes the case where the comparison between models M_{u_1} and $M_{u_1'}$ is a permuted form of the comparison between the models M_{u_2} and $M_{u_2'}$. There exists a permutation matrix P_m where either,

$$H_d^{(u_1,u_1')} = P_m H_d^{(u_2,u_2')} P_m', \tag{3.6.1}$$

$$H_d^{(u_1,u_1')} = -P_m H_d^{(u_2,u_2')} P_m'. \tag{3.6.2}$$

Theorem 3.3 For design d, if there is a permutation matrix that satisfies equation (3.6.1) or (3.6.2) for two different pairwise model comparisons $\{M_{u_1}, M_{u_1'}\}$ and $\{M_{u_2}, M_{u_2'}\}$, then the eigenvalues of the differenced projection matrix will be equivalent for both pairwise model comparisons. Assume equation (3.6.1) satisfies and let λ be an eigenvalue for the first model comparison $H_d^{(u_1,u_1')}$ with eigenvector $\underline{v}$. Using the definition of eigenvalues and eigenvectors,

$$H_d^{(u_1,u_1')}\underline{v} = \lambda\underline{v}.$$

The transpose of a permutation matrix is the inverse. We can multiply a form of the identity on both sides of the equation,

$$H_d^{(u_1,u_1')}P_m'P_m\underline{v} = \lambda P_m'P_m\underline{v},$$

pre-multiplying by the permutation matrix P_m on both sides to get the equation and using (3.6.1),

$$P_m H_d^{(u_1,u_1')}P_m'P_m\underline{v} = \lambda P_m P_m'P_m\underline{v},$$

$$\Rightarrow H_d^{(u_2,u_2')}(P_m\underline{v}) = \lambda P_m\underline{v},$$

Observe that λ is also an eigenvalue for $H_d^{(u_2,u_2')}$ with the permuted form of the eigenvector $P_m\underline{v}$. A similar proof will be shown for equation 3.6.2 but will need additional results to prove. A strategy that is used is to classify all $\binom{v_k}{2}$ pairwise model comparisons into classes where each comparison within a class satisfies equation (3.6.1) or equation (3.6.2). Theorem 3.3 states it is sufficient to understand the structure within each class and the comparisons for each class will have similar structures. For pairwise model comparisons within the same class, the differenced error sum of squares can be expressed as,

$$\underline{y}_d'\left(H_d^{(u_1,u_1')}\right)\underline{y}_d = \underline{y}_d'\left(P_m H_d^{(u_2,u_2')}P_m'\right)\underline{y}_d = \underline{y}_d^{*\prime}\left(H_d^{(u_2,u_2')}\right)\underline{y}_d^{*}.$$

This demonstrates the differenced error sum of squares within a class can be calculated by just permuting the observation vector $\underline{y}_d$.

If $Rank\left(H_d^{(u,u')}\right) = 2$, it is automatically seen that the eigenvalues for the differenced projection matrix for the comparison $\{M_u, M_{u'}\}$ must be $\pm\lambda$ from Theorem 3.2. For any design d that satisfies the rank conditions for model discrimination, want to infer the eigenvalues of the differenced projection matrix will have the form $EV\left(H_d^{(u,u')}\right) = \pm\lambda_1, \pm\lambda_2, \dots, \pm\lambda_r$. When comparing models M_u and $M_{u'}$ the overall mean, main effects, and $c_{uu'}$ two-factor interactions are in common with $2(k - c_{uu'})$ two-factor interactions are not in common. For any design d, the rank of the differenced projection matrix will be less than or equal to $2(k - c_{uu'})$ or that $Rank\left(H_d^{(u,u')}\right) \leq 2r$ for $r = k - c_{uu'}$, equality holds if the model discrimination rank conditions are satisfied. The number of $\binom{v_k}{2}$ pairwise model comparisons that have $c_{uu'} = i$ can be calculated by,

$$\frac{1}{2}\binom{\binom{m}{2}}{i}\binom{\binom{m}{2} - i}{k - i}\binom{\binom{m}{2} - i - (k - i)}{k - i}, \tag{3.7}$$

The logic of this formulation is to start with all $\binom{m}{2}$ two-factor interactions and remove the i common two-factor interactions for a total of $\binom{\binom{m}{2}}{i}$ combinations of doing this. Each model will have $(k - i)$ unique two-factor interactions. Model M_u will choose $k - i$ two-factor interactions of the $\binom{m}{2} - i$ remaining two-factor interactions after removing the common ones, and model $M_{u'}$ to receive $k - i$ two-factor interactions out the remaining $\binom{m}{2} - i - (k - i)$ two-factor interactions. Since assignment of model M_u and $M_{u'}$ does not matter, the number of unique combinations are halved. If $m = 5$, 10 two-factor interactions, and $k = 3$, then the total number of pairwise model comparisons will be 7140 with 2100 having $c_{uu'} = 0$, 3780 having $c_{uu'} = 1$, and 1260 having $c_{uu'} = 2$. Using this example, for a design d that satisfies the model discrimination properties, observe that there will be 2100 pairwise model comparisons where $Rank\left(H_d^{(u,u')}\right) = 6$, 3780 pairwise model comparisons where $Rank\left(H_d^{(u,u')}\right) = 4$, and 1260 pairwise model combinations where $Rank\left(H_d^{(u,u')}\right) = 2$. Theorem 3.4 gives us a necessary condition when $EV\left(H_d^{(u,u')}\right) = \pm\lambda_1, \pm\lambda_2, \dots, \pm\lambda_r$.

Theorem 3.4 Suppose that for any n x n matrix A and C where $CAC^{-1} = -A$. If λ is an eigenvalue with eigenvector $\underline{v}$ of A, then $-\lambda$ is another eigenvalue of A with eigenvector $C\underline{v}$.

If λ is an eigenvalue and $\underline{v}$ is an eigenvector of A, by the definition of eigenvalues and eigenvectors, we know that $A\underline{v} = \lambda\underline{v}$. For any invertible C we can multiply forms of the identity on both sides $AC^{-1}C\underline{v} = \lambda C^{-1}C\underline{v}$. By pre-multiplying C on both sides we obtain $CAC^{-1}C\underline{v} = \lambda CC^{-1}\underline{v}$. Substituting in $CAC^{-1} = -A$, our equation is then $-AC\underline{v} = \lambda C\underline{v}$ which can be written as $A(C\underline{v}) = (-\lambda)(C\underline{v})$. Hence, $-\lambda$ is also an eigenvalue and $C\underline{v}$ is an eigenvector of A. Recall from theorem 3.1 that the differenced projection matrices are Commutators and can be expressed in the form $H_d^{(u,u')} = AB - BA$. We see that if there exists a C such that,

$$C = C^{-1},$$

$$CABC = BA.$$

Theorem 3.4 can be used to show that, $CH_d^{(u,u')}C^{-1} = CABC - CBAC = BA - AB = -H_d^{(u,u')}$. When comparing designs in further chapters, examples of C will be given hence it is observed that $EV\left(H_d^{(u,u')}\right) = \pm\lambda_1, \pm\lambda_2, \dots, \pm\lambda_r$. In general, let X_1 and X_2 be any two real matrices, Ansderson et al. (1985) proves the eigenvalues for the differenced projection matrices using X_1 and X_2 will have the same eigenvalue structure. .Recall theorem 3.3, since the eigenvalues are symmetric about 0, if equation (3.5.2) is true instead of equation (3.6.1), let the eigenvalues for the first pairwise comparison be $EV\left(H_d^{(u_1,u_1')}\right) = \pm\lambda_1, \pm\lambda_2, \dots, \pm\lambda_r$ then the eigenvalues for the second pairwise comparison will be $EV\left(H_d^{(u_2,u_2')}\right) = -EV\left(H_d^{(u_1,u_1')}\right) = \pm\lambda_1, \pm\lambda_2, \dots, \pm\lambda_r$

3.5 Eigenvalue Decomposition

Consider design d, the observation vector is observed $\underline{y}_d$, then the differenced error sum of squares for model comparison $\{M_u, M_{u'}\}$ can be calculated using the observation vector $\underline{y}_d$ and the differenced projection matrix $H_d^{(u,u')}$

$$SSE_u(d) - SSE_{u'}(d) = \underline{y}_d'\left[H_d^{(u,u')}\right]\underline{y}_d$$

For $Rank\left(H_d^{(u,u')}\right) = 2r$ with eigenvalues $\pm\lambda_1^{(d,u,u')}, \pm\lambda_2^{(d,u,u')}, \dots, \pm\lambda_r^{(d,u,u')}$ there are orthonormal eigenvectors for each positive and negative eigenvalue, $\{\underline{a}_1, \underline{b}_1\}, \{\underline{a}_2, \underline{b}_2\}, \dots, \{\underline{a}_r, \underline{b}_r\}$. The eigenvectors depend on both the design d and the pairwise model comparison M_u and $M_{u'}$, notation has suppressed this for simplicity. $\{\underline{a}_1, \underline{b}_1\}, \{\underline{a}_2, \underline{b}_2\}, \dots, \{\underline{a}_r, \underline{b}_r\}$ The set of eigenvectors is an orthonormal basis for $H_d^{(u,u')}$ and satisfy the set of equations,

$$\underline{1}'\underline{a}_i = \underline{1}'\underline{b}_i = 0, \quad i = 1,2, \dots, r$$

$$\underline{a}_i{}'\underline{b}_i = 0, \quad i = 1,2, \dots, r$$

$$\underline{a}_i{}'\underline{a}_{i'} = \underline{b}_i{}'\underline{b}_{i'} = \begin{cases} 0 & \text{for } i \neq i' \\ \\ 1 & \text{for } i = i' \end{cases} \quad i, i' = 1, \dots, r$$

Using the properties of an orthonormal set of eigenvectors, an alternative form to the differenced error sum of squares that follows from equation (3.6),

$$SSE_u(d) - SSE_{u'}(d) = \sum_{i=1}^{r} \lambda_i^{(d,u,u')} \left[\left(\underline{a}_i + \underline{b}_i\right)' \underline{y}_d \right] \left[\left(\underline{a}_i - \underline{b}_i\right)' \underline{y}_d \right].$$

Define the random variables $S_{i,1}^{(d,u,u')}$ and $S_{i,2}^{(d,u,u')}$ to be $S_{i,1}^{(d,u,u')} = \left(\underline{a}_i + \underline{b}_i\right)' \underline{y}_d$ and $S_{i,2}^{(d,u,u')} = \left(\underline{a}_i - \underline{b}_i\right)' \underline{y}_d$. Observe that $S_{i,1}^{(d,u,u')}$ and $S_{i,2}^{(d,u,u')}$ are uncorrelated random variables since for a fixed i,

$$Cov\left(S_{i,1}^{(d,u,u')}, S_{i,2}^{(d,u,u')}\right) = Cov\left(\left(\underline{a}_i + \underline{b}_i\right)' \underline{y}_d, \left(\underline{a}_i - \underline{b}_i\right)' \underline{y}_d\right) = \left(\underline{a}_i' \, \underline{a}_i - \underline{b}_i' \, \underline{b}_i\right) * Cov\left(\underline{y}_d' \underline{y}_d\right).$$

From the properties of the orthonormal eigenvectors $\underline{a}_i$ and $\underline{b}_i$ $\underline{a}_i' \, \underline{a}_i = \underline{b}_i' \, \underline{b}_i = 1$,

$$Cov\left(S_{i,1}^{(d,u,u')}, S_{i,2}^{(d,u,u')}\right) = 0.$$

The differenced error sum of squares is a weighted sum of the product of uncorrelated random $S_{i,1}^{(d,u,u')}$ and $S_{i,2}^{(d,u,u')}$ variables with weights $\lambda_i^{(d,u,u')}$,

$$SSE_u(d) - SSE_{u'}(d) = \sum_{i=1}^{r} \lambda_i^{(d,u,u')} S_{i,1}^{(d,u,u')} S_{i,2}^{(d,u,u')}, \tag{3.8}$$

and the expected value of the differenced error sum of squares is expressed as,

$$E[SSE_u(d) - SSE_{u'}(d)] = \sum_{i=1}^{r} \lambda_i^{(d,u,u')} E\left[S_{i,1}^{(d,u,u')} S_{i,2}^{(d,u,u')}\right].$$

Since $S_{i,1}^{(d,u,u')}$ and $S_{i,2}^{(d,u,u')}$ are uncorrelated $\left(Cov\left(S_{i,1}^{(d,u,u')}, S_{i,2}^{(d,u,u')}\right) = 0\right)$. The expected value of the differenced error sum of squares can be written as a product of expectations,

$$E[SSE_u(d) - SSE_{u'}(d)] = \sum_{i=1}^{r} \lambda_i^{(d,u,u')} E\left[S_{i,1}^{(d,u,u')}\right] E\left[S_{i,2}^{(d,u,u')}\right]. \tag{3.9}$$

Recall from section 2.2 that the total number of parameters considered is $p = 1 + m + \binom{m}{2}$ for the one general mean, m main effects, and all $\binom{m}{2}$ two-factor interactions. Let X_d be the n x p matrix with n runs and the columns for the overall mean, main effects, and all two-factor interactions, $\underline{\beta}$ be the corresponding values for the unknown parameters which may or may not be zero. The means can be expressed in terms of $\underline{\beta}$,

$$E\left[S_{i,1}^{(d,u,u')}\right] = \left(\underline{a}_i + \underline{b}_i\right)' X_d \underline{\beta}, \quad i = 1,2,\dots,r \tag{3.10.1}$$

$$E\left[S_{i,2}^{(d,u,u')}\right] = \left(\underline{a}_i - \underline{b}_i\right)' X_d \underline{\beta}, \quad i = 1,2,\dots,r \tag{3.10.1}$$

The 1 x p vectors $\left(\underline{a}_i + \underline{b}_i\right)' X_d$ and $\left(\underline{a}_i - \underline{b}_i\right)' X_d$ are interpreted as the weights for the unknown parameters $\underline{\beta}$. This will be useful in understanding effect of the true model on the differenced error sum of squares for a design d and pairwise model comparison M_u and $M_{u'}$.

Consider any design d that satisfies the rank conditions for model discrimination. For the pairwise model comparison $\{M_u, M_{u'}\}$, let $X_d^{(uu')}$ be the columns of X_d that are in common between $X_d^{(u)}$ and $X_d^{(u')}$. These columns consist of the general mean, m main effects, and $c_{uu'}$ two-factor interactions. The differenced error sum of squares will not depend on the common parameters between models M_u and $M_{u'}$ or the following equations are satisfied,

$$\left(\underline{a}_i + \underline{b}_i\right)' X_d^{(uu')} = 0, \quad i = 1,2,..,r$$

$$\left(\underline{a}_i - \underline{b}_i\right)' X_d^{(uu')} = \underline{0}, \quad i = 1,2,\dots,r$$

Using the properties of the eigenvectors, the variances of the uncorrelated random variables $S_{i,1}^{(d,u,u')}$ and $S_{i,2}^{(d,u,u')}$ are calculated to be,

$$Var\left(\left(\underline{a}_i + \underline{b}_i\right)' \underline{y}_d\right) = \left(\underline{a}_i + \underline{b}_i\right)' \sigma^2 I_n \left(\underline{a}_i + \underline{b}_i\right) = 2\sigma^2$$

$$Var\left(\left(\underline{a}_i - \underline{b}_i\right)' \underline{y}_d\right) = \left(\underline{a}_i - \underline{b}_i\right)' \sigma^2 I_n \left(\underline{a}_i - \underline{b}_i\right) = 2\sigma^2$$

The variances of $S_{i,1}^{(d,u,u')}$ and $S_{i,2}^{(d,u,u')}$ are equal and proportional to the variance of the observations $\underline{y}$. This demonstrates only the means of the differenced error sum of squares depending on the unknown parameters $\underline{\beta}$ and the variances only depends on the unknown parameter σ^2.

3.6 Special Case: Orthogonal Designs

Let d^* be an orthogonal design with n runs, m main effects, with full estimation capacity for all two-factor interactions for a total of $p = 1 + m + \binom{m}{2}$ parameters. The design matrix X_{d^*} which has columns for the general mean, m main effects, and all $\binom{m}{2}$ two-factor interactions has dimensions n x p and satisfies the equation $X'_{d^*}X_{d^*} = nI_p$.

For a given k, an orthogonal design with no replicates will satisfy the rank conditions for model identification for all $v_k = \binom{\binom{m}{2}}{k}$ models if $n \geq 1 + m + k$ or will satisfy the rank conditions for model discrimination for all $\binom{v_k}{2}$ pairwise model comparisons if $n \geq 1 + m + 2k$. If an orthogonal design with no replicates satisfies the model identification or model discrimination rank conditions, then the resulting design with replicates will also satisfy the desired rank conditions. When estimating model M_u the projection matrix is simplified to be $H_{d^*}^{(u)} = \frac{1}{n}\left(X_{d^*}^{(u)} X_{d^*}^{(u)'}\right)$ and the differenced projection matrix can be calculated to be,

$$H_{d^*}^{(u,u')} = H_{d^*}^{(u')} - H_{d^*}^{(u)} = \frac{1}{n}\left(X_{d^*}^{(u')} X_{d^*}^{(u')'} - X_{d^*}^{(u)} X_{d^*}^{(u)'}\right), \tag{3.11}$$

For an orthogonal design, it is simple to show that the differenced projection matrix is only a function on the columns of $X_{d^*}^{(u)}$ and $X_{d}^{(u')}$ which are not in common with each other. Given the model comparison $\{M_u, M_{u'}\}$, denote $X_{d^*}^{(uu')}$ to be the columns of $X_{d^*}^{(u)}$ and $X_{d^*}^{(u')}$ that are in common, $X_{d^*}^{(u\bar{u}')}$ to be the columns of $X_{d^*}^{(u)}$ that are not in $X_{d^*}^{(u')}$, $X_{d^*}^{(u\bar{u}')}$ to be the columns of $X_{d^*}^{(u')}$ that are not in $X_{d^*}^{(u)}$ and $X_{d^*}^{(\underline{uu}')}$ are the remaining columns in X_{d^*} that are in not in $X_{d^*}^{(u)}$ and not in $X_{d^*}^{(u')}$. The number of two-factor interactions in common is $c_{uu'}$ $(0 \le c_{uu'} < k)$. Each matrix has n rows and the number of columns are summarized in the table 3.1,

Table 3.1 Summary of notations the number of columns for each matrix

$M_u \backslash M_{u'}$	$M_{u'}$	$M_{\bar{u}'}$
M_u	$X_{d^*}^{(uu')}$ $1 + m + c_{uu'}$	$X_{d^*}^{(u\bar{u}')}$ $k - c_{uu'}$
$M_{\bar{u}}$	$X_{d^*}^{(\bar{u}u')}$ $k - c_{uu'}$	$X_{d^*}^{(\bar{u}\bar{u}')}$ $p - (1 + m + 2k - c_{uu'})$

To separate out the columns that are in common and not in common, the design matrices for model M_u and $M_{u'}$ can then be expressed with the notation in table 3.1,

$$X_{d^*}^{(u)} = \left[X_{d^*}^{(uu')} | X_{d^*}^{(u\bar{u}')}\right],$$

$$X_{d^*}^{(u')} = \left[X_{d^*}^{(uu')} | X_{d^*}^{(\bar{u}u')}\right].$$

The projection matrix for model M_u and $M_{u'}$ can be expressed as functions of $X_{d^*}^{(uu')}$, $X_{d^*}^{(u\bar{u}')}$, and $X_{d^*}^{(\bar{u}u')}$,

$$H_{d^*}^{(u)} = \frac{1}{n}\left(X_{d^*}^{(u)} X_{d^*}^{(u)'}\right) = \frac{1}{n}\left(X_{d^*}^{(uu')} X_{d^*}^{(uu')'} + X_{d^*}^{(u\bar{u}')} X_{d^*}^{(u\bar{u}')'}\right),$$

$$H_{d^*}^{(u')} = \frac{1}{n}\left(X_{d^*}^{(u')} X_{d^*}^{(u')'}\right) = \frac{1}{n}\left(X_{d^*}^{(uu')} X_{d^*}^{(uu')'} + X_{d^*}^{(\bar{u}u')} X_{d^*}^{(\bar{u}u')'}\right).$$

This demonstrates that the differenced projection matrix presented in (3.11) can be expressed as only the columns that are not in common between $X_{d^*}^{(u)}$ and $X_{d^*}^{(u')}$,

$$H_{d^*}^{(u,u')} = \frac{1}{n}\left(X_{d^*}^{(\bar{u}u')} X_{d^*}^{(\bar{u}u')'} - X_{d^*}^{(u\bar{u}')} X_{d^*}^{(u\bar{u}')'}\right) \tag{3.12}$$

If λ is an eigenvalue with eigenvector $\underline{v}$ and there exists a C that satisfies $CH_{d^*}^{(u,u')}C^{-1} = -H_{d^*}^{(u,u')}$, then $-\lambda$ is an eigenvalue with eigenvector $C\underline{v}$. Using the expression of the differenced projection matrix for an orthogonal design, there are many ways of constructing a C that satisfies the equation in theorem 3.4. Let $P_m^{(1)}$ and $P_m^{(2)}$ be two different permutation matrices with non-zero elements being either 1 or -1 (-1 if the row/column to be multiplied by -1) that satisfy the equations,

$$P_m^{(1)} = P_m^{(1)'},$$

$$X_{d^*}^{(\bar{u}u')} = P_m^{(1)} X_{d^*}^{(u\bar{u}')} P_m^{(2)},$$

$$\Rightarrow X_{d^*}^{(u\bar{u}')} = P_m^{(1)} X_{d^*}^{(\bar{u}u')} P_m^{(2)'}.$$

The interpretation of these equations is that there exists a symmetric row permutation $P_m^{(1)}$ and some column permutation $P_m^{(2)}$ that can be used to permutate $X_{d^*}^{(u\bar{u}')}$ to obtain $X_{d^*}^{(\bar{u}u')}$. Then observe for an orthogonal design d^*,

$$H_{d^*}^{(u,u')} = \frac{1}{n}\left(P_m^{(1)} X_{d^*}^{(u\bar{u}')} P_m^{(2)} P_m^{(2)'} X_{d^*}^{(u\bar{u}')'} P_m^{(1)} - P_m^{(1)} X_{d^*}^{(\bar{u}u')} P_m^{(2)'} P_m^{(2)} X_{d^*}^{(\bar{u}u')'} P_m^{(1)}\right)$$

$$= \frac{1}{n}\left(P_m^{(1)} X_{d^*}^{(u\bar{u}')} X_{d^*}^{(u\bar{u}')'} P_m^{(1)} - P_m^{(1)} X_{d^*}^{(\bar{u}u')} X_{d^*}^{(\bar{u}u')'} P_m^{(1)}\right),$$

the equation in theorem 3.4 is satisfied with $C = P_m^{(1)}$,

$$P_m^{(1)} H_{d^*}^{(u,u')} P_m^{(1)} = \frac{1}{n}\left(X_{d^*}^{(u\bar{u}')} X_{d^*}^{(u\bar{u}')'} - X_{d^*}^{(\bar{u}u')} X_{d^*}^{(\bar{u}u')'}\right) = -H_{d^*}^{(u,u')}.$$

This result allows us to clearly see the eigenvalue structure of an orthogonal design. The number of orthogonal columns calculated in $H_{d^*}^{(u,u')}$ is $2(k - c_{uu'})$. The matrixes $X_{d^*}^{(u\bar{u}')}$ and $X_{d^*}^{(\bar{u}u')}$ are subsets of an orthogonal matrix hence $\frac{1}{n}\left(X_{d^*}^{(u\bar{u}')} X_{d^*}^{(u\bar{u}')'}\right)\left(X_{d^*}^{(u\bar{u}')} X_{d^*}^{(u\bar{u}')'}\right) = X_{d^*}^{(u\bar{u}')} X_{d^*}^{(u\bar{u}')'}$. The eigenvalues for an orthogonal design will always be sets of ± 1. For a value of k and design d^* that satisfies the model discrimination rank conditions, the eigenvalues of the differenced projection matrix will be in the form,

$$EV\left(H_{d^*}^{(u,u')}\right) = \pm 1_1, \dots, \pm 1_r \ ; \ r = 2(k - c_{uu'}).$$

This is intuitive since all two-factor estimates are orthogonal to each other hence an orthogonal design allows for the full capacity of model discrimination across all model comparisons.

