Optimal Design of Experiments

for Dual-Response Systems

by

Sarah Ellen Burke

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2016 by the
Graduate Supervisory Committee:

Douglas Montgomery, Co-Chair
Connie Borror, Co-Chair
Christine Anderson-Cook
Rong Pan
Rachel Silvestrini

ARIZONA STATE UNIVERSITY

August 2016

ABSTRACT

The majority of research in experimental design has, to date, been focused on designs when there is only one type of response variable under consideration. In a decision-making process, however, relying on only one objective or criterion can lead to oversimplified, sub-optimal decisions that ignore important considerations. Incorporating multiple, and likely competing, objectives is critical during the decision-making process in order to balance the tradeoffs of all potential solutions. Consequently, the problem of constructing a design for an experiment when multiple types of responses are of interest does not have a clear answer, particularly when the response variables have different distributions. Responses with different distributions have different requirements of the design.

Computer-generated optimal designs are popular design choices for less standard scenarios where classical designs are not ideal. This work presents a new approach to experimental designs for dual-response systems. The normal, binomial, and Poisson distributions are considered for the potential responses. Using the D-criterion for the linear model and the Bayesian D-criterion for the nonlinear models, a weighted criterion is implemented in a coordinate-exchange algorithm. The designs are evaluated and compared across different weights. The sensitivity of the designs to the priors supplied in the Bayesian D-criterion is explored in the third chapter of this work.

The final section of this work presents a method for a decision-making process involving multiple objectives. There are situations where a decision-maker is interested in several optimal solutions, not just one. These types of decision processes fall into one of two scenarios: 1) wanting to identify the best N solutions to accomplish a goal or

specific task, or 2) evaluating a decision based on several primary quantitative objectives along with secondary qualitative priorities. Design of experiment selection often involves the second scenario where the goal is to identify several contending solutions using the primary quantitative objectives, and then use the secondary qualitative objectives to guide the final decision. Layered Pareto Fronts can help identify a richer class of contenders to examine more closely. The method is illustrated with a supersaturated screening design example.

Dedicated in memory of Connie M. Borror

Thanks for getting me here!

ACKNOWLEDGMENTS

I would like to give special thanks to my committee, Doug Montgomery, Connie Borror, Christine Anderson-Cook, Rong Pan, and Rachel Silvestrini for their support and guidance as I completed my degree. Thanks to Dan McCarville for his encouragement and for serving as substitute at my defense. Thanks also to Jim Simpson who provided the motivating example for the work in Chapter 1. I would also like to acknowledge the Science of Test consortium for providing the funding for this work.

Thanks to my fellow students and friends, especially Edgar Hassler, Petek Yontay, and Derya Kilinc for their help and their distractions. Finally, thanks to my family: For my parents who supported me while I continued my education and for agreeing to store my stuff for a few extra years. And thanks to my sister Laura who provided me sanity breaks.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

CHAPTER 1

INTRODUCTION

Most real-world problems are complex systems with several input factors along with multiple responses of interest. Identifying the effects of active factors on these multi-response systems is critical in many industrial settings. There has been considerable research on appropriate designs for a response that is normally distributed. Factorial designs, fractional factorial designs, and central composite designs, for example, are all established design choices for a normally distributed response (Montgomery, 2013). Designs for a generalized linear model, for example the design for a logistic regression model of a binary response has received an increasing amount of research (Chaloner and Verdinelli, 1995), particularly as computation power continues to increase. The choice of an ideal design for multi-response systems when the responses have different distributions does not have a clear answer.

Computer-generated optimal designs are popular for less standard scenarios where classical designs are not ideal. For example, optimal designs can be tailored to match the sample size and/or design region required for a specific scenario. Optimal designs for linear models are typically straightforward to find using a coordinate-exchange algorithm (Meyer and Nachtsheim, 1995). Bayesian D-optimal designs were proposed by Chaloner and Larntz (1989) for generalized linear models and can also be found using coordinate-exchange. For multi-response systems, optimal designs can be used to combine optimal criteria for each response of interest.

Chapter 2 introduces a weighted optimality criterion to find optimal designs for dual response systems where one response is assumed to be normally distributed while the other response is binary. Several examples are demonstrated showing the relative efficiency of these designs compared to the D-optimal design for the normal theory model and the Bayesian D-optimal design for the logistic regression model. Several issues are discussed in the development of the criterion, including scaling the criteria, the choice of desirability function (DF), and the likelihood that separation in the logistic regression model occurs.

Chapter 3 extends the work developed in Chapter 2 to include a different distribution in the dual response system: the Poisson distribution. Examples include dual response systems where 1) one response is normally distributed while the other has a Poisson distribution and 2) one response is binary while the other has a Poisson distribution. The effect of the priors specified for the nonlinear models is explored in Chapter 3 as well.

