2012

Discriminating Between Optimal Follow-Up Designs

Kevin Donald Kelly
Virginia Commonwealth University

Discriminating Between Optimal Follow-Up Designs

Kevin D. Kelly

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

Masters of Science
in
Mathematical Sciences

David Edwards, Chair
D'arcy Mays
Roy Sabo

May 2nd, 2012
Richmond, Virginia

Keywords: Follow-Up Design, Foldover, Semifoldover, MD-Optimal, Bayesian Model Averaging, Preposterior Analysis

Acknowledgment

I would like to thank Dr. Edwards, Dr. Mays and Dr. Sabo for lending their time to serve on my committee. I want to give a special thanks to my adviser, Dr. Edwards, for always answering my phone calls and e-mails and meeting with me whenever I had questions. I would also like to thank all the faculty and staff in the SSOR department for putting together an outstanding graduate program. All my professors and the coursework effectively stimulated my interest in the subject and also helped me develop valuable analytical skills.

Dedication

I dedicate my thesis to Dee and Greg for supporting and encouraging my decision to return to school. You have always put my education first, and I will be forever thankful.

Contents

List of Figures

Chapter 3

Chapter 4

List of Tables

Chapter 2

Chapter 3

Chapter 4

Abstract

DISCRIMINATING BETWEEN OPTIMAL FOLLOW-UP DESIGNS

By Kevin D. Kelly, M.S.

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

Virginia Commonwealth University, 2012.

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

Sequential experimentation is often employed in process optimization wherein a series of small experiments are run successively in order to determine which experimental factor levels are likely to yield a desirable response. Although there currently exists a framework for identifying optimal follow-up designs after an initial experiment has been run, the accepted methods frequently point to multiple designs leaving the practitioner to choose one arbitrarily. In this thesis, we apply preposterior analysis and Bayesian model-averaging to develop a methodology for further discriminating between optimal follow-up designs while controlling for both parameter and model uncertainty.

Chapter 1

Introduction

When conducting an experiment, there exist two seemingly paradoxical goals of learning as much as possible about a particular process while simultaneously minimizing the total amount of resources dedicated to the experiment. Experiments generally have a fixed budget, and each experimental run often bears a high cost. As a result, a well-designed experiment is integral to satisfying the objective of obtaining knowledge about a process while maintaining efficiency. Prior to an experiment, the practitioner has a limited understanding of the process in question, so it would not be economical to run one large all-encompassing experiment at the onset. Instead, it is in the practitioner's best interest to conduct several smaller sets of experiments sequentially allowing the practitioner to consider the results of each experiment before proceeding.

Sequential experimentation can be employed as a method to identify at which levels input variables should be set in order to ensure that optimal values of the response variable or variables are achieved. In general, the practitioner has in mind a goal of maximizing, minimizing or reaching a target level for each response variable. Although a great deal of information can be gleaned from a small initial experiment, oftentimes there still remains a great deal of ambiguities and questions about the process. In these situations, a follow-up design should, therefore, be employed to resolve as many of these ambiguities as possible.

The existing methods for choosing follow-up designs may fail to fully consider the results of the initial experiment and can force the practitioner to arbitrarily choose between multiple equivalent follow-up designs deemed to be "optimal". It is the goal of this thesis to use Bayesian model averaging to discriminate among these equivalent designs, thus creating a methodology for selecting a follow-up plan.

The remainder of this thesis will proceed as follows. Section 2 provides a brief tutorial of experimental design and Bayesian optimization methods and also contains a summary of the existing literature concerning selecting follow-up designs. Next, Section 3 provides an in depth description of our proposed method of discriminating between follow-up designs through Bayesian model averaging and preposterior analysis. In Section 4, we illustrate this method by comparing optimal follow-up designs for three different examples where the commonly accepted

methodology fails to elect a single optimal follow-up design. Finally, Section 5 contains concluding remarks as well as suggestions for further research on this topic. Although the methodology explored in this thesis can be applied to multiple response optimization, we only consider the case of a single response variable for simplicity.

Chapter 2

Literature Review

Although a great deal of research exists regarding the choice of follow-up design, many of these methods point to multiple optimal choices which requires the experimenter to arbitrarily choose which follow-up design to use. The goal of this paper is to leverage prior knowledge and the results of an initial experiment using Bayesian analysis in order to make informed decisions when choosing between equivalent follow-up designs. The following three sections will outline the existing literature on follow-up designs and Bayesian optimization methods.

2.1 Two-Level Full and Fractional Factorial Designs

Two-level factorial designs are often employed by experimenters to examine the effects of multiple factors on a response variable. A "two-level" design is one in which the explanatory variables are set to one of two levels, generally a high and low level, and these levels are assigned a value of either "-1" or "1" indicating the low and high level, respectively. A two-level full factorial is when the experiment is run at every possible combination of each factor, so that a full factorial of k different factors requires 2^k experimental runs. An example of a 2^4 full factorial can be seen in Table 2.1.

Table 2.1 2^4 Full Factorial

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 number of runs needed for a full factorial increases exponentially as the number of factors increase, and as a result, a fraction of the full factorial is generally chosen as the initial experimental

design, which can make the analysis less straight-forward. By only looking at a fraction of the design locations in a full factorial, factorial effects will be aliased with each other leading to a great deal of ambiguities regarding the effect of each factor.

A fractional factorial design is created by using a full factorial in a portion of the factors and using the levels of those factors to generate the levels of the remaining factors. The aliasing structure of a particular design is dependent on how these remaining factors are generated and can be examined by looking at the defining relation of a design. The defining relation is a list of the factors and interactions that are aliased with the intercept and generally consists of third order interactions or greater. Each entity in the defining relation is known as a *word*, and the number of letters in each word, known as the word length, determines the aliasing structure of a design. The resolution of a fractional factorial design is defined as the shortest word length in the defining relation. Higher resolutions are preferred.

Fractional factorials are written in the form 2^{k-p} where k identifies the number of factors that are being examined and $\frac{1}{2^p}$ is the fraction of the full factorial that is being run. For example, the 2^{4-1} design in Table 2.2 has eight runs which is half as many as a 2^4 design. The resolution of the design can also be incorporated in this form as a Roman numeral sub-script, so the design in Table 2.2 can be written as 2_{IV}^{4-1}.

Table 2.2 2_{IV}^{4-1} Fractional Factorial

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

Notice that this is a full factorial in factors *A*, *B* and *C* and that the levels of D are the product of the values of the other three factors. The defining relation for the design in Table 2.2 is $I = ABCD$ which shows that the four-way interaction of *A*, *B*, *C* and *D* is confounded with the intercept. Note that as the product of these four factors results in a column of all ones. In this example, *ABCD* is the only word in the defining relation and has a length of four indicating that this is a resolution IV design. The aliasing structure of this fractional factorial design can be described as: $A = BCD$, $B = ACD$, $C = ABD$, $D = ABC$, $AB = CD$, $AC = BD$ and $AD = BC$.

The aliasing structure shows that the product of factors *B*, *C* and *D* results in a column that is identical to factor *A*. Thus, the main effect of factor *A* cannot be distinguished from the effect of the

interaction effect of *BCD*. It is generally assumed that interactions between three or more factors are negligible, so the only aliasing problem for this design is with the two-factor interactions. The defining relation and aliasing structure of most fractional factorials can be found in Montgomery (2009) or by using the design software JMP.

The aliasing structure of this 2_{IV}^{4-1} design shows which factorial effects cannot be differentiated when using the classical linear regression model written as $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$. In this equation, $\boldsymbol{y}$ is vector of response values, and X represents the model matrix which includes an intercept and the effects being estimated. For example, if we wanted to fit a model with all the main effects and two-factor interactions involving A for the 2_{IV}^{4-1} design previously described, then we would use the matrix described by Table 2.3 as our X. The vector $\boldsymbol{\beta}$ contains all the parameters that describe the linear relationship between each factorial effect and the response, and $\boldsymbol{\epsilon}$ is a vector of normally distributed random variables with a mean of zero and a variance of σ^2. The $\boldsymbol{\beta}$ vector is estimated using ordinary least squares.

Table 2.3 Model Matrix

I	A	B	C	D	AB	AC	AD
1	-1	-1	-1	-1	1	1	1
1	-1	-1	1	1	1	-1	-1
1	-1	1	-1	1	-1	1	-1
1	-1	1	1	-1	-1	-1	1
1	1	-1	-1	1	-1	-1	1
1	1	-1	1	-1	-1	1	-1
1	1	1	-1	-1	1	-1	-1
1	1	1	1	1	1	1	1

One of the major advantages of using full and 2^{k-p} fractional factorial designs is that all the columns in the model matrix are orthogonal to each other as long as there are no columns that are fully aliased with other columns. If all the columns in the matrix, X, are orthogonal in the linear regression equation, then the estimate of $\boldsymbol{\beta}$ will have minimum variance and the model will not suffer from the effects of multicollinearity.

2.2 Follow-Up Designs

There are a plethora of strategies for selecting follow-up designs when remaining in the same design location. In this thesis, we choose to consider three of the most common strategies (foldovers, semifoldovers and MD optimal designs).

2.2.1 Foldovers and Semifoldovers

A common strategy for following up two-level fractional factorial design is to employ a foldover design, which involves using a second fraction of equal size, obtained by reversing the signs of one or more columns of the initial design. This strategy eliminates a great deal of the aliasing between factorial effects and also decreases the error variance of parameter estimates. For every fractional factorial design, there are 2^k possible foldover choices where *k* is the number of experimental factor being studied which raises the question of which foldover plan to choose.

Li and Lin (2003) define core-foldover plans as foldovers where only the columns of the *p* generated factors have their signs reversed, and they show that every possible choice of foldover is equivalent to a core foldover plan, which reduces the number of possible choices. One common choice is a full foldover which is when the signs for all the factors are reversed, and an example of a full foldover of the 2_{IV}^{4-1} design discussed earlier can be found in Table 2.4. In this situation, the full foldover leads to an identical design as the one seen in Table 2.2, and therefore, does not break any aliasing chains. A foldover on only factor *D*, however, will lead to a full factorial in the combined design.

Table 2.4 Full Foldover of 2_{IV}^{4-1} Design

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 the realm of experimental design, it is preferable to have longer word lengths in a defining relation as shorter words lead to aliasing between lower order interactions and main effects, which is undesirable. As a result, the concept of aberration was developed as a method for discriminating between possible designs. Fries and Hunter (1980) describe *aberration* as:

> "When comparing two designs using resolution as the criterion, one considers the lengths of the shortest word in each defining relation. If these lengths are equal, the two designs are equivalent. With aberration as the criterion, however, one continues to examine the length of the next shortest word in each defining relation until one design is ranked superior to the other."

Consider several competing 2^{6-2} designs where each uses a different set of factors to generate the values of the last two factors. Table 2.5 shows the breakdown of word lengths for multiple 2^{6-2} designs when using different generators for each. According to the aberration criterion, designs 1 and 2 in the table are

superior as their shortest word length of four is larger than the other three designs. We would also deem design 3 to be superior to designs 4 and 5, because it only has one word of length three. Table 2.5 also provides an example where using aberration as a criterion leads to more than one optimal design which would force the practitioner to choose between designs 1 and 2. Aberration, therefore, can be used as a criterion for comparing potential foldovers and minimizing aberration should lead to the most preferable aliasing structure for the combined design.

Table 2.5 Competing 2^{6-2} Designs

Design	Generators	Defining Relation
1	E = ABC , F = BCD	*I = ABCE = BCDF = ADEF*
2	E = BCD, F = ACD	*I = BCDE = ACDF = ABEF*
3	E = CD, F = ACD	*I = CDE = ABDF = ABCEF*
4	E = BC, F = AD	*I = BCE = ADF =ABCDEF*
5	E = BC, F = CD	*I = BCE = CDF =BDEF*

Li and Lin (2003) develop an algorithm for identifying optimal foldovers in terms of minimizing aberration and use it to find the optimal foldovers for 16 and 32 run initial designs. Li and Mee (2002) show that although a foldover of all the factors in a resolution III design eliminates all aliasing between main effects and two-factor interactions, there are oftentimes better options that will lead to designs that are superior in terms of aberration. Montgomery and Runger (1996) identify optimal foldovers for several resolution IV designs. Although all of the aforementioned articles are able to develop a framework for choosing optimal foldovers, their methods generally lead to several equivalent designs with minimum aberration, which leaves the practitioner with the task of arbitrarily picking one of these designs.

