2013

Regularities in the Augmentation of Fractional Factorial Designs

Lisa Kessel
Virginia Commonwealth University

Regularities in the Augmentation of Fractional Factorial Designs

Lisa Kessel

Thesis submitted to the Faculty of
Virginia Commonwealth University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Mathematical Sciences

Advisor: David Edwards, Ph.D.
Assistant Professor of Statistics
Department of Statistical Sciences and Operations Research

Virginia Commonwealth University
Richmond, Virginia
May 3, 2013

Copyright 2013, Lisa Kessel

Acknowledgement

I would like to specially acknowledge Dr. David Edwards for his willingness to serve as my committee chair. I am so thankful to him for the countless questions he answered, the numerous copies of this thesis he read and commented on, and the time he has dedicated to helping me get to this point.

I would also like to thank Dr. D'Arcy Mays and Dr. Wen Wan for their willingness to serve as committee members. They have been an invaluable component of the completion of this thesis.

Abstract

REGULARITIES IN THE AUGMENTATION OF FRACTIONAL FACTORIAL DESIGNS

By Lisa Kessel

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2013

Major Director: Dr. David Edwards, Assistant Professor, Department of Statistical Sciences and Operations Research

Two-level factorial experiments are widely used in experimental design because they are simple to construct and interpret while also being efficient. However, full factorial designs for many factors can quickly become inefficient, time consuming, or expensive and therefore fractional factorial designs are sometimes preferable since they provide information on effects of interest and can be performed in fewer experimental runs. The disadvantage of using these designs is that when using fewer experimental runs, information about effects of interest is sometimes lost. Although there are methods for selecting fractional designs so that the number of runs is minimized while the amount of information provided is maximized, sometimes the design must be augmented with a follow-up experiment to resolve ambiguities.

Using a fractional factorial design augmented with an optimal follow-up design allows for many factors to be studied using only a small number of additional experimental runs, compared to the full factorial design, without a loss in the amount of information that can be gained about the effects of interest. This thesis looks at discovering regularities in the number of follow-up runs that are needed to estimate all aliased effects in the model of interest for 4-, 5-, 6-, and 7-factor resolution III and IV fractional factorial experiments.

From this research it was determined that for all of the resolution IV designs, four or fewer (typically three) augmented runs would estimate all of the aliased effects in the model of interest. In comparison, all of the resolution III designs required seven or eight follow-up runs to estimate all of the aliased effects of interest. It was determined that D-optimal follow-up experiments were significantly better with respect to run size economy versus fold-over and semi-foldover designs for (i) resolution IV designs and (ii) designs with larger run sizes.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

When doing any experiment, one must first consider the balance that exists between maximizing the information that can be gained from the experiment and minimizing the cost and the resources that are required for its successful execution. Excelling in either of these areas often comes with a penalty to the other, as experimental runs can be expensive and time consuming yet cutting them results in a loss of information. When information is gained for all main effects and interactions (especially for large numbers of factors), the size of the experiment grows exponentially, which is often times not affordable or the time required to complete all of the required runs is not feasible. Therefore, a compromise must be made between these two goals of the experiment, and one solution is to run a fractional factorial design instead of the possible full factorial design.

Whereas the full factorial design looks at all possible combinations of levels for all of the factors in the experiment, the fractional factorial designs only looks at a subset of these possible factor combinations and chooses a combination of runs that allows for the estimation of some of the lower order effects (main effects and two-factor interactions). For example, a full factorial design for a 7-factor experiment with two levels for each factor would require 128 runs. This design would allow for the estimation of all main effects and interactions (through the 7-factor interaction). In contrast, a quarter fraction of the same experiment only requires 32 runs and all of the main effects and some of the two-factor interactions are estimable. Only some of the two-

factor interactions will be estimable since with this design two-factor interactions are fully confounded (aliased) with other two-factor interactions, and therefore only one two-factor interaction per alias chain will be estimable.

Since a typical assumption is that not all multi-factor interactions will be necessary in the model of interest, a fractional factorial can be a good alternative to the full factorial design if the full factorial will be expensive or time consuming to execute. Due to the loss of information that is caused by the aliasing of effects, a follow-up design can be utilized to de-alias effects that may be present in the model of interest. This allows for some of the information about effects of interest that was previously lost due to aliasing to be regained at a smaller cost than if this same information had been found using the full factorial design.

There are many methods for choosing an optimal and appropriate follow-up experiment that balances the cost of the additional runs with the information they provide that is not given by the original fractional factorial experiment. It is the goal of this thesis to determine regularities in the minimum number of follow-up runs that are needed to estimate all of the effects of interest for 4-, 5-, 6-, and 7-factor resolution III and IV fractional factorial experiments.

The remainder of this thesis will have the following organizational structure. Chapter 2 will contain an overview of factorial and fractional factorial designs, follow-up designs, and strategies for choosing follow-up experiments. In addition, it will include a thorough discussion of the literature that is currently available on these topics. Chapter 3 will outline a plan to analyze the initial experiments used for this study as well as discuss how the optimal number of follow-up runs that are needed to estimate the effects in a model of interest will be determined. Chapter 4 will discuss the results of this study as well as summarize the study and its findings.

Chapter 5 will present recommendations, conclusions, and future work that is recommended in this area of research.

Chapter 2

Literature Review

2.1: Two-Level Full Factorial and Fractional Factorial Designs

Two-level factorial designs are widely used in the design of experiments. They are designs in which all factors in the experiment are explored at two levels, typically a high level and a low level. For example, to study the effect of sunlight on plant growth using a two-level factorial design, the experiment could be designed so that the plant receives either 4 hours of sunlight per day (the low level) or 8 hours of sunlight per day (the high level).

These two-level experiments are presented as 2^k designs, where k designates the number of factors in the experiment and 2^k is equivalent to the number of runs in the experiment. These designs have k degrees of freedom to estimate main effects and n-k-1 degrees of freedom for estimating two-factor interactions and higher order effects. For example, a two-level experiment with 4 factors would be written as a 2^4 design with 16 runs ($2^4 = 2 \times 2 \times 2 \times 2$). This full factorial design would have four degrees of freedom for estimating the main effects and eleven degrees of freedom for estimating the two-factor and higher order interaction terms ($16 - 4 - 1 = 11$). Table 2.1 on the next page is an example of a 2^4 full factorial experiment with the high and low levels of the factors coded as +1 and -1, respectively. In this table it can be seen that all combinations of the high and low levels are given for all four factors.

Table 2.1: 2^4 Full Factorial Experiment

A	B	C	D
1	1	1	1
1	1	1	-1
1	1	-1	1
1	1	-1	-1
1	-1	1	1
1	-1	1	-1
1	-1	-1	1
1	-1	-1	-1
-1	1	1	1
-1	1	1	-1
-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 disadvantage of using a full factorial experiment is that it carries a high cost as the number of factors increases. For 4, 5, 6, and 7 factors, a two-level full factorial experiment requires 16, 32, 64, and 128 runs, respectively. This illustrates how it can quickly become inefficient, time consuming, and/or expensive to use these designs. Full factorials become even more inefficient when only the main effects and two-factor interactions are of interest. When an experimenter is not interested in all of the possible interactions that are estimable with the full factorial design, it becomes more advantageous to use an experimental design that is less taxing on time and resources that will still provide information for the lower order effects using fewer experimental runs.

In these situations, a good option is to utilize a fractional factorial design. These are orthogonal designs that allow experimenters to study the main effects and select interaction terms of interest in fewer experimental runs than a full factorial design (Antony 2003). Although more efficient, the reduced number of runs in the fractional factorial design is a result of aliasing

lower order terms with higher order interactions. This aliasing allows for the estimation of the lower order (and more likely significant) effects, assuming that the higher order effects are negligible.

Aliasing refers to the scenario when columns in the design matrix representing different effects are equivalent, creating linear dependencies among effects. Designs in which lower order effects are aliased with three-factor and higher order effects are preferred, since this allows estimability of the lower order (and more likely significant) effects. Consider the following example, where Table 2.2 below is the one-half fraction of the experiment given previously in Table 2.1.

Table 2.2: Half Fraction of 2^4 Experiment

A	B	C	D
1	1	1	1
1	1	-1	-1
1	-1	1	-1
1	-1	-1	1
-1	1	1	-1
-1	1	-1	1
-1	-1	1	1
-1	-1	-1	-1

In this design, factor D is equivalent to A*B*C, which means that in the one-half fractional factorial design, factor D is aliased with the three-factor interaction ABC. This is written as I = ABCD and is known as the defining relation. The defining relation is all of the columns in the design matrix that are equal to the identity column *I*, where *I* is always the column of all +1s in the design matrix. Each element of the defining relation is called a *word*, and defining relations with longer words are preferable. From this defining relation, all aliased effects can be determined by multiplying the left side of the equation by all factor combinations. For this example, all of the aliased effects are presented in Table 2.3 on the next page.