Design selection using multiple criteria is considered in Chapter 4 for an application to select an appropriate 16-run supersaturated design for eight factors. The method extends the work developed in Lu et al. (2011) by implementing a two-phase layered Pareto front algorithm. The case study presented in Chapter 4 is an example of a broader class of problems where the goal is to select several optimal solutions, either to identify the best N solutions for a particular task or to combine both quantitative and qualitative criteria to make a decision. This approach incorporates multiple, and likely competing, objectives in the decision-making process. The goal is to balance the tradeoffs of all potential solutions in order to arrive at an informed decision.

A summary of the contributions of this work is provided in Chapter 5. In addition, several suggestions for future work on designs for multi-response systems are discussed.

CHAPTER 2

OPTIMAL DESIGNS FOR DUAL RESPONSE SYSTEMS: PART I

2.1 Introduction

The majority of research in experimental design has, to date, been focused on designs when there is only one type of response variable under consideration. Consequently, the problem of constructing a design does not have a clear answer when the multiple responses are of interest have different distributions. Consider, for example, planning an experiment for an air-to-ground missile simulation. In this scenario, the goal is to find appropriate settings of several factors in order to simultaneously model the miss distance of the missile from the target (a continuous response, which is not necessarily normally distributed), the time to acquire the target (a continuous response), and survival of the target (a binary response). With limited time and resources and only one experiment possible, the question of selecting an appropriate design to model all responses is critical. A computer-generated optimal design for a linear model for the continuous response may not be the best choice to fit the logistic regression model for the binary response. Johnson & Montgomery (2009), demonstrate that Bayesian D-optimal designs perform considerably better compared to standard designs when fitting the logistic regression model. Similarly, the optimal design for the logistic regression model is typically not an efficient design for the linear model.

In this chapter, the focus is on experiments where there are two responses of interest, one which is assumed to be normally distributed and the other which is binary. A

new design combined criterion is presented for creating computer-generated optimal designs that incorporate information from both the model for a normal response variable and the model for the binary response variable. The proposed method is demonstrated using various weighting schemes for combining the prioritization of the two models to show how the design changes when one response variable is emphasized over the other. Finally, a case study of an air-to-ground missile simulation is used to illustrate the efficiency of this new type of multi-response design in two ways. First we compare the design to a D-optimal design for the normal response model and then to a Bayesian D-optimal design for the binary response model.

2.2 Literature Review

Multi-objective optimization is not a new problem to design of experiments (DOE). Until recently, however, the focus has been on the analysis of multi-response problems. There are many scenarios where the expense of running an experiment suggests that it is reasonable to collect multiple responses, some of which must be optimized. When the number of responses and the number of input factors are not large, graphing techniques can provide solutions to this multiple-response optimization. Overlaid contour plots work well in these cases as was demonstrated by Lind et al. (1960). However, when the number of factors and/or the number of objectives increases, these methods become harder to implement, visualize, and interpret.

Traditional optimization techniques have also been used to identify the optimal solution using an algorithm. Myers and Carter (1973) use a constrained optimization approach for a dual response optimization problem by selecting a primary response for

the objective function and using the remaining response as a constraint. This method uses mathematical techniques to identify Lagrangian multipliers to find solutions that optimize the primary and secondary responses. Del Castillo and Montgomery (1993) demonstrate how to find the optimal settings of a dual response system using a generalized reduced gradient (GRG) algorithm, a common nonlinear programming algorithm. Del Castillo et al. (1997) use constrained optimization as well, but propose an alternative algorithm that guarantees a global optimum rather than a local optimum that can occur using the GRG algorithm in practice.

Variations using these optimization techniques involve minimizing a loss function based on target values for each response. The problem formulation of Tang and Xu (2006) require both the objective function and constraints to be quadratic forms; the GRG algorithm can then be used to find the optimal solution. Costa (2010) presents a method for optimizing the mean and standard deviation of a process simultaneously by combining both functions into one objective function, a weight function that quantifies the distance each response is from a target value. The author then recommends using the GRG algorithm with several random starting points from the designed experiment to avoid a local optimum. Ames et al. (1997) propose a quadratic quality loss function: $\sum_{r=1}^{R} w_r\left(y_r\left(x_1, \dots, x_p\right) - T_r\right)^2$ for R responses where $y_r\left(x_1, \dots, x_p\right)$ is the estimated model for response r, T_r is the target value for response r, and w_r is a weight added to each response to incorporate relative importance of each responsePignatiello (1993) and Vining (1998) also propose quadratic loss functions; however, their methods incorporate correlation structure between the responses. Pignatiello (1993) proposes minimizing an expected loss function where $loss\left(\boldsymbol{y}(\boldsymbol{x})\right) = (\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\tau})'\boldsymbol{C}(\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\tau})$, where $\boldsymbol{\tau}$ is a

vector of target values for each response and $\boldsymbol{C}$ is a positive-definite cost matrix. By minimizing the expected value of this loss function, the author incorporates any correlations between responses. Vining (1998) proposes a similar loss function as Pignatiello, however uses the covariance matrix of predicted responses to incorporate the quality of predictions from the response models.