In many cases, foldovers are an extremely inefficient follow-up strategy as they require more runs than necessary to break aliasing chains. Semifoldovers are half of the size of a foldover, and although they are not orthogonal, oftentimes they are able to de-alias equally as many two factor interactions as a full foldover. When further experimentation is prohibitively expensive and there are still ambiguities due to aliasing, semifoldovers provide a much cheaper option for eliminating these ambiguities as they only require half as many experimental runs.

Creating a semifoldover design requires the researcher to choose not only which factors to foldover but also which half of the full-foldover fraction to actually use. If the practitioner wanted to do a semi-fold of the full foldover seen in Table 2.4, for example, then they would need to choose which four runs from the full foldover to use. Mee and Peralta (2000) state that in choosing a semifoldover plan, the experimenter must consider which additional effects can be estimated as well as what precision these estimates will have. They also recommend keeping in mind whether there is a desired level of a particular factor or if a certain factor is difficult to change when choosing which half of the foldover to

include in the follow-up design. According to Mee and Peralta (2000) the choice of which factors to foldover is the primary driver of which effects will be de-aliased.

Edwards (2011) shows that using general minimum aberration (Deng and Tang (1999)) as a criterion for selecting a semifoldover plan is inappropriate and proposes using the concept of minimal dependent sets to find the optimal semifoldover for a given design. A minimal dependent set (MDS) is a set of linearly dependent vectors that will become linearly independent if any single vector is removed. Edwards (2011) looks at MDSs in the context of a model that is not estimable when containing a set of two factor interaction but becomes estimable when any single two factor interaction is removed from the model. This method identifies several equivalent optimal semifoldovers which, once again, requires the practitioner to choose which follow-up design to use.

For both foldovers and semifoldovers, choosing between optimal follow-up designs does not necessarily have to be an arbitrary decision. If there is a particular factor or interaction of interest to the researcher, then it would be advantageous to choose an optimal foldover/semifoldover that dealiases this factor or interaction. When this is not the case, choosing which follow-up design to use could affect the ability to optimize the response.

2.2.2 MD-Criterion

Oftentimes, after an initial experiment, there are several models that appear to fit the data equally well, and from an optimization standpoint, it is important to reduce the ambiguities around which model is most appropriate. Box and Hill (1967) describe a situation where the choice of follow-up runs is motivated by discriminating between rival models rather than aliased effects. They contend that in this situation the next experiment should be conducted at locations that maximize the ability to discriminate between models.

Meyer, et al. (1996) develop a criterion that measures the amount of discrimination that a follow-up design provides between all possible subset models. This criterion is called the model discrimination criterion, written as MD-criterion, and although a single calculated criterion value, alone, has no meaning, they can be used as a relative comparison tool between competing follow-up designs. For a given run size, the follow-up design with the largest MD-criterion is preferred, and this design maximizes the expected amount of model discrimination provided by a follow-up design of this size. The calculation of the MD-criterion requires an understanding of basic Bayesian statistical methods, so we will provide a background of these methods before providing an equation for the MD-criterion.

A factor is considered to be active when its effect on the response variable is significantly different than zero. If M_i is a model containing a combination of f_i active factors and the probability of

each factor being active is π, then the prior probability of a model M_i is

$$P(M_i) = \pi^{f_i}(1-\pi)^{k-f_i} \tag{2.2}$$

For equation (2.2), it is assumed that the probability of a particular factor being active is independent of all the other factors.

Using Bayes theorem, the posterior probability function of a vector of parameters $\boldsymbol{\theta}$ conditional on the vector of observed data $\mathbf{y}$ can be calculated using

$$P(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{P(\boldsymbol{\theta})P(\boldsymbol{y}|\boldsymbol{\theta})}{P(\boldsymbol{y})} \tag{2.3}$$

where

$$P(\boldsymbol{y}) = \int_{All\ \theta} P(\boldsymbol{\theta})P(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{\theta}, \tag{2.4}$$

$P(\boldsymbol{\theta})$ is the prior probability of $\boldsymbol{\theta}$ and $P(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function of $\boldsymbol{y}$ given $\boldsymbol{\theta}$**.** It can be seen in equation (2.3) that these methods also allow the practitioner to include prior information about $\boldsymbol{\theta}$ in the analysis by choosing a prior distribution P($\boldsymbol{\theta}$). A non-informative prior assumes that there is no prior knowledge about $\boldsymbol{\theta}$, so for example, it could considers all values between $-\infty$ and ∞ to be equally likely, which is equivalent to a uniform distribution along these bounds. A non-informative prior takes away the subjectivity often associated with Bayesian methods, but non-informative priors will often lead to results that are similar to the results of classical methods.

Once the posterior distribution of $\boldsymbol{\theta}$ has been found, Bayesian methods provide a method for making inferences about future observations, ỹ, by computing the posterior predictive distribution as follows

$$P(\tilde{y}|\boldsymbol{y}) = \int_{All\ \theta} P(\tilde{\boldsymbol{y}}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}. \tag{2.5}$$

Using equation (2.2) as the prior probability that a given model is the true model, the posterior density of a particular model, M_i, given the data $\mathbf{y}$ is

$$P(M_i|\boldsymbol{y}) = \frac{P(\boldsymbol{y}|M_i)P(M_i)}{\sum_i P(\boldsymbol{y}|M_i)P(M_i)} \tag{2.6}$$

where

$$P(\boldsymbol{y}|M_i) = \int_{\sigma^2}\int_{\boldsymbol{\beta}_i} P(\boldsymbol{y}|M_i,\sigma^2,\boldsymbol{\beta}_i)P(\sigma^2,\boldsymbol{\beta}_i|M_i)d\boldsymbol{\beta}_i d\sigma^2. \tag{2.7}$$

After observing a vector $\mathbf{y}$ of responses, it is possible to calculate a posterior probability for each model given the data, P(M_i|$\mathbf{y}$), using equation (2.6). Meyer, et. al (1996) show that the probability of a factor, *j*, being active can be calculated with

$$P_j = \sum P(M_i|\,\boldsymbol{y}), \tag{2.8}$$

for all M_i that contain factor *j*.

In order to calculate the model posterior probability, Meyer, et al. (1996) specify a g-prior, meaning all main effects and interactions in the model are assigned N(0,$\gamma^2\sigma^2$) prior distributions. The g-

prior is popular in Bayesian optimization, and a more in depth discussion of it will be presented in the following section.

Using the model posterior probabilities, Meyer, et al. (1996) developed their MD-criterion that can be calculated with

$$MD = \sum_{0 \le i \ne j \le m} P(M_i|\,\mathbf{y})P\left(M_j\middle|\mathbf{y}\right)I\left(p_i, p_j\right), \tag{2.9}$$

where p_i is the posterior predictive density for a new observation conditional on **y** and on M_i being the correct model. The $I(p_i,p_j)$ term is the Kullback-Leibler information (Kullback and Leibler (1951)) and is a measure of the mean information for discriminating in favor of M_i against M_j when M_i is true and can be computed with

$$I\left(p_i, p_j\right) = \int p_i(x)\ln(p_i(x)/p_j(x))dx. \tag{2.10}$$

It is assumed for our applications that neither p_i nor p_j will take a value of zero as we do not anticipate a posterior predictive probability of zero.

Choosing a design that maximizes the MD-criterion allows the practitioner to choose the set of follow-up runs that will provide the most discrimination between all the models under consideration. This criterion can be applied to follow-up designs of any size, which provides the researcher with a great deal of flexibility that is not permitted by foldover and semifoldover designs. The MD-criterion, unfortunately, does not always point to a single best choice of follow-up runs.

We have discussed that using some of the most commonly practiced follow-up design criterion, the researcher is often left to decide between multiple equivalent designs. There currently exists no framework for how to further discriminate between these design choices. Although this decision may seem inconsequential, one of these candidate designs could be more likely to provide the experimenter with an optimal level of the response variable. In other words, the design locations for a particular follow-up design may lead to responses that point to optimal factor settings that are more likely to produce the desired response level. Therefore, choosing a follow-up design haphazardly is unwise. The aim of this paper is to develop a method for further discrimination between follow-up designs using Bayesian analysis.

2.3 Bayesian Optimization Methods

In the classical approach to optimization, the model parameters are assumed to be unknown constants, and by using only standard errors to address parameter uncertainty, the frequentist approach fails to appropriately address all of the uncertainty in model parameters (Del Castillo (2007)). Bayesian methods, on the other hand, treat the parameters as random variables with their own probability distributions, which accounts for any uncertainties in the parameters (Del Castillo (2007)).

The posterior predictive distribution makes it possible to compute the probability that the response of interest lies within a desirable region for a particular design location. Integrating equation (2.5) over the desired range of response values will yield the probability that a given design location, $\mathbf{x}^*$, will produce a response in this optimal range.

Peterson (2004) shows how sampling from the predictive posterior can be used to optimize a process with multiple responses. This sampling is carried out by simulating from the posterior predictive distribution found with equation (2.5). Peterson (2004) introduces the idea of a preposterior analysis, which involves sampling from the posterior predictive distribution in order to determine whether collecting additional data will reduce parameter uncertainty. Miró-Quesada, et al. (2004) apply these same methods to a case where noise variables must be considered as well. Gilmour and Mead (1995) use preposterior analysis to determine whether further experimentation is beneficial and create a framework for deciding when an experiment should be ended. Bayarri and Mayoral (2002) consider the goals of replications and use these methods to determine whether replicate runs will accomplish these goals. Although all of these articles address parameter uncertainty, their analysis fails to take into account the ambiguities involved in the model selection.

After an initial experiment, there are oftentimes several different models that fit the data reasonably well. Classical optimization methods require that one of these models be selected as the "correct" model but fail to take into account the possibility that the chosen model is not the best choice. Each of the competing models may lead to different, conflicting optimal operating settings. Bayesian model averaging (BMA) provides a method for considering parameter and model selection uncertainty simultaneously by using a weighted average of the posterior predictive densities for all the candidate models.

The model posteriors that can be found using equation (2.6) are a logical choice for the weighting in Bayesian model averaging as they give the probability of a particular model being correct given the data. Therefore the model-averaged posterior (MAP) predictive density is

$$MAP = \sum_{All\ i} P(\tilde{\boldsymbol{y}}|\boldsymbol{x}^*, \boldsymbol{y}, M_i)P(M_i|\boldsymbol{y}). \qquad (2.11)$$

Equation (2.11) shows that the MAP density is a weighted average of the posterior predictive densities for each model where the $P(M_i|\mathbf{y})$ values are the weights, thus this equation can be used to compute the model-averaged probability of achieving a response in the desired range for a chosen design location $\mathbf{x}^*$. The MAP predictive density is similar to the posterior predictive density in equation (2.5), but the MAP density also takes model uncertainty into account.

Buckland, et al. (1997) proposes a weighting method that uses both the Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) as the model weights in order to avoid the intensive calculations required by equation (2.11), but advances in modern computing have made these

calculations much easier. The weighting method proposed by Buckland, et al. (1997) is heavily influenced by the number of parameters in the model so the MAP is preferred.

Rajagopal and Del Castillo (2005) illustrate how BMA can be used to optimize a process with multiple responses and also provide evidence that model-averaged predictions are more reliable than picking an assumed model. Both Rajagopal and Del Castillo (2005) and Ng (2010) sample from the MAP to optimize the process in a similar manner to the preposterior analysis discussed previously. Ng (2010) samples from the MAP not only to find the optimal setting for a process but to also inform the choice of follow up designs. It is the aim of this paper to utilize the methods proposed by Rajagopal and Del Castillo (2005) and Ng (2010) and take them a step further by applying them to break ties between equivalent optimal follow-up designs.