3.7 Analysis of a Fixed Design

For a chosen k, let a design d satisfy the model discrimination rank conditions. Our class of models will consist of $v_k = \binom{\binom{m}{2}}{k}$ identifiable models and $\binom{v_k}{2}$ pairwise model comparisons. Before data is collected, we have $\binom{v_k}{2}$ different matrices for the differenced projection matrix $H_d^{(u,u')}$. For a given design d,

1) Partition each pairwise model comparison $H_d^{(u,u')}$ into sets where any two pairwise model comparison M_{u_1} and $M_{u_1'}$ with M_{u_2} and $M_{u_2'}$ will satisfy either,

$$H_d^{(u_1,u_1')} = P_m H_d^{(u_2,u_2')} P_m',$$

$$H_d^{(u_1,u_1')} = -P_m H_d^{(u_2,u_2')} P_m'.$$

For model comparisons within each set, the eigenvalues will be the equivalent and the eigenvectors will be of some permuted form.

2) For a model comparison within a set, observe $\lambda_i^{(d,u,u')}$, $S_{i,1}^{(d,u,u')}$, and $S_{i,2}^{(d,u,u')}$ for $i = 1,2,\dots,r$ using the computed eigenvalues and eigenvectors. For a pairwise model

comparison $\{M_u, M_{u'}\}$, the values of the differenced error sum of squares can be characterized with the above quantities.

3) The expressions of $E\left[S_{i,1}^{(d,u,u')}\right]$ and $E\left[S_{i,2}^{(d,u,u')}\right]$ can be calculated as functions of the true model $\underline{\beta}$. For a true model and given values of $\underline{\beta}$, we can compare designs and determine which design will give larger values of the differenced sum of squares on average.

3.8 Example

Consider the design d_1 that illustrated the rank conditions for chapter 2 section 4. The design d_1 has $n = 12$ runs, 5 main effects, and $\binom{5}{2} = 10$ two-factor interactions. The design d_1 is capable of model discrimination for pairwise model comparisons with $k = 2$ two-factor interactions. For $k = 2$, there are $\binom{10}{2} = 45$ models to be estimated and $\binom{45}{2} = 990$ pairwise model comparisons. The differenced projection matrix $H_{d_1}^{(u,u')}$is considered for each of the 990 pairwise model comparisons $\{M_u, M_{u'}\}$.

For the model comparison $\{M_u, M_{u'}\}$ there is either one two-factor interaction in common $c_{uu'} = 1$ or no two-factor interactions in common $c_{uu'} = 0$. It is not possible to have $c_{uu'} = 2$ for $k = 2$ since $M_u \neq M_{u'}$. There are $\frac{1}{2}\binom{10}{1}\binom{9}{1}\binom{8}{1} = 360$ pairwise model combinations where $c_{uu'} = 1$ hence $Rank\left(H_{d_1}^{(u,u')}\right) = 2(k - c_{uu'}) = 2$ with eigenvalues $\pm\lambda_1$. There are $\frac{1}{2}\binom{10}{2}\binom{8}{2} = 630$ pairwise model comparisons where $c_{uu'} = 0$ and $Rank\left(H_{d_1}^{(u,u')}\right) = 4$ with eigenvalues $\pm\lambda_1, \pm\lambda_2$. The strategy is to classify the 990 pairwise model comparisons into unique sets where within a set, each $H_{d_1}^{(u,u')}$ are permutations of each other. There are a total of 17 unique sets with eigenvalues as follows,

Table 3.2 Classifications of the 990 Comparisons and their eigenvalues

Set #	Number of Comparisons	Eigenvalues
1	60	± 0.8211
2	60	± 0.8757
3	30	± 0.9374
4.1	60	± 0.9994
4.2	30	± 0.9994
5	120	± 0.9998
6	60	$\pm 0.9495, \pm 0.6045$
7	30	$\pm 1, \pm 0.569467$
8	60	$\pm 0.9267, \pm 0.7071$
9	60	$\pm 1, \pm 0.7184$
10	60	$\pm 0.9994, \pm 0.7485$
11	30	$\pm 0.885, \pm 0.866$
12	120	$\pm 0.9994, \pm 0.8206$
13	120	$\pm 0.9994, \pm 0.8751$
14	15	$\pm 1, \pm 0.9976$
15	60	$\pm 1, \pm 0.9984$
16	15	$\pm 1, \pm 0.9990$

Caution is advised when classifying into sets, there are two sets with eigenvalues of ± 0.9994 but the $H_{d_1}^{(u,u')}$ matrices are not permutations of each other across the sets. Each comparison within the set has been verified to be permutations of each other. Let $\{M_{u_1}, M_{u_1'}\}$ and $\{M_{u_2}, M_{u_2'}\}$ be two pairwise comparisons within a set, a naïve computational approach at this is to try all $12! = 479{,}001{,}600$ permutations until the permutation satisfies $H_{d_1}^{(u_1,u_1')} = P_m H_{d_1}^{(u_2,u_2')} P_m'$. Instead, a simple algorithm to verify that all $H_{d_1}^{(u,u')}$ matrices within a set are permutations of each other involves finding all permutations for the eigenvectors that satisfy $\underline{a}_1{}^{(d,u_1,u_1')} = P_m \underline{a}_1^{(d,u_2,u_2')}$

if equation (3.6.1) holds or $\underline{a}_1{}^{(d,u_1,u_1')} = P_m \underline{b}_1^{(d,u_2,u_2')}$ if equation (3.5.2) holds which usually only $6! = 720$ possible permutations since it is often that 6 elements of eigenvectors are already equivalent. Out of the possible 720 permutations, 6 to 12 of them is tested to find the one that satisfies $H_{d_1}^{(u_1,u_1')} = P_m H_{d_1}^{(u_2,u_2')} P_m'$.

Consider set 1 where the two eigenvalues are ± 0.8211, their corresponding eigenvectors $\underline{a}_1^{(d,u,u')}$ and $\underline{b}_1^{(d,u,u')}$. Two pairwise model comparisons,

$$\left\{M_{u_1}, M_{u_1'}\right\} = \{(x_1x_2, x_1x_3), (x_1x_2, x_4x_5)\},$$

$$\left\{M_{u_2}, M_{u_2'}\right\} = \{(x_1x_5, x_2x_4), (x_2x_4, x_3x_4)\}.$$

The elements of $H_{d_1}^{(u_1,u_1')}$ are negative of $H_{d_1}^{(u_2,u_2')}$ indicating equation (3.6.2) is the only possible equation to hold and must compare the eigenvector $\underline{a}_1^{(d_1,u_1,u_1')}$ with the eigenvector $\underline{b}_1^{(d_1,u_2,u_2')}$ as presented in table 3.3,

Table 3.3 Eigenvectors of length 12 for pairwise comparison 1 versus 2

1	0.075	-0.354	-0.234	0.468	-0.083	0.166	-0.075	0.112	0.597	-0.354	-0.234	-0.083
2	0.075	0.166	-0.075	0.468	-0.083	-0.354	-0.234	0.112	-0.083	-0.354	-0.234	0.597

Observe the first, fourth, fifth, eighth, tenth, and eleventh elements are the same there is only 720 possible permutations. The permutation that switches rows $2 \leftrightarrow 6$, $3 \leftrightarrow 7$, and $9 \leftrightarrow 12$ satisfy $P_m H_{d_1}^{(u_1,u_1)} P_m' = -H_{d_2}^{(u_2,u_2')}$ with the resulting matrix,

-0.042	-0.093	0.042	0.114	0.047	0.011	-0.057	-0.063	0.047	0.011	-0.057	0.042
-0.093	-0.207	0.093	0.253	0.103	0.023	-0.127	-0.140	0.103	0.023	-0.127	0.093
0.042	0.093	-0.042	-0.114	-0.047	-0.011	0.057	0.063	-0.047	-0.011	0.057	-0.042
0.114	0.253	-0.114	0.023	-0.127	-0.195	-0.012	0.171	-0.127	-0.195	-0.012	0.219
0.047	0.103	-0.047	-0.127	-0.052	-0.012	0.063	0.070	-0.052	-0.012	0.063	-0.047
0.011	0.023	-0.011	-0.195	-0.012	0.081	0.098	0.016	-0.012	0.081	0.098	-0.177
-0.057	-0.127	0.057	-0.012	0.063	0.098	0.006	-0.085	0.063	0.098	0.006	-0.110
-0.063	-0.140	0.063	0.171	0.070	0.016	-0.085	-0.094	0.070	0.016	-0.085	0.063
0.047	0.103	-0.047	-0.127	-0.052	-0.012	0.063	0.070	-0.052	-0.012	0.063	-0.047
0.011	0.023	-0.011	-0.195	-0.012	0.081	0.098	0.016	-0.012	0.081	0.098	-0.177
-0.057	-0.127	0.057	-0.012	0.063	0.098	0.006	-0.085	0.063	0.098	0.006	-0.110
0.042	0.093	-0.042	0.219	-0.047	-0.177	-0.110	0.063	-0.047	-0.177	-0.110	0.292

This method was used to understand that while set 4.1 and 4.2 had the same eigenvalues, it was not possible to construct a permutation between the two sets that satisfied (3.5.1) or (3.5.2). For this example, to fully understand the effect of the design and its pairwise model comparisons, it is sufficient to understand the eigenvalues, eigenvectors, and the random variables $S_{i,1}^{(d_1,u,u')}$ and $S_{i,2}^{(d_1,u,u')}$ for the 17 different sets. The elements of $\underline{a}_1, \underline{a}_1, \left(\underline{a_1} + \underline{b}_1\right)$, and $\left(\underline{a_1} + \underline{b}_1\right)$ can be interpreted as coefficients of the observation vector $\underline{y}_{d_1}$.

Table 3.4.1 Values for Set 1

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	0.075	-0.354	-0.234	0.468	-0.083	0.166	-0.075	0.112	0.596	-0.354	-0.234	-0.083
$\underline{b}_1$	0.238	-0.165	0.218	-0.437	-0.264	0.528	-0.238	0.357	-0.027	-0.165	0.218	-0.264
$\left(\underline{a}_1 + \underline{b}_1\right)$	0.312	-0.519	-0.016	0.031	-0.347	0.694	-0.312	0.468	0.569	-0.519	-0.016	-0.347
$\left(\underline{a}_1 - \underline{b}_1\right)$	-0.163	-0.19	-0.452	0.905	0.181	-0.363	0.163	-0.245	0.624	-0.19	-0.452	0.181

Table 3.4.2 Values for Set 2

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	-0.205	0.216	-0.573	0.227	0.227	0.005	0.205	-0.307	0.334	-0.243	0.346	-0.232
$\underline{b}_1$	0.205	-0.216	-0.005	-0.227	-0.227	0.573	-0.205	0.307	0.243	-0.334	0.232	-0.346
$\left(\underline{a}_1 + \underline{b}_1\right)$	0	0	-0.577	0	0	0.577	0	0	0.577	-0.577	0.577	-0.577
$\left(\underline{a}_1 - \underline{b}_1\right)$	-0.409	0.432	-0.568	0.455	0.455	-0.568	0.409	-0.614	0.091	0.091	0.114	0.114

Table 3.4.3 Values for Set 3

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	-0.176	-0.302	-0.302	0.176	0.301	0.196	0.196	-0.264	0.48	0.106	0.106	-0.516
$\underline{b}_1$	0.176	-0.106	-0.106	-0.176	0.516	-0.196	-0.196	0.264	-0.48	0.302	0.302	-0.301
$\left(\underline{a}_1 + \underline{b}_1\right)$	0	-0.408	-0.408	0	0.816	0	0	0	0	0.408	0.408	-0.816
$\left(\underline{a}_1 - \underline{b}_1\right)$	-0.352	-0.196	-0.196	0.352	-0.215	0.391	0.391	-0.528	0.959	-0.196	-0.196	-0.215

Table 3.4.4.1 Values for Set 4.1

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	-0.254	0.268	0.275	-0.134	-0.549	0.267	0.254	-0.38	0.26	-0.148	-0.141	0.282
$\underline{b}_1$	0.254	-0.268	0.134	-0.275	-0.267	0.549	-0.254	0.38	0.148	-0.26	0.141	-0.282
$\left(\underline{a}_1 + \underline{b}_1\right)$	0	0	0.408	-0.408	-0.816	0.816	0	0	0.408	-0.408	0	0
$\left(\underline{a}_1 - \underline{b}_1\right)$	-0.507	0.535	0.141	0.141	-0.282	-0.282	0.507	-0.761	0.113	0.113	-0.282	0.563

Table 3.4.4.2 Values for Set 4.2

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	-0.254	0.268	0.275	-0.549	-0.134	0.267	0.254	-0.38	-0.155	0.268	0.275	-0.134
$\underline{b}_1$	0.254	-0.268	0.134	-0.267	-0.275	0.549	-0.254	0.38	0.155	-0.268	0.134	-0.275
$(\underline{a}_1 + \underline{b}_1)$	0	0	0.408	-0.816	-0.408	0.816	0	0	0	0	0.408	-0.408
$(\underline{a}_1 - \underline{b}_1)$	-0.507	0.535	0.141	-0.282	0.141	-0.282	0.507	-0.761	-0.31	0.535	0.141	0.141

Table 3.4.5 Values for Set 5

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	-0.145	0.157	-0.558	0.161	0.161	0.145	0.153	-0.217	0.386	-0.317	0.39	-0.313
$\underline{b}_1$	-0.251	-0.137	0.257	0.278	0.278	0.251	-0.552	-0.376	-0.148	-0.142	0.267	0.274
$(\underline{a}_1 + \underline{b}_1)$	-0.395	0.02	-0.301	0.439	0.439	0.395	-0.4	-0.593	0.237	-0.459	0.657	-0.04
$(\underline{a}_1 - \underline{b}_1)$	0.106	0.294	-0.816	-0.118	-0.118	-0.106	0.705	0.159	0.534	-0.176	0.123	-0.587
$\underline{a}_2$	0.914	-0.042	-0.131	0.165	0.091	0.012	-0.308	0.002	0.049	0.003	0.109	-0.029
$\underline{b}_2$	0.285	0.094	0.365	-0.204	0.034	0.256	0.58	-0.446	-0.092	-0.295	0.083	0.176
$(\underline{a}_2 + \underline{b}_2)$	1.199	0.052	0.233	-0.039	0.125	0.268	0.272	-0.445	-0.043	-0.292	0.192	0.147
$(\underline{a}_2 - \underline{b}_2)$	0.629	-0.136	-0.496	0.369	0.057	-0.244	-0.888	0.448	0.142	0.297	0.025	-0.206

Table 3.4.6 Values for Set 6

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	0.1	-0.053	0.585	-0.328	-0.111	-0.033	-0.211	0.15	-0.522	0.313	-0.146	0.254
$\underline{b}_1$	-0.1	-0.254	0.211	0.146	0.111	0.522	-0.585	-0.15	0.033	-0.313	0.328	0.053
$(\underline{a}_1 + \underline{b}_1)$	0	-0.307	0.796	-0.182	0	0.489	-0.796	0	-0.489	0	0.182	0.307
$(\underline{a}_1 - \underline{b}_1)$	0.201	0.202	0.375	-0.475	-0.223	-0.554	0.375	0.301	-0.554	0.626	-0.475	0.202
$\underline{a}_2$	0.251	-0.441	0.111	0.253	-0.279	0.052	-0.072	0.377	0.359	-0.114	-0.543	0.048
$\underline{b}_2$	0.251	0.048	-0.072	-0.543	-0.279	0.359	0.111	0.377	0.052	-0.114	0.253	-0.441
$(\underline{a}_2 + \underline{b}_2)$	0.502	-0.393	0.039	-0.29	-0.558	0.41	0.039	0.753	0.41	-0.229	-0.29	-0.393
$(\underline{a}_2 - \underline{b}_2)$	0	-0.489	0.182	0.796	0	-0.307	-0.182	0	0.307	0	-0.796	0.489

Table 3.4.7 Values for Set 7

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	0.249	-0.277	-0.277	0.553	0.124	-0.249	-0.249	0.373	0.18	-0.277	-0.277	0.124
$\underline{b}_1$	0	-0.289	0.289	0	0	0.577	-0.577	0	0	-0.289	0.289	0
$\left(\underline{a}_1+\underline{b}_1\right)$	0.249	-0.565	0.012	0.553	0.124	0.328	-0.826	0.373	0.18	-0.565	0.012	0.124
$\left(\underline{a}_1-\underline{b}_1\right)$	0.249	0.012	-0.565	0.553	0.124	-0.826	0.328	0.373	0.18	0.012	-0.565	0.124
$\underline{a}_2$	0.083	0.15	0.15	-0.042	0.301	-0.342	-0.342	0.125	-0.685	0.15	0.15	0.301
$\underline{b}_2$	-0.435	0.019	0.019	0.457	0.277	-0.06	-0.06	-0.652	0.12	0.019	0.019	0.277
$\left(\underline{a}_2+\underline{b}_2\right)$	-0.352	0.169	0.169	0.415	0.578	-0.402	-0.402	-0.527	-0.565	0.169	0.169	0.578
$\left(\underline{a}_2-\underline{b}_2\right)$	0.518	0.132	0.132	-0.499	0.024	-0.283	-0.283	0.778	-0.805	0.132	0.132	0.024

Table 3.4.8 Values for Set 8

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	0.225	0.454	-0.311	-0.25	0.005	-0.118	-0.118	0.337	-0.373	-0.311	0.454	0.005
$\underline{b}_1$	0.309	-0.177	-0.028	-0.344	-0.393	0.239	0.239	0.464	0.289	-0.028	-0.177	-0.393
$\left(\underline{a}_1+\underline{b}_1\right)$	0.534	0.277	-0.339	-0.593	-0.388	0.121	0.121	0.801	-0.084	-0.339	0.277	-0.388
$\left(\underline{a}_1-\underline{b}_1\right)$	-0.085	0.631	-0.284	0.094	0.399	-0.357	-0.357	-0.127	-0.662	-0.284	0.631	0.399
$\underline{a}_2$	0	0.377	0.156	0	-0.533	-0.221	0.221	0	0	-0.156	-0.377	0.533
$\underline{b}_2$	0	0.156	-0.377	0	0.221	-0.533	0.533	0	0	0.377	-0.156	-0.221
$\left(\underline{a}_2+\underline{b}_2\right)$	0	0.533	-0.221	0	-0.312	-0.754	0.754	0	0	0.221	-0.533	0.312
$\left(\underline{a}_2-\underline{b}_2\right)$	0	0.221	0.533	0	-0.754	0.312	-0.312	0	0	-0.533	-0.221	0.754