Table 2.3: Aliased Effects in Half Fraction Example

Effect	Alias
A	$A^2BCD = BCD$
B	$AB^2CD = ACD$
C	$ABC^2D = ABD$
D	$ABCD^2 = ABC$
AB	$A^2B^2CD = CD$
AC	$A^2BC^2D = BD$
AD	$A^2BCD^2 = BC$

To determine which terms to include in a model for a fractional factorial experiment, the experimenter operates under the principle of effect hierarchy. This principle states that the main effects are more important than two-factor interactions and two-factor interactions are more important than three-factor and higher order interactions. Since the main effects and two-factor interactions are more likely to be significant in the model, three-factor and higher order interactions can be considered negligible, which allows for lower order terms that are aliased with higher order terms to be estimated.

Two-level fractional factorial designs are written as 2^{k-p}, where k is the number of factors and $\left(\frac{1}{2}\right)^p$ is the fraction of the full factorial. For example, a quarter fraction $\left(\frac{1}{2}\right)^2$ of a design with 6 factors is written as 2^{6-2} and would have 16 runs ($2^{6-2} = 2^4$), compared to a full factorial with 6 factors which would have 64 runs ($2^k = 2^6$). This fractional factorial experiment would have 15 degrees of freedom, of which 6 would be dedicated to estimating main effects and the other 9 degrees of freedom could be used for estimating two-factor interactions.

Consider the possible aliasing structure of E = BCD and F = ACD to create a 2^{6-2} design from a full 2^6 design. These aliased effects are the generators for the 2^{6-2} fractional factorial design and it can be seen from this aliasing structure that the main effects E and F would be aliased with three-factor interactions, which can be considered negligible, and therefore would be

7

estimable. The defining relation for this design would be I = BCDE = ACDF = ABEF where the last word in the defining relation is the product of the first two words (BCDE*ACDF = ABC^{2}D^{2}EF = ABEF).

From the words in the defining relation, word length is used to determine the resolution of the design. The word length is determined by counting the number of letters in each word in the defining relation and setting the length of the shortest word as the resolution of the design. In the example above, using the 2^{6-2} design with the defining relation I = BCDE = ACDF = ABEF, all the words in the defining relation have four letters and therefore have a word length of four. This means that this 2^{6-2} design is a resolution IV design.

Designs with the largest number of clear main effects and two-factor interactions are considered to be better designs, where clear effects are defined as those that are either not aliased or aliased with three-factor or higher interactions. This implies that higher resolution designs are preferable since with these designs the lower order effects are aliased with higher order terms that are less likely to be significant in the model of interest. It also implies that for a given fractional factorial design, if two resolution designs are possible, the higher resolution design will be chosen.

Consider the following example for a 2^{6-2} fractional factorial design. For this design, two sets of possible generators are E = ABC, F = ABD and E = AB, F = ACD. For the first set of generators, the defining relation would be I = ABCE = ABDF = CDEF and for the second set of generators, the defining relation would be I = ABE = ACDF = BCDEF. From the first defining relation, the shortest word has length four, which would result in a resolution IV design. The shortest word in the second defining relation has length three, which would result in a resolution

III design. The first set of generators would be preferred to create the design since they would results in a higher resolution design.

The disadvantage of using fractional factorial designs is that there is a loss of information that accompanies reducing the full number of needed runs. Two-level fractional factorial experiments do not possess all of the possible factor combinations; hence there are no longer enough degrees of freedom to estimate all of the main effects and interaction terms that could possibly be significant in the model. Aliasing sacrifices some higher order terms in order for the lower order terms to be estimable. In resolution III designs, aliasing can confound both main effects and two-factor interactions with other two-factor interactions. In resolution IV designs, aliasing can confound two-factor interactions with other two-factor interactions. Only in resolution V and higher designs are all aliased main effects and two-factor interactions estimable, as they can only be confounded with three-factor and higher order interactions.

When main effects or two-factor interactions are aliased with other two-factor interactions, only one of the effects in each alias chain can be estimated without augmenting the design with more experimental runs. The decision of which effect to estimate often depends on the opinion of the experimenter and which effect they consider to be more important in the model of interest. For an example, consider the 2^{5-2} fractional factorial design that is given on the next page in Table 2.4. All three-factor and higher order interactions are omitted in the aliasing pattern.

Table 2.4: 2^{5-2} Fractional Factorial Experiment and Aliasing Pattern

2^{5-2} Design					Aliasing Pattern
A	**B**	**C**	**D**	**E**	**I = ABCD = BCE = ADE**
-1	-1	-1	-1	1	A = DE
-1	-1	1	1	-1	B = CE
-1	1	-1	1	-1	C = BE
-1	1	1	-1	1	D = AE
1	-1	-1	1	1	E = AD = BC
1	-1	1	-1	-1	AB = CD
1	1	-1	-1	-1	AC = BD
1	1	1	1	1	

The main effects and two-factor interactions that are aliased with other two-factor interactions in the resolution III 2^{5-2} fractional factorial design can be seen above in Table 2.4. Since aliasing results in design columns being shared by effects, without additional runs only one effect from each alias chain can be included in the model. This means that of the 15 possible main effects and two-factor interactions, only seven can be included in the model. Effect hierarchy suggests that the main effects are more significant than the two-factor interactions, but without prior information it is unknown which two-factor interaction(s) from each alias chain should be included in the model. It could also be possible that the two-factor interactions aliased with the main effects should be included in the model, or that both two-factor interactions in an alias chain should be included.

In situations such as those presented above when main effects and two-factor interactions are confounded with other two-factor interactions and the experimenter does not have knowledge that would allow them to choose between aliased effects, follow-up runs are used to allow for the aliased effects to be de-aliased so that all of the confounded terms of interest can be estimated.

For more information on the construction and analysis of full and fractional factorial designs, please reference Box and Hunter (1961).

2.2: Follow-Up Designs

     Follow-up designs are often needed when there is a need to resolve ambiguity that results from the use of a fractional factorial design (Meyer et al. 1996). Occasions often arise when there is not a clear picture of which effects are significant from the original experiment, and in those instances additional runs help clarify which of the aliased effects of interest are indeed significant. There are three methods for choosing follow-up designs that will be discussed in this thesis: fold-over designs for resolution III and IV fractional factorial experiments, semi-foldover designs for resolution IV fractional factorial experiments, and optimal follow-up designs.

2.2.1: Fold-Over Follow-Up Designs

     Fold-over designs can be used when either (i) all of the main effects or (ii) one main effect and all the two-factor interactions involving that main effect need to be de-aliased (Wu and Hamada 2009). The idea behind these designs is that in every 2^k full factorial experiment there exist 2^p-1 additional fractions from the same grouping as the 2^{k-p} fractional factorial experiment that was chosen. If an additional fraction from this group is added, it will break the alias chains in half and allow for one or more of the confounded effects to become estimable.

     There are several strategies for folding-over a fractional factorial. There are 2^k ways to generate a fold-over design, with various reasons behind the method of construction for these designs. If the analysis of the initial design reveals a set of main effects and two-factor interactions that are significant, a fold-over design should be chosen to resolve confounding issues with these effects (Li and Lin 2003). For resolution III designs, reversing the sign of all the factors increases the resolution of the experiment. For resolution IV designs, reversing one

factor allows for the estimation of the $k - 1$ two-factor interactions for that factor (Mee and Peralta 2000).

Consider the 2^{5-2} fractional factorial design previously presented in Table 2.4 in Section 2.1. A full fold-over on all factors of this design is given below in Table 2.5.

Table 2.5: Full Fold-Over of 2^{5-2} Design

	A	B	C	D	E	Aliasing Pattern (Initial)
Original Design	-1	-1	-1	-1	1	**I = ABCD = BCE = ADE**
	-1	-1	1	1	-1	A = DE
	-1	1	-1	1	-1	B = CE
	-1	1	1	-1	1	C = BE
	1	-1	-1	1	1	D = AE
	1	-1	1	-1	-1	E = AD = BC
	1	1	-1	-1	-1	AB = CD
	1	1	1	1	1	AC = BD
Fold-Over Design	1	1	1	1	-1	**Aliasing Pattern (Combined)**
	1	1	-1	-1	1	AD = BC
	1	-1	1	-1	1	AC = BD
	1	-1	-1	1	-1	AB = CD
	-1	1	1	-1	-1	
	-1	1	-1	1	1	
	-1	-1	1	1	1	
	-1	-1	-1	-1	-1	

After the design is augmented with the eight fold-over runs, only six effects remain aliased. All of the main effects have been de-aliased, which means that now those effects as well as the two-factor interactions they were aliased with are now estimable. The remaining aliased effects are the confounded two-factor interactions AB = CD and AC = BD as well as the two-factor interactions that were previously aliased with main effect E. In the original design, only seven effects could be included in the model, but after the fractional factorial is augmented with the fold-over runs, 12 effects can be included in the model. In addition, the original design was resolution III but after the eight fold-over runs are added, the new design is resolution IV.