Chapter 3

Illustration of Methodology

We will now develop the follow-up strategy for a small initial experiment in order to provide an in-depth illustration of our follow-up design discrimination technique. We will look at a 2_{IV}^{8-4} injection molding experiment from Mee (2009). The data for this experiment in coded units can be found in Table 3.1. The experiment examined the effects of 8 experimental factors on shrinkage with the goal of minimizing this response variable. For the sake of this analysis, we are assuming that a shrinkage value of 14.5 or less is desired. The eight factors considered in this experiment are *A* = screw speed, *B* = moisture content, *C* = holding pressure, *D* = cavity thickness, *E* = booster pressure, *F* = cycle time, *G* = gate size and *H*= mold temperature. The generating rules for this design are *E* = *-ACD*, *F* = *-BCD*, *G* = *ABC* and *H* = *-ABD*.

Table 3.1 Injection Molding Experiment

A	B	C	D	E	F	G	H	Y
-1	-1	-1	-1	1	1	-1	1	14
1	-1	-1	-1	-1	1	1	-1	16.8
-1	1	-1	-1	1	-1	1	-1	15
1	1	-1	-1	-1	-1	-1	1	15.4
-1	-1	1	-1	-1	-1	1	1	27.6
1	-1	1	-1	1	-1	-1	-1	24
-1	1	1	-1	-1	1	-1	-1	27.4
1	1	1	-1	1	1	1	1	22.6
-1	-1	-1	1	-1	-1	-1	-1	22.3
1	-1	-1	1	1	-1	1	1	17.1
-1	1	-1	1	-1	1	1	1	21.5
1	1	-1	1	1	1	-1	-1	17.5
-1	-1	1	1	1	1	1	-1	15.9
1	-1	1	1	-1	1	-1	1	21.9
-1	1	1	1	1	-1	-1	1	16.7
1	1	1	1	-1	-1	1	-1	20.3

3.1 Follow-Up Strategies

This experiment is one sixteenth of the size of a full factorial which means that there are 240 design points in the design region of the full factorial that have not yet been run. Oftentimes the choice of follow-up design is motivated by the practitioner's desire to break specific aliasing chains, but when this

is not the case, the practitioner usually would prefer to choose the follow-up design that is most likely to identify the operating conditions that produce the desired response reliably.

According to Li and Lin (2003), foldovers on the following groups of factors *EF*, *EG*, *EH*, *FG*, *FH*, *GH* and *EFGH* are optimal in terms of minimum aberration which leaves us with many options to choose from. All of these foldovers will lead to the same word length pattern of the combined design, but each one will break different aliasing chains. Table 3.2 shows that every optimal foldover leads to a defining relation with six 4-letter words and one 8-letter word. Although each defining relation contains some words in common with the others, they are all unique, however, and will lead to different aliasing structures.

Table 3.2 Minimum Aberration Foldovers

Foldover Factors	Defining Relation
EF	*I = ABEF = ABCG = CEFG = -ABDH = -DEFH = -CDGH = -ABCDEFGH*
EG	*I = -BCDF = -BDEG = CEFG = -ABDH = ACFH = AEGH = -ABCDEFGH*
EH	*I = -BCDF = ABCG = -ADFG = BCEH = DEFH = AEGH = -ABCDEFGH*
FG	*I = -ACDE = -ADFG = CEFG = -ABDH = BCEH = BFGH = -ABCDEFGH*
FH	*I = -ACDE = ABCG = -BDEG = ACFH = -DEFH = BFGH = -ABCDEFGH*
GH	*I = -ACDE = -BCDF = ABEF = -CDGH = AEGH = BFGH = -ABCDEFGH*
EFGH	*I = ABEF = -BDEG = -ADFG = BCEH = ACFH = -CDGH = -ABCDEFGH*

Edwards (2011) identifies multiple optimal semifoldover choices that could follow up this eight factor experiment based on his MDS criterion. Each of these optimal semifoldovers leads to a combined minimum MDS-aberration design with 18 minimal dependent sets, all of which are of size two. The first step in creating these semifoldovers is to create a foldover of the initial experiment by reversing the signs of factors A and B which can be seen in Table 3.3. For this design, choosing a subset of runs from Table 3.3 based on any of the eight factors will lead to a minimum MDS-aberration design. Subsetting by a particular factor means selecting the runs where the chosen factor is either set at the high level or the low level which means that there are 16 optimal semifoldovers that can be chosen to follow this experiment.

The MD criterion can also be used to determine the optimal follow-up plan for this experiment, and it can be used to construct follow-up designs of any run size. In order to calculate the MD-criterion, the practitioner must first set the values of the hyperparameters π and γ which are required by the process. We chose the values of $\pi = .25$ and $\gamma = .984$, and the reasoning behind these choices will discussed in the next sub-section.

We used the BsMD statistical package in R to determine which four-run design would be the best follow-up for this experiment based on the MD-criterion. Although the output from R was able to identify a single model with the highest MD-criterion value of 3.947, the MD-criterion values of the top

ranked follow-up designs are all within .06 units of each other. Practically speaking, these top models are negligibly different from each other, so it may not be in the practitioner's best interest to simply select the four-run design with the highest MD-Criterion.

Table 3.3 Foldover on Factors A and B

A	B	C	D	E	F	G	H
1	1	-1	-1	1	1	-1	1
-1	1	-1	-1	-1	1	1	-1
1	-1	-1	-1	1	-1	1	-1
-1	-1	-1	-1	-1	-1	-1	1
1	1	1	-1	-1	-1	1	1
-1	1	1	-1	1	-1	-1	-1
1	-1	1	-1	-1	1	-1	-1
-1	-1	1	-1	1	1	1	1
1	1	-1	1	-1	-1	-1	-1
-1	1	-1	1	1	-1	1	1
1	-1	-1	1	-1	1	1	1
-1	-1	-1	1	1	1	-1	-1
1	1	1	1	1	1	1	-1
-1	1	1	1	-1	1	-1	1
1	-1	1	1	1	-1	-1	1
-1	-1	1	1	-1	-1	1	-1

We have discussed that three of the most commonly practiced follow-up design discrimination techniques fail to identify a single optimal design for injection molding experiment. Although choosing between these equivalent designs seems like a trivial decision, the selection of follow-up runs may ultimately have a direct effect on the optimal operating conditions that are chosen based on the results of this experiment. It is possible that one of these optimal designs will lead to a better and more reliable optimum. In this situation, we recommend that the practitioner leverage all the available information about the process to make this decision rather than choosing an optimal follow-up design arbitrarily.

3.2 Discrimination Methodology

The first step of our proposed discrimination technique is to determine a set of competing models to consider in the analysis. Although it would not be completely unreasonable to include all possible models when there are very few experimental factors, this tactic, however, would be unwieldy for larger experiments. A more practical approach is to consider a set of the most probable models based on model posterior probabilities calculated using equation (2.6). We believe that it is reasonable to include only

models with posterior probabilities greater than .01 in our set of competing models as the inclusion of all other models would have a negligible impact on the analysis.

As seen in equation (2.6), the model posterior probability is a function of the model prior, $P(M_i)$, and the marginal likelihood of the model given the data, $P(\boldsymbol{y}|M_i)$. The model prior can be found using equation (2.2) but requires the selection of the hyperparameter π, which is the probability that an individual experimental factor is active. The widely accepted sparsity of effects principle states that only a few experimental factors are likely to be significant in a fractional factorial, and as a result, we believe that $\pi = .25$ is a reasonable hyperparameter choice. Rajagopal and Del Castillo (2005) provide evidence that the calculated model posteriors are fairly robust to the selection of the π hyperparameter.

The marginal likelihood of a model for a given data set is shown in equation (2.7) and is considerably more complex than the model prior. The first part of this equation is called the likelihood function and when assuming that the error terms are normally distributed, the likelihood function can be written as:

$$P(\boldsymbol{y}|M_i, \sigma^2, \boldsymbol{\beta}_i) \propto \sigma^{-n}\exp[-\tfrac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}_i\boldsymbol{\beta}_i)'(\boldsymbol{y} - \boldsymbol{X}_i\boldsymbol{\beta}_i)] \tag{3.1}$$

The second part of equation (2.7) is the joint prior of the model parameters, $\boldsymbol{\beta}_i$ and σ^2, which are considered to be independent a priori. Using a g-prior, the priors are chosen to be

$$P(\sigma^2) \propto k_1, \tag{3.2}$$

$$P(\beta_0) \propto k_2, \tag{3.3}$$

and

$$\boldsymbol{\beta}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_i\sigma^2)_{\boldsymbol{i>0}}, \tag{3.4}$$

where k_1 and k_2 are some constants,

$$\boldsymbol{\Sigma}_i^{-1} = \tfrac{1}{\gamma^2}\begin{pmatrix} 0 & \boldsymbol{0} \\ \boldsymbol{0} & I_{t_i} \end{pmatrix}, \tag{3.5}$$

t_i is the number of parameters in the model excluding the intercept, and I_{t_i} is the identity matrix of order t_i.

Meyer et. al (1996) suggest choosing the value of γ that minimizes the posterior probability of the null model which is a model with the intercept as its only parameter. They are able to empirically illustrate that choosing this value of γ maximizes the posterior density $P(\gamma|\boldsymbol{y})$. We found that a value of $\gamma = .984$ minimizes the posterior probability of the null model in our analysis of the injection molding experiment, and this value was therefore chosen for the hyperparameter.

Rajagopal and Castillo (2005) show that plugging equation (3.1) and the assumed priors into equation (2.7) allows the marginal likelihood of the model given the data to be rewritten as

$$P(\boldsymbol{y}|M_i) \propto \gamma^{-t_i}\left|\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{X}_i'\boldsymbol{X}_i\right|^{-1/2} S_i^{-(n-1)/2}. \tag{3.6}$$

Then, omitting the constant denominator in equation (2.6), we now find that

$$P(M_i|\boldsymbol{y}) \propto \pi^{f_i}(1-\pi)^{k-f_i}\gamma^{-t_i}\left|\boldsymbol{\Sigma}_i^{-1}+\boldsymbol{X}_i'\boldsymbol{X}_i\right|^{-1/2}S_i^{-(n-1)/2}, \tag{3.7}$$

where

$$S_i = \left(\boldsymbol{y}-\boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_i\right)'\left(\boldsymbol{y}-\boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_i\right)+\widehat{\boldsymbol{\beta}}_i'\boldsymbol{\Sigma}_i^{-1}\widehat{\boldsymbol{\beta}}_i \tag{3.8}$$

$$= \boldsymbol{y}'\boldsymbol{y}-\boldsymbol{y}'\boldsymbol{X}_i(\boldsymbol{\Sigma}_i^{-1}+\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{y} \tag{3.9}$$

and

$$\widehat{\boldsymbol{\beta}}_i = (\boldsymbol{\Sigma}_i^{-1}+\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{y}. \tag{3.10}$$

We calculated the model posterior probabilities for all subset models in the injection molding experiment and found only three models with a probability greater than .01. These three models will be the only models that we consider in our analysis, and these models and their posterior probabilities are displayed in Table 3.4. Each model includes all the main effects listed and their two-factor interactions.

Table 3.4 Injection Molding Model Posterior Probabilities

Factors	$P(M\|\boldsymbol{y})$
C, *D* and *E*	0.654
A, *C* and *E*	0.189
A, *C*, *D* and *E*	0.134

Once a competing set of models has been selected, the next step is to develop a set of follow-up experiments to compare. We showed that for the injection molding experiment three popular follow-up strategies lead to multiple optimal designs which creates our set of competing follow-up plans. We will employ our discrimination technique on each of these candidate designs in order to determine which design will identify optimal operating conditions that maximize the model-averaged posterior (MAP) probability of producing a response in the desired range.