Ortiz et al. (2004) provide an alternative to nonlinear programming techniques using a genetic algorithm. The authors argue that for large problems (large number of responses and/or factors), traditional nonlinear programming methods are more likely to lead to locally optimal solutions. For these more complex cases, the genetic algorithm performs better than the GRG approach.

Many of the analysis-focused optimization methods consider continuous optimization problems. In cases like design of experiments where we are searching for points within the design space, optimization methods are more challenging. For example, we can use genetic algorithms to find designs (see, for example, Lin et al., 2015); however, these methods can be difficult to implement effectively and are computationally intensive. Using a genetic algorithm successfully requires development of the relevant chromosomal representation which can be challenging. In addition, deciding on an advantageous method for crossover and mutation operations in the genetic algorithm that matches how we expect improvements in the design to occur affect the both the quality of the resulting design and the computation time (Lin et al., 2015).

A common method for combining objectives into one overall score is through desirability functions (DFs), introduced by Derringer and Suich (1980). When the responses are on different measuring scales, this approach allows each criterion to be

scaled to values in the range [0,1] and then more fairly combined using one of several different functional forms. Many of the common choices for these functions used in practice, which include the additive and multiplicative DF, contain points where the first derivative does not exist. In these cases, gradient optimization methods, which require the objective function to have a continuous first derivative, cannot be used. Del Castillo et al. (1996) propose a differentiable DF approach so that gradient-based optimization methods can be used. The DF approach for combining multiple objectives is incorporated in the design criterion for this application, comparing the resulting designs from both the additive and multiplicative DF.

Khuri and Conlon (1981) propose a distance metric for multiple responses and implement an algorithm that minimizes the maximum distance to identify optimal settings. However, while this approach accounts for potential linear dependencies between the multiple responses, it does not account for differences in the measurement scales of the criteria. The method further assumes that the multiple responses are fit using a multivariate normal model. Being able to account for responses or criteria that are not continuous or do not follow the same distribution needs further consideration.

In the previous research discussed, no recommendations are given on the experimental design used to perform these experiments and to fit the multiple response models. The emphasis in these papers is to discuss techniques to find the optimal settings of a system that meet requirements of all the responses after the data have been collected. Much of the literature available on the experimental designs for multi-response systems focuses on scenarios where there is only one factor in the experiment. Draper and Hunter (1967) discuss the problem for multi-response systems for nonlinear models. The authors

use a two-stage, iterative approach to incorporate prior distributions of the parameters. However, only one factor is in the model. Similarly, Chang et al. (2001) propose an optimality criterion for a system with k responses, but only one experimental factor. Their optimality criterion further assumes a known covariance matrix of the responses. Additional methods for multi-response optimization can be found in Carlyle et al. (2000) and in Chapter 7 of Myers et al. 2016).

One area of research where designs for multi-response systems has received more consideration is in clinical trials where efficacy and toxicity of a drug are typically the two responses of interest. Several authors have discussed designs for dose-response studies in the case where two or more responses are considered. The goal of these types of studies is to determine the levels of the drug doses in a clinical trial. Therefore, these types of designs generally only have one factor. Heise and Myers (1996) consider D-optimal and I-optimal designs for dose-response studies where the responses are assumed to have a bivariate logistic model. This model allows the authors to incorporate correlations between the responses into the design for the dose-response experiment. More recently, Fedorov et al. (2012) consider two binary variables as the responses in a dose-response design. The authors assume that the responses initially have a bivariate normal distribution and can therefore incorporate the correlation between the two responses into the optimality criterion. Gueorguieva et al. (2006) present another application of multi-response problems in the pharmacokinetic field. Their study was longitudinal and they created a design to find the optimal times to collect data on patients using a locally optimal design.

Recently, another approach for creating designs using multiple objectives is through the use of a Pareto front. Lu et al. (2011) and Lu et al. (2014) demonstrate a two-stage procedure to construct designs optimal for multiple criteria. In the first stage, a collection of non-inferior Pareto-optimal designs is found. This stage does not take into account any prioritization of design criteria. The second stage consists of graphical tools to guide users to a design choice that provides acceptable design performance and simultaneously meets all the desired criteria. The algorithm uses a Pareto Aggregating Point Exchange (PAPE) algorithm to search through potential designs. The authors consider a scenario where two criteria are used for constructing the designs that have precise estimation of model parameters and protection from model misspecification. The two criteria used, therefore, are to maximize the determinant of the information matrix ($|\mathbf{X}'\mathbf{X}|$) and minimize $tr(\mathbf{A}\mathbf{A}')$, where **X** is the model matrix and **A** is the alias matrix of the design. The literature does not consider the case for multi-response systems with multiple factors of interest. This work addresses this problem using a weighted optimality criterion.