12

The advantage of using these fold-over designs is that they de-alias lower order terms that are more likely to be active than higher order interactions. The full factorial design also allows for the estimation of the main effects and two-factor interactions; however, it dedicates many degrees of freedom to the estimation of higher order effects that are considered negligible under the hierarchy principle. The fold-over design allows for the experiment to be conducted in fewer runs yet still gain information about the main effects and two-factor interactions that the experimenter prefers to estimate.

Another advantage of this design is the simplicity of its construction. Since building this follow-up design only involves reversing the signs of the design column for the factor(s) on which the design is being folded-over, an experimenter can easily decide which levels of the factors to use for the follow-up design without complicated methods or the use of software.

The disadvantage of using fold-over designs is that not all of the alias chains can be broken with one fold-over. In addition, two-factor interactions cannot always be de-aliased with a fold-over design. For a resolution III design, folding over on all factors will increase the resolution, but Li and Mee (2002) found that this is often an inferior method of fold-over since it only estimates odd-length words from the defining relation. In addition, they also found that while this standard fold-over design for resolution III experiments will de-alias main effects from two-factor interactions, it will not resolve aliasing among two-factor interactions and it will only provide fewer than 2^{k-p} degrees of freedom for estimating two-factor interactions. For resolution IV designs, folding over on all factors is not applicable since the augmented design will have the same number of length four words (Li and Lin 2003). When folding over a resolution IV design, the fold-over will only add at most $k - l$ degrees of freedom for two-factor interactions (Mee and Peralta 2000).

2.2.1.1: Fold-Over Designs for Resolution III Fractions

Since a resolution III fractional factorial design confounds some of the main effects with two-factor interactions, it is common practice that to de-alias the confounded effects the original fractional factorial is augmented with another fractional factorial design in such a way that the final augmented design becomes a resolution IV design. To accomplish this, the original resolution III fractional factorial design is augmented with its mirror image, which is constructed by reversing the signs of the elements in the design matrix for all of the factors. By conducting another 2^{k-p} experiment using the same number of factors and factor levels, the combined experiments now form a $\frac{2}{2^p} = \frac{1}{2^{p-1}}$ fraction of the original 2^k full factorial experiment (Mee 2009). There are several benefits of augmenting a resolution III fractional factorial with its mirror-image fractional factorial:

1. All confounded main effects are de-aliased from two-factor interactions.

2. Two-factor interactions previously confounded with main effects are now estimable.

3. The precision of the estimates is increased. Assuming the error variance is unchanged, the standard error for coefficients decreases by a factor of $\left(\frac{1}{2}\right)^{1/2}$ (Mee 2009).

For most resolution III designs, the mirror-image fraction is the only design that will increase the resolution by reversing the sign of length three words in the defining relation. However, in instances when the number of factors *k* is larger than *N/2* where *N* is the number of experimental runs, there exist other follow-up designs that perform better than the mirror-image follow-up design (Mee and Peralta 2000).

14

2.2.1.2: Fold-Over and Semi-Foldover Designs for Resolution IV Fractions

Although foldover designs are simple to construct and analyze, these designs are degree-of-freedom inefficient for 2_{IV}^{k-p} factorial experiments. Mee and Peralta (2000) state that for these designs of size N, augmenting the design with a full foldover adds fewer than $N/2$ additional degrees of freedom for estimating two-factor interactions in the following cases:

1. Any design in which the foldover is obtained by reversing a single factor.

2. Any 2_{IV}^{k-p} design of size $N \leq 32$.

3. Any 2_{IV}^{k-p} design with only even-length words.

4. Any design with $k < 12$ for $N = 64$ or in general $\binom{k}{2} < N$.

In addition, the number of two-factor interactions that these designs typically estimate is less than half the size of the follow-up design (Mee and Peralta 2000). They are a good choice when the goal is increasing precision but an inefficient choice of follow-up design when the goal is estimation of effects of interest. For this reason, a better option to consider is semi-fold designs, which generally allow for the estimation of the same number of effects as the full fold-over design. Assuming three-factor interactions are negligible, resolution IV fractional factorial designs typically have only a few alias chains that include effects of interest. For that reason, it is often not necessary to add another fraction of the same size as the original fractional factorial experiment.

When a full fold-over would require more experimental runs than needed, a more practical method to gain information on confounded effects is to use semi-folding, which is augmenting the original experiment with a second fraction that is half the size of the original fraction. The two decisions that must be made for this follow-up design are (i) on which columns to reverse the sign and (ii) which half of the new fraction to use as the follow-up design.

Reversing a single factor is sufficient for designs with alias chains of length two or three, although for designs with alias chains of length four or more, more than one factor must be reversed.

Generally, the choice of sub-setting level of a factor depends on the preferred level for a highly significant effect. For even resolution IV designs, using any of the factors will result in the semi-fold fraction performing as well as the full fold-over would. The designs for which semi-folding is most appropriate are even designs with $k > (5/16)N$ where k is the number of parameters and N is the number of experimental runs. The resolution IV designs that were used in this study are summarized by these criteria below in Table 2.6.

Table 2.6: Resolution IV Designs

Design	k	(5/16)N
2^{4-1}	4	$(5/16)(8) = 2.5$
2^{6-2}	6	$(5/16)(16) = 5$
2^{7-2}	7	$(5/16)(32) = 10$
2^{7-3}	7	$(5/16)(16) = 5$

Although from Table 2.6 it can be seen that semi-folding would be a good option for the majority of these designs, it would add 4, 8, 16, and 8 runs, respectively, to the original fractional factorial experiments. It is the purpose of this thesis to determine if the number of follow-up runs could be reduced to less than that, especially since resolution IV designs have no aliased main effects and typically have a relatively low number of aliased two-factor interactions.

Semi-fold designs have the following advantages over D-optimal designs (which will be discussed in section 2.2.2.2):

1. They are simple to construct and don't require the use of software.

2. They produce an irregular $[\left(\frac{3}{2}\right) 2^{k-p}]$ design for which the analysis is well-understood.

16

3. They can be followed by the remaining *N/2* foldover to create a full foldover design if the semi-foldover is not sufficient (Mee and Peralta 2000).

Semi-folded resolution IV designs are recommended in cases where there are several aliased two-factor interactions and the significant effects need to be determined. When more precision is needed, it is better to use the full fold-over design.

2.2.2: Optimal Follow-Up Designs

After the initial fractional factorial experiment is completed, there are often many equivalent models that can explain the results. This ambiguity is a result of the decreased number of run sizes that is a characteristic of fractional factorial designs. When there are multiple equivalent models, adding follow-up designs can help to resolve some of this ambiguity. The goal of these follow-up designs, however, is not just gaining information on effects that were identified as significant in the original fractional factorial experiment, but also acquiring this information with as few experimental runs as possible, so as not to unduly tax the resources and time needed to complete the follow-up design. Optimal follow-up designs are a good choice of follow-up experiment because they allow an experimenter to fit the model of interest without drastically increasing the total run size (Goos and Jones 2011).

2.2.2.1: Optimal Follow-Up Methodology

An optimal follow-up design starts with a model that includes significant effects from the original experiment. This model also includes any effects that are aliased with any of the significant effects but are also determined to be potentially significant in the model of interest. Since the augmented optimal runs are done after the original experiment, a block term is also included in the model to account for any significant differences in the average response between the two sets of experimental runs. After this model is chosen, factor settings can be chosen to optimize a specified design criterion. The design criterion provides a way to measure the performance of the combined factor settings for the given model. For these reasons, the results achieved with these follow-up design methods are very much dependent on the model that is chosen based on the initial experiment as well as the optimal design criterion that is selected to determine the augmented experimental runs (Wu and Hamada 2009).

There are several different design criteria that can be used to determine the factor settings for the optimal follow-up design. They are presented below in Table 2.7.

Table 2.7: Optimal Design Criteria

Design	Criteria
A – Optimality	Minimizes the trace of the inverse of the information matrix X'X. This minimizes the average variance of the coefficient estimates.
C – Optimality	Minimizes the variance of the best linear unbiased estimator (BLUE) of a predetermined linear combination of model parameters.
D – Optimality	Minimizes $|(X'X)^{-1}|$, which is the determinant of the inverse of the information matrix. This maximizes the differential Shannon information content of the parameter estimates.
E – Optimality	Maximizes the minimum eigenvalue of the information matrix.
T – Optimality	Maximizes the trace of the information matrix.
G – Optimality	Minimizes the maximum entry in the diagonal of the matrix $X(X'X)^{-1}X'$. This minimizes the maximum variance of the predicted values.
I – Optimality	Minimizes the average prediction variance over the design space.
V – Optimality	Minimizes the average prediction variance over a set of m specific points.

From all of the possible optimal criteria, D-optimality will be the focus of this section as this optimality criterion was used to construct the follow-up runs in the analyses done for this thesis. For more information on the other optimal design criteria, please reference Abd El-Monsef and Seyam (2011), Galil and Kiefer (1977), and Mandal et al. (2012).