In order to calculate an MAP probability, we must first find the probability of achieving a response in the desired range when only considering a single model, which according to Rajagopal and Del Castillo (2005) can be defined as

$$P(L \le \tilde{y} \le U|M_i, \boldsymbol{x}^*, \boldsymbol{y}) = \int_L^U P(\tilde{y}|M_i, \boldsymbol{x}^*, \boldsymbol{y})d\tilde{y}, \tag{3.11}$$

where $\mathbf{x}^*$ is a given set of experimental factor levels. They go even further to prove that when g-priors are used that the predictive density in equation (3.11) has a t_{n-1} distribution with a cumulative predictive density of

$$P\left(\frac{\tilde{y}-\boldsymbol{x}^{*\prime}\widehat{\boldsymbol{\beta}}_i}{\widehat{\sigma}_i\sqrt{1+\boldsymbol{x}^{*\prime}\left(\boldsymbol{\Sigma}_i^{-1}+\boldsymbol{X}_i'\boldsymbol{X}_i\right)^{-1}\boldsymbol{x}^*}} < t|M_i, \boldsymbol{x}^*, \boldsymbol{y}\right). \tag{3.12}$$

Therefore, the probability of some design point, $\mathbf{x}^*$, producing a response between our upper and lower bounds for a given model can be calculated by finding the cumulative probability for the upper bound and

subtracting the cumulative probability of the lower bound. It follows, then, that when using g-priors, the MAP probability that a design location, **x***, will produce a future response, ỹ, between upper and lower bounds, U and L, respectively can be computed with

$$MAP = \sum_{All\ i}(P(\tilde{y} < U|M_i, \boldsymbol{x}^*, \boldsymbol{y}) - P(\tilde{y} < L|M_i, \boldsymbol{x}^*, \boldsymbol{y}))P(M_i|\boldsymbol{y}). \quad (3.13)$$

Since we have decided to only use models that have posterior probabilities greater than .01, the model posterior probabilities must be rescaled, in order to ensure that they sum to one.

Once a set of competing designs and models to be considered has been created and the posterior probabilities of these models have been calculated, then our proposed design discrimination can be run for each candidate design as follows:

1. Select a model at random using the model posterior probabilities, $P(M_i|\boldsymbol{y})$, as weights.
2. Create the X_2 model matrix based on the follow-up design and the chosen model.
3. Simulate a response, y, for every experimental run, $\mathbf{x}^*$, in X_2 using the t_{n-1} predictive density described by equation (3.12).
4. Recalculate the new model posterior probabilities that consider the initial and simulated responses combined.
5. Determine which factor settings maximize the MAP probability defined by equation (3.13) using the Nelder-Mead (Nelder and Mead (1965)) simplex method carried out by the fmincon function in MATLAB.
6. Record the MAP probability that was found at these optimum factor settings.
7. Run 1,000 iterations of steps 1-6.

Once this process is repeated 1,000 times for a given design, we are able to analyze the distribution of all the MAP probabilities that have been recorded by calculating the mean, standard deviation and five number summary of the data set. We also recommend creating histograms of the model-averaged probabilities as a visual comparison tool. The distribution of these probabilities can be used to further discriminate between optimal follow-up designs, and although this approach does not always point to a clear "winner" amongst a group of follow-up designs, it provides valuable information that should be considered before selecting a design.

3.3 Injection Molding Analysis

We considered all optimal foldovers and semifoldovers, as well as the top five 4-run MD-Optimal designs from the injection molding experiment. In addition, we also included a 16-run and an 8-run MD-Optimal design in our analysis in order to compare an MD-Optimal design to both the foldover and semifoldover designs. Prior to our analysis, we used the optimization process on the initial experiment and found that

at the optimum operating conditions, the MAP probability of producing a response in the desired range is .4985.

Tables 3.5 through 3.8 show the results from applying our discrimination technique to all the follow-up designs being considered. In Table 3.5, the listed factors indicate which factors had their signs reversed as part of the foldover procedure. The notation for the semifoldovers in Tables 3.5 and 3.6 provides the factor that was used to subset Table 3.3 and a positive or negative sign which indicates whether the high or low value of that factor was chosen. For example, the column titled A+ is all the runs from Table 3.3 where factor A is set at the high level.

Table 3.5 Foldovers and Top 16-Run MD-Optimal Design

Design	EF	EG	EH	FG	FH	GH	EFGH	MD-16run
Mean	0.4978	0.495	0.5005	0.5005	0.492	0.496	0.4961	0.5107
Min	0.1567	0.1369	0.1229	0.1488	0.0529	0.1555	0.1793	0.1514
Max	0.9341	0.9011	0.8869	0.8635	0.8888	0.871	0.8722	0.9938
Q1	0.4054	0.3998	0.401	0.4059	0.3974	0.4028	0.4014	0.4072
Q3	0.5973	0.5869	0.6009	0.5863	0.5796	0.5882	0.5868	0.6026
Median	0.4953	0.4971	0.4916	0.4992	0.4913	0.4987	0.4905	0.5076
Std Dev	0.1358	0.132	0.1379	0.1271	0.1332	0.1306	0.1319	0.1433

Table 3.6 Semifoldovers

Design	A-	A+	B-	B+	C-	C+	D-	D+
Mean	0.5088	0.5061	0.5132	0.5097	0.5077	0.5077	0.5087	0.5054
Min	0.2063	0.1692	0.168	0.1834	0.1633	0.3777	0.1709	0.1929
Max	0.9963	0.9603	0.9794	0.9892	0.9936	0.951	0.964	0.8939
Q1	0.4567	0.4184	0.4144	0.413	0.4078	0.4637	0.4239	0.4575
Q3	0.5446	0.5857	0.6091	0.6047	0.6025	0.5432	0.5912	0.5397
Median	0.4988	0.5193	0.511	0.5114	0.5072	0.5026	0.5217	0.4943
Std Dev	0.0918	0.1229	0.1364	0.1337	0.1394	0.0592	0.1249	0.0833

Table 3.7 Additional Semifoldovers and Top 8-Run MD Optimal Design

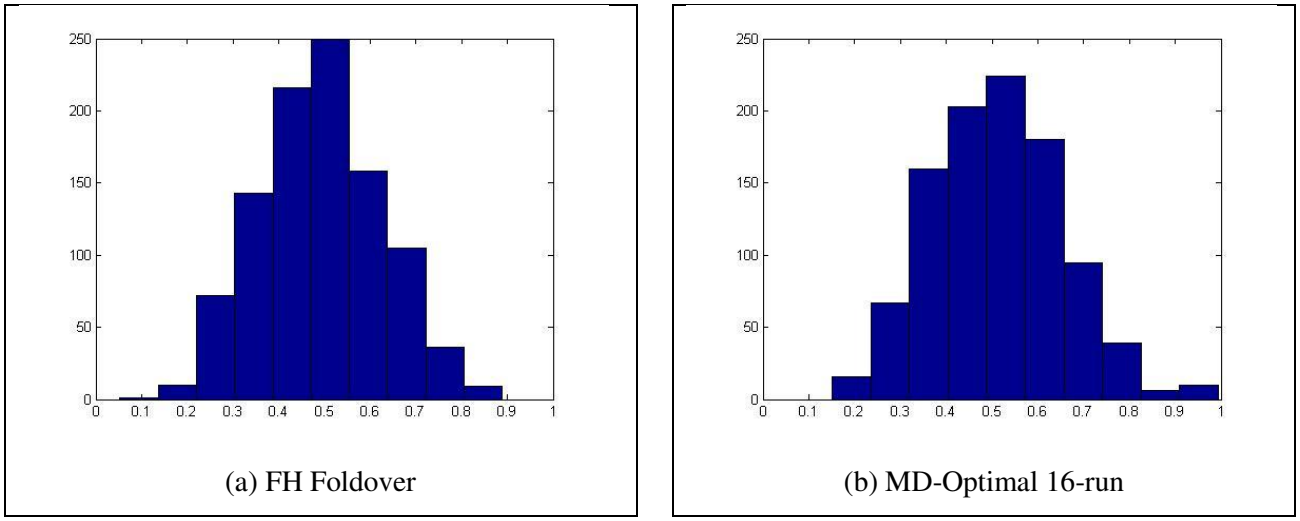
Design	E-	E+	F-	F+	G-	G+	H-	H+	MD-8run
Mean	0.5076	0.5077	0.5066	0.5054	0.5106	0.5139	0.5048	0.5061	0.5285
Min	0.3891	0.1641	0.1817	0.1628	0.1717	0.1745	0.1837	0.1302	0.1776
Max	0.8435	0.9372	0.9546	0.9718	0.9718	0.9834	0.9895	0.977	0.9996
Q1	0.4627	0.4171	0.41	0.4086	0.4182	0.4223	0.4102	0.413	0.4185
Q3	0.5445	0.599	0.5965	0.5884	0.597	0.6071	0.5938	0.5998	0.6251
Median	0.5014	0.5031	0.5049	0.5	0.5074	0.512	0.5064	0.5042	0.5232
Std Dev	0.0636	0.1331	0.1322	0.1359	0.1308	0.1342	0.1331	0.1372	0.1485

Table 3.8 Top Five 4-Run MD Optimal Designs

Design	MD1	MD2	MD3	MD4	MD5
Mean	0.5013	0.5205	0.5143	0.5159	0.4999
Min	0.1696	0.267	0.2741	0.3019	0.4241
Max	0.863	0.9832	0.9962	0.9871	0.6689
Q1	0.4076	0.4319	0.4196	0.4644	0.4749
Q3	0.5953	0.5943	0.5895	0.5299	0.5251
Median	0.5016	0.5104	0.5118	0.4923	0.5006
Std Dev	0.127	0.1232	0.1211	0.0963	0.0335

The mean, minimum, maximum, 1st quartile, 3rd quartile, median and standard deviation of the simulated MAP probabilities are found under each design, and these values provide a summary of the distribution of these probabilities for all the potential follow-up designs. Each individual MAP probability is the probability that the optimum found after the simulated follow-up design will lead to a response in the desired region, thus we have found a distribution that shows the capability of each design for finding a reliable optimum.

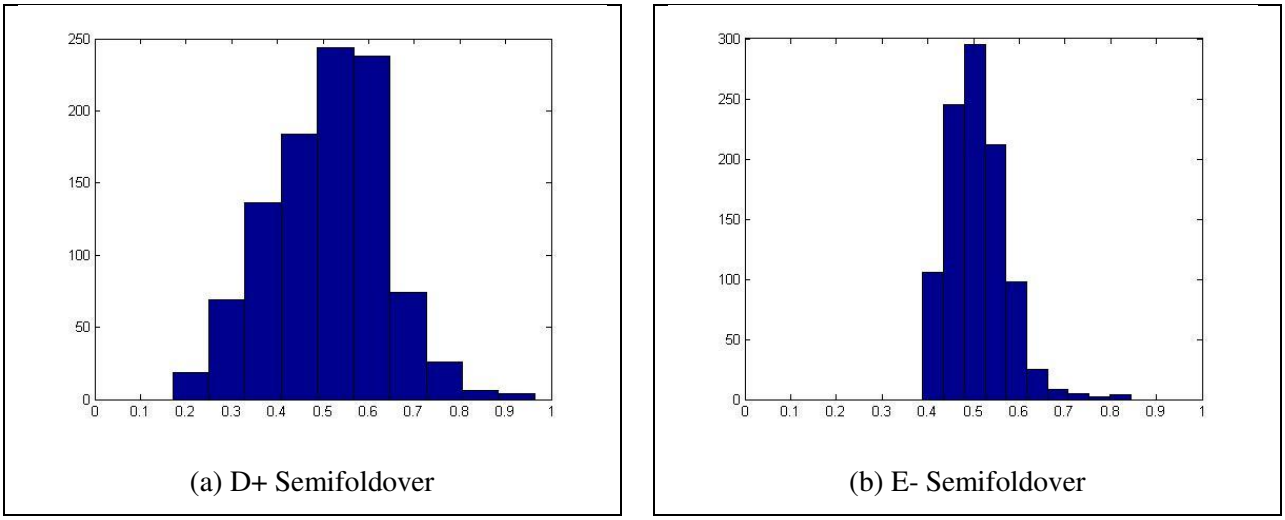
Table 3.5 shows that the MAP probabilities of all the foldovers are similar for the most part, but the results do suggest that the FH foldover is least preferable. The minimum for this foldover is lower than the other foldovers, and its maximum, Q1 and Q3 values are also among the lowest. The 16-run MD-Optimal design appears to have an advantage over all the foldovers as its maximum is much higher than the rest of the design. Figure 3.1 shows the histogram for the FH foldover and the 16-run MD Optimal design, and it appears that MD-Optimal design appears to have a slight advantage.