Table 3.4.9 Values for Set 9

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	0	-0.289	0.577	0	-0.289	0	0	0	0	0.289	-0.577	0.289
$\underline{b}_1$	0.258	0.129	-0.258	-0.286	-0.286	-0.229	0.544	0.386	0.157	0.129	-0.258	-0.286
$\left(\underline{a}_1+\underline{b}_1\right)$	0.258	-0.16	0.32	-0.286	-0.575	-0.229	0.544	0.386	0.157	0.417	-0.835	0.002
$\left(\underline{a}_1-\underline{b}_1\right)$	-0.258	-0.417	0.835	0.286	-0.002	0.229	-0.544	-0.386	-0.157	0.16	-0.32	0.575
$\underline{a}_2$	0.326	0.031	0.192	-0.362	0.023	-0.025	-0.347	0.489	-0.571	0.031	0.192	0.023
$\underline{b}_2$	0.118	-0.306	0.091	-0.131	-0.287	0.626	-0.17	0.177	0.384	-0.306	0.091	-0.287
$\left(\underline{a}_2+\underline{b}_2\right)$	0.444	-0.276	0.283	-0.494	-0.264	0.601	-0.516	0.666	-0.187	-0.276	0.283	-0.264
$\left(\underline{a}_2-\underline{b}_2\right)$	0.208	0.337	0.1	-0.231	0.31	-0.651	-0.177	0.312	-0.955	0.337	0.1	0.31

Table 3.4.10 Values for Set 10

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	-0.249	0.263	-0.554	0.256	0.263	-0.132	0.277	-0.374	0.277	-0.139	0.27	-0.159
$\underline{b}_1$	0.249	-0.277	-0.263	0.159	0.554	-0.27	-0.277	0.374	-0.263	0.139	0.132	-0.256
$(\underline{a}_1 + \underline{b}_1)$	0	-0.013	-0.816	0.415	0.816	-0.402	0	0	0.013	0	0.402	-0.415
$(\underline{a}_1 - \underline{b}_1)$	-0.498	0.54	-0.291	0.097	-0.291	0.138	0.554	-0.747	0.54	-0.277	0.138	0.097
$\underline{a}_2$	-0.105	-0.475	0.126	-0.14	0.113	0.417	0.117	-0.157	0.341	0.303	0.003	-0.542
$\underline{b}_2$	-0.105	0.341	0.113	-0.542	0.126	0.003	0.117	-0.157	-0.475	0.303	0.417	-0.14
$(\underline{a}_2 + \underline{b}_2)$	-0.21	-0.134	0.239	-0.682	0.239	0.42	0.233	-0.315	-0.134	0.607	0.42	-0.682
$(\underline{a}_2 - \underline{b}_2)$	0	-0.816	0.013	0.402	-0.013	0.415	0	0	0.816	0	-0.415	-0.402

Table 3.4.11 Values for Set 11

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	-0.286	-0.256	-0.256	0.318	0.318	0.152	0.152	-0.429	0.48	0.072	0.072	-0.337
$\underline{b}_1$	0.286	-0.152	-0.152	-0.318	-0.318	0.256	0.256	0.429	0.337	-0.072	-0.072	-0.48
$(\underline{a}_1 + \underline{b}_1)$	0	-0.408	-0.408	0	0	0.408	0.408	0	0.816	0	0	-0.816
$(\underline{a}_1 - \underline{b}_1)$	-0.573	-0.103	-0.103	0.636	0.636	-0.103	-0.103	-0.859	0.143	0.143	0.143	0.143
$\underline{a}_2$	0	0.558	-0.558	0	0	-0.149	0.149	0	0	-0.408	0.408	0
$\underline{b}_2$	0	-0.149	0.149	0	0	0.558	-0.558	0	0	-0.408	0.408	0
$(\underline{a}_2 + \underline{b}_2)$	0	0.408	-0.408	0	0	0.408	-0.408	0	0	-0.816	0.816	0
$(\underline{a}_2 - \underline{b}_2)$	0	0.707	-0.707	0	0	-0.707	0.707	0	0	0	0	0

Table 3.4.12 Values for Set 12

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	-0.253	-0.127	-0.175	-0.12	0.281	0.253	0.267	-0.38	0.274	0.247	0.281	-0.548
$\underline{b}_1$	0.253	0.141	-0.26	-0.274	-0.281	-0.255	0.55	0.379	0.147	0.135	-0.268	-0.267
$(\underline{a}_1 + \underline{b}_1)$	0	0.014	-0.436	-0.394	0	-0.002	0.817	0	0.421	0.381	0.013	-0.815
$(\underline{a}_1 - \underline{b}_1)$	-0.506	-0.268	0.085	0.154	0.562	0.508	-0.283	-0.759	0.127	0.112	0.549	-0.282
$\underline{a}_2$	-0.061	0.362	-0.588	0.361	0.067	-0.18	0.068	-0.091	0.219	-0.483	0.219	0.105
$\underline{b}_2$	-0.234	0.167	0.031	0.168	0.26	-0.532	0.258	-0.351	-0.223	0.432	-0.222	0.248
$(\underline{a}_2 + \underline{b}_2)$	-0.295	0.528	-0.557	0.529	0.327	-0.712	0.326	-0.442	-0.004	-0.051	-0.004	0.354
$(\underline{a}_2 - \underline{b}_2)$	0.174	0.195	-0.618	0.194	-0.193	0.352	-0.19	0.26	0.443	-0.914	0.441	-0.143

Table 3.4.13 Values for Set 13

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	0.262	-0.283	-0.235	0.124	0.54	-0.262	-0.276	0.393	-0.283	0.159	0.125	-0.263
$\underline{b}_1$	-0.262	-0.124	0.262	0.283	0.276	0.235	-0.54	-0.393	-0.152	-0.132	0.256	0.29
$\left(\underline{a_1}+\underline{b}_1\right)$	0	-0.408	0.027	0.408	0.816	-0.027	-0.816	0	-0.435	0.027	0.381	0.027
$\left(\underline{a_1}-\underline{b}_1\right)$	0.523	-0.159	-0.497	-0.159	0.264	-0.497	0.264	0.785	-0.131	0.292	-0.131	-0.553
$\underline{a}_2$	0.191	-0.208	0.586	-0.227	-0.243	0.01	-0.204	0.286	-0.326	0.235	-0.346	0.246
$\underline{b}_2$	0.191	-0.227	0.01	-0.208	-0.204	0.586	-0.243	0.286	0.231	-0.342	0.25	-0.33
$\left(\underline{a_2}+\underline{b}_2\right)$	0.382	-0.436	0.596	-0.436	-0.447	0.596	-0.447	0.573	-0.095	-0.107	-0.095	-0.084
$\left(\underline{a_2}-\underline{b}_2\right)$	0	0.019	0.577	-0.019	-0.038	-0.577	0.038	0	-0.558	0.577	-0.596	0.577

Table 3.4.14 Values for Set 14

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	0	-0.289	-0.289	0	0.577	0	0	0	0	0.289	0.289	-0.577
$\underline{b}_1$	0	0.289	-0.289	0	0	-0.577	0.577	0	0	0.289	-0.289	0
$\left(\underline{a_1}+\underline{b}_1\right)$	0	0	-0.577	0	0.577	-0.577	0.577	0	0	0.577	0	-0.577
$\left(\underline{a_1}-\underline{b}_1\right)$	0	-0.577	0	0	0.577	0.577	-0.577	0	0	0	0.577	-0.577
$\underline{a}_2$	-0.359	0.09	0.09	-0.199	-0.199	0.378	0.378	-0.539	0.378	0.09	0.09	-0.199
$\underline{b}_2$	0.359	-0.09	-0.09	-0.378	-0.378	0.199	0.199	0.539	0.199	-0.09	-0.09	-0.378
$\left(\underline{a_2}+\underline{b}_2\right)$	0	0	0	-0.577	-0.577	0.577	0.577	0	0.577	0	0	-0.577
$\left(\underline{a_2}-\underline{b}_2\right)$	-0.718	0.18	0.18	0.18	0.18	0.18	0.18	-1.078	0.18	0.18	0.18	0.18

Table 3.4.15 Values for Set 15

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	0	0	0	-0.408	0.408	0	0	0	-0.408	0.408	0.408	-0.408
$\underline{b}_1$	0	0.289	-0.289	0	0	-0.577	0.577	0	0	0.289	-0.289	0
$\left(\underline{a_1}+\underline{b}_1\right)$	0	0.289	-0.289	-0.408	0.408	-0.577	0.577	0	-0.408	0.697	0.12	-0.408
$\left(\underline{a_1}-\underline{b}_1\right)$	0	-0.289	0.289	-0.408	0.408	0.577	-0.577	0	-0.408	0.12	0.697	-0.408
$\underline{a}_2$	0.29	-0.314	-0.314	0.402	0.402	-0.305	-0.305	0.435	-0.064	-0.072	-0.072	-0.081
$\underline{b}_2$	-0.355	0.095	0.095	0.374	0.374	-0.203	-0.203	-0.532	-0.21	0.089	0.089	0.388
$\left(\underline{a_2}+\underline{b}_2\right)$	-0.065	-0.218	-0.218	0.776	0.776	-0.509	-0.509	-0.098	-0.274	0.016	0.016	0.307
$\left(\underline{a_2}-\underline{b}_2\right)$	0.645	-0.409	-0.409	0.028	0.028	-0.102	-0.102	0.967	0.146	-0.161	-0.161	-0.468

Table 3.4.16 Values for Set 16

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
$\underline{a}_1$	0	0	-0.408	0	0	0.408	0	0	0.408	-0.408	0.408	-0.408
$\underline{b}_1$	0	0	0	0.408	-0.408	0	0	0	0.408	-0.408	-0.408	0.408
$\left(\underline{a}_1 + \underline{b}_1\right)$	0	0	-0.408	0.408	-0.408	0.408	0	0	0.816	-0.816	0	0
$\left(\underline{a}_1 - \underline{b}_1\right)$	0	0	-0.408	-0.408	0.408	0.408	0	0	0	0	0.816	-0.816
$\underline{a}_2$	0.286	-0.318	0.405	-0.302	-0.302	0.405	-0.318	0.429	-0.072	-0.072	-0.072	-0.072
$\underline{b}_2$	0.286	-0.318	-0.302	0.405	0.405	-0.302	-0.318	0.429	-0.072	-0.072	-0.072	-0.072
$\left(\underline{a}_2 + \underline{b}_2\right)$	0.573	-0.636	0.103	0.103	0.103	0.103	-0.636	0.859	-0.143	-0.143	-0.143	-0.143
$\left(\underline{a}_2 - \underline{b}_2\right)$	0	0	0.707	-0.707	-0.707	0.707	0	0	0	0	0	0

The information in these tables can be used to understand the weight of each observation in the observation vector $\underline{y}_{d_1}$. When considering the pairwise model comparison $\{M_u, M_{u'}\}$, the value of the differenced error sum of squares is dependent $\lambda_i^{(d_1,u,u')}$, $S_{i,1}^{(d_1,u,u')}$, and $S_{i,1}^{(d_1,u,u')}$. For the referenced model comparison in set 16 the value of $S_{i,1}^{(d_1,u,u')}$ will only depend on $y_3, y_4, y_5, y_6, y_{11}$, and y_{12}. Observe that y_{11} and y_{12} will have twice the weight on the value of $S_{i,1}^{(d_1,u,u')}$ than y_3 to y_6. Similarly, only y_3 to y_6 will affect the value of $S_{i,2}^{(d_1,u,u')}$. Overall, the differenced error sum of squares is not dependent on the observation values of y_1, y_2, and y_7 to y_{10}. Any other model comparison in set 16 are permutations of the above values, all leading to there are six observations where the differenced error sum of squares is not dependent on. The observation values are contrasts and can be used to understand what values of the observation vector results in a positive differenced error sum of squares or when model M_u will be chosen.

The expected values of $S_{i,1}^{(d,u,u')}$ and $S_{i,2}^{(d,u,u')}$ is calculated using (3.10.1) and (3.10.2)

$$E\left[S_{i,1}^{(d,u,u')}\right] = \left(\underline{a}_i + \underline{b}_i\right)' X_d \underline{\beta}, \;\; i = 1,2,\dots,r$$

$$E\left[S_{i,2}^{(d,u,u')}\right] = \left(\underline{a}_i - \underline{b}_i\right)' X_d \underline{\beta}, \;\; i = 1,2,\dots,r$$

The values of $\left(\underline{a}_i + \underline{b}_i\right)' X_d$ and $\left(\underline{a}_i - \underline{b}_i\right)' X_d$ can be interpreted as weights to the parameter coefficients $\underline{\beta}$. Since the values of $\left(\underline{a}_i + \underline{b}_i\right)' X_d$ and $\left(\underline{a}_i - \underline{b}_i\right)' X_d$ depend on the model comparison, there are 990 different tables of these weights, three examples of pairwise model comparisons are presented with their non-zero eigenvalues,

$$\left\{M_{u_1}, M_{u_1'}\right\} = \{(x_1x_2, x_1x_3), (x_1x_2, x_1x_4)\} \;\; ; \;\; NZEV(H_d^{(u_1,u_1')}) = \pm 0.9994,$$

$$\left\{M_{u_2}, M_{u_2'}\right\} = \{(x_1x_2, x_1x_3), (x_1x_2, x_2x_4)\} \;\; ; \;\; NZEV(H_d^{(u_2,u_2')}) = \pm 0.8758,$$

$$\left\{M_{u_3}, M_{u_3'}\right\} = \{(x_1x_3, x_1x_5), (x_1x_3, x_2x_5)\} \;\; ; \;\; NZEV(H_d^{(u_3,u_3')}) = \pm 0.9998.$$

To understand the effect of a true model, the table of weights for $\underline{\beta}$ are presented in table 3.5. While the differenced error sum of squares does not depend on the common parameters between model M_u and $M_{u'}$ (including general mean and main effects), table 3.5 displays those parameter weights for illustration purposes,

Table 3.5 Weights for the unknown parameters $\underline{\boldsymbol{\beta}}$

β_j	$\{M_{u_1}, M_{u_1'}\}$		$\{M_{u_2}, M_{u_2'}\}$		$\{M_{u_3}, M_{u_3'}\}$	
$j =$	$(a+b)'X_d$	$(a-b)'X_d$	$(a+b)'X_d$	$(a-b)'X_d$	$(a+b)'X_d$	$(a-b)'X_d$
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
1,2	0	0	0	0	-0.1757	0.0471
1,3	3.266	-3.1552	-3.9101	2.3094	0	0
1,4	-3.266	-3.1552	0.1819	0	1.7395	1.9249
1,5	0	0.2254	-1.1821	-2.3094	3.1793	-3.2425
2,3	-1.633	-1.4649	0.1819	0	1.5018	-1.5977
2,4	1.633	-1.4649	-3.9101	-2.3094	0.6321	-2.5602
2,5	0	-3.1552	-1.1821	2.3094	2.7849	2.8402
3,4	0	0.1127	0.0909	0	3.417	0.2801
3,5	-1.633	-1.5776	-1.2731	-2.3094	-0.1757	0.0471
4,5	1.633	-1.5776	-1.2731	2.3094	-1.0454	-0.9154

Table 3.5 demonstrates that $E\left[S_{i,1}^{(d,u,u')}\right]$ and $E\left[S_{i,2}^{(d,u,u')}\right]$ do not depend on the unknown parameters in common $\underline{\beta}^{(uu')}$. For the first comparison, the expected value of $S_{i,1}^{(d,u,u')}$ does not depend on the values of $\beta_{1,5}$, $\beta_{2,5}$, and $\beta_{3,4}$. If the true model consisted two non-zero two-factor interaction parameters $\beta_{1,2}$ and $\beta_{1,3}$, then the resulting consequence on the mean of the differenced sum of squares can be assessed.

Chapter 4

Model Comparisons – Balanced Fractional Factorial Designs

4.1 Introduction

The design d_1 presented in chapter 2 and chapter 3 considers $m = 5$ main effects, $n = 12$ runs, has full estimation capacity for $0 \leq k \leq 5$, two-factor interactions, and the capability to discriminate between all pairwise model comparisons for $k = 1$ or $k = 2$. The design d_1 is a balanced fractional design in main effects and two-factor interactions with the 12 runs consisting of the $\binom{5}{5} = 1$ treatment combination where the level of all factors are low, $\binom{5}{2} = 10$ treatment combinations where 2 of the 5 factors are low, and the $\binom{5}{0} = 1$ treatment combination where none of the factors are low.

Consider a new design d_2 with $m = 5$ main effects, $n = 12$ runs, has full estimation capacity for $0 \leq k \leq 3$, two-factor interactions, and can discriminate between all pairwise model comparisons for $k = 1$. d_2 is also a balanced fractional design in main effects and two-factor interactions consisting of the $\binom{5}{5} = 1$ treatment combination where all factors are low, $\binom{5}{4} = 5$ treatment combinations where 4 of the 5 factors are low, $\binom{5}{1} = 5$ treatment combinations where 1 of the 5 factors are low, and the $\binom{5}{0} = 1$ treatment combination where no factors are low. The value of $k = 1$ is chosen to compare the designs d_1 and d_2, the class of models consisting of $v_k = \binom{10}{1} = 10$ models with one two-factor interaction, and a total of $\binom{10}{2}$ pairwise model comparisons. The 12 run treatment combinations for both designs are presented for comparisons,

d_1

Run	Treatment	$\underline{x}_1$	$\underline{x}_2$	$\underline{x}_3$	$\underline{x}_4$	$\underline{x}_5$
1	(1)	-1	-1	-1	-1	-1
2	125	1	1	-1	-1	1
3	135	1	-1	1	-1	1
4	145	1	-1	-1	1	1
5	235	-1	1	1	-1	1
6	245	-1	1	-1	1	1
7	345	-1	-1	1	1	1
8	12345	1	1	1	1	1
9	123	1	1	1	-1	-1
10	124	1	1	-1	1	-1
11	134	1	-1	1	1	-1
12	234	-1	1	1	1	-1

d_2

Run	Treatment	$\underline{x}_1$	$\underline{x}_2$	$\underline{x}_3$	$\underline{x}_4$	$\underline{x}_5$
1	(1)	-1	-1	-1	-1	-1
2	1	1	-1	-1	-1	-1
3	2	-1	1	-1	-1	-1
4	3	-1	-1	1	-1	-1
5	4	-1	-1	-1	1	-1
6	5	-1	-1	-1	-1	1
7	2345	-1	1	1	1	1
8	1345	1	-1	1	1	1
9	1245	1	1	-1	1	1
10	1235	1	1	1	-1	1
11	1234	1	1	1	1	-1
12	12345	1	1	1	1	1

The six optimality criterion functions discussed in section 1.4.3 can be calculated for both designs d_1 and d_2 for the class of $v_k = 10$ models. The smaller the value for the criterion function, the better the performance for a design d,

Table 4.1 Calculated optimality criterion functions

Design\Criterion	AD	AT	AMCR	GD	GT	GMCR
d_1	3.77E-08	0.6375	0.140825	3.77E-08	0.6375	0.140825
d_2	6.81E-08	0.723214	0.125	6.81E-08	0.723214	0.125

4.2 Model Comparisons

For $k = 1$, there are $v_k = 10$ models M_u in the class of models and 45 pairwise model comparisons $\{M_u, M_{u'}\}$ considered. For each pairwise model comparison, the differenced projection matrix $H_d^{(u,u')}$ is calculated for each design $d = d_1, d_2$. The number of two-factor interactions in common between models M_u and $M_{u'}$ is always $0 \leq c_{uu'} \leq k - 1$ hence no model comparisons will have any two-factor interactions in common ($c_{uu'} = 0$). The rank of the differenced projection matrix for a design d that satisfies the model discrimination rank conditions is calculated to be $Rank\left(H_d^{(u,u')}\right) = 2(k - c_{uu'})$, observe for both designs d_1 and d_2, the value of $c_{uu'} = 0$ and $Rank\left(H_d^{(u,u')}\right) = 2$ for any model comparison $\{M_u, M_{u'}\}$ and design $d = d_1, d_2$. In this simple case, each model comparison will only have the non-zero eigenvalues of $\pm\lambda$. The positive eigenvalues for each design and comparison is presented,

Table 4.2 Positive eigenvalues for each comparison

#	M_u	$M_{u'}$	d_1	d_2
1	x1x2	x1x3	0.9994	0.9682
2	x1x2	x1x4	0.9994	0.9682
3	x1x2	x1x5	0.9994	0.9682
4	x1x2	x2x3	0.9994	0.9682
5	x1x2	x2x4	0.9994	0.9682
6	x1x2	x2x5	0.9994	0.9682
7	x1x2	x3x4	0.8754	0.866
8	x1x2	x3x5	0.8754	0.866
9	x1x2	x4x5	0.8754	0.866
10	x1x3	x1x4	0.9994	0.9682
11	x1x3	x1x5	0.9994	0.9682
12	x1x3	x2x3	0.9994	0.9682
13	x1x3	x2x4	0.8754	0.866
14	x1x3	x2x5	0.8754	0.866
15	x1x3	x3x4	0.9994	0.9682
16	x1x3	x3x5	0.9994	0.9682
17	x1x3	x4x5	0.8754	0.866
18	x1x4	x1x5	0.9994	0.9682
19	x1x4	x2x3	0.8754	0.866
20	x1x4	x2x4	0.9994	0.9682
21	x1x4	x2x5	0.8754	0.866
22	x1x4	x3x4	0.9994	0.9682
23	x1x4	x3x5	0.8754	0.866
24	x1x4	x4x5	0.9994	0.9682
25	x1x5	x2x3	0.8754	0.866
26	x1x5	x2x4	0.8754	0.866
27	x1x5	x2x5	0.9994	0.9682
28	x1x5	x3x4	0.8754	0.866
29	x1x5	x3x5	0.9994	0.9682
30	x1x5	x4x5	0.9994	0.9682
31	x2x3	x2x4	0.9994	0.9682
32	x2x3	x2x5	0.9994	0.9682
33	x2x3	x3x4	0.9994	0.9682
34	x2x3	x3x5	0.9994	0.9682
35	x2x3	x4x5	0.8754	0.866
36	x2x4	x2x5	0.9994	0.9682
37	x2x4	x3x4	0.9994	0.9682
38	x2x4	x3x5	0.8754	0.866
39	x2x4	x4x5	0.9994	0.9682
40	x2x5	x3x4	0.8754	0.866
41	x2x5	x3x5	0.9994	0.9682
42	x2x5	x4x5	0.9994	0.9682
43	x3x4	x3x5	0.9994	0.9682
44	x3x4	x4x5	0.9994	0.9682
45	x3x5	x4x5	0.9994	0.9682