2.2.2.2: D-Optimal Follow-Up Designs

The D-optimal design criterion is the most important and popular design criterion for several reasons. Firstly, an experimenter can apply the criterion and check the optimality of a design much more easily with the D-optimal criterion than with any of the other optimal follow-up techniques, as many statistical software packages have the capacity to construct these follow-up experiments. Secondly, this optimality criterion is the most effective in optimizing the parameter efficiency and model robustness via minimization of the variance-covariance matrix. Thirdly, this design provides reliable results since it puts an emphasis on the quality of the parameter estimates (Abd El-Monsef and Seyam 2011).

The purpose of the additional D-optimal experimental runs is to resolve ambiguities that exist between effects in the model of interest. In addition to adding sufficient runs to allow for the estimation of all the effects in this model, a block effect needs to be added to the model. If the average response has shifted between the original fractional factorial experiment and the follow-up experiment, this block effect will capture that change in the mean response.

A D-optimal follow-up experiment is a design in which the augmented runs minimize the determinant of the variance-covariance matrix. Thus, it can be said that these designs minimize the determinant of $(X'X)^{-1}$, where X is the model matrix. If the determinant is positive, this

suggests that the design contains sufficient information to estimate all of the effects in the model

(Goos and Jones 2011).

The interpretation of the D-optimality criterion is easily understood in terms of the confidence region. The confidence region for a set of parameters can be derived from the variance-covariance matrix much like a confidence interval is calculated for a single parameter in the model of interest. For a fixed number of treatments, the volume of this confidence region is proportional to the determinant of the variance-covariance matrix. Thus, the smaller the determinant of the variance-covariance matrix, the smaller the volume of the confidence region and the smaller the error associated with our parameter estimates (Mead et al. 2012).

To minimize this confidence region, the determinant of the variance-covariance matrix, which is also known as the generalized variance, must also be minimized (Mead et al. 2012). For $y = \alpha + \beta x$, the least squares matrix is

$$X'X = \begin{bmatrix} N & \sum x \\ \sum x & \sum x^2 \end{bmatrix}$$

and the generalized variance is

$$G = \sigma^2 / \{N \sum x^2 - (\sum x)^2\}$$

Minimizing G is equivalent to minimizing the variance of $\hat{\beta}$, which can be written as

$$VAR(\hat{\beta}) = \sigma^2 / \sum_i (x_i - \bar{x})^2$$

From this it follows that by using D-optimal follow-up experiments, the error that is associated with the parameter estimates in our model of interest is minimized.

To construct the follow-up design, the first step is to create the model matrix that corresponds to the new design, which has the same number of rows as the model matrix for the

original fractional factorial design but instead of only having columns for the main effects, also includes columns for the effects of interest and the block effect that need to be estimated in the fitted model. The next step in the augmentation process is to add one row to the model matrix for each additional experimental run and find values for the levels of each factor that will maximize the determinant of X'X for the augmented design.

Let X_1 designate the model matrix for the original experiment and X_2 designate the model matrix for the follow-up experiment. Then the model matrix for the augmented design can be expressed as:

$$X'X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}' \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} X_1' & X_2' \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = X_1'X_1 + X_2'X_2$$

It can be seen from the equation above that the information matrix for the augmented design is simply the sum of the information matrix for the original fractional factorial design and the information matrix for the D-optimal follow-up experiment. The values chosen for the levels of the factors must therefore maximize the following determinant of the information matrix across all levels of X_2:

$$|X'X| = |X_1'X_1 + X_2'X_2|$$

A coordinate exchange algorithm is used to find the values for the levels of the factors that will maximize the determinant of the information matrix for the augmented design. This algorithm starts by generating random variables on the interval [-1, 1] for each of the elements in the X_2 model matrix. For n follow-up runs in k factors, the coordinate exchange algorithm will need to calculate nk elements in the overall model matrix. The algorithm goes through each element individually, repeatedly selecting a value from the interval [-1,1] that optimizes the D-

optimality criterion. If the D-optimality criterion improves with a successive selected value, the algorithm exchanges the previous value that was used for the new value. This process continues until no more coordinates need exchanging in place of a value that better maximizes the determinant of the augmented design model matrix (Goos and Jones 2011). For this thesis, JMP was used to compute all of the D-optimal follow-up designs in the analyses.

Chapter 3

Design and Research Methods

The purpose of this thesis is to use optimal follow-up experiments to augment fractional factorial designs to discover regularities that exist in the number of follow-up runs that are needed to estimate all aliased effects in the model of interest for 4-, 5-, 6-, and 7-factor resolution III and IV fractional factorial experiments. If regularities can be identified, this thesis will provide insight to researchers as to when an optimal follow-up design could be a better alternative to other follow-up methods, such as fold-over and semi-foldover designs, for estimating aliased effects in the fractional factorial experiments presented in this thesis.

3.1: Full Factorial Experiments

The data for this thesis involved a collection of 26 published engineering experiments that included 59 responses, referenced in the paper *Regularities in Data from Factorial Experiments* by Li et al. (2006). This paper conducted an analysis of 113 published factorial experiments to quantify regularities among main effects and multi-factor interactions. In addition to observing the regularities of effect sparsity, hierarchy, and heredity, their analyses suggested the existence of a fourth regularity, titled "Asymmetric Synergistic Interaction Structure", which determined the degree to which the sign of main effects determined the likely sign of interactions.

The data for this thesis consisted of 4-, 5-, 6-, and 7-factor two-level full factorial experiments that came from a variety of disciplines, including biology, chemistry, materials science, mechanical engineering, and manufacturing. There were 22 four-factor experiments with a total of 51 responses, 2 five-factor experiments with a total of five responses, 1 six-factor experiment with two responses, and 1 seven-factor experiment with one response.

All of the data sets had coded explanatory variables so that the low level of the factor was represented as -1 and the high level was represented as +1. The coding of the variables as ± 1 allows for the comparison of the variables on an even scale, so that the magnitude of the effect is not affected by the scale on which the variable is measured. All of the experiments had continuous response variables.

Consider Table 3.1 on the next page, which gives a 2^4 full factorial design that was analyzed for this thesis. This table begins an example that will continue through section 3.4 to demonstrate the methodology that was used for the analyses completed for this thesis. Note that all sixteen possible factor combinations are represented in this design.

Table 3.1: 2^4 Full Factorial Experiment

A	B	C	D	Response
1	1	-1	-1	3.55
1	1	1	-1	5.52
1	-1	-1	-1	2.17
1	-1	1	-1	3.94
-1	1	-1	-1	5.91
-1	1	1	-1	8.67
-1	-1	-1	-1	3.94
-1	-1	1	-1	7.29
1	1	-1	1	2.37
1	1	1	1	3.94
1	-1	-1	1	1.97
1	-1	1	1	2.76
-1	1	-1	1	2.76
-1	1	1	1	5.72
-1	-1	-1	1	2.76
-1	-1	1	1	3.94

3.2: Fractional Factorial Construction

The 59 original full factorial experiments were used to construct 203 fractional factorial experiments. Table 3.2 below shows how many fractional factorial designs were analyzed for each number of factors, as well as which fractions were analyzed.

Table 3.2: Fractional Factorial Analyses

# of Factors	Fraction				
	1/2	1/4	1/8	1/16	Total
4	51	-	-	-	51
5	-	40	-	-	40
6	-	16	32	-	48
7	-	16	16	32	64
Total	51	72	48	32	203

In Table 3.2, there are nine factor sub-setting levels (fractions) that are missing from the analyses. The one-half fraction for 5-, 6-, and 7-factor designs was not included in the analyses as these would be resolution V or higher designs, which would not require follow-up runs for the estimation of the main effects or two-factor interactions. In the cases of the other six excluded fractions, these designs would have been too small for the number of factors that were being considered. For example, a one-quarter fraction of a 2^4 design would not even allow for estimation of the main effects, and would never be a design chosen in practice. For these six fractions, all would have had only four, two, or one runs, and would be inefficient and impractical designs for fractional factorial experiments.

For these analyses, only resolution III and IV designs were considered. Resolution III designs have no main effects confounded with other main effects, but main effects and two-factor interactions can be aliased with other two-factor interactions. Resolution IV designs have no main effects confounded with other main effects or two-factor interactions, but two-factor interactions are aliased with each other (Antony 2003).

The purpose behind this was that higher resolution designs would have significant degrees of freedom to estimate all main effects and two-factor interactions, and since the purpose of this thesis was to determine how many follow-up runs are needed to estimate these effects when they are aliased with other two-factor interactions, higher resolution designs need not be considered. The generators that were used to construct the fractional factorials came from three sources:

1. JMP 10 Software

2. *A Catalogue of Two-Level and Three-Level Fractional Factorial Designs with Small Runs* by Chen et al. (1993)

3. *Experiments: Planning, Analysis, and Optimization* by Wu and Hamada (2009)

The design generators that were used for constructing the fractional factorial experiments are presented below in Table 3.3.