(a) FH Foldover (b) MD-Optimal 16-run

Figure 3.1 Injection Molding Experiment 16-Run Follow-Up Design Histograms

Tables 3.6 and 3.7 show that there is a much starker contrast between the distributions of the semifoldovers. The diagnostics of most the semifoldovers are alike, but the C+, D+ and E- semifoldovers stand out with their low standard deviations and differing distributions. The E- semifoldover, for example, has the highest minimum and lowest maximum MAP probabilities by far. Figure 3.2 shows the drastic distributional difference between the D+ and E- semifoldovers, and it appears that the E- semifoldover is a more conservative option. Each of the semifoldovers that exhibits a tight distribution involves an experimental factor that is in two or more of the models used in this simulation.



(a) D+ Semifoldover

(b) E- Semifoldover

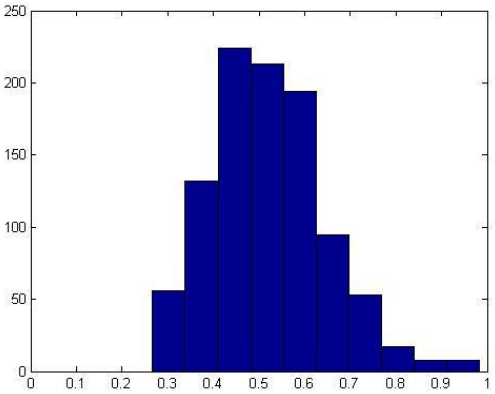
Figure 3.2 Injection Molding Experiment 8-Run Follow-Up Design Histograms

Table 3.8 shows that designs MD2, MD3 and MD4 have much higher maximum and minimum simulated MAP probabilities than MD1 which gives them an advantage. Designs MD2 and MD3 are preferable as their 3rd quartile is much higher than that of MD4. This provides an example where the highest MD-Criterion value is not necessarily the best option. MD5 sticks out with a relatively low standard deviation. The histograms in Figure 3.3 show that the MD2 design produced a wider range of MAP probabilities and has much greater potential to produce high MAP probabilities at the optimum factor setting than MD5. I would recommend choosing design MD2 over MD5, because it appears that there are a higher proportion of large simulated MAP probabilities.

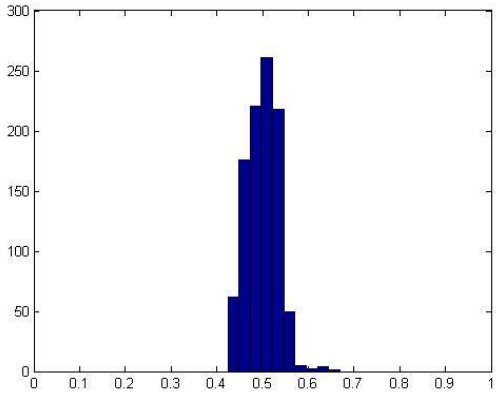
Although the summary statistics and histograms of the MAP probabilities are able to provide a great deal of discrimination between the candidate follow-up designs, there is a great deal of subjectivity in this method. Ultimately the goal of the practitioner is to choose the follow-up design that points to optimal factor settings with the highest MAP probability among all follow-up designs. As a result, we are proposing a maximax criterion where the follow-up design with the largest simulated maximum MAP

probability is preferred. This criterion leads to the selection of the design that shows the most potential for improving the maximum MAP probability of the combined design.

We recommend only using this criterion to compare follow-up designs of the same size, because we do not know how well this criterion will perform when comparing the maximum simulated MAP probability of two designs of different sizes. When there are multiple follow-up designs with maximum simulated MAP probabilities that are close to the highest, it would be beneficial for the practitioner to inspect the distribution visually through the histograms in order to further discriminate. According to our maximax criterion 16-run and 8-run MD-Optimal designs are the best choice among all 16-run and 8-run follow-up designs that we have considered. We also conclude that the third ranked 4-run MD-Optimal design is preferred among all 4-run designs considered.



(a) 4-Run MD-Optimal Design 2



(a) 4-Run MD-Optimal Design 5

Figure 3.3 Injection Molding Experiment 4-run Follow-Up Design Histograms

Chapter 4

Examples

4.1 Acid Leaching Experiment

This example uses a 12-run Plackett-Burman (Plackett and Burman (1946)) design with 7 factors taken from Mee (2009). The goal of this experiment is to optimize the acid leaching step for determining trace amounts of manganese in seafood products, and for our analysis, we set the desired range of the response variable to 90 or more units of manganese. The seven experimental factors are A = nitric acid concentration, B = hydrochloric acid concentration, C = hydrogen peroxide concentration, D = acid solvent volume, E = ultrasonic water-bath temperature, F = ultrasound exposure time and G = mussel particle size. The data from this experiment can be found in Table 4.1.

Table 4.1 Acid Leaching Experiment

A	B	C	D	E	F	G	MN
1	-1	1	-1	-1	-1	1	100.4
1	1	-1	1	-1	-1	-1	86.9
-1	1	1	-1	1	-1	-1	105
1	-1	1	1	-1	1	-1	59.8
1	1	-1	1	1	-1	1	107.2
1	1	1	-1	1	1	-1	87.8
-1	1	1	1	-1	1	1	88
-1	-1	1	1	1	-1	1	67.5
-1	-1	-1	1	1	1	-1	34.6
1	-1	-1	-1	1	1	1	106.6
-1	1	-1	-1	-1	1	1	104.2
-1	-1	-1	-1	-1	-1	-1	46.6

After the initial experiment, the optimum factor setting led an MAP probability of producing a response in the desired range of 90 or more units of manganese is .6791. The hyperparameters were set to π=*.25* and γ = *.832*, and we found that 17 different models had model posterior probabilities greater than .01. The top five models and their posterior probabilities are displayed in Table 4.2. All 17 models are used in the follow-up design analysis.

Table 4.2 Acid Leaching Experiment Model Posterior Probabilities

Factors	$P(M\|\mathbf{y})$
A, *B* and *G*	0.324
B and *G*	0.212
None	0.098
B	0.07
A and *B*	0.056

According the minimum aberration criterion, the full foldover is the only optimal foldover for a 12-run Plackett-Burman design. The minimal dependents sets criterion shows that folding over on all factors and subsetting on factor *A*, *B*, *C*, *D*, *E* or *F* will lead to an optimal semifoldover. We consider the full foldover, the 12 optimal semifoldovers, the top five 3-run MD-Optimal designs and an MD-Optimal designs with 12 and 6 runs in our analysis. Tables 4.3 through 4.5 present the simulated MAP probabilities for each of these candidate follow-up designs.

Table 4.3 Full Foldover, Top 12-Run MD-Optimal and Top Five 3-Run MD-Optimal Designs

	Full	MD-12run	MD1	MD2	MD3	MD4	MD5
Mean	0.7034	0.7096	0.6977	0.6972	0.6965	0.6899	0.694
Min	0.2982	0.3095	0.3964	0.3825	0.45	0.3609	0.3858
Max	0.9931	0.9752	0.9101	0.9097	0.891	0.8512	0.8941
Q1	0.5606	0.576	0.5965	0.587	0.5884	0.6043	0.6066
Q3	0.8554	0.8636	0.8036	0.808	0.7962	0.7861	0.7857
Median	0.717	0.7234	0.7093	0.7114	0.727	0.721	0.7054
Std Dev	0.173	0.1712	0.1282	0.1315	0.1175	0.1157	0.1092

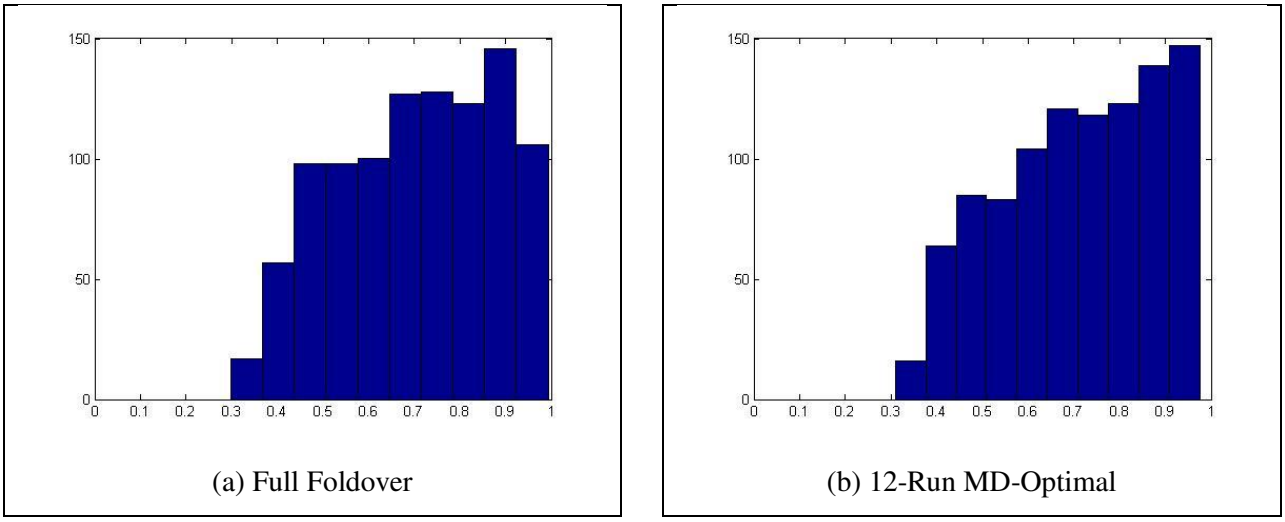
Table 4.4 Semifoldovers

	A-	A+	B-	B+	C-	C+
Mean	0.6965	0.6969	0.6928	0.7081	0.7115	0.6896
Min	0.3617	0.3411	0.4279	0.3169	0.3347	0.365
Max	0.963	0.9739	0.9622	0.9633	0.9614	0.9647
Q1	0.5905	0.5766	0.5585	0.5997	0.604	0.5626
Q3	0.8098	0.8141	0.8132	0.8295	0.8366	0.8166
Median	0.7155	0.6972	0.6939	0.7306	0.7333	0.6939
Std Dev	0.1391	0.1488	0.1463	0.1457	0.1504	0.149

Table 4.5 Additional Semifoldovers and Top 6-Run MD-Optimal Design

	D-	D+	E-	E+	F-	F+	MD-6run
Mean	0.7023	0.7065	0.7119	0.7076	0.7014	0.7088	0.7037
Min	0.3142	0.3824	0.3215	0.3139	0.3316	0.34	0.3518
Max	0.9613	0.9657	0.9655	0.9597	0.9573	0.9599	0.9445
Q1	0.5852	0.5913	0.5871	0.5984	0.5939	0.5926	0.5884
Q3	0.8281	0.8326	0.8494	0.8275	0.8333	0.8378	0.827
Median	0.7261	0.7157	0.7219	0.7299	0.7191	0.7156	0.7302
Std Dev	0.1548	0.1534	0.1574	0.1483	0.1561	0.1511	0.144

The results in Table 4.3 indicate that there is not much difference between the MAP probability distributions of the full foldover and the 12-run MD-Optimal design, but the descriptive statistics do not always paint the clearest picture. The histograms in Figure 4.1 show that the 12-run MD-Optimal design has a greater proportion of MAP probabilities above .8 and is therefore a more preferable choice. Although the difference looks subtle, the histograms show that the MD design has around 35 more MAP probabilities in the highest grouping of around .9 or greater. Both 12-run designs dominate the 3-run designs in both the maximum and 3rd quartile values and show a slight advantage over the 6-run designs. Based on our maximax criterion, I would use the full foldover rather than the 12-run MD-Optimal design.