Let the positive eigenvalue for design d and pairwise model comparison $\{M_u, M_{u'}\}$ be $\lambda^{(d,u,u')}$, observe that $\lambda^{(d_1,u,u')} > \lambda^{(d_2,u,u')}$ for every model comparison $\{M_u, M_{u'}\}$. When $\lambda^{(d_1,u,u')} = 0.8754$ it is the case that $\lambda^{(d_2,u,u')} = 0.866$ and when $\lambda^{(d_1,u,u')} = 0.9994$ it is the case that $\lambda^{(d_2,u,u')} = 0.9682$. For each design, the 45 pairwise model comparisons can be classified into one of two sets. Descriptive statistics for the positive eigenvalues of the 45 model comparisons for each design is presented,

Table 4.3 Descriptive statistics for 45 positive eigenvalues

Design	Mean	Median	Min	Max
d_1	0.9581	0.9994	0.8754	0.9994
d_2	0.9341	0.9682	0.866	0.9682

For the 45 pairwise model comparisons, d_1 has larger mean, median, minimum, and maximum for the positive eigenvalues than d_2. Since $Rank\left(H_d^{(u,u')}\right) = 2$ for all model comparisons $\{M_u, M_{u'}\}$ and $d = d_1, d_2$, the differenced error sum of squares can be expressed as,

$$SSSE_u(d) - SSE_{u'}(d) = \lambda^{(d,u,u')} S_1^{(d,u,u')} S_2^{(d,u,u')},$$

$$S_1^{(d,u,u')} = (\underline{a} + \underline{b})' \underline{y}_d,$$

$$S_2^{(d,u,u')} = (\underline{a} - \underline{b})' \underline{y}_d.$$

The 45 pairwise model comparisons $\{M_u, M_{u'}\}$ can be classified into two sets unique to their eigenvalues,

Table 4.4 Classification of the 45 pairwise model comparisons

Set #	Number of Comparisons	Eigenvalues $(\boldsymbol{d_1})$	Eigenvalues $(\boldsymbol{d_2})$
1	30	± 0.8754	± 0.866
2	15	± 0.9994	± 0.9682

Any of the 45 model comparisons can be analyzed by permuting these reference comparisons, one from each set,

$$\left\{M_{u_1}, M_{u_1'}\right\} = \{x_1x_2, x_1x_3\} \;;\; NZEV\left(H_{d_1}^{(u_1,u_1')}\right) = \pm 0.8754 \;;\; NZEV\left(H_{d_2}^{(u_1,u_1')}\right) = \pm 0.866,$$

$$\left\{M_{u_2}, M_{u_2'}\right\} = \{x_1x_2, x_3x_4\} \;;\; NZEV\left(H_{d_1}^{(u_2,u_2')}\right) = \pm 0.9994 \;;\; NZEV\left(H_{d_2}^{(u_2,u_2')}\right) = \pm 0.9682.$$

The values for $\underline{a}$, $\underline{b}$, $(\underline{a}+\underline{b})$, $(\underline{a}-\underline{b})$ are provided for each set, any comparison within the set are permutations of these values,

Table 4.5.1 Observation weights for set 1

Set #1		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
d_1	$\underline{a}$	0.245	-0.265	0.143	-0.272	-0.272	0.558	-0.258	0.368	0.15	-0.265	0.143	-0.272
	$\underline{b}$	-0.245	-0.143	0.265	0.272	0.272	0.258	-0.558	-0.368	-0.15	-0.143	0.265	0.272
	$\underline{a}+\underline{b}$	0	-0.408	0.408	0	0	0.816	-0.816	0	0	-0.408	0.408	0
	$\underline{a}-\underline{b}$	0.49	-0.123	-0.123	-0.545	-0.545	0.3	0.3	0.735	0.3	-0.123	-0.123	-0.545
d_2	$\underline{a}$	-0.183	0.365	-0.262	0.445	-0.183	-0.183	0.365	-0.262	0.445	-0.183	-0.183	-0.183
	$\underline{b}$	-0.183	0.365	0.445	-0.262	-0.183	-0.183	0.365	0.445	-0.262	-0.183	-0.183	-0.183
	$\underline{a}+\underline{b}$	-0.365	0.73	0.183	0.183	-0.365	-0.365	0.73	0.183	0.183	-0.365	-0.365	-0.365
	$\underline{a}-\underline{b}$	0	0	-0.707	0.707	0	0	0	-0.707	0.707	0	0	0

Table 4.5.2 Observation weights for set 2

Set #2		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
d_1	$\underline{a}$	0.198	0.58	-0.22	-0.22	-0.22	-0.22	0.003	0.297	-0.338	0.198	0.58	-0.22
	$\underline{b}$	0.198	0.003	-0.22	-0.22	-0.22	-0.22	0.58	0.297	0.239	0.198	0.003	-0.22
	$\underline{a}+\underline{b}$	0.396	0.583	-0.44	-0.44	-0.44	-0.44	0.583	0.594	-0.099	0.396	0.583	-0.44
	$\underline{a}-\underline{b}$	0	0.577	0	0	0	0	-0.577	0	-0.577	0	0.577	0
d_2	$\underline{a}$	-0.289	-0.106	-0.106	0.394	0.394	-0.289	-0.106	-0.106	0.394	-0.289	-0.106	-0.106
	$\underline{b}$	0.289	-0.394	-0.394	0.106	0.106	0.289	-0.394	-0.394	0.106	0.289	-0.394	-0.394
	$\underline{a}+\underline{b}$	0	-0.5	-0.5	0.5	0.5	0	-0.5	-0.5	0.5	0	-0.5	-0.5
	$\underline{a}-\underline{b}$	-0.577	0.289	0.289	0.289	0.289	-0.577	0.289	0.289	0.289	-0.577	0.289	0.289

Table 4.5.1 and 4.5.2 illustrate the weights of the observations y_i have on the differenced error sum of squares. For the first set of comparisons, there are six observations which will not influence the distribution of S_1 for d_1 and there are eight observations that will not influence the distribution of S_2 for d_2. For d_1, the value of y_6 and y_7 will have the largest weights on the differenced error sum of squares for comparison 1 than the other observations. For d_2, the observations y_3, y_4, y_8, and y_9 have the highest weights. Similar observations can be made for comparisons in set #2. Comparisons within these sets will be similar but permuted weights. Using (3.10.1) and (3.10.2), observe the effect of the two-factor interaction parameters on the expected values of $S_1^{(d,u,u')}$ and $S_2^{(d,u,u')}$

Table 4.6.1 Parameter weights for set 1

	Set #1	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{2,5}$	$\beta_{3,4}$	$\beta_{3,5}$	$\beta_{4,5}$
d_1	$(\underline{a}+\underline{b})'X_d$	-3.266	3.266	0	0	0	1.633	1.633	-1.633	-1.633	0
	$(\underline{a}-\underline{b})'X_d$	3.159	3.159	-0.218	-0.218	-0.218	1.471	1.471	1.471	1.471	3.159
d_2	$(\underline{a}+\underline{b})'X_d$	-3.651	-3.651	-1.461	-1.461	-1.461	0.73	0.73	0.73	0.73	2.921
	$(\underline{a}-\underline{b})'X_d$	2.828	-2.828	0	0	0	2.828	2.828	-2.828	-2.828	0

Table 4.6.2 Parameter weights for set 2

	Set #2	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{2,5}$	$\beta_{3,4}$	$\beta_{3,5}$	$\beta_{4,5}$
d_1	$(\underline{a}+\underline{b})'X_d$	3.913	-0.176	-0.176	1.187	-0.176	-0.176	1.187	3.913	1.187	1.187
	$(\underline{a}-\underline{b})'X_d$	-2.309	0	0	2.309	0	0	2.309	2.309	-2.309	-2.309
d_2	$(\underline{a}+\underline{b})'X_d$	4	0	0	2	0	0	2	-4	-2	-2
	$(\underline{a}-\underline{b})'X_d$	-2.309	-2.309	-2.309	1.155	-2.309	-2.309	1.155	-2.309	1.155	1.155

To interpret tables 4.6.1or 4.6.2, consider the value of $SSE_{u_1}(d) - SSE_{u_1'}(d)$ for design $d = d_1, d_2$. Recall from section 3.5, $S_1^{(d,u,u')}$ and $S_2^{(d,u,u')}$ are uncorrelated random variables since,

$$Cov\left(S_1^{(d,u,u')}, S_2^{(d,u,u')}\right) = Cov\left(\left(\underline{a} + \underline{b}\right)' \underline{y}_d, \left(\underline{a} - \underline{b}\right)' \underline{y}_d\right) = \left(\underline{a}' \underline{a} - \underline{b}' \underline{b}\right) * Cov\left(\underline{y}_d' \underline{y}_d\right) = 0$$

The expected value of the differenced error sum of squares is calculated using (3.9),

$$E[SSE_u(d) - SSE_{u'}(d)] = \lambda^{(d,u,u')} E\left[S_1^{(d,u,u')}\right] E\left[S_2^{(d,u,u')}\right],$$

When comparing between the two-factor interactions $x_1 x_2$ and $x_1 x_3$, the expected value does not depend on the parameters $\beta_{1,4}$, $\beta_{1,5}$, $\beta_{2,3}$, or $\beta_{4,5}$. If true model has $\beta_{1,2} \neq 0$ while all other two-factor interactions are zero, observe that the expected value for the differenced error sum of squares for d_1 is calculated to be $0.9994(-3.266)(3.159)\beta_{1,2}^2 = -10.3111\beta_{1,2}^2$ and d_2 to be $0.9682(-3.651)(2.828)\beta_{1,2}^2 = -10\beta_{1,2}^2$. The expected value of both designs are negative indicating the correct model is more likely to be selected with respect to this comparison. If the true model had $\beta_{1,2} \neq 0$, the design d_1 has a higher expected value (in magnitude) and is more likely to pick the correct model than d_2 with this comparison indicating in this case, d_1 would have higher discrimination ability.

4.3 Simulated Data

The discussion at the end of section 4.2 demonstrates if the true model only had non-zero parameter of $\beta_{1,2} \neq 0$, design d_1 would have a larger differenced error sum of squares (in magnitude) on average when comparing $\{M_{u_1}, M_{u_1'}\} = \{x_1x_2, x_1x_3\}$. It is understood the performance of a design depends on the true model. To compare design d_1 and d_2 data will be simulated from the true model,

$$\underline{y} = 5\left(\underline{1} + \sum_{j=1}^{5} \underline{x}_j\right) + 5\underline{x_1x_2} + \underline{\epsilon}\,;\ \underline{\epsilon} \sim Normal\left(\underline{0}, \sigma^2 I\right),$$

100,000 datasets were generated using the true model for different values of σ. Consider the example comparisons in section 4.2, $\{M_{u_1}, M_{u_1'}\} = \{x_1x_2, x_1x_3\}$ and $\left\{M_{u_2}, M_{u_2'}\right\} = \{x_1x_2, x_3x_4\}$, using the simulated datasets we can plot the distributions of $S_1^{(d,u,u')}$ and $S_2^{(d,u,u')}$.

Figure 4.1.1 Histograms for $\boldsymbol{\sigma = 0.5}$

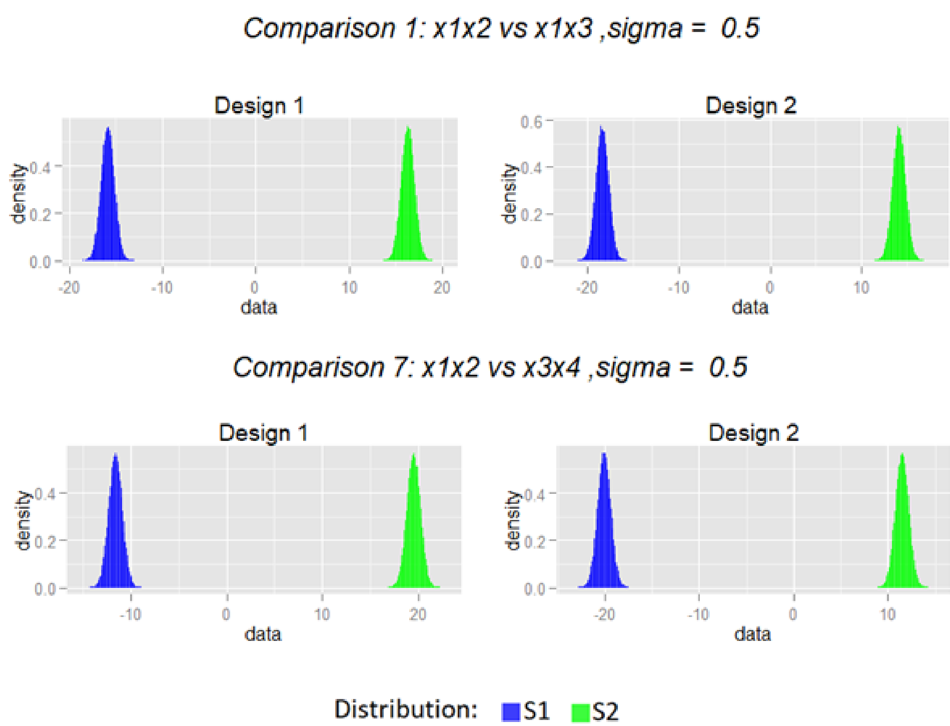


Figure 4.1.2 Histograms for $\sigma = 5$

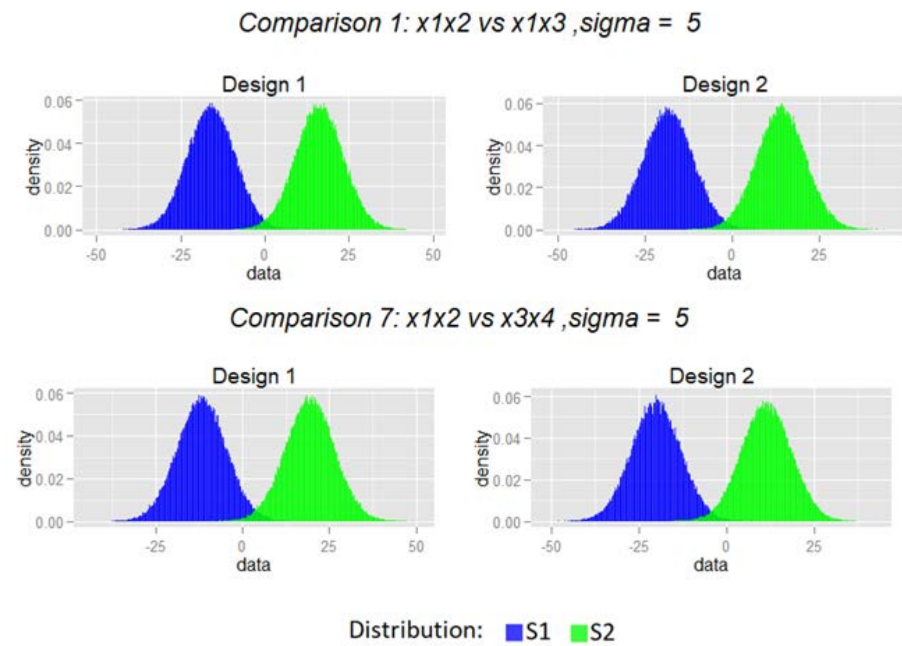
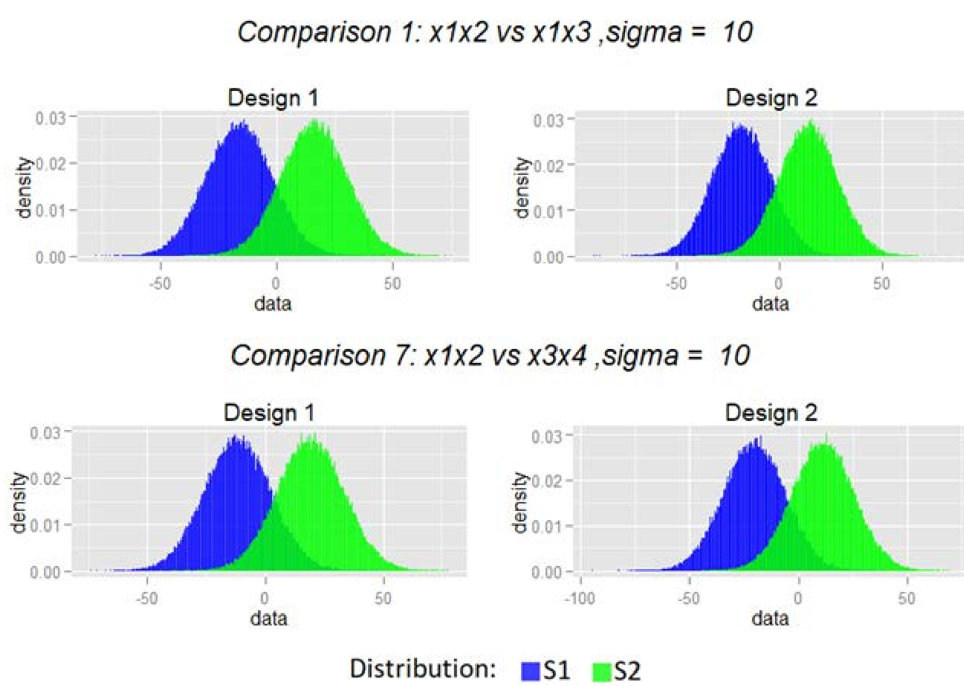


Figure 4.1.3 Histograms for $\sigma = 10$



To interpret these figures, the decision rule to choose model M_u or $M_{u'}$ using the realized values of $S_1^{(d,u,u')}$ and $S_2^{(d,u,u')}$ can be summarized by table 4.7,

Table 4.7 Summary of decision rules

S_1	S_2	$SSE_u(d_i) - SSE_{u'}(d_i)$	Model Selected
$+$	$+$	$+$	$M_{u'}$
$+$	$-$	$-$	M_u
$-$	$+$	$-$	M_u
$-$	$-$	$+$	$M_{u'}$

Observe for both designs and smaller values of σ, the distributions of $S_1^{(d,u,u')}$ and $S_2^{(d,u,u')}$ will almost always generate values which are opposite of each other hence the correct model M_u will be selected. As the level of noise increases, the distributions of $S_1^{(d,u,u')}$ and $S_2^{(d,u,u')}$ start overlapping and the realized values of $S_1^{(d,u,u')}$ and $S_2^{(d,u,u')}$ are more likely to give the same signs and the incorrect model $M_{u'}$ will be selected. By choosing σ^2 too large, the distributions of $S_1^{(d,u,u')}$ and $S_2^{(d,u,u')}$ start to look identical with complete overlapping and the correct model will be chosen about 50% of the time.