2.3 Methods

2.3.1 Linear Models and Generalized Linear Models

In this approach to constructing computer-generated optimal designs for a dual response system, one response variable (y_1) is assumed to follow a normal distribution while the second response (y_2) follows a binomial distribution. For the designs considered, higher-order terms (third order or higher) are assumed to be negligible so that

the models may include main effects, two-factor interactions, or quadratic terms. For the linear model and k factors, the model at most will have the form:

$$\boldsymbol{y_1} = \mathbf{X}\boldsymbol{\beta} = \beta_0 + \sum_{i=1}^{k} \beta_i \boldsymbol{x}_i + \sum_{i<j} \sum_{j=2}^{k} \beta_{ij} \boldsymbol{x}_i \boldsymbol{x}_j + \sum_{i=1}^{k} \beta_{ii} \boldsymbol{x}_i^2, \tag{2.1}$$

where $\mathbf{X}$ is an $(n \times p)$-model matrix (the design matrix expanded to model form), $\boldsymbol{\beta}$ is a $(p \times 1)$-vector containing the model coefficients, n is the number of runs in the design, and p is the number of parameters in the model. If the model contains only main effects, $p = k + 1$; if the model contains main effects plus all two-factor interactions, $p = 1 + \frac{k(k-1)}{2}$; and if the model is a full quadratic model, $p = 1 + 2k + \frac{k(k-1)}{2}$.

Generalized linear models (GLMs) extend from normal-theory linear models and are made up of three key components: 1) a response variable in the exponential family, 2) a linear predictor $\boldsymbol{\eta} = \boldsymbol{x}'\boldsymbol{\beta}$, and 3) a link function g such that $\boldsymbol{\eta} = g(\boldsymbol{\mu})$ (Myers et al., 2012 p. 206). In the case that the response variable $\boldsymbol{y_2}$ is binary, we assume that it follows a binomial distribution with parameters n and $\pi(\boldsymbol{x})$. The canonical link function, the logistic link, is defined as:

$$\eta = \ln\left(\frac{\pi(\boldsymbol{x})}{1 - \pi(\boldsymbol{x})}\right). \tag{2.2}$$

Therefore,

$$\mathrm{E}[\boldsymbol{y_2}] = \frac{e^{\boldsymbol{x}'\boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}'\boldsymbol{\beta}}}. \tag{2.3}$$

A GLM for a binary response variable and logistic link function is also called a logistic regression model.

2.3.2 D-Optimal Designs

Computer generated optimal designs are popular choices for designs when classical designs (e.g. fractional factorial, central composite designs, etc.) are not ideal. Optimal designs have greater flexibility and can be tailored to match the sample size and/or design region required for a specific scenario. The D-optimality criterion minimizes the generalized variance of the estimated model (Montgomery, 2013). This criterion is a popular choice for optimal experimental designs because of its advantageous properties regarding parameter estimation. It is therefore ideal for screening designs where the goal of the experiment is to identify all active effects on the response. Equivalently, a D-optimal design is one in which the determinant of the information matrix is maximized. When we consider a linear model, the information matrix reduces to $\mathbf{X}'\mathbf{X}$. The D-optimality criterion is then defined as:

$$\varphi_N = |\mathbf{X}'\mathbf{X}|. \tag{2.4}$$

This criterion is not dependent on the model parameters and, therefore, constructing a D-optimal design for a linear model with known design region is fairly straightforward. One of the most widely used algorithms for constructing optimal designs is the coordinate-exchange algorithm (Meyer & Nachtsheim, 1995) which is implemented in many software packages to create D-optimal designs.

For this dual response system, we must also consider the D-criterion for the binary response model. As shown in section 2.3.1, the logistic regression model is nonlinear. The D-criterion for nonlinear models is often based on the asymptotic information matrix **I** for this nonlinear model, where **I** is defined as:

$$\mathbf{I} = \mathbf{X}'\mathbf{V}\mathbf{X}. \tag{2.5}$$

$\mathbf{V}$ is an $n \times n$ diagonal matrix that is a function of the mean μ and depends on the distribution of the response variable. For the binary response model, this weight matrix is defined as (Myers et al., 2012):

$$\mathbf{V}_{\mathrm{B}_{ii}} = \frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{\left(1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\right)^2}. \tag{2.6}$$

The information matrix for this nonlinear model is dependent on the model coefficients, making the D-optimality criterion ($|\mathbf{X}'\mathbf{V}\mathbf{X}|$) more difficult to determine. Chernoff (1953) introduced the concept of locally optimal designs for GLMs, where historical information or a best guess for the values of the unknown model parameter values are used to estimate the optimality criterion. However, these locally-optimal designs are sensitive to the parameter values selected. A natural extension of using a single point estimator is to place a prior distribution on the unknown parameters. These types of designs are called Bayesian D-optimal designs and were first introduced by Chaloner and Larntz (1989) for a single factor logistic regression model. The design criterion for a Bayesian D-optimal design is:

$$\varphi_{\mathrm{B}} = \int \log|\mathbf{X}'\mathbf{V}_{\mathrm{B}}\mathbf{X}|\, f(\boldsymbol{\beta}) d\boldsymbol{\beta}, \tag{2.7}$$