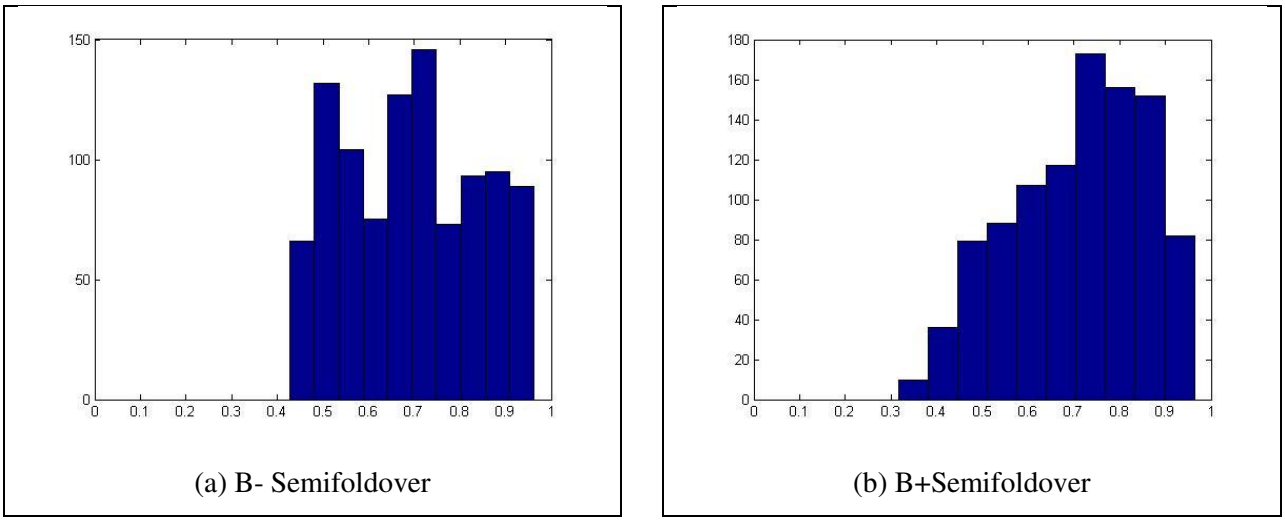
(a) Full Foldover (b) 12-Run MD-Optimal

Figure 4.1 Acid Leaching Experiment 12-Run Follow-Up Design Histograms

Much like the results of the 12-run follow-up designs, the descriptive statistics for the semifoldovers in Tables 4.4 and 4.5 do very little to differentiate these follow-up designs. The histograms, once again, are able to detect differences in the distributions of the MAP probabilities that

were not apparent in summary statistics. The A+, B- and C+ semifoldovers all lead to distributions with a lower proportion of MAP probabilities on the higher end of the spectrum. These semifoldovers produce the three lowest medians, and the B- semifoldover has by far the highest minimum. Figure 4.2 shows that although the B- semifoldover has a much higher minimum, the B+ semifoldover has a definitively larger proportion of high MAP probabilities. The maximax criterion points to the A+ semifoldover as the preferred 6-run follow up design, but most of the maximum simulated MAP probabilities for these designs are close to that of the A+ semifoldover. This suggests that the choice of 6-run follow-up design may not influence the reliability of the optimum that the combined design locates.

Neither the descriptive statistics nor histograms are able to further discriminate between the top five 3-run MD-Optimal designs. The maximax criterion points to the top ranked 3-run MD-Optimal (MD1) design, but once again, the maximum simulated MAP probabilities of the other designs of this size are fairly close. In this situation, it may be in the practitioner's best interest to use the MD3 optimal instead design as it has a high maximum simulated MAP probability, and a much higher minimum simulated MAP probability than the other 3-run designs.



(a) B- Semifoldover

(b) B+Semifoldover

Figure 4.2 Acid Leaching Experiment 6-Run Follow-Up Design Histograms

4.2 Heat Treating Experiment

Next we analyze a 2_{IV}^{6-2} heat treating experiment from Montgomery (2009) that investigates the carbonization of metal parts. The response variable of interest is pitch thickness, and a thickness between 160 and 180 is desired by the practitioners. The experimental factors are *A* = carbon concentration, *B* = cycle time, *C* = furnace temperature, *D* = duration of carbonization cycle, *E* = carbon concentration of the

diffuse cycle and F = duration of the diffuse cycle, and the defining relation for this design is $I = A*B*C*E = A*B*D*F = C*D*E*F$. The data for this experiment is in Table 4.6.

Table 4.6 Heat Treating Experiment

A	B	C	D	E	F	Y
-1	-1	-1	-1	-1	-1	74
-1	-1	-1	1	-1	1	121
-1	-1	1	-1	1	-1	190
-1	-1	1	1	1	1	188
-1	1	-1	-1	1	1	133
-1	1	-1	1	1	-1	135
-1	1	1	-1	-1	1	127
-1	1	1	1	-1	-1	170
1	-1	-1	-1	1	1	115
1	-1	-1	1	1	-1	126
1	-1	1	-1	-1	1	101
1	-1	1	1	-1	-1	175
1	1	-1	-1	-1	-1	54
1	1	-1	1	-1	1	126
1	1	1	-1	1	-1	144
1	1	1	1	1	1	193

The maximum MAP probability of producing a pitch with a thickness between 160 and 180 is only .4325 after the initial experiment which means that there is a great deal of room for improvement. We chose $\pi = .25$ and $\gamma = 1$ as our hyperparameters and calculated that there 8 models with a posterior probability greater than .01. All these models and their posteriors can be found in Table 4.7. In this example, the most probable model has a posterior probability of .667 which means it will most likely dominate the analysis.

Table 4.7 Heat Treating Experiment Model Posterior Probabilities

Factors	$P(M\|\mathbf{y})$
C, D and E	0.667
A, C, D and E	0.086
C	0.082
C and D	0.042
C, D, E and F	0.027
C and E	0.024
None	0.024
D	0.011

A foldover on factors E, F and both factors simultaneously will lead to an optimal foldover which includes every single core foldover. This means that the minimum aberration criterion has not pared down the candidate list of foldovers at all. There are 12 optimal semifoldovers for this design according to the minimal dependent sets criterion, and they can be formed by folding over on factor A and subsetting on any of the six factors. We applied our methodology to the 3 foldovers, 12 semifoldovers and several different MD-Optimal designs as well, and the results can be seen in Tables 4.8 through 4.10.

Table 4.8 Foldovers, Top 16-Run MD-Optimal and Top Five 4-Run MD-Optimal Designs

	E	F	EF	MD-16Run	MD1	MD2	MD3	MD4	MD5
Mean	0.4323	0.4229	0.438	0.4133	0.4213	0.4206	0.4249	0.4261	0.4199
Min	0.1166	0.1165	0.1166	0.1199	0.1203	0.1474	0.1236	0.1519	0.142
Max	0.7328	0.7053	0.7327	0.6982	0.5627	0.562	0.5625	0.5643	0.5618
Q1	0.3994	0.3732	0.4095	0.3654	0.3723	0.3711	0.3904	0.3816	0.3748
Q3	0.5059	0.5015	0.5061	0.487	0.4953	0.4941	0.4949	0.4975	0.4917
Median	0.461	0.4544	0.4632	0.4418	0.457	0.4567	0.4566	0.4602	0.4547
Std Dev	0.1185	0.1184	0.1137	0.1102	0.1005	0.1003	0.0977	0.0983	0.0991

Table 4.9 Semifoldovers

	A-	A+	B-	B+	C-	C+	D-	D+
Mean	0.4417	0.4314	0.4439	0.4398	0.4344	0.4296	0.4301	0.4277
Min	0.1206	0.1193	0.1069	0.114	0.1314	0.1182	0.1297	0.1131
Max	0.6324	0.6569	0.6481	0.6446	0.6517	0.6819	0.6786	0.6196
Q1	0.411	0.3915	0.4112	0.4111	0.3952	0.3831	0.3977	0.3786
Q3	0.5096	0.5047	0.5139	0.5072	0.5018	0.5097	0.5007	0.5065
Median	0.4715	0.466	0.4751	0.4699	0.4597	0.4643	0.461	0.4573
Std Dev	0.104	0.1125	0.1096	0.1055	0.1026	0.119	0.1121	0.1097

Table 4.10 Additional Semifoldovers and Top 8-Run MD-Optimal Design

	E-	E+	F-	F+	MD8-Run
Mean	0.4298	0.4372	0.4317	0.4324	0.4148
Min	0.1085	0.1174	0.1128	0.1175	0.1011
Max	0.6576	0.6713	0.6516	0.6761	0.6699
Q1	0.3901	0.3869	0.3836	0.3903	0.3631
Q3	0.503	0.5108	0.507	0.508	0.4881
Median	0.4589	0.4678	0.463	0.4678	0.447
Std Dev	0.1094	0.1111	0.1128	0.1135	0.1074

The descriptive statistics in Tables 4.8 through 4.10 show that the larger follow-up designs have a greater potential for finding an optimum with a high MAP probability of falling in the desired range, but

these summary statistics fail to discriminate between designs of the same size. The histograms in Figure 4.3 show that the E+ semifoldover may have a slight advantage over the D+ semifoldover, but for the most part, the histograms fail to show any noticeable differences among designs of the same size.

The maximax criterion selects the E foldover, the C+ semifoldover and the MD4 design as the best choice of 16-run, 8-run and 4-run follow-up designs respectively. The maximum simulated MAP probabilities are all very close together for follow-up designs of the same run size, so there may not be a major advantage to picking these designs over the others. As a result, we conclude that for a given run size, the selection of follow-up design should not impact the reliability of the optimum.

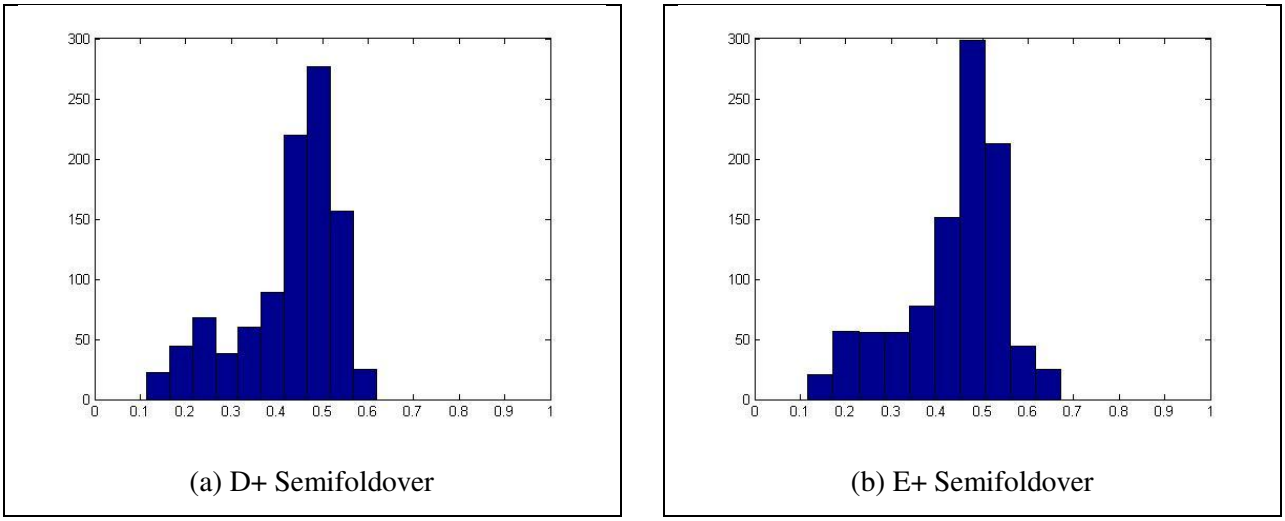


Figure 4.3 Heat Treating Experiment 8-Run Follow-Up Design Histograms

It should be noted that the maximum MAP probabilities for all the designs in this example are relatively low. A follow-up design with 16 runs will at best find an optimum with an MAP probability of .7328 which would be unacceptable in any industrial setting. These results make the argument that the range of acceptable response values may be too narrow and that this range needs to be extended.