There are 45 different comparisons between M_u and $M_{u'}$, the average value of the differenced error sum of square are calculated for the 100,000 different datasets. When $\beta_{1,2} = 5$ and the remaining two-factor interactions are 0, the expected value for the model comparison $\{M_{u_1}, M_{u_1'}\} = \{x_1x_2, x_1x_3\}$ is $-10.3111\beta_{1,2}^2 = -257.78$ for d_1 and $-10\beta_{1,2}^2 = 250$ for d_2, this is observed by the simulated datasets,

Table 4.8.1 Average differenced SSE values for d_1 ($\sigma = 0.05$)

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	-257.78	-257.78	-257.78	-257.78	-257.78	-257.78	-197.78	-197.78	-197.78
x_1x_3		0.00	0.00	0.00	0.00	0.00	60.00	60.00	60.00
x_1x_4			0.00	0.00	0.00	0.00	60.00	60.00	60.00
x_1x_5				0.00	0.00	0.00	60.00	60.00	60.00
x_2x_3					0.00	0.00	60.00	60.00	60.00
x_2x_4						0.00	60.00	60.00	60.00
x_2x_5							60.00	60.00	60.00
x_3x_4								0.00	0.00
x_3x_5									0.00

Table 4.8.2 Average differenced SSE Values for d_2 ($\sigma = 0.05$)

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	-250.00	-250.00	-250.00	-250.00	-250.00	-250.00	-200.00	-200.00	-200.00
x_1x_3		0.00	0.00	0.00	0.00	0.00	50.00	50.00	50.00
x_1x_4			0.00	-0.01	0.00	0.00	50.00	50.00	49.99
x_1x_5				-0.01	0.00	0.00	50.00	50.00	49.99
x_2x_3					0.00	0.00	50.01	50.00	50.00
x_2x_4						0.00	50.00	50.00	50.00
x_2x_5							50.00	50.00	50.00
x_3x_4								0.00	-0.01
x_3x_5									0.00

Table 4.8.1 and 4.8.2 demonstrate with this true model, there are four different expected values of the differenced error sum of squares. If model M_u contains the correct term x_1x_2, the average differenced error sum of square is either -257.78 or -197.78 for d_1 or either -250 or -200 depending on which set the comparison was classified under. Otherwise the average differenced error sum of square is either 60 or 0 for d_1 or either 50 or 0 depending on which set the comparison was classified under (lower values for lower eigenvalues). Tables 4.9.1.1 presents the proportion of times model M_u (rows) is chosen over model $M_{u'}$ (columns) of the 100,000 simulated datasets for $\sigma^2 = 0.05$ and design d_1,

Table 4.9.1.1 Proportion of times M_u is chosen for d_1 ($\sigma = 0.05$)

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	1	1	1	1	1	1	1	1	1
x_1x_3		0.50028	0.50038	0.50151	0.50035	0.5016	0	0	0
x_1x_4			0.49937	0.50099	0.49815	0.50003	0	0	0
x_1x_5				0.50125	0.49943	0.50161	0	0	0
x_2x_3					0.49945	0.49852	0	0	0
x_2x_4						0.49935	0	0	0
x_2x_5							0	0	0
x_3x_4								0.50017	0.49955
x_3x_5									0.49985

In the presence of low noise, d_1 had a 100% success rate at choosing the correct two-factor interaction over the other 9 incorrect two-factor interactions. The model comparisons that have the expected differenced error sum of squares as 0 had 50% probability of picking model M_u or $M_{u'}$. The positive expected differenced error sum of squares had a 0% probability of picking model M_u over $M_{u'}$ for this true model. Similar results are seen with d_2

Table 4.9.1.2 Proportion of times M_u is chosen for d_2 ($\sigma = 0.05$)

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	1	1	1	1	1	1	1	1	1
x_1x_3		0.49658	0.49678	0.50285	0.49978	0.50086	0	0	0
x_1x_4			0.49717	0.50377	0.50285	0.50204	0	0	0
x_1x_5				0.50437	0.50315	0.50285	0	0	0
x_2x_3					0.49658	0.49678	0	0	0
x_2x_4						0.49717	0	0	0
x_2x_5							0	0	0
x_3x_4								0.50283	0.50322
x_3x_5									0.50342

Table 4.9.2.1 and 4.9.2.2 presents the proportion of times model M_u is chosen over $M_{u'}$ when the noise increases from $\sigma = 0.05$ to $\sigma = 5$,

Table 4.9.2.1 Proportion of times M_u is chosen for d_1 ($\sigma = 5$)

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	0.9771	0.97747	0.97752	0.97712	0.97695	0.97613	0.94539	0.94485	0.94713
x_1x_3		0.49908	0.50067	0.50335	0.50067	0.49912	0.23639	0.2382	0.23315
x_1x_4			0.50046	0.50193	0.50271	0.49996	0.23612	0.23129	0.23731
x_1x_5				0.50096	0.50039	0.50243	0.23075	0.23702	0.23618
x_2x_3					0.49952	0.49737	0.23431	0.23656	0.2315
x_2x_4						0.49821	0.23658	0.23185	0.23696
x_2x_5							0.2322	0.23611	0.23773
x_3x_4								0.50308	0.50221
x_3x_5									0.49961

Table 4.9.2.2 Proportion of times M_u is chosen for d_2 ($\sigma = 5$)

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	0.97236	0.97278	0.97337	0.9725	0.97232	0.97281	0.94793	0.94662	0.9463
x_1x_3		0.49959	0.49843	0.49826	0.49733	0.49918	0.31065	0.30974	0.2529
x_1x_4			0.4997	0.50095	0.49719	0.50159	0.31007	0.25167	0.31232
x_1x_5				0.50127	0.49917	0.49995	0.25224	0.31154	0.31307
x_2x_3					0.49935	0.49961	0.31164	0.31134	0.25395
x_2x_4						0.50199	0.31256	0.2532	0.31352
x_2x_5							0.254	0.31231	0.31293
x_3x_4								0.50092	0.50131
x_3x_5									0.50029

When the noise is increased to $\sigma = 5$, the model comparisons that have 0 expected differenced error sum of squares still pick model M_u 50% of the time. The positive differenced error sum of squares no longer only pick model $M_{u'}$ over M_u and is moving towards 50% of the time picking model M_u over $M_{u'}$. The important information is that when the expected difference error sum of squares was larger, we star to see a higher proportion of times model M_u is picked over model

$M_{u'}$. d_1 is performing slightly better than d_2 for $M_{u'} = \{x_1x_3, \dots, x_2x_5\}$ and slightly worse than d_2 when $M_{u'} = \{x_3x_4, x_3x_5, x_4x_5\}$ as anticipated by their differenced expected error sum of square values. Tables 4.9.3.1 and 4.9.3.2 presents the results when the noise is increased from $\sigma = 5$ to $\sigma = 10$,

Table 4.9.3.1 Proportion of times M_u is chosen for d_1 $(\sigma = 10)$

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	0.77439	0.77398	0.77497	0.77532	0.77435	0.7758	0.74165	0.74263	0.74409
x_1x_3		0.49926	0.49779	0.50096	0.50045	0.49948	0.41317	0.41204	0.40418
x_1x_4			0.50106	0.50234	0.50182	0.50333	0.41344	0.40414	0.41557
x_1x_5				0.50012	0.50018	0.50237	0.40594	0.41526	0.41469
x_2x_3					0.50085	0.50054	0.41166	0.41117	0.40582
x_2x_4						0.49996	0.41299	0.40525	0.41477
x_2x_5							0.40442	0.41316	0.41353
x_3x_4								0.49984	0.50123
x_3x_5									0.50084

Table 4.9.3.2 Proportion of times M_u is chosen for d_2 $(\sigma = 10)$

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	0.7755	0.77412	0.77619	0.77531	0.77378	0.77369	0.74634	0.74752	0.74931
x_1x_3		0.4995	0.50166	0.50084	0.50135	0.49952	0.42981	0.43121	0.41862
x_1x_4			0.5001	0.50122	0.5029	0.50086	0.42939	0.4194	0.43065
x_1x_5				0.49918	0.49893	0.50056	0.41619	0.4292	0.43084
x_2x_3					0.50132	0.49936	0.42895	0.43002	0.41657
x_2x_4						0.49986	0.42831	0.41685	0.43131
x_2x_5							0.4182	0.42937	0.43205
x_3x_4								0.50165	0.50226
x_3x_5									0.49816

Observe that when the value of σ gets larger, the proportion of times M_u is picked over $M_{u'}$ approaches 50%. In conclusion, both d_1 and d_2 have similar discrimination abilities when $\beta_{1,2} = 5$ and $\beta_{1,3} = \cdots = \beta_{4,5} = 0$ with slight differences depending on the value of $\beta_{1,2}$.

Chapter 5

Model Comparisons – Unbalanced Fractional Factorial Designs

5.1 Introduction

Chapter 4 presented the model comparisons for $k = 1$ with the two designs d_1 and d_2 where d_1 and d_2 had a balanced covariance structure for all $v_k = \binom{\binom{5}{2}}{1} = 10$ models. Section 4.2 demonstrated that the balanced design in main effects and two-factor interactions had similar eigenvalue structures which simplified the comparisons of both designs. The procedure is identical to compare unbalanced designs but will not have the structure as presented in chapter 4. Two new designs are presented in chapter 5, d_3 and d_4 which are both Plackett-Burman designs (1946). These designs are orthogonal and balanced with the main effects, but not necessarily balanced for the two-factor interaction terms. Both designs have $m = 5$ main effects, $n = 12$ runs, and has full estimation capacity for $0 \leq k \leq 3$ two-factor interactions, and can discriminate between all pairwise model comparisons for $k = 1$. Presented are designs d_3 and d_4,

d_3

Run	Treatment	$\underline{x}_1$	$\underline{x}_2$	$\underline{x}_3$	$\underline{x}_4$	$\underline{x}_5$
1	1245	1	1	-1	1	1
2	23	-1	1	1	-1	-1
3	134	1	-1	1	1	-1
4	24	-1	1	-1	1	-1
5	35	-1	-1	1	-1	1
6	45	-1	-1	-1	1	1
7	15	1	-1	-1	-1	1
8	125	1	1	-1	-1	1
9	123	1	1	1	-1	-1
10	2345	-1	1	1	1	1
11	345	-1	1	1	1	-1
12	(1)	-1	-1	-1	-1	-1

d_4

Run
1
2
3
4
5
6
7
8
9
10
11
12

Treatment
1245
235
134
245
35
4
15
12
123
234
1345
(1)

$\underline{x}_1$	$\underline{x}_2$	$\underline{x}_3$	$\underline{x}_4$	$\underline{x}_5$
1	1	-1	1	1
-1	1	1	-1	1
1	-1	1	1	-1
-1	1	-1	1	1
-1	-1	1	-1	1
-1	-1	-1	1	-1
1	-1	-1	-1	1
1	1	-1	-1	-1
1	1	1	-1	-1
-1	1	1	1	-1
1	-1	1	1	1
-1	-1	-1	-1	-1

The six optimality criterion functions discussed in chapter 1 can be calculated for both designs d_1 and d_2 and the class of $v_k = 10$ models.

Table 5.1 Calculated optimality criterion functions

Design\Criterion	AD	AT	AMCR	GD	GT	GMCR
d_3	4.19E-08	0.6666667	0.1971688	4.19E-08	0.6666667	0.1971688
d_4	4.19E-08	0.6666667	0.1971688	4.19E-08	0.6666667	0.1971688

Both design d_3 and design d_4 have the same optimality criterion values unlike the design comparisons of design d_1 and design d_2 in chapter 4 where d_1 had stronger performance for the determinant and trace criterion functions.

5.2 Model Comparisons

For all $\binom{10}{1} = 45$ model comparisons, the differenced error sum of squares is calculated using the differenced projection matrix $H_d^{(u,u')}$. Table 5.2 presents the eigenvalues of $H_d^{(u,u')}$ for both designs and all 45 model comparisons,

Table 5.2 Positive eigenvalues for each comparison

#	M_u	$M_{u'}$	d_3	d_4
1	x1x2	x1x3	0.9428	1
2	x1x2	x1x4	0.9428	0.9428
3	x1x2	x1x5	0.9428	1
4	x1x2	x2x3	0.9428	0.9428
5	x1x2	x2x4	0.9428	1
6	x1x2	x2x5	0.9428	1
7	x1x2	x3x4	0.9428	0.7454
8	x1x2	x3x5	0.9428	0.9428
9	x1x2	x4x5	0.9428	0.9428
10	x1x3	x1x4	0.9428	1
11	x1x3	x1x5	0.9428	0.9428
12	x1x3	x2x3	0.9428	1
13	x1x3	x2x4	0.9428	0.9428
14	x1x3	x2x5	0.9428	0.9428
15	x1x3	x3x4	0.9428	0.9428
16	x1x3	x3x5	0.9428	1
17	x1x3	x4x5	0.9428	0.7454
18	x1x4	x1x5	0.9428	1
19	x1x4	x2x3	0.9428	0.9428
20	x1x4	x2x4	0.9428	1
21	x1x4	x2x5	0.9428	0.7454
22	x1x4	x3x4	0.9428	1
23	x1x4	x3x5	0.9428	0.9428
24	x1x4	x4x5	0.9428	0.9428
25	x1x5	x2x3	0.9428	0.7454
26	x1x5	x2x4	0.9428	0.9428
27	x1x5	x2x5	0.9428	0.9428
28	x1x5	x3x4	0.9428	0.9428
29	x1x5	x3x5	0.9428	1
30	x1x5	x4x5	0.9428	1
31	x2x3	x2x4	0.9428	1
32	x2x3	x2x5	0.9428	1
33	x2x3	x3x4	0.9428	1
34	x2x3	x3x5	0.9428	0.9428
35	x2x3	x4x5	0.9428	0.9428
36	x2x4	x2x5	0.9428	0.9428
37	x2x4	x3x4	0.9428	0.9428
38	x2x4	x3x5	0.9428	0.7454
39	x2x4	x4x5	0.9428	1
40	x2x5	x3x4	0.9428	0.9428
41	x2x5	x3x5	0.9428	1
42	x2x5	x4x5	0.9428	1
43	x3x4	x3x5	0.9428	1
44	x3x4	x4x5	0.9428	1
45	x3x5	x4x5	0.9428	0.9428

For design d_3, every model comparison have non-zero eigenvalue equal to ± 0.9428. Consider design d_3, take any two model comparisons $\{M_{u_1}, M_{u_1'}\}$ and $\{M_{u_2}, M_{u_2'}\}$, there is a permutation matrix P_m such that $H_{d_3}^{(u_1,u_1')} = P_m H_{d_3}^{(u_2,u_2')} P_m'$. Also observe the eigenvalues for d_4, 5 comparisons have non-zero eigenvalues ± 0.7454 and are one set, 20 comparisons have non-zero eigenvalues ± 0.9428 and another set, and the remaining 20 comparisons have non-zero eigenvalues as ± 1 forming three sets depending on their eigenvalues. Both d_3 and d_4 have comparisons with eigenvalues ± 0.9428. Table 5.3 presents the descriptive statistics for the eigenvalues of designs d_3 and d_4,

Table 5.3 Descriptive statistics for 45 positive eigenvalues

Design	Mean	Median	Min	Max
d_3	0.9428	0.9428	0.9428	0.9428
d_4	0.9462	0.9428	0.7453	1

For the 45 pairwise model comparisons, d_3 and d_4 have approximately the same mean and have equivalent medians. The minimum value for d_3 is greater than d_4 but d_4 has the largest positive eigenvalue of the 45 model comparisons. Since $Rank\left(H_d^{(u,u')}\right) = 2$ for all model comparisons $\{M_u, M_{u'}\}$ and $d = d_3, d_4$, the differenced error sum of squares can be expressed as,

$$SSSE_u(d) - SSE_{u'}(d) = \lambda^{(d,u,u')} S_1^{(d,u,u')} S_2^{(d,u,u')},$$

$$S_1^{(d,u,u')} = \left(\underline{a} + \underline{b}\right)' \underline{y}_d,$$

$$S_2^{(d,u,u')} = \left(\underline{a} - \underline{b}\right)' \underline{y}_d.$$

In chapter 4, the two balanced designs d_1 and d_2 classified the 45 model comparisons into two equivalent sets. Design d_3 will have one set and design d_4 will have three sets based on their eigenvalues. Table 5.4.1 and 5.4.2 summarizes the sets for both designs,

Table 5.4.1 Classification of the 45 pairwise model comparisons for d_3

Set #	Number of Comparisons	Eigenvalues
1	45	± 0.9428

Table 5.4.2 Classification of the 45 pairwise model comparisons for d_4

Set #	Number of Comparisons	Eigenvalues
1	5	± 0.7453
2	20	± 0.9428
3	20	± 1

Any of the 45 model comparisons can be analyzed by permuting a reference model comparison for each set, the following three model comparisons are chosen to be reference model comparisons,

$$\left\{M_{u_1}, M_{u_1'}\right\} = \{x_1x_2, x_3x_4\} \ ; \ NZEV\left(H_{d_3}^{(u_1,u_1')}\right) = \pm 0.9428 \ ; \ NZEV\left(H_{d_4}^{(u_1,u_1')}\right) = \pm 0.7453,$$

$$\left\{M_{u_2}, M_{u_2'}\right\} = \{x_1x_2, x_1x_4\} \ ; \ NZEV\left(H_{d_3}^{(u_2,u_2')}\right) = \pm 0.9428 \ ; \ NZEV\left(H_{d_4}^{(u_2,u_2')}\right) = \pm 0.9428,$$

$$\left\{M_{u_3}, M_{u_3'}\right\} = \{x_1x_2, x_1x_3\} \ ; \ NZEV\left(H_{d_3}^{(u_3,u_3')}\right) = \pm 0.9428 \ ; \ NZEV\left(H_{d_4}^{(u_3,u_3')}\right) = \pm 1.$$

The values for $\underline{a}$, $\underline{b}$, $(\underline{a} + \underline{b})$, $(\underline{a} - \underline{b})$ are provided for each set, any comparison within the set are permutations of these values. Design d_3 is presented for all three model comparisons to demonstrate the permuted forms,

Table 5.5.1 Observation weights for set 1

Set #1		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
d_3	$\underline{a}$	-0.204	-0.289	0	-0.289	-0.204	-0.204	0.12	0.289	-0.204	0.697	0	0.289
	$\underline{b}$	0.204	-0.289	0	-0.289	0.204	0.204	-0.697	0.289	0.204	-0.12	0	0.289
	$\underline{a}+\underline{b}$	0	-0.577	0	-0.577	0	0	-0.577	0.577	0	0.577	0	0.577
	$\underline{a}-\underline{b}$	-0.408	0	0	0	-0.408	-0.408	0.816	0	-0.408	0.816	0	0
d_4	$\underline{a}$	-0.07	-0.113	-0.113	-0.113	-0.07	-0.478	0.07	0.295	-0.478	0.478	0.295	0.295
	$\underline{b}$	0.478	-0.295	-0.295	-0.295	0.478	0.07	-0.478	0.113	0.07	-0.07	0.113	0.113
	$\underline{a}+\underline{b}$	0.408	-0.408	-0.408	-0.408	0.408	-0.408	-0.408	0.408	-0.408	0.408	0.408	0.408
	$\underline{a}-\underline{b}$	-0.548	0.183	0.183	0.183	-0.548	-0.548	0.548	0.183	-0.548	0.548	0.183	0.183

Table 5.5.2 Observation weights for set 2

Set #2		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
d_3	$\underline{a}$	-0.697	-0.289	0	0.204	-0.204	0.289	0.12	0.289	0.289	0.204	0	-0.204
	$\underline{b}$	0.12	-0.289	0	-0.204	0.204	0.289	-0.697	0.289	0.289	-0.204	0	0.204
	$\underline{a}+\underline{b}$	-0.577	-0.577	0	0	0	0.577	-0.577	0.577	0.577	0	0	0
	$\underline{a}-\underline{b}$	-0.816	0	0	0.408	-0.408	0	0.816	0	0	0.408	0	-0.408
d_4	$\underline{a}$	-0.408	-0.289	-0.289	0.204	0.085	0.289	0.408	0	0.289	0.204	0	-0.493
	$\underline{b}$	0.408	-0.289	-0.289	-0.204	0.493	0.289	-0.408	0	0.289	-0.204	0	-0.085
	$\underline{a}+\underline{b}$	0	-0.577	-0.577	0	0.577	0.577	0	0	0.577	0	0	-0.577
	$\underline{a}-\underline{b}$	-0.816	0	0	0.408	-0.408	0	0.816	0	0	0.408	0	-0.408

Table 5.5.3 Observation weights for set 3

Set #3		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
d_3	$\underline{a}$	-0.289	-0.204	0	0.289	-0.289	0.204	-0.12	-0.289	0.697	-0.204	0	0.204
	$\underline{b}$	0.289	-0.204	0	-0.289	0.289	0.204	-0.697	0.289	0.12	-0.204	0	0.204
	$\underline{a}+\underline{b}$	0	-0.408	0	0	0	0.408	-0.816	0	0.816	-0.408	0	0.408
	$\underline{a}-\underline{b}$	-0.577	0	0	0.577	-0.577	0	0.577	-0.577	0.577	0	0	0
d_4	$\underline{a}$	0.236	0	0	-0.471	0.236	0	0.236	0.236	-0.471	0.471	-0.236	-0.236
	$\underline{b}$	0.471	-0.236	-0.236	-0.236	0.471	0.236	-0.471	0	0.236	-0.236	0	0
	$\underline{a}+\underline{b}$	0.707	-0.236	-0.236	-0.707	0.707	0.236	-0.236	0.236	-0.236	0.236	-0.236	-0.236
	$\underline{a}-\underline{b}$	-0.236	0.236	0.236	-0.236	-0.236	-0.236	0.707	0.236	-0.707	0.707	-0.236	-0.236

Tables 5.5 illustrate the weights of the observations y_i have on the differenced error sum of squares. The differenced error sum of squares is not influenced by two observations y_i when $\lambda = \pm 0.9428$ for both designs d_3 and d_4 as observed in table 5.5.2. For design d_3, the random variable S_1 does not depend on six observed values, similar with S_2 (observations may be different). Using (3.10.1) and (3.10.2), observe the effect of the two-factor interaction parameters on the expected values of $S_1^{(d,u,u')}$ and $S_2^{(d,u,u')}$

Table 5.6.1 Parameter weights for set 1

	Set #1	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{2,5}$	$\beta_{3,4}$	$\beta_{3,5}$	$\beta_{4,5}$
d_3	$(\underline{a}+\underline{b})'X_d$	2.309	0	0	-2.309	0	0	2.309	2.309	2.309	2.309
	$(\underline{a}-\underline{b})'X_d$	-3.266	-1.633	-1.633	0	1.633	1.633	0	3.266	0	0
d_4	$(\underline{a}+\underline{b})'X_d$	1.633	-1.633	1.633	1.633	-1.633	1.633	-1.633	1.633	1.633	1.633
	$(\underline{a}-\underline{b})'X_d$	-3.651	-0.73	-0.73	0.73	0.73	0.73	-0.73	3.651	-0.73	-0.73

Table 5.6.2 Parameter weights for set 2

	Set #2	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{2,5}$	$\beta_{3,4}$	$\beta_{3,5}$	$\beta_{4,5}$
d_3	$(\underline{a}+\underline{b})'X_d$	2.309	2.309	-2.309	-2.309	0	-2.309	0	0	2.309	0
	$(\underline{a}-\underline{b})'X_d$	-3.266	0	-3.266	0	1.633	0	-1.633	1.633	0	-1.633
d_4	$(\underline{a}+\underline{b})'X_d$	2.309	0	-2.309	0	0	0	-2.309	-2.309	0	0
	$(\underline{a}-\underline{b})'X_d$	-3.266	0	-3.266	0	1.633	0	-1.633	1.633	-1.633	-1.633

Table 5.6.3 Parameter weights for set 3

	Set #3	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{2,5}$	$\beta_{3,4}$	$\beta_{3,5}$	$\beta_{4,5}$
d_3	$(\underline{a}+\underline{b})'X_d$	3.266	3.266	0	0	0	-1.633	1.633	-1.633	1.633	0
	$(\underline{a}-\underline{b})'X_d$	-2.309	2.309	-2.309	2.309	2.309	0	0	0	0	-2.309
d_4	$(\underline{a}+\underline{b})'X_d$	2.828	-2.828	0.943	0.943	-0.943	0.943	-0.943	-0.943	0.943	-0.943
	$(\underline{a}-\underline{b})'X_d$	-2.828	-2.828	-0.943	0.943	0.943	0.943	-0.943	2.828	-0.943	-2.828

The expected value of the differenced error sum of squares is calculated using tables 5.6.1 to 5.6.3 and (3.9),

$$E[SSE_u(d) - SSE_{u'}(d)] = \lambda^{(d,u,u')} E\left[S_1^{(d,u,u')}\right] E\left[S_2^{(d,u,u')}\right].$$

If true model has $\beta_{1,2} \neq 0$ while all other two-factor interactions are zero, observe that the expected value of the differenced error sum of squares for the three referenced model comparisons will be in the form $c * \beta_{1,2}^2$, the values of c are presented,

Table 5.7 Expected differenced error sum of squares

$E[SSE_u(d) - SSE_{u'}(d)]$	$\{M_{u_1}, M_{u_1'}\}$	$\{M_{u_2}, M_{u_2'}\}$	$\{M_{u_3}, M_{u_3'}\}$
d_3	-7.1099	-7.1099	-7.1099
d_4	-4.4439	-7.1099	-7.9976

Table 5.7 demonstrates the relationship between the eigenvalues and the expected differenced error sum of squares. For d_3, the expected differenced error sum of squares is $-7.1099\beta_{1,2}^2$ regardless of the model comparison. For d_4, the expected differenced error sum of squares increases as the eigenvalue of the model comparison increases. Observe that when the eigenvalues of the two designs are the same, both designs have the expected differenced error sum of squares to be $-7.1099\beta_{1,2}^2$.