where $f(\boldsymbol{\beta})$ is a specified prior distribution of the model parameters. Bayesian D-optimal designs were historically difficult to implement in practice, particularly for designs with more than one factor because of the computational intensive nature of the design criterion. The p-dimensional integral in φ_{B} must be calculated many times in a typical coordinate exchange algorithm when constructing the optimal design. Gotwalt et al. (2009) introduced a radial-spherical integration technique to approximate φ_{B} with high

numerical accuracy. Combining several integral numerical approximation techniques, the algorithm allows for faster numerical approximations of the optimality criterion for a GLM. This method is implemented in JMP 12 and users can specify uniform, normal, lognormal, and exponential priors on the model parameters. For the remainder of this paper, we say D-optimal design to refer to the optimal design for the linear, normal-theory model. We say binary optimal design to refer to the Bayesian D-optimal design for the binomial response model.

2.3.3 Proposed Criterion

To create a computer-generated optimal design that can be used simultaneously to estimate a linear model and a nonlinear, logistic regression model, we propose a weighted optimality criterion that combines both φ_N and φ_B. For this weighted criterion, we consider the resulting designs for varying values of w_N, the weight placed on the criterion for the linear model. This weight may represent the priority the experimenter has of one model over the other. Because φ_{N} and φ_{B} are on different measurement scales, we first scale the criteria to make them more comparable when they are combined into one overall optimality score. The scaled criteria are determined by first finding the optimal design for each of the models of responses y_1 and y_2 independently. The design criteria φ_N and φ_B are then used to scale the criteria for a design with a given value of w_N. For the linear model, the scaled criterion $\tilde{\varphi}_N$ for a given value of the weight w_N is divided by the value of the criterion for a D-optimal design for a linear model (i.e., the D-criterion when $w_N = 1$):

$$\tilde{\varphi}_N = \frac{\varphi_{N,w_N}}{\varphi_{N,w_N=1}}. \quad (2.8)$$

For the Bayesian D-optimal criterion, the scaled criterion $\tilde{\varphi}_{\mathrm{B}}$ is similarly defined by scaling by the value of the Bayesian D-optimal criterion for the binary model. Because the values are negative, we use the following ratio so that the scaled criterion $\tilde{\varphi}_B$ is optimized by maximizing:

$$\tilde{\varphi}_B = \frac{\varphi_{\mathrm{B},w_N=0}}{\varphi_{B,w_N}}. \quad (2.9)$$

These scaled criteria in (2.8) and (2.9) are now both in the range $[0,1]$ and can be more fairly combined using a DF so that the proposed criterion incorporates information from both the normal response model and the binary response model. The form of the DF could take on one of several forms. We define our optimality criterion using the multiplicative DF such that:

$$\varphi_{NB} = \left(\tilde{\varphi}_N^{w_N}\right)\left(\tilde{\varphi}_{\mathrm{B}}^{(1-w_N)}\right), \quad (2.10)$$

where w_N is a weight between 0 and 1. This weight allows the user to customize the design in the case where one response is deemed more important than the other. Note that a weight value of 1 places all of the weight on the linear response and the resulting design is a standard D-optimal design. Conversely, a weight value of 0 places all of the weight on the criterion for the binary response and the resulting design is a binary optimal design. Other forms of the DF considered include the additive DF such that:

$$\varphi_A = w_N\tilde{\varphi}_N + (1-w_N)\tilde{\varphi}_{\mathrm{B}}. \quad (2.11)$$

We also consider another form of the additive DF, but scale the criterion for the linear model differently. Note that in (2.7), the Bayesian D-criterion is defined on a log scale.

To account for this scaling, we consider an additive DF, but perform a log transformation on φ_{N} and do not perform further scaling on φ_B. This criterion, therefore, has the form:

$$\varphi_L = w_N \log \varphi_N + (1 - w_N)\varphi_B. \tag{2.12}$$

The results of using the proposed criterion in (2.10) are briefly compared to the alternative criteria (2.11) and (2.12) in section 2.4. The choice of DF and scaling may have important effects on the resulting designs. In general, the multiplicative DF tends to penalize solutions where one criterion has worse performance compared to the other criterion. This DF is naturally advantageous in this setting as the goal is to find a design that works well simultaneously for both the linear and nonlinear models. A design that performs well for one model, but not the other would not be an ideal solution. The additive DF, on the other hand, is more forgiving when one criterion is worse than the other. A high criterion value for one model can compensate poor performance of the other criterion. We demonstrate differences between the optimal designs using these different choices in scaling and DF (results using φ_A and φ_L are shown in Appendix A).