Table 3.3: Design Generators

Design	Generators			
2^{4-1}	D = ABC			
2^{5-2}	D = ± ABC	E = ± BC		
	D = ± AB	E = ± AC		
2^{6-2}	E = ± BCD	F = ± ACD		
	E = ± ABC	F = ± ABD		
2^{6-3}	D = ± ABC	E = ± BC	F = ± AC	
	D = ± AB	E = ± AC	F = ± BC	
2^{7-2}	F = ± CDE	G = ± ABDE		
	F = ± ABC	G = ± ABDE		
	F = ± ABC	G = ± ADE		
	F = ± ABC	G = ± ABD		
2^{7-3}	E = ± BCD	F = ± ACD	G = ± ABD	
	E = ± ABC	F = ± ABD	G = ± ACD	
2^{7-4}	D = ± ABC	E = ± BC	F = ± AC	G = ± AB
	D = ± AB	E = ± AC	F = ± BC	G = ± ABC

For any design generator, there exist 2^p-1 other generator options for selecting the fractional factorial experiment. Therefore, in addition to the generators recommended by the three given sources, the negative generators were also used in order to create more possible fractional factorial experiments. For example, for the 2^{5-2} design, JMP recommended D = ABC and E = BC as design generators. This led to also using the following generators:

1. D = ABC and E = -BC

2. D = -ABC and E = BC

3. D = -ABC and E = -BC

Thus it is more appropriate to say that instead of just using D = ABC and E = BC as generators, D = ± ABC and E = ± BC were used as generators for this design, which is how the

generators are presented in Table 3.3. This method was used only for the 5-, 6-, and 7-factor experiments since there were only 5, 2, and 1 full factorial experiments at these levels, respectively. There were 51 four-factor experiments in the original data set, which was deemed a sufficient number of analyses using only one generator per experiment.

Additionally, for each fraction in each number of factors, only generators that would allow for the best resolution design were used. For example, if a certain fractional factorial design had a generator that would produce a resolution IV design, generators that would produce a resolution III design were not considered.

To continue with the example from the analysis that was presented in Table 3.1, the resolution IV one-half fractional factorial design that was extracted from that original data set is given below in Table 3.4. The generator that was used for this 2^{4-1} fractional factorial design is $D = ABC$, which results in the following aliasing pattern: $AB = CD$, $AC = BD$, and $AD = BC$.

Table 3.4: 2^{4-1} Fractional Factorial Experiment

A	B	C	D	Response
1	-1	-1	1	1.97
-1	1	-1	1	2.76
-1	-1	-1	-1	3.94
-1	-1	1	1	3.94
-1	1	1	-1	8.67
1	-1	1	-1	3.94
1	1	1	1	3.94
1	1	-1	-1	3.55

3.3: Model of Interest Determination

      Once all of the fractional factorial designs and responses from the original full factorial experiments had been extracted, the effects in the model of interest needed to be determined. This was done by specifying a model that included all of the main effects and two-factor interactions as candidate terms in JMP. All possible subsets regression was then utilized to fit all of the models that could be built using the specified main effects and two-factor interactions. These models were built under the principle of effect heredity, which implies that interactions are only significant if one of the parent (included) factors is significant. This significantly reduced the number of models that could be built since models were only included in the all subsets regression if at least one of the parent factors for each of the interaction terms in the model was also included in the model. Additionally, this principle is advantageous to use in situations where aliasing is present as it enables experimenters to more easily identify likely two-factor interactions (Li et al. 2006).

      The model of interest for a follow-up experiment was considered to be the union of the models from the all-subsets regression that best fit certain model selection criteria. This was done under the assumption that including all of the terms from a set of equivalent models would reveal the best of many competing models. The model selection criteria that were used to rank the possible best models were the Coefficient of Determination (R^2), Root Mean Square Error (RMSE), corrected Akaike Information Criterion (AIC_C), and Bayesian Information Criterion (BIC). The candidate models that best fit these criterions were considered to be the models that maximized R^2 while minimizing RMSE, AIC_C, and BIC.

The Akaike Information Criterion (AIC) is one of the most popular strategies used for model selection. However, Shibata (1980) showed that while AIC is efficient, it is not consistent. Many researchers have also found that for small samples, AIC tends toward over fitting the model. To improve AIC, Sugiura (1978) and Hurvich and Tsai (1989) found AIC_C by estimating the Kullbach-Leibler distance directly in regression models (Rao and Wu 2001). For this study, corrected AIC (AIC_C) was used since it is AIC with a correction for finite sample size. The formulas for AIC and AIC_C are given below:

$$AIC(\theta) = -2\log(\mathcal{L}) + 2k$$

and

$$AIC_C(\theta) = AIC + \frac{2k(k+1)}{n-k-1}$$

where *n* is the sample size and *k* is the number of independent parameters in the model.

From this equation it can be seen that AIC is a penalized log-likelihood criterion that offers a balance between good fit (which would be seen with high values of the log-likelihood) and complexity (more complex models will have a higher penalty than simple models). A model that minimizes AIC_C will also be the model that minimizes information loss (Claeskens and Hjort 2008).

The reasoning behind using AIC_C instead of just using the maximum likelihood to select the best model is that more complex models need to be penalized. Under the effect sparsity principle, the number of active effects in a factorial experiment is relatively small compared to the number of possible model effects. Using maximum likelihood would result in choosing the model with the most parameters. However, AIC_C punishes models for being too complex (having too many parameters) and therefore allows for simpler models to be considered that still do an equally sufficient job of fitting the data (Li et al. 2006).

The Bayesian Information Criterion (BIC) is another technique very similar to AIC that is used for model selection. The equation for BIC is given below.

$$BIC(\theta) = -2\log(\mathcal{L}) + k\log(n)$$

where *n* is the sample size and *k* is the number of independent parameters in the model.

It can be seen that BIC is very similar to AIC in that it balances good fit with model complexity. Like AIC, there is still a term that penalizes a model for having more parameters. The "best" simple model with good fit will have a large maximum likelihood and a smaller penalty term, which will minimize both BIC and AIC. It can also be seen from this equation that in cases where $n \geq 8$, the BIC imposes a harsher penalty for model complexity than AIC does.

In addition to AIC and BIC, R^2 and RMSE were used to determine the best models. RMSE measures the differences between the model's predicted and actual values so that it can be determined how good a model's predictive capabilities are. Small RMSE indicates that the model is accurately predicting new responses. R^2 is also a measure of the predictive capabilities of a model. It can be used to determine how much of the variation in the response can be explained by the regression model. Higher values of R^2 indicate better predictive capabilities and therefore a better model. However, the disadvantage of R^2 is that it carries no penalty for model complexity and thus will increase until the model is saturated. Both RMSE and R^2 were used in determining the best model as they are some of the most common goodness-of-fit measures used in model selection.

Table 3.5 on the next page is a continuation of the analysis example in Table 3.4, with Table 3.5 showing the "best" candidate models from the all subsets regression. Note that the four candidate models are equivalent in terms of the model selection criteria R^2, RMSE, AIC_C, and BIC.

Table 3.5: Results of All Subsets Regression

Model	# of Terms	R^2	RMSE	AIC_C	BIC
X1, X2, X3, X4, X1*X3, X1*X4	6	0.9824	0.6965	-127.72	16.9160
X1, X2, X3, X4, X1*X3, X2*X3	6	0.9824	0.6965	-127.72	16.9160
X1, X2, X3, X4, X1*X4, X2*X4	6	0.9824	0.6965	-127.72	16.9160
X1, X2, X3, X4, X2*X3, X2*X4	6	0.9824	0.6965	-127.72	16.9160

From the all subsets regression, the "best" candidate models are unioned together to determine the model of interest for the follow-up experiment. This is done under the assumption that including all of the terms from a set of equivalent models will reveal the best of many competing models. In the example given in Table 3.5, four models stood out as "best" models in terms of R^2, RMSE, AIC_C, and BIC.

However, the unioned model of interest in this example would include the terms X1, X2, X3, X4, X1*X3, X1*X4, X2*X3, and X2*X4. Since X1*X3 is confounded with X2*X4 and X1*X4 is confounded with X2*X3, without additional follow-up runs the model of interest cannot be fit while including all of these aliased terms. Using only the original fractional factorial experiment, there are seven degrees of freedom to estimate main effects and two-factor interactions. However, the model of interest includes four main effects and four two-factor interactions, which would require at least eight degrees of freedom to estimate. Without adding follow-up runs, the fitted model can only include one effect from each alias chain. Since there are four aliased effects (two alias chains) in the model of interest that need to be estimated and none of them can be considered negligible, this example clearly demonstrates a scenario when there is a need for follow-up runs to collect important information that was lost due to aliasing.

3.4: Follow-Up Experiments

After determining and unioning the best candidate models, a follow-up experiment was constructed in such a way that the aliased terms in the model of interest could be estimated. This was done by using JMP 10 software to augment the original fractional factorial designs with D-optimal follow-up experiments.

The number of required follow-up runs was considered to be the minimum number of experimental runs that would eliminate all linear dependencies among factors as well as allow for the estimation of a block effect in the model of interest. In determining the number of runs for the follow-up experiment, it was permitted for a saturated model to be fit, since the purpose of this thesis was maximizing the estimability of effects, not the precision of the estimates.