5.3 Simulated Data

To compare design d_3 and d_4 data will be simulated from the true model,

$$\underline{y} = 5\left(\underline{1} + \sum_{j=1}^{5} \underline{x}_j\right) + 5\underline{x_1 x_2} + \underline{\epsilon}\ ;\ \underline{\epsilon} \sim Normal\left(\underline{0}, \sigma^2 I\right),$$

100,000 datasets were generated using the true model for different values of σ. Table 5.8 presents the expected values of the different reference model comparisons when $\beta_{1,2} = 5$ and will be compared to the average simulated differenced error sum of squares,

Table 5.8 Expected differenced error sum of squares

$E[SSE_u(d) - SSE_{u'}(d)]$	$\{M_{u_1}, M_{u_1'}\}$	$\{M_{u_2}, M_{u_2'}\}$	$\{M_{u_3}, M_{u_3'}\}$
d_3	-177.7476	-177.7476	-177.7476
d_4	-111.0969	-177.7476	-199.9396

Table 4.8 demonstrates that in this case design d_3 will have much better discrimination abilities than d_4 for model comparisons in set #1, will have the same performance as d_4 for model comparisons in set #2, and will do marginally worse with model comparisons in set #3. Table2 5.9.1 and 5.9.2 presents the average differenced SSE values for designs d_3 and d_4,

Table 5.9.1 Average differenced SSE values for $\boldsymbol{d_3}$ $(\boldsymbol{\sigma = 0.05})$

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	-177.8	-177.83	-177.84	-177.86	-177.8	-177.82	-177.81	-177.85	-177.82
x_1x_3		-0.03	-0.04	-0.05	0	-0.02	-0.01	-0.05	-0.02
x_1x_4			-0.01	-0.02	0.03	0.01	0.02	-0.02	0.01
x_1x_5				-0.01	0.04	0.02	0.03	-0.01	0.02
x_2x_3					0.05	0.03	0.04	0.01	0.03
x_2x_4						-0.02	-0.01	-0.05	-0.02
x_2x_5							0.01	-0.02	0
x_3x_4								-0.04	-0.01
x_3x_5									0.03

Table 5.9.2 Average differenced SSE Values for d_4 ($\sigma = 0.05$)

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	-199.9	-177.72	-199.9	-177.71	-199.9	-199.9	-111.09	-177.73	-177.66
x_1x_3		22.18	0	22.19	0	0	88.81	22.17	22.24
x_1x_4			-22.17	0.01	-22.18	-22.18	66.63	-0.01	0.06
x_1x_5				22.19	0	0	88.81	22.17	22.23
x_2x_3					-22.19	-22.19	66.62	-0.02	0.05
x_2x_4						0	88.81	22.17	22.24
x_2x_5							88.81	22.17	22.24
x_3x_4								-66.64	-66.57
x_3x_5									0.07

When comparing models $M_u = x_1x_2$ and other models $M_{u'}$, the average value of the simulations are approximately the calculated expected values in table 5.8. Like the designs in chapter 4, when model M_u is not the correct model, the average differenced error sum of squares depends on which set the model comparison is in. For design d_4 and model $M_u \neq x_1x_2$, the average differenced values of $M_{u'} = x_3x_4$ have all been shifted up by 88.81 from the calculated values on table 5.8. Table 5.10.1.1 and table 5.10.1.2 presents the proportion of times model M_u is chosen for designs d_3 and d_4 for a low level of noise,

Table 5.10.1.1 Proportion of times M_u is chosen for d_3 ($\sigma = 0.05$)

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	1	1	1	1	1	1	1	1	1
x_1x_3		0.49745	0.50138	0.50262	0.50092	0.50055	0.49885	0.50042	0.50115
x_1x_4			0.50052	0.49962	0.49762	0.50115	0.49945	0.50485	0.49682
x_1x_5				0.50318	0.49958	0.50142	0.49718	0.49908	0.50038
x_2x_3					0.49648	0.49962	0.49958	0.50115	0.49832
x_2x_4						0.50228	0.50318	0.50288	0.50055
x_2x_5							0.49832	0.49962	0.50092
x_3x_4								0.50278	0.49738
x_3x_5									0.49832

Table 5.10.1.2 Proportion of times M_u is chosen for d_4 ($\sigma = 0.05$)

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	1	1	1	1	1	1	1	1	1
x_1x_3		0	0.49942	0	0.50235	0.50222	0	0	0
x_1x_4			1	0.50018	1	1	0	0.50148	0.49438
x_1x_5				0	0.50098	0.50045	0	0	0
x_2x_3					1	1	0	0.49992	0.49662
x_2x_4						0.50172	0	0	0
x_2x_5							0	0	0
x_3x_4								1	1
x_3x_5									0.49335

In the presence of low noise, d_3 and d_4 have a 100% success rate at choosing the correct two-factor interaction over the other 9 incorrect two-factor interactions. The model comparisons that have the expected differenced error sum of squares as 0 had 50% probability of picking model M_u or $M_{u'}$. If the expected differenced error sum of squares is positive, model $M_{u'}$ is chosen else model $M_{u'}$ is chosen in the model selection process. Table 5.10.2.1 and 5.10.2.2 presents the proportion of times M_u is chosen over $M_{u'}$ when the noise increases from $\sigma = 0.05$ to $\sigma = 5$.

Table 5.10.2.1 Proportion of times M_u is chosen for d_3 ($\sigma = 5$)

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	0.93878	0.93875	0.94115	0.94178	0.94048	0.93981	0.94118	0.93961	0.93831
x_1x_3		0.49612	0.50075	0.49982	0.50362	0.49092	0.50028	0.49025	0.49908
x_1x_4			0.50362	0.50248	0.50598	0.49682	0.50412	0.49792	0.50442
x_1x_5				0.49788	0.50465	0.49585	0.50095	0.49625	0.50002
x_2x_3					0.50482	0.49565	0.50658	0.49368	0.50078
x_2x_4						0.49432	0.49705	0.49089	0.49532
x_2x_5							0.50228	0.49732	0.50622
x_3x_4								0.49608	0.49838
x_3x_5									0.50375

Table 4.10.2.2 Proportion of times M_u is chosen for d_4 ($\sigma = 5$)

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	0.95601	0.93831	0.95521	0.93918	0.95768	0.95681	0.87376	0.94478	0.94128
x_1x_3		0.38224	0.50028	0.37425	0.49822	0.49955	0.16056	0.38024	0.35232
x_1x_4			0.62012	0.49742	0.62439	0.64825	0.26117	0.50192	0.50155
x_1x_5				0.34892	0.50262	0.49822	0.1605	0.37861	0.38221
x_2x_3					0.62402	0.62432	0.26207	0.50588	0.50821
x_2x_4						0.49798	0.158	0.35235	0.37954
x_2x_5							0.16216	0.37998	0.37938
x_3x_4								0.74116	0.74129
x_3x_5									0.50212

When the noise is increased to $\sigma = 5$, the model comparisons that have 0 expected differenced error sum of squares still pick model M_u 50% of the time. When model $M_u = x_1x_2$, design d_3 has a consistent about 94% of the time picking the correct model M_u over the remaining 9 models for $M_{u'}$. The performance of picking the correct model for design d_4 depends which set the model comparison belongs to and performs identical to d_3 when the eigenvalues are the same. For the comparison $\{M_u, M_{u'}\} = \{x_1x_2, x_3x_4\}$, observe that design d_3 performs significantly better than design d_4. Table 5.10.3.1 and 5.10.3.2 presents the proportion of times M_u is chosen over $M_{u'}$ when the noise increases from $\sigma = 5$ to $\sigma = 10$.

Table 5.10.3.1 Proportion of times M_u is chosen for d_3 ($\sigma = 10$)

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	0.7183	0.72383	0.72033	0.72067	0.7225	0.72137	0.7175	0.7194	0.7232
x_1x_3		0.50228	0.49748	0.49932	0.50052	0.50715	0.49868	0.50305	0.50215
x_1x_4			0.49848	0.49822	0.49785	0.50218	0.49578	0.49862	0.50322
x_1x_5				0.49802	0.49832	0.49848	0.49352	0.50062	0.50118
x_2x_3					0.49855	0.50402	0.49642	0.49705	0.50375
x_2x_4						0.50312	0.50012	0.50438	0.50512
x_2x_5							0.49578	0.49762	0.50188
x_3x_4								0.50232	0.50132
x_3x_5									0.50142

Table 5.10.3.2 Proportion of times M_u is chosen for d_4 $(\sigma = 10)$

$M_u \backslash M_{u'}$	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
x_1x_2	0.73839	0.71907	0.7336	0.72107	0.7351	0.73063	0.67721	0.72566	0.72226
x_1x_3		0.46129	0.50328	0.46629	0.49828	0.49458	0.37221	0.46686	0.45519
x_1x_4			0.53741	0.50455	0.53651	0.54371	0.4144	0.50785	0.50735
x_1x_5				0.45703	0.50205	0.49688	0.36915	0.46959	0.47022
x_2x_3					0.53674	0.53118	0.41127	0.50195	0.50028
x_2x_4						0.49468	0.37165	0.45853	0.46786
x_2x_5							0.37721	0.47006	0.46799
x_3x_4								0.58913	0.59116
x_3x_5									0.50038

When the value of σ gets larger, the proportion of times M_u is picked over $M_{u'}$ approaches 50%. The difference between design d_3 and d_4 starts reducing as the value of σ increases. In conclusion, d_3 and d_4 have their advantages and disadvantages. If no information is known about the true model, design d_3 is a safe design to choose as it has relatively good performance for all model comparisons. Design d_4 has 20 model comparisons which have better discrimination abilities than d_3, if some information about the true model is known, it may be possible to strategically use design d_4 to maximize the discrimination ability of the true model with the other models.

Chapter 6

Ridge Regression

6.1 Introduction

Consider the data of size n, $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_n, y_n)$ where $\boldsymbol{x}_i$ is data collected over m fixed variables is denoted as $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})$ with the mean of observation y_i depending on $\boldsymbol{x}_i$ denoted by $E[y_i|\boldsymbol{x}_i]$. Each of size n, denote the observation vector $\underline{y} = (y_1, y_2, \ldots, y_n)'$, main effects as $\underline{x}_j = \left(x_{1j}, x_{2j} \ldots x_{nj}\right)'$, two-factor interactions as $\underline{x_j x_{j'}} = \left(x_{1j}x_{1j'}, x_{2j}x_{2j'}, \ldots, x_{nj}x_{nj'}\right)'$, quadratic terms $\underline{x}_j^2 = \left(x_{1j}^2, x_{2j}^2, \ldots, x_{nj}^2\right)'$, or higher order terms in the same manner. The columns of the design matrix X are constructed using a column of ones $\underline{1}$ and selected columns of main effects, interaction terms, and higher order terms for a total of p columns and n rows. For a constructed design matrix X, let $\underline{\beta} = \left(\beta_1, \beta_2, \ldots, \beta_p\right)'$ be the set of unknown p parameters, one unknown parameter for each column of X, and the linear model is expressed by,

$$M: E\left[\underline{y}|X\right] = X\underline{\beta} \quad , \quad Var\left(\underline{y}\right) = \sigma^2 I_n.$$

The least squares estimator minimizes the least squares objective function,

$$\underline{\hat{\beta}}^{(LS)} = argmin_{\underline{\beta}}\left\{\left(\underline{y} - X\underline{\beta}\right)'\left(\underline{y} - X\underline{\beta}\right)\right\},$$

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

If $Rank(X) = Rank(X'X) = p$, then the matrix $X'X$ is invertible, the least squares estimator has a unique estimate for the unknown parameters $\underline{\beta}$. If $Rank(X) < p$, there are infinite many solutions of $\underline{\hat{\beta}}^{(LS)}$ that attain the minimum value. Ridge regression is another method at estimating the unknown parameters $\underline{\beta}$ for a specified value of a tuning parameter $\lambda \geq 0$. For a given value of the tuning parameter λ, the ridge regression estimates minimize the objective function,

$$\underline{\hat{\beta}}^{(R)} = argmin_{\underline{\beta}} \left\{ \left(\underline{y} - X\underline{\beta}\right)' \left(\underline{y} - X\underline{\beta}\right) + \lambda \underline{\beta}' \underline{\beta} \right\},$$

$$\underline{\hat{\beta}}^{(R)} = \left(X'X + \lambda I_p\right)^{-1} X' \underline{y}. \tag{6.1}$$

For $\lambda > 0$, the inverse of $X'X + \lambda I_p$ is unique guaranteeing that a unique estimate of $\underline{\hat{\beta}}^{(R)}$ can be calculated (Hoerl, et al., 1970). The resulting estimates are biased and critiqued by Smith and Campbell (1980). Alternatively, ridge regression can be interpreted as a constrained least squares optimization objective, for a given $c > 0$,

$$\min\left(\left(\underline{y} - X\underline{\beta}\right)' \left(\underline{y} - X\underline{\beta}\right)\right),$$

$$\text{subject to } \underline{\beta}' \underline{\beta} \leq c.$$

The constraint on the objective function forces the estimates within a hypersphere of radius $\sqrt{c}$. For a given c, if the least squares estimator $\underline{\hat{\beta}}^{(LS)'} \underline{\hat{\beta}}^{(LS)} \leq c$, then the least squares estimator already minimizes the objective function $\left(\underline{y} - X\underline{\beta}\right)' \left(\underline{y} - X\underline{\beta}\right)$ hence using value of $\lambda = 0$ in

(6.1) solves the optimization objective, else there is a unique value of λ that satisfies the equality constraint $\underline{\hat{\beta}}^{(R)'}\underline{\hat{\beta}}^{(R)} = c$. The inequality constraint $\underline{\hat{\beta}}^{(R)'}\underline{\hat{\beta}}^{(R)} \leq c$ can be written in terms of the design matrix X and observation vector $\underline{y}$,

$$\underline{y}'X\left(X'X + \lambda I_p\right)^{-2}X'\underline{y} \leq c. \tag{6.2}$$

Observe the relationship as c decreases, the value of λ must increases to satisfy the inequality constraint. As λ increases, the ridge regression estimates $\hat{\beta}_j^{(R)}$ approaches 0 but do not attain 0 unless $\hat{\beta}_j^{(LS)} = 0$. This makes ridge regression an undesirable approach at model selection because it is not directly clear how to separate out negligible effects from important effects. In practice, the choice of the tuning parameter λ is often selected to minimize the generalized cross validation (Golub, et al., 1979).

6.2 Special Case: Orthogonal Designs

Let d^* be an orthogonal design with n runs, m main effects, where the design matrix with n rows and p columns satisfies the property $X'_{d^*}X_{d^*} = nI_p$. The estimates given in (6.1) simplify to,

$$\underline{\hat{\beta}}^{(R)} = \left(X'X + \lambda I_p\right)^{-1}X'\underline{y} = \frac{X'\underline{y}}{n+\lambda}.$$

The least squares estimation for an orthogonal design is calculated to be $\frac{1}{n}X'\underline{y}$. For conceptual purposes, express the ridge regression estimates in terms of the least squares estimates,

$$\underline{\hat{\beta}}^{(R)} = \frac{n}{n+\lambda}\underline{\hat{\beta}}^{(LS)}.$$

For any least square estimate $\hat{\beta}_j^{(LS)}$ the ridge regression estimate will be the least square estimate for $\lambda = 0$ or have a shrinkage coefficient of $n(n+\lambda)^{-1} \leq 1$. The shrinkage coefficient is the same for all least squares estimates, hence if $\hat{\beta}_j^{(LS)} \geq \hat{\beta}_{j'}^{(LS)}$ then $\hat{\beta}_j^{(R)} \geq \hat{\beta}_{j'}^{(R)}$ the ordering of the ridge regression estimates will be consistent for all values of $\lambda \geq 0$. To demonstrate the relationship between c and λ, (6.2) can be simplified for an orthogonal design,

$$0 \leq \frac{\underline{y}'XX'\underline{y}}{c} \leq (n+\lambda)^2,$$

$$\Rightarrow \lambda \geq -n + \sqrt{\frac{\underline{y}'XX'\underline{y}}{c}}.$$

For any value of c , design matrix X, and observation vector $\underline{y}$ of n runs, the value of λ could be calculated by $\lambda = -n + \sqrt{\frac{\underline{y}'XX'\underline{y}}{c}}$. If $-n + \sqrt{\frac{\underline{y}'XX'\underline{y}}{c}}$ is a negative value, then the value of $\lambda = 0$ is used and the least squares estimate is returned. This also demonstrates the inverse relationship between c and λ.

Consider the fractional orthogonal design d^* with $m = 5$ main effects and $n = 2^{5-1} = 16$ runs determined by the identity relation $A_1 + A_2 + A_3 + A_4 + A_5 = 1$. This design has full estimation capacity for the overall mean, $m = 5$ main effects, and all $\binom{5}{2} = 10$ main effects for a total of $p = 16$ parameters. The observation vector $\underline{y}$ is given by the Design and Analysis of Experiments – 5th edition, Montgomery,

d^*

Run	**Treatment**	$\underline{x}_1$	$\underline{x}_2$	$\underline{x}_3$	$\underline{x}_4$	$\underline{x}_5$	$\underline{y}$
1	1	1	-1	-1	-1	-1	6
2	2	-1	1	-1	-1	-1	16
3	3	-1	-1	1	-1	-1	34
4	4	-1	-1	-1	1	-1	9
5	5	-1	-1	-1	-1	1	8
6	123	1	1	1	-1	-1	44
7	124	1	1	-1	1	-1	21
8	125	1	1	-1	-1	1	15
9	134	1	-1	1	1	-1	50
10	135	1	-1	1	-1	1	30
11	145	1	-1	-1	1	1	10
12	234	-1	1	1	1	-1	60
13	235	-1	1	1	-1	1	45
14	245	-1	1	-1	1	1	22
15	345	-1	-1	1	1	1	52
16	12345	1	1	1	1	1	63

For this orthogonal design d^*, let the design matrix X_{d^*} have the columns for the general mean, 5 main effects, and 10 two-factor interactions for a total of $p = 16$ parameters. The $Rank(X_{d^*}) = p = n = 16$ and the least squares estimation is unique and given by $\frac{1}{16}X'\underline{y}$.