To evaluate these weighted designs, we first compare the proposed designs in terms of design efficiency. We compare the new design to the performance of both D-optimal designs and binary optimal designs. The relative D-efficiency of the proposed design for weight w_N to the normal theory D-optimal design is defined as:

$$D_{e,w}^{N} = \left(\frac{|\mathbf{X}_w' \mathbf{X}_w|}{|\mathbf{X}_{w=1}' \mathbf{X}_{w=1}|} \right)^{1/p}, \tag{2.13}$$

where $\mathbf{X}_w$ denotes the model matrix for the proposed design for a weight w and $\mathbf{X}_{w=1}$ denotes the model matrix for a D-optimal design for a linear model. We use the design

created in the algorithm when the weight w is equal to one for $\mathbf{X}_{w=1}$. The relative D-efficiency of the proposed design to the binary optimal design is defined as:

$$D_{e,w}^{B} = \left(\frac{|\,\mathbf{X}_{w}'\mathbf{V}\mathbf{X}_{w}|}{|\mathbf{X}_{w=0}'\mathbf{V}\mathbf{X}_{w=0}|} \right)^{1/p}, \tag{2.14}$$

where $\mathbf{X}_{w=0}$ is the model matrix for the binary optimal design. For this model matrix, we use the design created in the algorithm when the weight w is equal to 0. We use the asymptotic information matrix in this design efficiency such that the mean of the supplied priors for the model parameters are used in the diagonal elements of $\mathbf{V}$.

We also compare these designs in terms of the predicted variance using fraction of design space (FDS) plots (Ozol-Godfrey et al., 2008 and Zahran et al., 2003). FDS plots display the prediction variance of the response for a given model across regions of the design space. Although the designs are intended to be screening designs to identify active effects for the dual-response system, FDS plots can provide additional insight into different properties among designs. These plots provide a simple way to compare designs in their potential ability for prediction. To create the FDS plot, thousands of random locations within the design space are sampled. The prediction variance is calculated at each of these locations. For the logistic regression model, we use the mean of the specified parameter estimates. These values are then ordered from smallest to largest and plotted against the estimated percentile. To ensure good coverage within the design space, 10,000 random locations are recommended (Myers et al., 2016, p. 479). FDS plots are useful to visualize the minimum, median, and maximum prediction variance across the design space. Ideally, the curve in the FDS plot is relatively flat and low so that the prediction variance does not change drastically across the design space.

2.3.4 Algorithm

Given user-specified values of the number of runs n, the number of factors k, the desired weight on the linear model w, the form of the models (main effects only, main effects plus two factor interactions, or full quadratic), and the prior distribution of the model parameters, the algorithm implements a coordinate exchange algorithm adapted from Meyer and Nachtsheim (1995) and Gotwalt et al. (2009) to construct the optimal design based on the criterion in (2.10). The design space is assumed to be cuboidal so that all factors lie (in coded form) in the range [-1, 1]. Factor values in experimental units can be converted to coded units using the method as described in Myers et al. (2016, p. 86). The algorithm begins by creating a random starting design with points in the range [-1, 1]. The algorithm then searches through the design space, one coordinate in the design matrix at a time, maximizing the criterion in equation (2.10). For continuous factors, the method uses Brent's optimization method (Press et al., 1986) for each coordinate exchange. For categorical factors, the value of the criterion is calculated for each level of the factor and the level that yields the higher criterion value is used. To calculate the Bayesian D-criterion $\tilde{\varphi}_{\mathrm{B}}$ for the binary response, the algorithm implements the quadrature scheme developed by Gotwalt et al. (2009). The final step of the algorithm is one iteration of point exchange, performed using the final set of points in order to make the design more practical. This final step, as recommended by Gotwalt et al. (2009), allows any clusters of points to potentially be combined. Fewer factor levels can be helpful for practitioners when performing the experiment. We assume that the linear predictor in the GLM is of the same maximal form as the model for the normal response

variable (i.e. $\boldsymbol{x}'\boldsymbol{\beta}$ is the same for both models). We force this requirement in order to avoid missing any important effects that may occur in either model. These designs are screening designs in that the goal is to identify active effects in both models. This restriction could be relaxed in future work to allow for more flexibility in the model forms. Of the three types of models considered (main effects, main effects plus two-factor interactions, or quadratic), we assume all are full models. The algorithm is implemented in the JMP Scripting Language (JSL, V. 12) and is available upon request by the author. A JMP Add-In with user interface is currently under development.

2.4 Results

To present this method, we first demonstrate this design optimality criterion with two hypothetical factors and return to this larger experiment in a section 2.5.

2.4.1 Example: Main Effects Only Model

Johnson & Montgomery (2009) demonstrate results of binary optimal designs for full quadratic logistic regression models. They compare the efficiency of these designs to a linear D-optimal design. As a comparison, we choose the same priors for the model parameters of the binary response for the following test cases. The priors for the model parameters of the binary model are assumed to follow a normal distribution with the given $\pm 2\sigma$ limits. This framework is analogous to the nonlinear design procedure available in JMP.