After the fractional factorial experiment was augmented with the D-optimal follow-up design, linear regression was used to fit the model of interest to ensure that all aliased terms had been de-aliased and that all effects, including the block effect, were estimable. These linear regression analyses were performed under the normal assumptions of linear regression, including:

1. X_i are nonrandom and observed without error

2. $E(\varepsilon_i) = 0$

3. $Var(\varepsilon_i) = \sigma^2$ for all i

4. ε_i are uncorrelated

5. Y_i are normally distributed

Table 3.6 below shows a 2^{4-1} fractional factorial experiment augmented with a D-optimal follow-up experiment as a continuation of the example given in Table 3.5 in section 3.3. Note that the last three rows of Table 3.6 show the D-optimal follow-up design.

Table 3.6: 2^{4-1} Fractional Factorial Augmented with D-Optimal Follow-Up Design

A	B	C	D	Block	Response
1	-1	-1	1	1	1.97
-1	1	-1	1	1	2.76
-1	-1	-1	-1	1	3.94
-1	-1	1	1	1	3.94
-1	1	1	-1	1	8.67
1	-1	1	-1	1	3.94
1	1	1	1	1	3.94
1	1	-1	-1	1	3.55
-1	1	-1	-1	2	5.91
-1	-1	1	-1	2	7.29
-1	1	1	1	2	5.72

Linear regression showed that the additional runs effectively de-aliased the two alias chains of X1*X3 = X2*X4 and X1*X4 = X2*X3. All main effects and two-factor interactions were estimable, as well as the included block effect. The aliased terms could have also been completely de-aliased with a full fold-over or a semi-foldover of any one of the factors. However, a full fold-over would require eight follow-up runs and a semi-foldover would require four follow-up runs, whereas this D-optimal follow-up design allowed for the estimation of the same effects of interest but only required three additional runs. Thus, this example clearly demonstrates that there are situations where an optimal follow-up design can be an efficient method for gaining information about aliased effects of interest without significantly increasing the total run size of the experiment.

Chapter 4

Results

The purpose of this thesis was to use optimal follow-up experiments to augment fractional factorial designs to discover regularities that existed in the number of follow-up runs that were required to estimate all of the aliased effects in the model of interest for 4-, 5-, 6-, and 7-factor resolution III and IV fractional factorial experiments. The following section outlines the results for the 203 analyses that were completed for this thesis, identifies the significant regularities that were discovered through the analyses, and demonstrates how the D-optimal follow-up designs performed compared to other follow-up design methods such as fold-over and semi-foldover.

The first significant result was that all 203 analyses resulted in fitted models with all confounding resolved and estimates for all effects, including the block effect. Only three of these analyses required a follow-up design that was larger than the initial fractional factorial design. In all three of these cases, the initial design was a 2^{5-2} design with eight runs where the model of interest was the saturated model that included all 15 possible main effects and two-factor interactions. All of these designs had eight alias chains to de-alias as well as a block effect to estimate, which required a follow-up design of nine experimental runs.

Excluding these three analyses, the other 200 analyses all required no more than eight follow-up runs to de-alias all of the effects of interest from the original experiment and to

estimate a block effect. This suggests that D-optimal follow-up runs have better run size economy on average than fold-over designs, while guaranteeing the estimability of the effects of interest.

The second significant result from the analyses provided insight on determining the number of required follow-up runs for resolution III and IV designs. It was expected that resolution IV designs would require fewer follow-up runs than resolution III designs since these designs have no main effects aliased with two-factor interactions and often have relatively few aliased two-factor interactions. Figure 4.1 below presents the number of runs required of a follow-up experiment to estimate the model of interest as categorized by factor size and resolution. It is significant to note that fewer follow-up runs were required, on average, for the resolution IV designs than for the resolution III augmented fractional factorial designs.
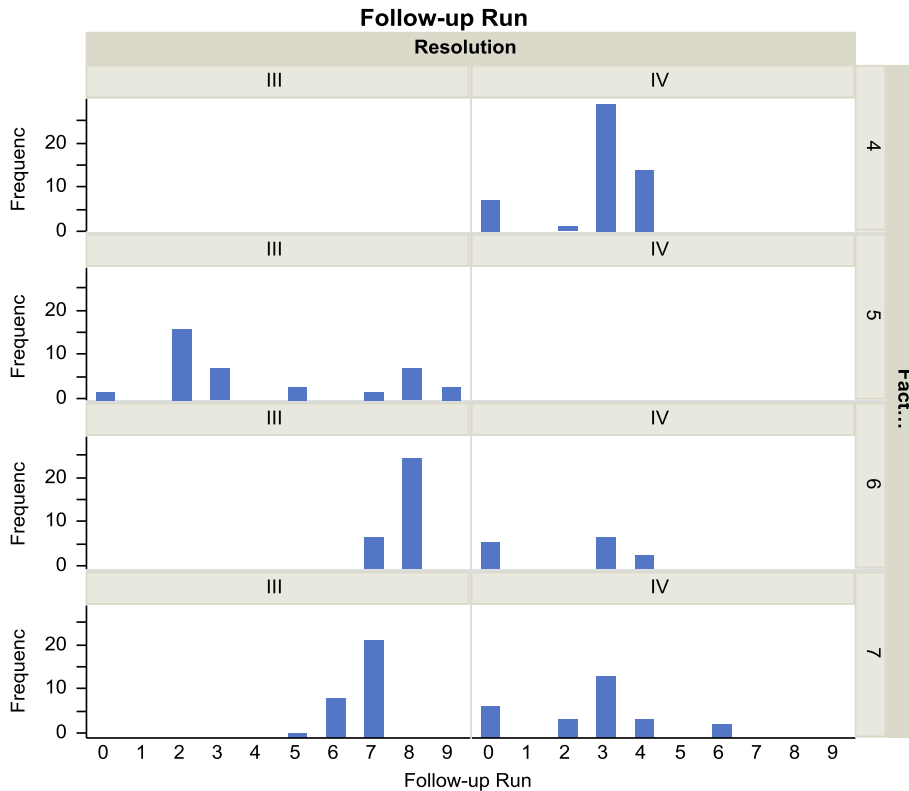

Figure 4.1: Follow-Up Runs by Resolution and Number of Factors

From Figure 4.1, it can clearly be seen that typically the resolution IV designs require fewer follow-up runs to de-alias all of the effects than resolution III designs. For resolution IV designs, the 4- and 6-factor experiments required at most four follow-up runs, with the 4-factor experiments typically requiring only three follow-up experimental runs. For the 7-factor resolution IV designs, typically three follow-up runs are needed to resolve any issues with confounding, but up to six needed runs were seen in these analyses. Comparatively, the 5-factor resolution III experiments typically required either two or eight follow up runs, with three of the experiments requiring nine follow-up runs, which were discussed previously. The 6- and 7-factor resolution III experiments typically required seven or eight and seven follow-up runs, respectively.

The model of interest was the cause of the discrepancy in the number of needed runs for the 5-factor experiments. Of the forty 5-factor experiments that were analyzed, 15 experiments had eleven or more terms in the model of interest, which required five or more follow-up runs to resolve all of the confounding issues. Twenty-five of the experiments had eight or fewer terms in the model, which required three or fewer follow-up runs to de-alias all confounded effects.

Thus it can be seen that the resolution IV designs primarily required fewer follow-up runs than the resolution III designs, which was the expected result since resolution IV designs have no main effects aliased with two-factor interactions and relatively low confounding among two-factor interactions. There were no situations where the same number of follow-up runs could be used to de-alias all effects of interest in both the resolution III and IV designs for the same factor except for the 7-factor experiments, where the upper bound of six required follow-up runs for the resolution IV design is greater than the lower bound of five follow-up runs for the resolution III

design. However, the 7-factor resolution IV design typically only needs three follow up runs, whereas the resolution III design typically requires seven follow-up runs.

Conclusions can also be made from Figure 4.1 about the number of required follow-up runs for each of the different fractional factorial designs that were analyzed. Additionally, consider Table 4.1 below, which shows the fractional factorial designs that are represented for each resolution and factor size.

Table 4.1: Fractional Factorial Designs by Resolution and Number of Factors

Resolution	# of Factors	Fraction	Design	# of Runs
III	4	-	-	-
	5	$1/4$	2^{5-2}	8
	6	$1/8$	2^{6-3}	8
	7	$1/16$	2^{7-4}	8
IV	4	$1/2$	2^{4-1}	8
	5	-	-	-
	6	$1/4$	2^{6-2}	16
	7	$1/8$	2^{7-3}	16
	7	$1/4$	2^{7-2}	32

From Table 4.1, it can be seen that for the graphs presented in Figure 4.1, all but the graph representing the 7-factor designs had only one fraction per factor. This allows for Figure 4.1 to also be used for identifying regularities about the number of follow-up runs needed for each fractional factorial design.