Table 6.1 (next page) presents the values of the least squares estimation. If it is desired for the ridge regression estimates to be within $\underline{\hat{\beta}}^{(R)'}\underline{\hat{\beta}}^{(R)} \leq 100 = c$, the value of the tuning parameter λ that corresponds with that value of c is $\lambda = 42.23915$. The shrinkage coefficient is calculated to be $16(16 + 42.23915)^{-1} = 0.2747$. Figure 6.1 demonstrates the values of the ridge regression estimate for various values of λ (intercept was not included),

Figure 6.1 Values of ridge regression estimates for different values of λ

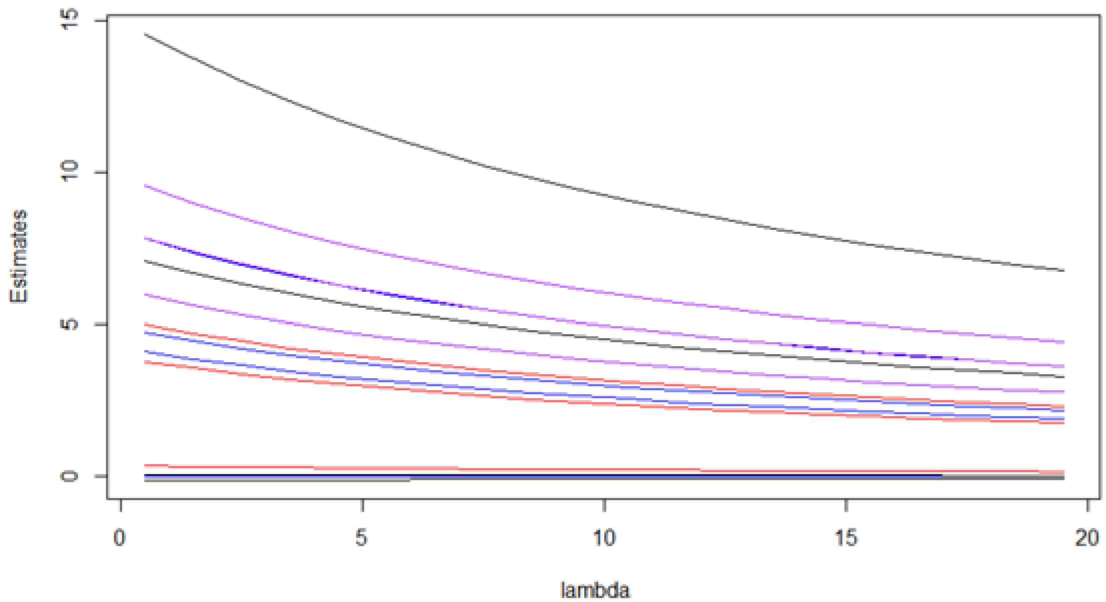


Table 6.1 Least squares and ridge regression estimates

j	$\hat{\beta}_j^{(LS)}$ $\lambda = 0$	$\hat{\beta}_j^{(R)}$, $\lambda = 42.23915$
0	30.3125	8.4732
1	5.5625	1.5548
2	16.9375	4.7345
3	5.4375	1.5199
4	-0.4375	-0.1222
5	0.3125	-0.0873
1,2	3.4375	0.9608
1,3	0.1875	0.0524
1,4	0.5625	0.1572
1,5	0.5625	-0.1572
2,3	0.3125	0.0873
2,4	-0.0625	-0.0174
2,5	-0.0625	0.0174
3,4	0.4375	0.1222
3,5	0.1875	-0.0524
4,5	-0.6875	0.1921
$\underline{\hat{\beta}}'\underline{\hat{\beta}}$	1279.812	100

Observe that by ordering the least squares estimates by magnitude in descending order, for any value of λ, the ordering will remain the same as the same shrinkage constant will be applied to each $\hat{\beta}_j^{(LS)}$ estimate which is demonstrated by both figure 6.1 and table 6.1. Denote the ordered estimates $\hat{\beta}_{(j)}$ and satisfies condition,

$$\hat{\beta}_{(1)} \leq \hat{\beta}_{(2)} \leq \cdots \leq \hat{\beta}_{(p)}.$$

Using the value of $\lambda = 110000$, table 6.2 demonstrates the ordering of the estimates will remain the same and the large value of λ shrinks the estimates close to 0, where < 0.0001 is defined to be close to 0 (general mean has been excluded),

Table 6.2 Ordered least squares and ridge regression estimates

$(\boldsymbol{j})$	$\boldsymbol{j}$	$\hat{\boldsymbol{\beta}}_{(\boldsymbol{j})}^{(\boldsymbol{LS})}$ $\lambda = 0$	$\hat{\boldsymbol{\beta}}_{\boldsymbol{j}}^{(\boldsymbol{R})}$, $\lambda = 1.1 * 10^5$
(15)	2	16.938	0.00246
(14)	1	5.5625	0.00081
(13)	3	5.4375	0.00079
(12)	1,2	3.4375	0.0005
(11)	4,5	0.6875	<0.0001
(10)	1,4	0.5625	<0.0001
(9)	1,5	0.5625	<0.0001
(8)	4	0.4375	<0.0001
(7)	3,4	0.4375	<0.0001
(6)	5	0.3125	<0.0001
(5)	2,3	0.3125	<0.0001
(4)	1,3	0.1875	<0.0001
(3)	3,5	0.1875	<0.0001
(2)	2,4	0.0625	<0.0001
(1)	2,5	0.0625	<0.0001

For a non-orthogonal design, as λ increases, the ordering of the estimates may change as the shrinkage constant will be different per estimate.

6.3 Model Selection Procedure

Ridge regression rarely provides estimates exactly 0 which adds further complications to a model selection procedure. One proposition is to pick an arbitrary threshold t so for a given λ, if the magnitude of a ridge regression estimate is equal or below t, then the effect is considered negligible. The example in section 6.2, if $t = 0.0001$, all but the top 4 estimates would be chosen. The model selection process would be as follows,

1) Start out estimating a full model (general mean, linear terms, and all desired higher order terms)
2) For a chosen value of λ and t, if the magnitude of the ridge regression estimate is below the threshold t $\left(\left|\hat{\beta}_j^{(R)}\right| \leq t\right)$ then the parameter is negligible $\left(\beta_j = 0\right)$ and not chosen in the model selection process.

This model process should cause concern because selecting a specific value of the tuning parameter λ is often challenging and adding another quantity t as that only increases the difficulty of the model selection process.

The argument will be that for a low value of t^*, the performance of the model selection process will be approximately consistent for all $t \leq t^*$. Consider design d_1 that had $n = 12$ runs, $m = 5$ main effects, and $\binom{5}{2} = 10$ two-factor interactions as described in the earlier chapters.

Using the true model presented in chapter 4 or chapter 5 for simulation purposes, a sample of observations were generated for $\sigma = 5$. The design d_1 and observation vector $\underline{y}$ are presented,

d_1

Run	$\underline{x}_1$	$\underline{x}_2$	$\underline{x}_3$	$\underline{x}_4$	$\underline{x}_5$	$\underline{y}$
1	-1	-1	-1	-1	-1	-19.935
2	1	1	-1	-1	1	11.392
3	1	-1	1	-1	1	6.014
4	1	-1	-1	1	1	2.028
5	-1	1	1	-1	1	5.997
6	-1	1	-1	1	1	13.205
7	-1	-1	1	1	1	21.415
8	1	1	1	1	1	28.032
9	1	1	1	-1	-1	12.759
10	1	1	-1	1	-1	17.094
11	1	-1	1	1	-1	4.532
12	-1	1	1	1	-1	9.384

Starting with the full model consisting of the general mean, $m = 5$ main effects, and all 10 two-factor interactions for a total of $p = 16$ parameters. This design has estimation capacity for 5 two-factor interactions so the least squares estimation will not be unique. However, for $\lambda > 0$, the ridge regression estimates are unique. A word of caution, while theoretically any value of $\lambda > 0$ the inverse of $(X'X + \lambda I)$ exists, for values close to $\lambda = 0$, the inverse may still be computationally singular and unstable. Table 6.2 presents two-factor interaction ridge regression estimates for $\lambda = 10, 100, 1000,$ and 10000 where the ridge regression estimates are in descending order of magnitude,

Table 6.2 Ridge regression estimates (magnitude) for different λ

j	$\lambda = 10$	j	$\lambda = 10^2$	j	$\lambda = 10^3$	j	$\lambda = 10^4$
1,2	1.413946	1,5	0.31172	1,5	0.037258	1,5	0.0038
1,5	1.189305	1,4	0.297346	1,4	0.035617	1,4	0.003634
1,4	1.125172	2,3	0.288432	2,3	0.034599	2,3	0.003531
2,3	1.085401	1,2	0.271767	1,2	0.029361	1,2	0.002957
2,5	0.704044	2,5	0.202955	2,5	0.02484	2,5	0.002541
1,3	0.597518	1,3	0.179079	1,3	0.022114	1,3	0.002264
3,5	0.480158	3,5	0.06247	3,5	0.005465	3,5	0.000533
4,5	0.394459	4,5	0.043261	2,4	0.00403	2,4	0.00043
3,4	0.187959	2,4	0.020693	4,5	0.003272	4,5	0.000311
2,4	0.109127	3,4	0.003023	3,4	0.002013	3,4	0.000225

The ordering of the ridge regression estimates $\hat{\beta}_j^{(R)}$ for $\lambda = 10$ and the ranking changes for $\lambda = 100$. The estimate $\hat{\beta}_{1,2}^{(R)}$ went from being the largest estimate to the fourth largest estimate. The estimate $\hat{\beta}_{3,4}^{(R)}$ switched rankings with $\hat{\beta}_{2,4}^{(R)}$. For $\lambda = 100$ to $\lambda = 1000$ only the last two estimates switched rankings. The rankings between $\lambda = 10^3$ and $\lambda = 10^4$ remained unchanged and the implication is for any value $\lambda \in (10^3, 10^4)$ would have the same ranking. As λ becomes larger, the rate of the estimates decreasing is approximately same relative to the size of the ridge regression estimate so that the ranks of the estimates remain the same, similar to the orthogonal design case. To study the effect of the threshold run a simulation as follows,

1) Generate observations $\underline{y}$ from the true model for $\sigma = 5$,

$$\underline{y} = 5\left(\underline{1} + \sum_{j=1}^{5} \underline{x}_j\right) + 5\underline{x_1 x_2} + 7\underline{x_2 x_3} + \underline{6x_4 x_5} + \underline{\epsilon}\,;\ \underline{\epsilon} \sim Normal\left(\underline{0}, \sigma^2 I\right).$$

2) Start out estimating the full model (general mean, linear terms, and all desired higher order terms).

3) For a chosen value of t, find a value of λ such that 3 two-factor estimates are above the threshold of t while all other two-factor estimates are equal or below t.

4) For each dataset, the number of correct two-factor interactions can either be 0, 1, 2, or 3. Repeat 100,000 times for $t = 0.01, 0.001, 0.00001, 0.000001$ and present the proportion of times the number of correct two-factor interaction terms were chosen.

This model selection procedure can be computationally performed by starting with two values of λ: one low value ($\lambda_L = 0.001$) and one high value ($\lambda_H = 10^{12}$). Each iteration calculates the midpoint of the low value and high value $\lambda_m = \frac{1}{2}(\lambda_L + \lambda_H)$, calculates the ridge regression estimate the midpoint λ_m, and calculates the number of ridge regression estimates above the threshold t. If there are less than 3 two-factor interaction estimates below t, then a smaller value of λ_m is needed, else a larger value of λ_m is needed. To obtain a smaller value of λ_m, set $\lambda_H = \lambda_m$, else set $\lambda_L = \lambda_m$ then repeat. It is important to observe that there are a range of λ values that will satisfy the condition where only 3 two-factor interactions are above the threshold t and the remaining two-factor interactions are below the threshold t, making this algorithm to find a suitable value of λ that satisfies the condition very quick and effective. Instead of searching over 3 two-factor interactions, this model procedure could generalize to a positive integer k two-factor interactions and the methodology is comparable to the search design methodology in chapter 3-5.

This model selection procedure only requires a specification of t and k. Table 6.3 presents the model selection performance for this model selection procedure for $k = 3$ and $t = 0.01, 0.001, 0.0001, 0.00001$, and 0.000001,

Table 6.3 Model selection performance for different values of t

# correct	$t = 10^{-2}$	$t = 10^{-3}$	$t = 10^{-4}$	$t = 10^{-5}$	$t = 10^{-6}$
0	0	0	0	0	0
1	0.00001	0.00001	0	0.00001	0.00001
2	0.03487	0.03511	0.03548	0.03564	0.03585
3	0.96512	0.96488	0.96452	0.96435	0.96414
$mean(\lambda)$	5727.15	55856.64	523294.58	5725980.46	56534305.12

The model selection procedure has consistent results for the selected values of t, approximately 96.5% of the time recovering the true model. The idea is that by selecting a small enough threshold t, a large value of λ will be selected and the rankings of the estimates are fixed in for larger values of λ giving a consistent model selection approach. To select a threshold t, pick a small value of t so that the model selection procedure results is consistent, but not too small as the value of λ required can get unnecessarily large. When using ridge regression as a model selection technique, most consideration needs to be emphasized on the choice of the positive integer k similar to the search design methodology. Chapter 8 compares the ridge regression model selection procedure with the search design methodology and LASSO model selection procedure (chapter 7).

Chapter 7

Least Absolute Shrinkage and Selection Operator (LASSO)

7.1 Introduction

Consider the data of size n, $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_n, y_n)$ where $\boldsymbol{x}_i$ is data collected over m fixed variables is denoted by $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})$ with the mean of observation y_i depending on $\boldsymbol{x}_i$ denoted by $E[y_i|\boldsymbol{x}_i]$. Each of size n, denote the observation vector $\underline{y} = (y_1, y_2, \ldots, y_n)'$, main effects $\underline{x}_j = \left(x_{1j}, x_{2j} \ldots x_{nj}\right)'$, two-factor interactions as $\underline{x_j x_{j'}} = \left(x_{1j}x_{1j'}, x_{2j}x_{2j'}, \ldots, x_{nj}x_{nj'}\right)'$, quadratic terms $\underline{x}_j^2 = \left(x_{1j}^2, x_{2j}^2, \ldots, x_{nj}^2\right)'$, or higher order terms in the same manner. The columns of the design matrix X are constructed using a column of ones $\underline{1}$ and selected columns of main effects, interaction terms, and higher order terms for a total of p columns and n rows. For a constructed design matrix X, let $\underline{\beta} = \left(\beta_1, \beta_2, \ldots, \beta_p\right)'$ be the set of unknown p parameters, one unknown parameter for each column of X, and the linear model is expressed by,

$$M: E\left[\underline{y}|X\right] = X\underline{\beta} \;\; , \;\; Var\left(\underline{y}\right) = \sigma^2 I_n.$$

The least squares estimator are the estimates that minimize the least squares objective function,

$$\underline{\hat{\beta}}^{(LS)} = argmin_{\underline{\beta}} \left\{\left(\underline{y} - X\underline{\beta}\right)' \left(\underline{y} - X\underline{\beta}\right)\right\},$$

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

If $Rank(X) = Rank(X'X) = p$, then the matrix $X'X$ is invertible, the least squares estimator has a unique estimate of unknown parameters $\underline{\beta}$. If $Rank(X) < p$, there are infinite many solutions of $\underline{\hat{\beta}}^{(LS)}$ that attain the minimum value. LASSO is another method at estimating the unknown parameters $\underline{\beta}$ for a specified value of a tuning parameter $\lambda \geq 0$. For a given value of the tuning parameter λ, the LASSO estimates minimize the objective function,

$$\underline{\hat{\beta}}^{(L)} = argmin_{\underline{\beta}} \left\{ \left(\underline{y} - X\underline{\beta}\right)' \left(\underline{y} - X\underline{\beta}\right) + \lambda \sum_{j=1}^{p} \left|\beta_j\right| \right\}.$$

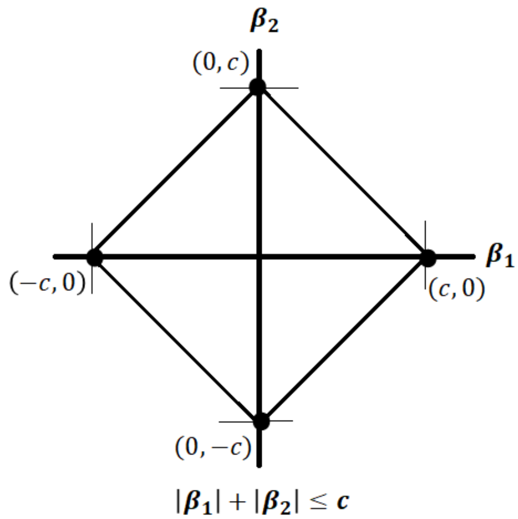
LASSO estimation is usually completed by computational methods, section 7.2 describes a procedure at finding LASSO estimates and what conditions must be necessary for the LASSO estimates to be unique. LASSO can be interpreted as a constrained least squares optimization objective, for a given $c > 0$,

$$\min\left(\left(\underline{y} - X\underline{\beta}\right)' \left(\underline{y} - X\underline{\beta}\right)\right),$$

$$\text{subject to } \sum_{j=1}^{p} \left|\beta_j\right| \leq c.$$

The constraint on the objective function forces the estimates within a hypercube where the vertices are a distance c away from the origin. A graphical representation of the constraint function for two-dimensions is presented in figure 7.1,

Figure 7.1 Graphical representation of the LASSO constraint function



For a given c, if the least squares estimator $\sum_{j=1}^{p} \left|\hat{\beta}_j^{(LS)}\right| \leq c$ then the least squares estimates simultaneously minimizes the objective function and is within the constraint so the value of $\lambda = 0$ is chosen. Otherwise, there is a value of λ that finds a minimum of the LASSO objective function for a given c. LASSO is different than ridge regression in a sense that as λ increases, LASSO estimates can attain exactly 0. When using LASSO for model selection, a value of the tuning parameter is chosen by some criterion function, and the non-zero estimates are the chosen in the model selection process.

7.2 LASSO Estimation

Consider the situation where $Rank(X) = p$ so that the least squares estimates are unique, in this situation LASSO estimates will be unique for all $\lambda > 0$. When considering the least squares or ridge regression objective functions, the analytical forms of the estimates can be found by taking the first derivative with respect to $\underline{\beta}$. Estimation for LASSO is not analytically straight forward as the objective function involves absolute values which are not continuously differentiable. One strategy to calculate the first derivative is to define a sign function $z(\beta_j)$ to be 1 if $\beta_j > 0$ or -1 if $\beta_j \leq 0$,

$$z(\beta_j) = z_j = \begin{cases} 1 & if\ \beta_j > 0 \\ -1 & if\ \beta_j \leq 0 \end{cases}$$

Denote $\underline{z}' = (z_1, z_2, \ldots, z_p)$, the objective function can be expressed as,

$$\left(\underline{y} - X\underline{\beta}\right)' \left(\underline{y} - X\underline{\beta}\right) + \lambda \underline{z}' \underline{\beta},$$

taking the first derivative, setting $\underline{\beta} = \underline{\hat{\beta}}^{(L)}$, and the set of equations to be 0,

$$\begin{aligned} \underline{\hat{\beta}}^{(L)} &= (X'X)^{-1}X'\underline{y} - \frac{\lambda}{2}(X'X)^{-1}\underline{\hat{z}}, \\ &= \underline{\hat{\beta}}^{(LS)} - \frac{\lambda}{2}(X'X)^{-1}\underline{\hat{z}}. \end{aligned} \tag{7.1}$$

Where $z\left(\hat{\beta}_j^{(L)}\right) = \hat{z}_j$ and $\underline{\hat{z}} = (\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_p)$. Since the sign of the LASSO estimates will be identical to the sign of the least squares estimates, the values for $z\left(\hat{\beta}_j^{(L)}\right) = z\left(\hat{\beta}_j^{(LS)}\right)$. Observe that $\frac{\lambda}{2}(X'X)^{-1}\underline{\hat{z}}$ is a p x 1 vector that determines the rate at which the least squares estimate $\underline{\hat{\beta}}^{(LS)}$ reaches 0. The j^{th} row of $\frac{\lambda}{2}(X'X)^{-1}\underline{\hat{z}}$ is the rate for which $\hat{\beta}_j^{(LS)}$ reaches 0. From the LASSO objective function when that $\lambda = 0$ results in the least squares estimates. As $\lambda \to \infty$ all LASSO estimates will attain exactly 0. There needs to be caution when using (7.1) since for large values of λ, the formula does not account for the number of LASSO estimates which are exactly 0. Before discussing how to use (7.1) to obtain the LASSO estimates, some notation and concepts need to be discussed.

Denote $p^*(\lambda)$ to be the number of non-zero LASSO estimates and denote $\lambda_{(j)}$ to be the smallest value of λ such that there are j LASSO estimates that are exactly 0 for $j = 0,1,2, \dots, p$. Observe that $p^*\left(\lambda_{(j)}\right) = p - j$ and the sequence of $\lambda_{(j)}$ values satisfy,

$$0 = \lambda_{(0)} \leq \lambda_{(1)} \leq \cdots \leq \lambda_{(j)} \leq \lambda_{(p)} < \infty.$$

If the sequence of $\lambda_{(j)}$ is known and which of the estimates are 0, (7.1) can be used to calculate the non-zero LASSO estimates and the remaining j estimates are exactly 0 for a given λ. The procedure to calculate the sequence of $\lambda_{(j)}$ and which of the j estimates will be 0 is as follows,

1) Calculate the least squares estimates $\underline{\hat{\beta}}^{(LS)}$ and the rate of shrinkage for each estimate $\frac{1}{2}(X'X)^{-1}\underline{\hat{z}}$. Let $A_j = \hat{\beta}_j^{(LS)}$ and B_j the j^{th} row of $\frac{1}{2}(X'X)^{-1}\underline{\hat{z}}$, there are p equations in the form of,

$$A_j - \lambda_j B_j = 0.$$

2) Calculate λ_j for all p equations by calculating $\lambda_j = \frac{A_j}{B_j}$

3) The first lambda value in the sequence is the minimum of the positive values of λ_j $\lambda_{(1)} = \min\{\lambda_j | \lambda_j > 0\}$

4) The value of $\lambda_{(1)}$ is known along with the corresponding estimate that attained 0. Remove that column from the design matrix X and repeat steps 1-3 for all p predictors.

This procedure is only valid when $Rank(X) = p$ and the least squares estimates are unique. For demonstration purposes, consider a random generated 4 x 4 design matrix X that is full rank and random generated $\underline{y}$ vector,

Table 7.1 Sample data

X				$\underline{y}$
1	1.510767	2.458121	1.678338	6
1	7.200772	2.676358	5.710503	16
1	4.840743	4.109823	4.98994	34
1	-0.70587	8.475614	2.501209	9

The least squares estimates, the rate of shrinkages, and the values of λ_j are calculated from step 1 presented in table 7.2,

Table 7.2 Calculated values of A_j, B_j, and λ_j

j	A_j	B_j	λ_j
1	19.88324	8.27873	2.401726
2	-38.2974	-11.2851	3.393629
3	-21.4974	-6.59623	3.259043
4	57.68706	15.40609	3.744432

For the first iteration $\lambda_{(1)} = 2.017$ is the minimum value of λ for the intercept estimate $j = 1$.