We first consider several cases for a two-factor dual response system. The first case has twelve runs where each model is assumed to have main effects only. The prior information for the three parameters of the logistic regression model are as follows:

$$1 \leq \beta_0 \leq 3$$

$$1.5 \leq \beta_1 \leq 4.5$$

$$-3 \leq \beta_2 \leq -1$$

Figure 2.1 displays the resulting designs for the optimal designs using the criterion in (2.10). The binary optimal design and the D-optimal design (using weights $w_N = 0$ and $w_N = 1$, respectively, in the algorithm) are shown in Figure 1a and e, respectively. Note that these designs are quite different, although each have four distinct design points. The D-optimal design is a replicated 2^2 factorial design while the binary optimal design does not have any design points with $x_1 > 0.5$. Figure2.b, c, and d show designs for three increasing weight values. For weights $w_N = 0.25$ and 0.50, the designs

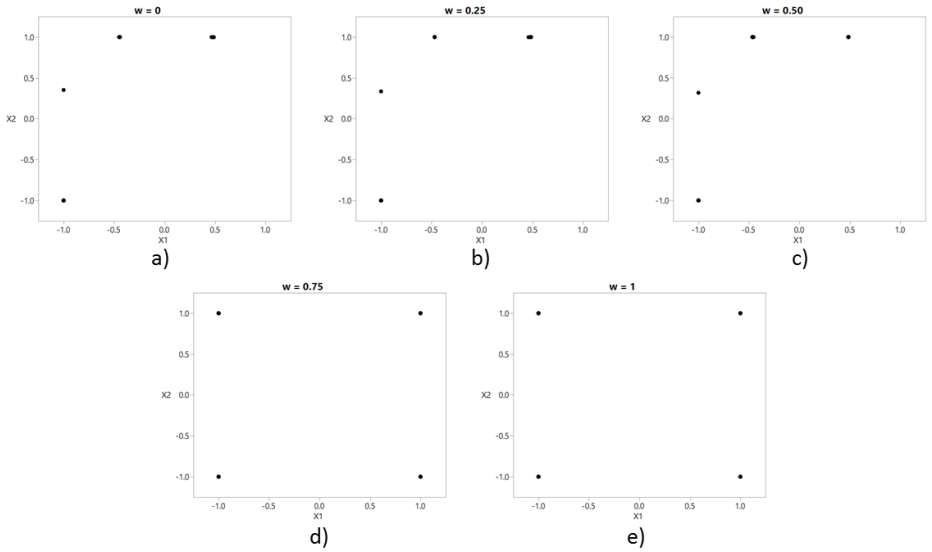

Figure 2.1. Designs for main effect only models with weights (a) $w_N = 0$ (binary optimal design), (b) $w_N = 0.25$, (c) $w_N = 0.50$, (d) $w_N = 0.75$, and (e) $w_N = 1$ (D-optimal design)

are exactly the same or very close to the binary optimal design. The design at weight $w_N = 0.75$ is equivalent to the D-optimal design.

We compare the efficiency of this proposed design to designs for different values of the weight w using the formulas for efficiency defined in (2.13) and (2.14). Figure 2.2 shows the relative D-efficiency of the proposed designs with respect to the standard D-optimal design for linear models (solid line) and to the binary optimal design for the logistic regression model (dashed line). The efficiency plot highlights the disadvantages of using a pure D-optimal design or pure binary optimal design to model both responses. If we used a D-optimal design to fit both the linear model and the logistic regression model (i.e. $w_N = 1$), while the design is 100% efficient for the linear model, it is only approximately 54% as efficient as the binary optimal design for the binary model. Conversely, if we used a binary optimal design to fit both models ($w_N = 0$), the Bayesian D-optimal is also only 52% as efficient as a D-optimal design for the linear model. This type of figure is also informative to experimenters deciding on the appropriate value of the weight for their scenario. This tool allows the experimenter to compare the advantages of one design over another. Note that with a weight of $w_N = 0.65$, the efficiency for the binary model improves to 72% while the efficiency for the linear model only drops to 92%. The design for $w_N = 0.65$ is shown in Figure 2.3. Note how this design has six design points. Compared to the binary optimal design, this design has pushed more design points to the corners of the design region.

Figure 2.4 shows the poor prediction variance of the mean response of the logistic regression model when the D-optimal design is used to fit the binary model. By increasing the weight w, the prediction variance improves. Similarly, the prediction

variance is much higher when the binary optimal design is used to fit the normal theory model.