For the resolution III 2^{5-2} designs, the number of follow-up runs needed in these analyses heavily depended on the number of aliased terms in the model, which resulted in a bi-modal distribution of required follow-up runs. Thus, these designs will typically require two or eight

follow-up runs. At best, they perform better than a semi-foldover or full fold-over design, and at their worst they perform equivalently with a full fold-over design. Depending on the aliasing structure, they may even perform better than a full fold-over since the fold-over design will not necessarily resolve confounding among all two-factor interactions.

For the resolution III 2^{6-3} and 2^{7-4} designs, the number of required follow-up runs was at most 8 and 7, respectively. Both of these designs performed equally well or better than a full fold-over design in terms of run size economy and estimability of confounded effects of interest. A full fold-over would require eight follow-up runs, and may not resolve aliasing among two-factor interactions included in the model. While the 2^{6-3} design typically needed eight follow-up runs, all of the analyses for this design completely resolved the confounding between all of the effects of interest. The 2^{7-4} design performed better than a full fold-over would in terms of run size economy and estimability of aliased effects as it completely resolved all of the confounding issues with only seven follow-up experimental runs. If there was a semi-foldover design that could resolve the aliasing for these designs, it would be the preferable design as it would only require four follow-up runs instead of the seven or eight required by these two designs. However, the only design for which a single semi-foldover design could potentially resolve all of the confounding issues among the effects of interest would be for a 2^{7-4} fractional factorial design.

For the resolution IV 2^{4-1} and 2^{6-2} experiments, Figure 4.1 shows that both of these designs would require no more than four follow-up runs to completely resolve any confounding issues, with the 2^{4-1} design typically needing only three additional runs. The 2^{4-1} design would require four runs for a semi-foldover design and eight additional runs for a full fold-over design. Thus, the optimal follow-up design is much more efficient for this design in terms of run size

economy than the other two methods discussed. The 2^{6-2} design would require eight runs for a semi-foldover design or sixteen additional runs for a fold-over design, so with a maximum of four additional runs to completely resolve confounding issues, this is a significant improvement in run size economy over both the fold-over and semi-foldover methods.

For the graph of the resolution IV 7-factor experiments in Figure 4.1, the represented designs are the 2^{7-2} and 2^{7-3} fractional factorial designs. This graph shows that typically three follow-up runs are needed to resolve any confounding issues, but as many as six additional runs could be needed. For full fold-over designs, the 2^{7-2} experiments would require 32 additional runs and the 2^{7-3} experiments would require 16 additional runs. For semi-foldover designs, the 2^{7-2} designs would require 16 additional runs and the 2^{7-3} designs would require 8 additional runs. This suggests that in terms of run size economy, the optimal follow-up design would be more efficient than both the full fold-over and semi-foldover designs for these 7-factor experiments.

Another interesting perspective on the needed number of follow-up runs is in terms of the different run sizes that were used in the analyses. Figure 4.2 on the next page presents the distributions for the number of needed follow-up runs as classified by fractional factorial run size and resolution.
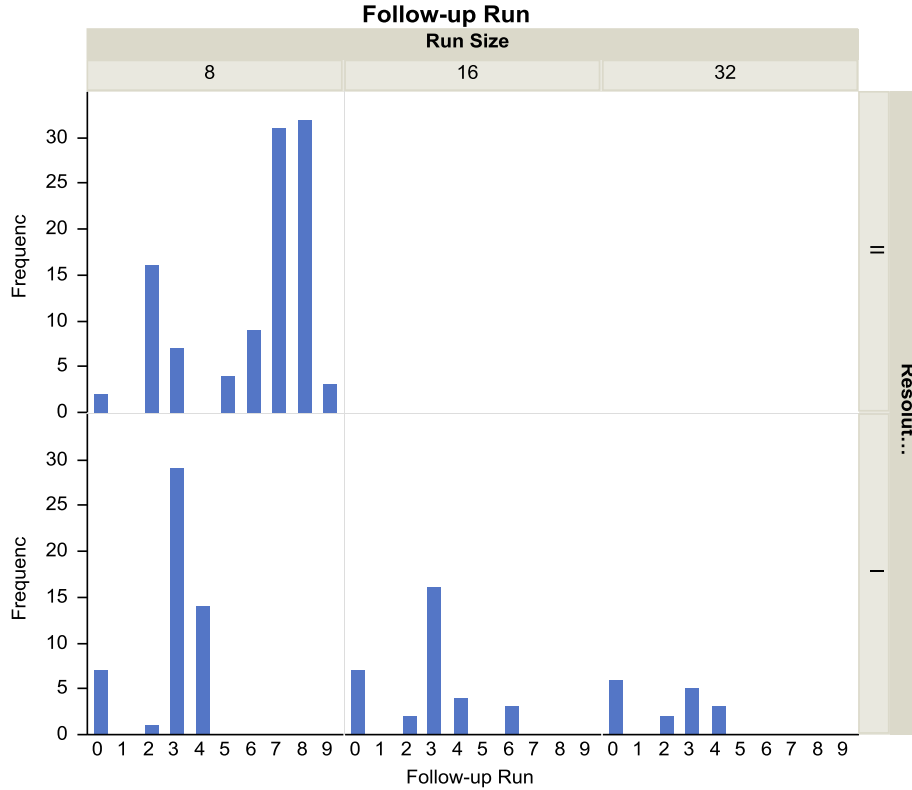
Figure 4.2: Follow-Up Runs by Run Size and Resolution

In Figure 4.2, the graph of the 8-run resolution III experiments shows that typically either

two, seven, or eight additional runs are needed to resolve any confounding issues among the

effects of interest. The 8-run resolution III designs represented in Figure 4.2 are the 2^{5-2}, 2^{6-3},

and 2^{7-4} experiments, and from the previous discussion it was determined that these designs

would typically require two or eight, seven or eight, and seven follow-up runs, respectively, to

estimate all aliased main effects and two-factor interactions. It was determined that the 2^{5-2}

design is responsible for the discrepancy, which is a result of the optimal follow-up method

being heavily dependent on the model of interest that is determined from the original fractional

factorial design. Other than this discrepancy, it can be seen that an experiment with a run size of

eight typically needs approximately seven or eight follow-up runs to estimate all of the main effects and two-factor interactions in the model of interest.

Since a full fold-over design for these experiments would require eight additional runs and would not necessarily resolve the confounding among two-factor interactions, it can be concluded that the D-optimal follow up designs were more efficient in terms of run size economy and potentially allow for the estimability of more effects of interest than the full fold-over method would. A semi-foldover for these designs would only require four follow-up runs, and therefore if a semi-foldover pattern could be identified that would completely resolve any confounding issues, it would have better run size economy than the D-optimal follow-up experiment. However, it may be difficult to find one fold-over or semi-foldover design that completely resolves any confounding issues in the model of interest, and in these cases the D-optimal design would also be preferable because of its estimation capabilities.

As an example, consider a 2^{5-2} experiment from the analyses. The model of interest was determined to include X1, X2, X3, X4, X5, X1*X2, X1*X3, X1*X4, X1*X5, X2*X4, X2*X5, X3*X4, and X3*X5 and had the following alias structure: X2 = X3*X5, X3 = X2*X5, X4 = X1*X5, X5 = X1*X4, X1*X2 = X3*X4, and X1*X3 = X2*X4. For this design, there are six alias chains that need broken, suggesting that a semi-foldover design will not provide a sufficient number of runs to estimate all of the aliased effects. Also, there is no fold-over on any number of factors that will completely de-alias all of the confounded effects of interest. Therefore, this example suggests that in scenarios such as these, the optimal follow-up design would be a better alternative since with only seven additional runs (one for each alias chain plus one for the block effect) it would allow for the estimation of all of the aliased effects in the model of interest.

Continuing with the results from Figure 4.2, the graph of the 8-run resolution IV experiments shows that typically only three follow-up runs are needed and no more than four additional runs were ever needed to de-alias the confounded effects of interest. Since the only 8-run resolution IV designs analyzed in this thesis were 2^{4-1} fractional factorial experiments, the interpretation of this graph follows with the interpretation of 4-factor resolution IV designs discussed from Figure 4.1. It was determined that these designs have better estimation capabilities and run size economy than the full fold-over designs and in these terms also perform as well as or better than the semi-foldover designs, since the optimal follow-up designs require no more than four follow-up runs and guarantee the estimation of all aliased effects in the model of interest.

In Figure 4.2, the graph of the 16-run resolution IV experiments represents 2^{6-2} and 2^{7-3} fractional factorial designs. This graph shows that three follow-up runs are required, although as many as six were sometimes needed to completely de-alias the effects of interest. Whenever more than four additional runs were required, they were needed for the 2^{7-3} design, which is intuitive as this design has one more factor than the 2^{6-2} design yet both designs have the same number of experimental runs. This causes additional aliasing since the 2^{7-3} designs have seven more possible effects than the 2^{6-2} designs.

Compared to the full fold-over designs, the optimal follow-up designs perform better in terms of effect estimability and run size economy. Whereas the fold-over design would require sixteen additional runs, the optimal follow-up requires, on average, three additional runs and at most, six additional runs. Compared to the semi-foldover designs, the optimal follow-up designs would still perform better, since the semi-foldover design would require eight additional runs. In addition, the optimal follow-up design guarantees estimability of the aliased effects from the

fractional factorial design, whereas that guarantee cannot be made for all fold-over and semi-foldover designs.