The second iteration starts by deleting the intercept column and repeating the procedure,

Table 7.3 Design matrix *X* without the fourth column

$\boldsymbol{X}$			$\underline{y}$
1.510767	2.458121	1.678338	6
7.200772	2.676358	5.710503	16
4.840743	4.109823	4.98994	34
-0.70587	8.475614	2.501209	9

The least squares estimates, rate of shrinkage, and new λ_j values are re-calculated.

Table 7.4 Calculated values of $\boldsymbol{A_j}, \boldsymbol{B_j}$, and $\boldsymbol{\lambda_j}$

$\boldsymbol{j}$	$\boldsymbol{A_j}$	$\boldsymbol{B_j}$	$\boldsymbol{\lambda_j}$
2	-24.4285	-5.51054	4.433051
3	-12.5563	-2.87344	4.369774
4	40.16657	8.111132	4.95203

For the second iteration $\lambda_{(2)} = 4.3697$ is the minimum value for the estimate that corresponds to $j = 3$. The remaining iterations can be completed to obtain the lambda sequence,

Table 7.5 Lambda sequence

$\boldsymbol{j}$	$(\boldsymbol{j})$	$\boldsymbol{\lambda_{(j)}}$
1	1	2.401726
3	2	4.369774
2	3	8.764383
4	4	587.21384

Table 7.5 presents the values and which estimates are 0. If $\lambda = 1$ then all 4 estimates are non-zero, if $\lambda = 3$, then only $\hat{\beta}_1^{(L)} = 0$ and the remaining three estimates can be calculated using (7.1), the design matrix X in (7.1) consists only columns corresponding to the non-zero estimates. To verify this procedure is correct, the optim function in R using the BFGS algorithm will do a computational optimization of the LASSO objective function. Table 7.6 presents values of λ, the estimates using equation (7.1), the results from the optim function, and the value of the LASSO objective function.

Table 7.6.1 LASSO and optim estimates for $\boldsymbol{\lambda = 2}$

Method	$\hat{\boldsymbol{\beta}}_1^{(L)}$	$\hat{\boldsymbol{\beta}}_2^{(L)}$	$\hat{\boldsymbol{\beta}}_3^{(L)}$	$\hat{\boldsymbol{\beta}}_4^{(L)}$	**Objective Function**
Equation (7.1)	3.3257	-15.7272	-8.3049	26.8749	191.5979
Optim	3.3276	-15.7296	-8.3064	26.8781	191.5979

Table 7.6.2 LASSO and optim estimates for $\boldsymbol{\lambda = 3}$

Method	$\hat{\boldsymbol{\beta}}_1^{(L)}$	$\hat{\boldsymbol{\beta}}_2^{(L)}$	$\hat{\boldsymbol{\beta}}_3^{(L)}$	$\hat{\boldsymbol{\beta}}_4^{(L)}$	**Objective Function**
Equation (7.1)	0	-7.8969	-3.9360	15.8332	229.5345
Optim	0	-7.8992	-3.9380	15.8373	229.5345

Table 7.6.3 LASSO and optim estimates for $\boldsymbol{\lambda = 6}$

Method	$\hat{\boldsymbol{\beta}}_1^{(L)}$	$\hat{\boldsymbol{\beta}}_2^{(L)}$	$\hat{\boldsymbol{\beta}}_3^{(L)}$	$\hat{\boldsymbol{\beta}}_4^{(L)}$	Objective Function
Equation (7.1)	0	-0.2193	0	4.5822	260.0035
Optim	-0.0001	-0.2177	0.0002	4.5777	260.0052

Table 7.6.4 LASSO and optim estimates for $\boldsymbol{\lambda = 300}$

Method	$\hat{\boldsymbol{\beta}}_1^{(L)}$	$\hat{\boldsymbol{\beta}}_2^{(L)}$	$\hat{\boldsymbol{\beta}}_3^{(L)}$	$\hat{\boldsymbol{\beta}}_4^{(L)}$	Objective Function
Equation (7.1)	0	0	0	2.1568	1219.263
Optim	0	0.0019	0.0081	2.1571	1219.458

Table 7.6.1 to table 7.6.4 all indicate the procedure outlines through section 7.2 is giving approximately the same LASSO estimates as the optim function in R. The BFGS algorithm performs worse at minimizing the LASSO objective function when the estimates are exactly 0 as indicated by table 7.6.3 and table 7.6.4 where the analytical approach achieves a smaller minimum.

The estimation technique outlined in section 7.2 only is possible when $Rank(X) = p$. When $Rank(X) < p$ computational approaches can be used to find a unique LASSO estimates for a given λ only in the case where the rank of the design matrix is equal to the effective degrees (Zou, et al., 2007) of freedom, $Rank(X) = p^*(\lambda) \leq p$.

7.3 Special Case: Orthogonal Designs

Let d^* be an orthogonal design with n runs, m main effects, where the design matrix with n rows and p columns satisfies the property $X'_{d^*}X_{d^*} = nI_p$. The estimates given in (7.1) simplify to,

$$\underline{\hat{\beta}}^{(L)} = \underline{\hat{\beta}}^{(LS)} - \frac{\lambda}{2n}\underline{\hat{z}}.$$

The interpretation is that all least squares estimates have a constant linear rate of shrinkage towards zero of $\frac{\lambda}{2n}$. This simplifies the procedure to find the corresponding λ sequence, since all effects are uncorrelated with each other, there is no need to redefine the X matrix every iteration. The procedure to calculate the lambda sequence $\lambda_{(j)}$,

1) Order the least squares estimates by magnitude to satisfy the condition,

$$0 \leq \hat{\beta}_{(1)}^{(LS)} \leq \hat{\beta}_{(2)}^{(LS)} \leq \cdots \leq \hat{\beta}_{(p)}^{(LS)}.$$

2) The lambda sequence is calculated by $\lambda_{(j)} = 2n\hat{\beta}_{(j)}^{(LS)}$

The procedure to calculate the LASSO estimates is identical as described in section 7.2 once the sequence of lambdas are calculated. Consider the orthogonal design d^* in chapter 6 with $m = 5$ main effects where the $n = 2^{5-1}$ runs using the identity relation, $A_1 + A_2 + A_3 + A_4 + A_5 = 1$.

d^*

Run	Treatment	$\underline{x}_1$	$\underline{x}_2$	$\underline{x}_3$	$\underline{x}_4$	$\underline{x}_5$	$\underline{y}$
1	1	1	-1	-1	-1	-1	6
2	2	-1	1	-1	-1	-1	16
3	3	-1	-1	1	-1	-1	34
4	4	-1	-1	-1	1	-1	9
5	5	-1	-1	-1	-1	1	8
6	123	1	1	1	-1	-1	44
7	124	1	1	-1	1	-1	21
8	125	1	1	-1	-1	1	15
9	134	1	-1	1	1	-1	50
10	135	1	-1	1	-1	1	30
11	145	1	-1	-1	1	1	10
12	234	-1	1	1	1	-1	60
13	235	-1	1	1	-1	1	45
14	245	-1	1	-1	1	1	22
15	345	-1	-1	1	1	1	52
16	12345	1	1	1	1	1	63

For this orthogonal design d^*, let the design matrix X_{d^*} have the columns for the general mean, 5 main effects, and 10 two-factor interactions for a total of $p = 16$ parameters. The $Rank(X_{d^*}) = p = n = 16$ and the least squares estimation is unique and given by $\frac{1}{16}X'\underline{y}$. To calculate the LASSO estimates, first calculate the least squares estimates, order the least squares estimates by magnitude, and calculate the shrinkage rate (shrinkage rate is constant since d^* is an orthogonal design). Table 7.7 presents the ordered least squares estimates is ascending order, the shrinkage rates for each estimate, and the corresponding $\lambda_{(j)}$ value to shrink the $\hat{\beta}_{(J)}^{(LS)}$ to 0 (intercept not included),

Table 7.7 Ordered least squares, shrinkage rates, lambda sequence values

(j)	j	$\hat{\boldsymbol{\beta}}_{(j)}^{(LS)}$	Shrinkage Rate	$\lambda_{(j)}$
(1)	2,4	0.0625	0.03125	2
(2)	2,5	0.0625	0.03125	2
(3)	1,3	0.1875	0.03125	6
(4)	3,5	0.1875	0.03125	6
(5)	5	0.3125	0.03125	10
(6)	2,3	0.3125	0.03125	10
(7)	4	0.4375	0.03125	14
(8)	3,4	0.4375	0.03125	14
(9)	1,4	0.5625	0.03125	18
(10)	1,5	0.5625	0.03125	18
(11)	4,5	0.6875	0.03125	22
(12)	1,2	3.4375	0.03125	110
(13)	3	5.4375	0.03125	174
(14)	1	5.5625	0.03125	178
(15)	2	16.938	0.03125	542.016

None of the LASSO estimates will attain 0 for selecting $\lambda = 1$, eight of the LASSO estimates will attain 0 for selecting $\lambda = 8$, all but one estimate will attain 0 for selecting $\lambda = 300$. Table 7.8 provides LASSO estimates for different values of λ (top 10 least squares estimates provided, intercept excluded by calculated in $\underline{\hat{z}}'\underline{\hat{\beta}}^{(L)}$),

Table 7.8 LASSO estimates for different values of $\boldsymbol{\lambda}$

(j)	j	$\hat{\boldsymbol{\beta}}_{(j)}^{(L)}$ $\lambda = 0$	$\hat{\boldsymbol{\beta}}_{(j)}^{(L)}$ $\lambda = 2$	$\hat{\boldsymbol{\beta}}_{(j)}^{(L)}$ $\lambda = 14$	$\hat{\boldsymbol{\beta}}_{(j)}^{(L)}$ $\lambda = 22$
(15)	2	16.938	16.875	16.5	16.25
(14)	1	5.5625	5.5	5.125	4.875
(13)	3	5.4375	5.375	5	4.75
(12)	1,2	3.4375	3.375	3	2.75
(11)	4,5	0.6875	0.625	0.25	0
(10)	1,4	0.5625	0.5	0.125	0
(9)	1,5	0.5625	0.5	0.125	0
(8)	4	0.4375	0.375	0	0
(7)	3,4	0.4375	0.375	0	0
(6)	5	0.3125	0.25	0	0
$\underline{\hat{z}}'\underline{\hat{\beta}}^{(L)}$		65.5	64.5	59.625	58.25

7.4 Model Selection Procedure

For ridge regression, the estimation step is straight forward but the model selection procedure requires attention. For LASSO, the estimation step requires attention but the model selection procedure is straight forward since LASSO estimates can be exactly 0. A model selection procedure that is analogous to the search design methodology for a given positive integer k,

1) Start out estimating a full model (general mean, linear terms, and all desired higher order terms)
2) For a given positive integer k, find a value of λ that returns k non-zero two-factor estimates. The value of λ is not unique but the selected model for that choice of k is unique.

Using this model selection procedure with LASSO, most of the emphasis on the choice of k. An alternative model selection approach is to specify a criterion function, such as a cross validation (Tibshirani and Robert, 1976), generalized cross validation (Jansen, 2014). or AIC/BIC (Hsu, et al., 2008), to select the value of λ instead of selecting a value of k. Zou proposes an alternative to LASSO, the adaptive LASSO, as another form of model selection using LASSO estimation (2006)

Chapter 8 compares the LASSO model selection procedure with the ridge regression model selection procedure (chapter 6) and the search design methodology.

Chapter 8

Model Selection Procedure Performances

8.1 Introduction

A total of 4 non-orthogonal designs (d_1, d_2, d_3, and d_4) and three model selection procedures (search design methodology, ridge regression, and LASSO), was discussed in earlier chapters. Each design has $m = 5$ main effects and $n = 12$ runs. The purpose of this chapter is to compare the performance of each model selection procedure across the four designs. A true model will be known to generate datasets where each design will use each model selection procedure to determine if the correct model was selected.

8.2 Simulation

For the search design methodology, design d_1 has full estimation capacity for positive integer $0 < k \leq 5$ and full model discrimination capabilities for $0 \leq k \leq 5$. Designs d_2, d_3, and d_4 has full estimation capacity for $0 < k \leq 3$ and full model discrimination capabilities for $0 \leq k \leq 1$. For $k = 2$, there are $\binom{m}{2}$ =While d_2, d_3, and d_4 do not have full model discrimination capabilities for $k = 2$, there are $\binom{5}{2} = 10$ two-factor interactions and $\binom{10}{2} = 45$ models. A careful choice of the true model with 2 non-zero two-factor interactions is still possible if the rank conditions are satisfied between the model comparisons involving the true model with the other 45 models. Section 2.4 illustrates a design $d_1^{(-i)}$ where the model discrimination rank conditions are not satisfied but there are still models M_u that satisfy the rank conditions for all other models $M_{u'}$. The procedure to compare the model selection procedure performances across the four designs,

1) Generate data from the true model with non-zero two-factor interaction terms $x_1 x_3$ and $x_2 x_3$,

$$\underline{y}_d = 3(\underline{1}) + 5\underline{x}_{1,d} + 10\underline{x}_{2,d} + 15\underline{x}_{3,d} + 4\underline{x}_{4,d} + 8\underline{x}_{5,d} + 7\underline{x_1 x}_{3,d} + 8\underline{x_2 x_{3,d}} + \underline{\epsilon}$$

$$\underline{\epsilon} \sim Normal(\underline{0}, \sigma^2 I) \;;\; \sigma = 0.5, 1, 5, 10, 15$$

2) For each design, perform the model selection procedure,

   a) Search design methodology – Find the model that minimizes the error sum of squares for $k = 2$ two-factor interactions
   b) Ridge regression – Find a value of λ that returns $k = 2$ two-factor interaction estimates above the threshold $t = 0.00001$ and the remaining 8 two-factor interaction estimates below or equal the threshold $t = 0.00001$.
   c) LASSO – Find a value of λ that returns $k = 2$ two-factor interaction estimates that are non-zero while the remaining 8 two-factor interaction estimates equal to 0.

3) For each design and model selection procedure, report back either 0 or 1 depending whether it found the correct two-factor interaction terms $x_1 x_3$ and $x_2 x_3$

4) Generate 10,000 datasets for each design, each value of σ, and perform the model selection procedure for each dataset to estimate the model selection procedure performance for each design.

Table 8.1 Search design methodology performance results

Design	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 5$	$\sigma = 10$	$\sigma = 15$
d_1	1	1	0.9136	0.3681	0.1595
d_2	1	1	0.912	0.3757	0.1661
d_3	1	1	0.832	0.2743	0.1134
d_4	1	1	0.8639	0.3335	0.1469

Table 8.2 Ridge regression performance results

Design	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 5$	$\sigma = 10$	$\sigma = 15$
d_1	1	1	0.9801	0.5892	0.2873
d_2	1	1	0.9909	0.6945	0.3896
d_3	0	0	0	0.0005	0.0038
d_4	0	0	0	0.0007	0.0032

Table 8.3 LASSO performance results

Design	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 5$	$\sigma = 10$	$\sigma = 15$
d_1	1	1	0.7364	0.3425	0.1589
d_2	1	1	0.9936	0.6796	0.3857
d_3	0	0	0	0.0068	0.0107
d_4	0	0	0.0021	0.0104	0.0175

8.3 Design and Method Comparisons

For the search design methodology, the results in table 8.1 indicates that d_1 had similar performance to d_2 across all levels of σ, d_4 performed better than d_3 but slightly worse than d_1 or d_2. For the ridge regression model selection procedure, design d_2 performed better than design d_1 while design d_3 and d_4 was not at all compatible with the model selection procedure, only performing better with higher levels of noise σ^2. For the LASSO model selection procedure, d_2 had better performance than d_1 and similar to ridge regression, d_3 and d_4 were not compatible with the model selection procedure.

The ridge regression model selection procedure had better performance than the search design methodology for both d_1 and d_2. LASSO model selection procedure had the same performance as the ridge regression model selection procedure for d_2 but the LASSO model selection procedure performed worse than the search design methodology for d_1. Table 8.4 summarizes the choice of model selection procedure for each design based on the simulation results.

Table 8.4 Recommended method for each design

Design	SDM	Ridge Regression	LASSO
$\boldsymbol{d_1}$		X	
$\boldsymbol{d_2}$		X	X
$\boldsymbol{d_3}$	X		
$\boldsymbol{d_4}$	X		

Design d_2 had the best performance for all three methodologies and would be recommended if the true model was the data generating model for the simulation.

Chapter 9

Conclusions

When searching over k two-factor interactions, model discrimination is the capability of a design d to separate two competing models within a class of v_k models. For each of the $\binom{v_k}{2}$ pairwise model comparisons $\{M_u, M_{u'}\}$, the value of the differenced error sum of squares is dependent on the observed values $\underline{y}_d$ and the difference of two projection matrices $H_d^{(u,u')}$.

It is desired for the true model to have the smallest error sum of squares, an incorrect model to have a larger error sum of squares, and a large distance between the two error sum of squares to maximize the success for model selection. By dividing the differenced error sum of squares with the sum of squares for the observations, the range of possible values is completely determined by the eigenvalues of the differenced projection matrix. For each model comparison $\{M_u, M_{u'}\}$, we examined the eigenvalues of the differenced projection matrices of multiple designs to determine which design is more capable of having a larger differenced error sum of squares for a general observation vector $\underline{y}_d$.

Using eigenvalue decomposition, we have calculated the expected difference error sum of squares as a function of the parameters $\underline{\beta}$. An example was given for any value of $\beta_{1,2} \neq 0$ and the remaining two-factor interactions are equal to zero, design d_1 will have higher expected difference for the error sum of squares for most of the model comparisons and have better model discrimination capabilities than design d_2. We have also observed for the two 12-run Plackett-Burman designs and $k = 1$, comparing the eigenvalues between the designs is sufficient to determine which of the two designs which will have better model selection performance capabilities. Ridge regression and LASSO as model selection procedures were also examined for the four non-orthogonal designs. It has shown there is a strong interaction effect between the model selection methodology and the design used. Ridge regression or LASSO had better performance on the two balanced designs but had low probability of success on the Plackett-Burman designs. The search design methodology had a good success rate for all designs.

Bibliography

[1] Akaike, Hirotugu. "A new look at the statistical model identification." Automatic Control, IEEE Transactions on 19.6 (1974): 716-723.

[2] Anderson Jr, William N., E. James Harner, and George E. Trapp. "Eigenvalues of the difference and product of projections." *Linear and Multilinear Algebra* 17.3-4 (1985): 295-299.

[3] Ghosh, Subir, Hongjie Deng, and Ying Luan. "Effects of noise in performance comparisons of designs for model identification and discrimination." Journal of Statistical Planning and Inference 137.12 (2007): 3871-3881.

[4] Ghosh, Subir, Hongjie Deng, and Ying Luan. "Error sum of squares and power comparisons for model search, identification and discrimination." Journal of statistical planning and inference 137.11 (2007): 3475-3482.

[5] Ghosh, Subir, and Lance Teschmacher. "Comparisons of search designs using search probabilities." Journal of statistical planning and inference 104.2 (2002): 439-458.

[6] Ghosh, Subir, and Ying Tian. "Optimum two level fractional factorial plans for model identification and discrimination." Journal of multivariate analysis 97.6 (2006): 1437-1450.

[7] Golub, Gene H., Michael Heath, and Grace Wahba. "Generalized cross-validation as a method for choosing a good ridge parameter." *Technometrics* 21.2 (1979): 215-223.

[8] Hinkelmann, Klaus, ed. *Design and analysis of experiments, special designs and applications*. Vol. 3. John Wiley & Sons, 2011.

[9] Hoerl, Arthur E., and Robert W. Kennard. "Ridge regression: Biased estimation for nonorthogonal problems." *Technometrics* 12.1 (1970): 55-67.

[10] Hsu, Nan-Jung, Hung-Lin Hung, and Ya-Mei Chang. "Subset selection for vector autoregressive processes using lasso." *Computational Statistics & Data Analysis* 52.7 (2008): 3645-3657.

[11] James, G., Witten, D., Hastie, T., and Tibshirani, R. *The elements of statistical learning*. Vol. 2. No. 1. New York: Springer, 2009. DOI: 10.1007/978-1-4614-7138-7

[12] Jansen, Maarten. "Generalized Cross Validation in variable selection with and without shrinkage." *Journal of Statistical Planning and Inference* (2014).

[13] Plackett, Robin L., and J. Peter Burman. "The design of optimum multifactorial experiments." *Biometrika* 33.4 (1946): 305-325.

[14] Smith, Gary, and Frank Campbell. "A critique of some ridge regression methods." *Journal of the American Statistical Association* 75.369 (1980): 74-81.

[15] Srivastava, J. N. "Designs for searching non-negligible effects." *A survey of statistical design and linear models* (1975): 507-519.

[16] Theobald, C. M. "Generalizations of mean square error applied to ridge regression." *Journal of the Royal Statistical Society. Series B (Methodological)*(1974): 103-106.

[17] Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* (1996): 267-288.

[18] Ye, Jianming. "On measuring and correcting the effects of data mining and model selection." *Journal of the American Statistical Association* 93.441 (1998): 120-131.

[19] Zhang, Cun-Hui, and Jian Huang. "The sparsity and bias of the lasso selection in high-dimensional linear regression." *The Annals of Statistics* (2008): 1567-1594.

[20] Zou, Hui. "The adaptive lasso and its oracle properties." *Journal of the American statistical association* 101.476 (2006): 1418-1429.

[21] Zou, Hui, Trevor Hastie, and Robert Tibshirani. "On the “degrees of freedom” of the lasso." *The Annals of Statistics* 35.5 (2007): 2173-2192.