The heat treating experiment provides an example where our technique does not necessarily provide a great deal of discrimination between these designs. The homogeneity in the results could be the result of having one model with a dominant posterior probability or perhaps due to setting a response range that is too conservative.

4.3 Amino Acid Experiment

We pull our last example from a nine-factor full factorial experiment taken from Mee (2008) that investigates the binding process of the neuropeptide substance P systematically by replacing native L-amino acids with D-amino acids in nine positions. The nine experimental factors in this experiment are

categorical variables, but for the sake of our analysis, we assume that they are continuous. The response variable being studied is the percentage of inhibition of a reagent. Higher inhibition values are preferred, so we select a desired response range of 90% inhibition or greater. We filtered out the 32 runs from the full factorial that create the minimum aberration 2_{IV}^{9-4} design with the following generators: $F = BCDE$, $G = ACDE$, $H = ABDE$ and $J = ABCF$. We treat this 32-run design, seen in Table 4.11, as our initial experiment for this analysis.

Table 4.11 Amino Acid Experiment

A	B	C	D	E	F	G	H	J	Y
-1	-1	-1	-1	-1	1	1	1	1	96
1	-1	-1	-1	-1	1	-1	-1	-1	8
-1	1	-1	-1	-1	-1	1	-1	-1	14
1	1	-1	-1	-1	-1	-1	1	1	9
-1	-1	1	-1	-1	-1	-1	1	-1	0
1	-1	1	-1	-1	-1	1	-1	1	23
-1	1	1	-1	-1	1	-1	-1	1	5
1	1	1	-1	-1	1	1	1	-1	29
-1	-1	-1	1	-1	-1	-1	-1	1	8
1	-1	-1	1	-1	-1	1	1	-1	14
-1	1	-1	1	-1	1	-1	1	-1	0
1	1	-1	1	-1	1	1	-1	1	43
-1	-1	1	1	-1	1	1	-1	-1	27
1	-1	1	1	-1	1	-1	1	1	18
-1	1	1	1	-1	-1	1	1	1	75
1	1	1	1	-1	-1	-1	-1	-1	18
-1	-1	-1	-1	1	-1	-1	-1	-1	21
1	-1	-1	-1	1	-1	1	1	1	0
-1	1	-1	-1	1	1	-1	1	1	62
1	1	-1	-1	1	1	1	-1	-1	20
-1	-1	1	-1	1	1	1	-1	1	34
1	-1	1	-1	1	1	-1	1	-1	22
-1	1	1	-1	1	-1	1	1	-1	0
1	1	1	-1	1	-1	-1	-1	1	16
-1	-1	-1	1	1	1	1	1	-1	81
1	-1	-1	1	1	1	-1	-1	1	3
-1	1	-1	1	1	-1	1	-1	1	22
1	1	-1	1	1	-1	-1	1	-1	9
-1	-1	1	1	1	-1	-1	1	1	17
1	-1	1	1	1	-1	1	-1	-1	10
-1	1	1	1	1	1	-1	-1	-1	0
1	1	1	1	1	1	1	1	1	99

The hyperparameters were set to $\pi = .25$ and $\gamma = .365$ in our analysis, and we found that the maximum MAP probability is only .2585 after the initial experiment. There are 20 models with posterior probabilities greater than .01 for the data in Table 4.11 that will be included in our model averaging process. The top five models and their posterior probabilities are displayed in Table 4.12. The top two models appear to dominate although their posterior probabilities are relatively low.

Table 4.12 Amino Acid Experiment Model Posterior Probabilities

Factors	$P(M\|\mathbf{y})$
F, G, H and J	0.206
F, G and H	0.169
F and G	0.072
G	0.049
D, F, G, H and J	0.044

Foldovers on the combination of factors FG, FH, FJ, GH, GJ and HJ are all minimum aberration and therefore optimal. Based on the minimally dependent sets criteria, there are 16 optimal semifoldovers that can be produced by first folding over on factors F and G then subsetting on any factor other than E. We also included the top five MD-Optimal designs with only 8 runs, as well as a 16-run and 32-run MD-Optimal design in our analysis. The descriptive statistics from all these simulations are displayed in Tables 4.13 through 4.16.

Table 4.13 Foldovers and Top 32-Run MD-Optimal Design

	FG	FH	FJ	GH	GJ	HJ	MD 32-Run
Mean	0.2612	0.2695	0.2603	0.2665	0.2597	0.2688	0.2499
Min	0.0131	0.0121	0.014	0.0115	0.0098	0.0115	0.013
Max	0.8419	0.8336	0.8339	0.8306	0.8579	0.8521	0.7559
Q1	0.0824	0.0872	0.0839	0.0793	0.08	0.0992	0.0879
Q3	0.4224	0.4291	0.4026	0.4274	0.4053	0.4154	0.3888
Median	0.2089	0.2355	0.2205	0.2218	0.2152	0.2245	0.2161
Std Dev	0.2004	0.1973	0.1954	0.2023	0.1978	0.1942	0.1803

Table 4.14 Semifoldovers

	A-	A+	B-	B+	C-	C+	D-	D+
Mean	0.2718	0.2664	0.2668	0.2652	0.2655	0.2617	0.2595	0.2656
Min	0.0119	0.016	0.015	0.0153	0.0172	0.015	0.0204	0.0137
Max	0.7686	0.7068	0.8423	0.7193	0.7733	0.8303	0.7186	0.7999
Q1	0.1072	0.1036	0.104	0.1132	0.1035	0.1001	0.1154	0.0962
Q3	0.4216	0.4144	0.4004	0.4044	0.4099	0.3975	0.3939	0.4187
Median	0.2385	0.2321	0.2319	0.2499	0.2403	0.236	0.2309	0.2339
Std Dev	0.1838	0.1797	0.1811	0.1691	0.1762	0.1796	0.1679	0.183

Table 4.15 Additional Semifoldovers and Top 16-run MD-Optimal Design

	F-	F+	G-	G+	H-	H+	MD 16-run
Mean	0.2548	0.2578	0.2573	0.2624	0.2588	0.2594	0.2588
Min	0.0552	0.0136	0.0469	0.0127	0.0416	0.0124	0.0313
Max	0.6727	0.7958	0.6006	0.8174	0.6334	0.9115	0.6257
Q1	0.1326	0.0746	0.1412	0.0908	0.1483	0.0837	0.1383
Q3	0.3647	0.4125	0.3597	0.4039	0.3547	0.4043	0.3689
Median	0.2256	0.2126	0.2405	0.2238	0.2516	0.2154	0.2395
Std Dev	0.1421	0.1959	0.1395	0.1888	0.1323	0.1926	0.1396

Table 4.16 Top Five 8-run MD-Optimal Designs

	MD 1	MD 2	MD 3	MD 4	MD 5
Mean	0.2596	0.2626	0.2524	0.2528	0.2702
Min	0.0488	0.0499	0.0513	0.0237	0.0242
Max	0.5436	0.526	0.524	0.651	0.6979
Q1	0.1566	0.1685	0.1627	0.1128	0.1288
Q3	0.3534	0.3533	0.3366	0.3784	0.3984
Median	0.2562	0.2619	0.2475	0.2385	0.2608
Std Dev	0.1167	0.1116	0.1097	0.154	0.1595

The descriptive statistics show an intriguing relationship between the optimal foldovers and semifoldovers in that several of the semifoldovers preform equally as well as or even better than the much larger foldover designs. The B-, C+ and G+ semifoldovers have maximum MAP probabilities on par with the maximums of the foldovers while the H+ semifoldover's maximum is the largest among all follow-up designs being compared. The histograms of the EF foldover and G+ semifoldover displayed in Figure 4.4 show that the two distributions are virtually indistinguishable which implies that the extra 16 runs in the foldover may be unnecessary.

According to the maximax criterion, the GJ foldover and the H+ semifoldover are the preferred choice among designs of their size. While the H+ semifoldover has a much higher maximum MAP probability than the other 16-run designs, the HJ foldover has a maximum simulated MAP probability that is very close to that of the GJ foldover. As a result both of these foldovers would be good choices.

The summary statistics for the semifoldovers in Tables 4.14 and 4.15, on the other hand, indicate that there are noticeable differences in the MAP probability distributions of these designs. Several of the semifoldovers clearly dominate the rest in terms of their maximum and 3rd quartile MAP probability values, such as B-, C+, D+, F+ and H+. The F-, G- and H- semifoldovers have lower standard deviations which lead to unusually low maximums, while the minimums are only slightly higher than the rest of the semifoldover designs.

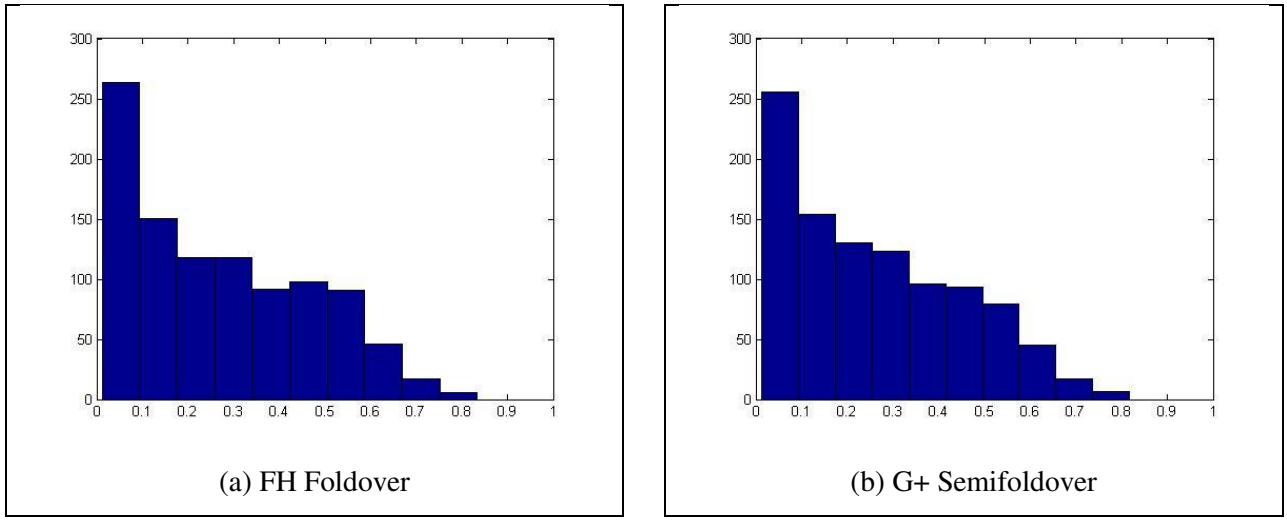
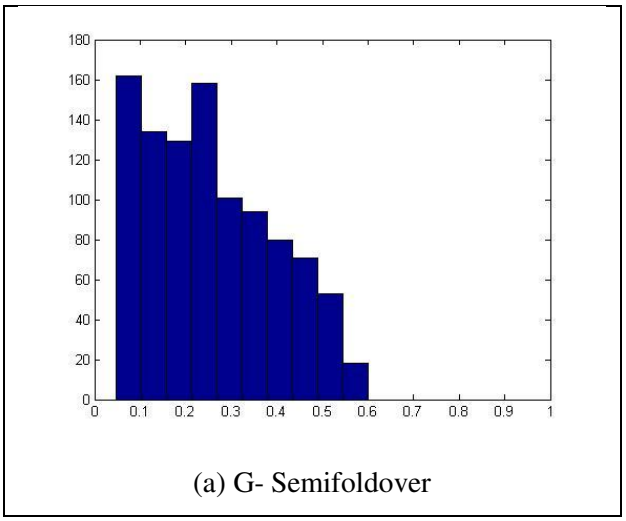
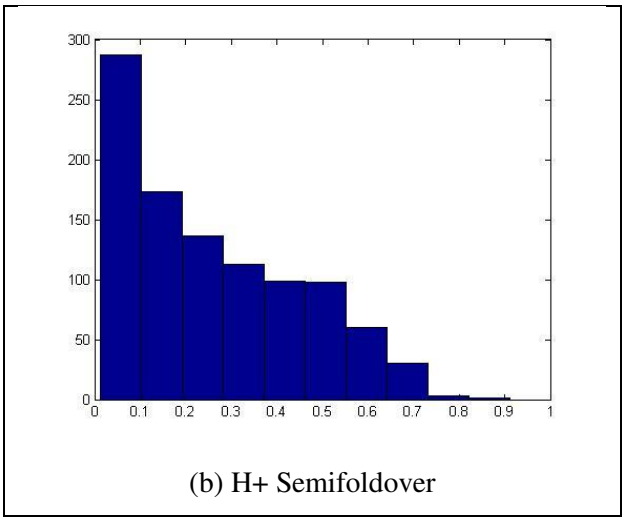