In Figure 4.2, the final graph depicts the 32-run resolution IV designs, which only represents the 2^{7-2} experiments. For these designs, no more than four additional runs are needed to estimate all of the effects of interest, which supports the fact that resolution IV designs have no aliased main effects and relatively few aliased two-factor interactions. Optimal follow-up experiments for these designs are significantly better in terms of run size economy than the other discussed follow-up methods, as a fold-over design would require 32 additional runs and a semi-foldover design would require 16 additional runs.

From these analyses, it can be seen that optimal follow-up designs often perform better than fold-over designs and semi-foldover designs in terms of estimability of effects and run size economy. In addition, several regularities were discovered from the analyses conducted in this thesis. First, it was shown that across number of factors and run size, resolution III designs require more follow-up runs than resolution IV designs. The specific regularities found in this thesis are tabulated in Table 4.2 on the next page. Note that the given numbers of required D-optimal follow-up runs are the typical number needed as seen in the analyses.

Table 4.2: Regularities in Augmented Fractional Factorial Designs

Resolution	Design	Typical Number of Augmented Runs by Follow-Up Method			
		Range	D-Optimal	Fold-Over	Semi-Foldover
III	2^{5-2}	2 – 9	2 or 8	8	4
	2^{6-3}	7 – 8	8	8	4
	2^{7-4}	5 – 7	7	8	4
IV	2^{4-1}	2 – 4	3	8	4
	2^{6-2}	3 – 4	4 or less	16	8
	2^{7-3}	2 – 6	3	16	8
	2^{7-2}	2 – 4	3	32	16
Run Size					
8 runs (Resolution III)	-		7 or 8*	8	4
8 runs (Resolution IV)	-		3	8	4
16 runs (Resolution IV)	-		3	16	8
32 runs (Resolution IV)	-		4 or less	32	16

*Fewer runs possible if using a 5-factor experiment.

From Table 4.2, it should be noted that optimal follow-up designs have the most improvement over fold-over and semi-foldover designs for (i) resolution IV designs and (ii) designs with greater run sizes. Although there are situations in the above analyses where the optimal designs appears to perform at or below the level of the fold-over and semi-foldover designs, it should also be noted that only the D-optimal follow-up experiments guarantee estimability of all of the aliased effects in the model of interest.

Chapter 5

Conclusions

5.1: Recommendations

Although there are a plethora of follow-up designs and techniques, for each experiment that is done there will be a goal which will be satisfied by some methods better than others. There may be issues with resources, time, cost or different goals such as effect estimability, run size economy, or estimation precision. The purpose of this thesis was to use optimal follow-up experiments to augment fractional factorial designs to discover regularities that existed in the number of follow-up runs that were required to estimate all of the aliased effects in the model of interest for 4-, 5-, 6-, and 7-factor resolution III and IV fractional factorial experiments.

From this research and as presented in Table 4.2, it was determined that for all of the resolution IV designs, four or fewer (typically three) augmented runs would estimate all of the aliased effects in the model of interest. In comparison, all of the resolution III designs required seven or eight follow-up runs to estimate all of the aliased effects. It was determined that D-optimal follow-up experiments were significantly better than fold-over and semi-foldover designs for (i) resolution IV designs and (ii) designs with larger run sizes.

While fold-over and semi-foldover designs are simple for researchers to construct and analyze, they often provide many more additional runs that what is needed to estimate the aliased effects in the model of interest. In addition, these designs are often degree of freedom inefficient

and do not necessarily resolve all of the confounding issues among the effects of interest. In comparison, D-optimal follow-up designs require software to construct but are still simple to analyze and interpret. In addition, their run size is based on the number of aliased effects in the model of interest from the original fractional factorial design and guarantees the estimability of all of the aliased effects of interest.

Given the information in this thesis, it is suggested that D-optimal follow-up designs therefore be considered as a simple and efficient method for constructing follow-up designs to resolve confounding among effects in the model of interest when the original design does not allow for the complete estimation of these effects.

5.2: Future Research

Future research that could be done in this field would be to see if regularities such as these found for regular fractional designs can be found for non-regular fractional designs, such as Plackett-Burman designs. These designs have complex aliasing and are often used for estimating main effects only, but can be used to estimate some two-factor interactions when the number of main effects is low.

In addition, it would be of interest to not only compare the performance of D-optimal follow-up designs to fold-over and semi-foldover designs in terms of run size economy and estimability of effects, but to also consider the performance of D-optimal designs in comparison to semi-foldover designs in terms of D-efficiency of the designs.

5.3: Conclusion

In the field of experimental design, a wide spread goal of experimenters is to be able to maximize the amount of information they can receive from an experiment while minimizing the costs, resources, and time needed to complete it. While this is often a difficult balance to maintain, fractional factorial designs augmented with optimal follow-up experiments are a useful method to achieve this balance. They allow for large experiments to be done in far fewer experimental runs than a full factorial would permit, and while there is a loss of information at the initial stage of the process, that can easily be corrected by performing a simple, optimal follow-up experiment to regain the lost information without requiring a huge strain on resources, time, or money. These designs are easy to construct, simple to analyze, efficient, and full of needed information about the effects in the model of interest. While these designs are not always the best alternative, they should definitely be considered as a viable option when considering an experimental design.

References

Abd El-Monsef, M.M.E., and M.M. Seyam. "CDT-Optimum Designs for Model Discrimination, Parameter Estimation and Estimation of a Parametric Function." *Journal of Statistical Planning and Inference*. 131.2 (2011): 639-643. Print.

Antony, Jiju. *Design of Experiments for Engineers and Scientists*. Burlington: Butterworth-Heinemann, 2003. Print.

Box, G.E.P., and J.S. Hunter. "The 2^{k-p} Fractional Factorial Designs Part I." *Technometrics*. 3.3 (1961): 311-357. Print.

Chen, Jiahua, D.X. Sun, and C.F.J. Wu. "A Catalogue of Two-Level and Three-Level Fractional Factorial Designs wtih Small Runs." *International Statistical Institute*. 61.1 (April 1993): 131-145. Print.

Claeskens, Gerda, and Nils Lid Hjort. *Model Selection and Model Averaging*. Cambridge University Press, 2008. Print.

Galil, Z., and J. Kiefer. "Comparison of Simplex Designs for Quadratic Mixture Models." *Technometrics*. 19.4 (1977): 445-453. Print.

Goos, Peter, and Bradley Jones. *Optimal Design of Experiments: A Case Study Approach*. 1st. John Wiley & Sons, Ltd., 2011. 47-67. Print.

Hurvich, C.M., and C.L. Tsai. "Regression and Time Series Model Selection in Small Samples." *Biometrika*. 76. (1989): 297-307. Print.

Li, Huo, and Robert W. Mee. "Better Foldover Fractions for Resolution III 2^{k-p} Designs." *Technometrics*. 44.3 (2002): 278-283. Print.

Li, William, and Dennis K.J. Lin. "Optimal Fold-Over Plans for Two-Level Fractional Factorial Designs." *Technometrics*. 45.2 (2003): 142-149. Print.

Li, Xiang, Nandan Sudarsanam, and Daniel D. Frey. "Regularities in Data from Factorial Experiments." *Complexity*. 11. (2006): 32-45. Print.

Mandal, N.K., Manisha Pal, and M.L. Aggarwal. "Pseudo-Bayesian A-Optimal Designs for Estimating the Point of Maximum in Component-Amount Darroch-Waller Mixture Model." *Statistics and Probability Letters*. 82.6 (2012): 1088-1094. Print.

Mead, R, S.G. Gilmour, and A. Mead. *Statistical Principles for the Design of Experiments*. New York: Cambridge University Press, 2012. Print.

Mee, Robert W. *A Comprehensive Guide to Factorial Two-Level Experimentation*. Springer Science Business Media, LLC, 2009. Print.

Mee, Robert W., and Marta Peralta. "Semifolding 2^{k-p} Designs." *Technometrics*. 42.2 (May 2000): 122 - 123. Print.

Meyer, R. Daniel, David M. Steinberg, and George E.P. Box. "Follow-Up Designs to Resolve Confounding in Multifactor Experiments." *Technometrics*. 38.4 (1996): 303-313. Print.

Rao, C.R., and Y. Wu. "On Model Selection." Trans. Array *Model Selection*. Institute of Mathematical Statistics, 2001. 1-64. Print.

Shibata, R. "Asymptotic Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process." *Annals of Statistics*. 8. (1980): 147-164. Print.

Sugiura, N. "Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections." *Communications in Statistics-Theory and Methods*. 7.1 (1978): 13-26. Print.

Wu, C.F. Jeff, and Michael S. Hamada. *Experiments: Planning, Analysis, and Optimization*. 2nd edition. John Wiley & Sons, Inc., 2009. Print.