Figure 4.4 Amino Acid Experiment 32 and 16-Run Follow-Up Design Histograms

In this example, the histograms do not highlight the differences in these distributions nearly as well as the descriptive statistics do. Although the statistics in Table 4.15 indicate that the H+ semifoldover is clearly a better option than the G- semifoldover, this conclusion is much more questionable when considering their histograms seen in Figure 4.5. These histograms show that while the H+ semifoldover has a greater proportion of MAP probabilities at the higher end of the spectrum, this semifoldover also has a much larger proportion of low probabilities as well. Despite the questionable results seen in the histograms, we still feel that the H+ semifoldover is a better option due to the presence of much higher MAP probabilities in the results. Fortunately, this design was sampled from a full factorial, so we are able to test our hypothesis empirically by calculating the maximum MAP probability when the actual response values that were produced for these follow-up runs are considered.



(a) G- Semifoldover



(b) H+ Semifoldover

Figure 4.5 Amino Acid Experiment 16-Run Follow-Up Design Histograms

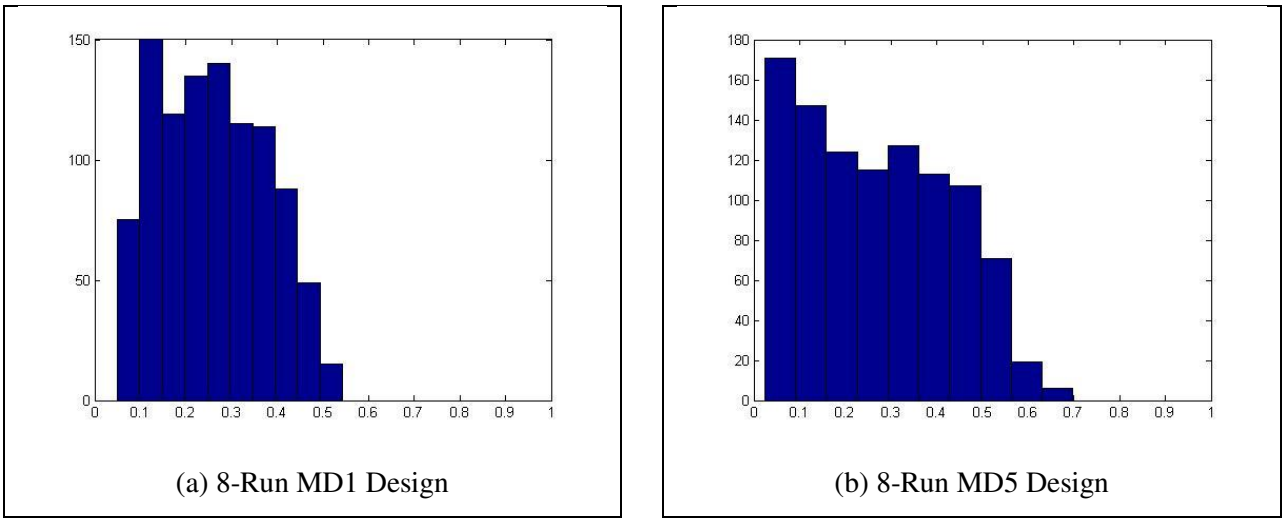
First we combine the results of the initial and follow-up designs from the full factorial experiment and then recalculate the model posterior probabilities for each combined design. Finally, we use these new model posteriors to locate the factor settings associated with the maximum MAP probability of producing a desirable response.

Table 4.17 Actual MAP Probabilities

Design	MAP Probability	Simulated Maximum
Foldover FJ	0.6187	0.8339
Foldover GJ	0.5579	0.8579
Semifoldover A-	0.6503	0.7686
Semifoldover A+	0.5236	0.7068
Semifoldover B-	0.612	0.8423
Semifoldover C+	0.5993	0.8303
Semifoldover F-	0.4451	0.6727
Semifoldover G-	0.4687	0.6006
Semifoldover G+	0.6174	0.8174
Semifoldover H-	0.4416	0.6334
Semifoldover H+	0.6274	0.9115
MD-Optimal 1	0.3792	0.5436
MD-Optimal 2	0.3899	0.526
MD-Optimal 4	0.4748	0.651
MD-Optimal 5	0.5022	0.6979

The actual MAP probabilities for several of the follow-up designs are shown in Table 4.17, and they appear to confirm our hypothesis about the G- and H+ semifoldovers. The output in Table 4.17 also verifies that the best semifoldovers find an equally reliable optimum as the foldovers. These results suggest that the maximax criterion that we have introduced may be a powerful tool for discriminating optimal follow-up designs. We found a highly significant Pearson correlation coefficient (Montgomery (2009)) of $r = .919$ between the real maximum MAP probability and the maximum simulated MAP probability. This provides evidence that the maximax criterion maybe be good predictor of the actual maximum MAP probability that will be found after a follow-up experiment is run.

The summary statistics for the 8-run MD-Optimal designs in Table 4.16 show that the MD4 and MD5 designs have slightly more spread out MAP probabilities with much higher maximums and 3rd quartile values and only slightly lower minimums. The histograms for the 8-run MD-Optimal designs confirm this conclusion, as it can be seen in Figure 4.6 that the MD5 design has a greater proportion of higher MAP probabilities than the MD1 design.



(a) 8-Run MD1 Design

(b) 8-Run MD5 Design

Figure 4.6 Amino Acid Experiment 8-Run Follow-Up Design Histograms

We have shown, once again, that the design with the highest MD-criterion value is not always the best option. The descriptive statistics for the MD-Optimal designs also show that the MD4 and MD5 designs are competitive with the worst of the semifoldover designs. The results in Table 4.17 validate that the MD4 and MD5 designs are much better options than the MD1 follow-up and also that the best 8-run MD-Optimal designs lead to optimums that are competitive with the worst semifoldover choices.

This example demonstrates the ability of our discrimination technique to differentiate between follow-up designs of all run sizes and also provides empirical evidence that this technique can be a powerful instrument for informing the follow-up design selection process.

Chapter 5

V. Conclusion

Sequential experimentation is a widely accepted and practiced process optimization method, and follow-up design selection is vital to the success of the optimization process. We have shown that three widely used follow-up design selection criteria will oftentimes lead to multiple optimal designs which requires that the practitioner selects one arbitrarily. The choice of design can greatly influence the optimization process, but there is currently no framework in place for further discriminating between these optimal follow-up designs.

In this paper, we are able to develop a technique for distinguishing between follow-up designs that are otherwise considered to be equivalent. Our discrimination technique uses Bayesian optimization methods to control for parameter and model uncertainty, and through simulations we are able to identify which follow-up designs are more likely to lead to a reliable optimum. The descriptive statistics and histograms of the simulated optimum MAP probabilities can both be used to differentiate between follow-up designs of the same and different sizes, but we recommend using a maximax criterion to reduce the subjectivity involved in design selection.

The examples in this paper show that our proposed methods do allow further discrimination between optimal follow-up designs, and in the final example, we are even able to demonstrate that our preferred designs tend to lead to better MAP probabilities at their optimum. Our results create a great deal of opportunities in the future. We hope to apply our methodology to other optimal follow-up designs, such as A-optimal and D-optimal designs, and we would also like to adapt out method for three-level designs as well as experiments with multiple response variables. Also, we would like to develop a methodology that does not assume a point estimation of the γ hyperparameter and instead uses a density curve for possible values of γ instead. We feel that it would be extremely beneficial to sample from more full factorials in order to test the efficacy of this discrimination technique in greater depth.

Bibliography

Bayarri, M. J. and Mayoral, A.M. (2002), "Bayesian Designs of "Successful" Replications," *The American Statistician*, 56, 207-214.

Box, G. E. P. and Hill, W. J. (1967), "Discrimination among Mechanistic Models," *Technometrics*, 9, 57-71.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997), "Model Selection: An Integral Part of Inference," *Biometrics*, 53, 603-618.

Del Castillo, E. (2007), *Process Optimization, A Statistical Approach*, New York, NY: Springer Science+Business Media, LLC, 1st ed.

Deng and Tang. (1999), "Minimum G_2-Aberration For Nonregular Fractional Factorial Designs," *Annals of Statistics*, 27, 1,914-1,925.

Edwards, David J. (2011), "Optimal Semifoldover Plans for Two-Level Orthogonal Designs," *Technometrics*, 53, 274-284.

Fries, A. and Hunter, William G. (1980), "Minimum Aberration 2^{k-p} Designs," *Technometrics*, 22, 601-608.

Gilmour, S. G. and Mead, R. (1995), "Stopping Rules for Sequence of Factorial Designs," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44, 343-355.

Kullback, S. and Leibler, R. A. (1951). "On Information and Sufficiency," *Annals of Statistics,* 22, 79-86.

Li, H. and Mee, R. W. (2002), "Better Foldover Fractions for Resolution III 2^{k-p} Designs," *Technometrics*, 44, 278-283.

Li, W. and Lin, Dennis K. J. (2003), "Optimal Foldover Plans for Two Level Fractional Factorial Designs," *Technometrics*, 45, 142-149.

Mee, R. W. (2009), *A Comprehensive Guide to Factorial Two-Level Experimentation*, New York, NY: Springer Science+Business Media, LLC, 1st ed.

Mee, R. W. and Peralta, M. (2000), "Semifolding 2^{k-p} Designs," *Technometrics*. 42, 122-134.

Meyer, R. D., Steinberg, D. M., and Box, G. (1996), "Follow-Up Designs to Resolve Confounding in Multifactor Experiments," *Technometrics*, 38, 303-313.

Miro-Quesada, G., Del Castillo, E., and Peterson, J. J. (2004), "A Bayesian Approach for Multiple Response Surface Optimization in the Presence of Noise Variables," *Journal of Applied Statistics*, 31, 251-270.

Montgomery, D. C. (2009), *Design and Analysis of Experiments*, Hoboken, NJ: John Wiley and Sons, Inc., 7th ed.

Montgomery, D. C. and Runger, G. C. (1996), "Foldovers of 2^{k-p} Resolution IV Experimental Designs," *Journal of Quality Technology*, 28, 446-450.

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

Ng, Szu Hui (2010), "A Bayesian Model Averaging Approach for Multiple-Response Optimization," *Journal of Quality Technology*, 42, 52-68.

Peterson, John J. (2004), "A Posterior Predictive Approach to Multiple Response Surface Optimization," *Journal of Quality Technology*, 36, 139-153.

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

Rajagopal, Ramkumar and Del Castillo, Enrique (2005), "Model-Robust Process Optimization Using Bayesian Model Averaging," *Technometrics*, 47, 152-163.