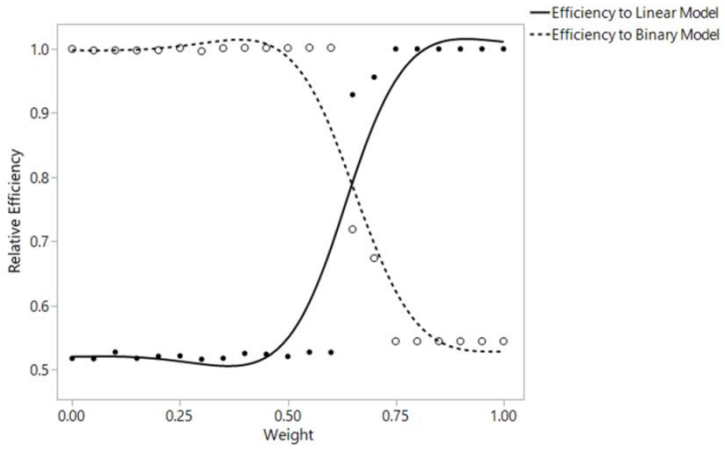

Figure 2.2. Relative D-efficiency for both responses (main effects model)

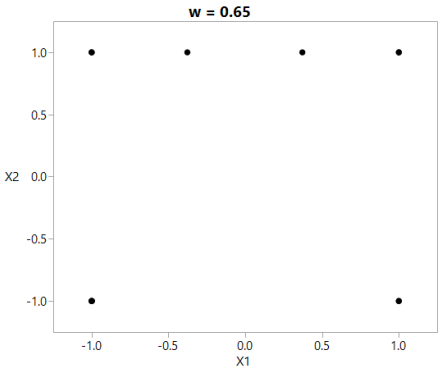

Figure 2.3. Design for weight w = 0.65 (main effects model)

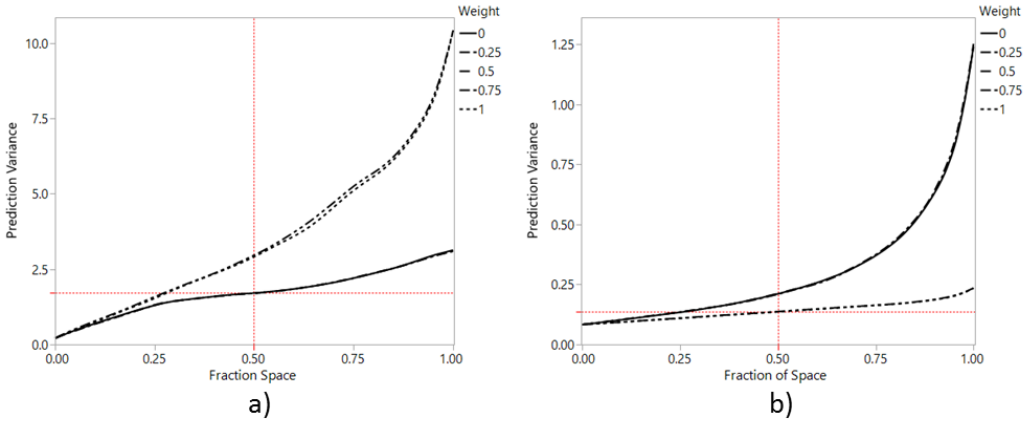

Figure 2.4. FDS plots (main effects model) for (a) Logistic regression model and (b) Normal model

2.4.2 Example: Main Effects Plus Two-Factor Interactions

Next we consider the same two-factor study with 12 runs, but now include the interaction term between x_1 and x_2 into both models. The prior information for the four parameters of the logistic regression model, assumed to be normally distributed, are:

$$1 \leq \beta_0 \leq 3$$

$$1.5 \leq \beta_1 \leq 4.5$$

$$-3 \leq \beta_2 \leq -1$$

$$-1.5 < \beta_{12} < -0.5$$

Figure 2.5 summarizes the designs for this scenario under various weights w_N. Figure 2.5a and e show the binary optimal design and D-optimal design, respectively. The binary optimal design has 4 distinct design points with replicated points in the corners of the design. The D-optimal design is again the replicated 2^2 factorial design. Including the interaction term has added higher levels of x_1 in these design compared to the main effects only model from section 2.4.1. Figure 2.5 shows how the designs change as the weight w_N increases, where there is a larger effect of the weight w_N on the resulting designs. The design points are pushed out to the corners of the design space as the weight w_N approaches 1. The efficiency plot in Figure 2.6 shows clusters of similar designs as the weight increases. For weights w_N between 0.25 and 0.50, the designs have similar design efficiency for both the linear model and binary model. With $w_N = 0.40$ (design shown in Figure 2.7), the relative efficiency is 82% and 88% to the linear model and binary model, respectively. For weights less than 0.25, there is considerable drop-off in the D-efficiency for the linear model; similarly, for weights greater than 0.5, the relative

efficiency to the binary optimal designs decreases quickly. For extreme values of w_N, there is a considerable difference between the efficiency of the proposed design to the D-optimal and binary optimal designs.

Figure 2.7 shows the designs for weights $w_N = 0.40$ and 0.45. The design in Figure 2.7a which had favorable design efficiencies for both models has 6 distinct design points, four of which are corner points. This design also contains two design points in the interior of the design space, maintaining part of the design geometry for the binary optimal design (Figure 2.5a). Figure 2.7b shows the design for $w_N = 0.45$, which has 92% and 78% relative efficiency to the linear and binary model, respectively. This design has 5 distinct design points, and has more runs in common with the D-optimal design.

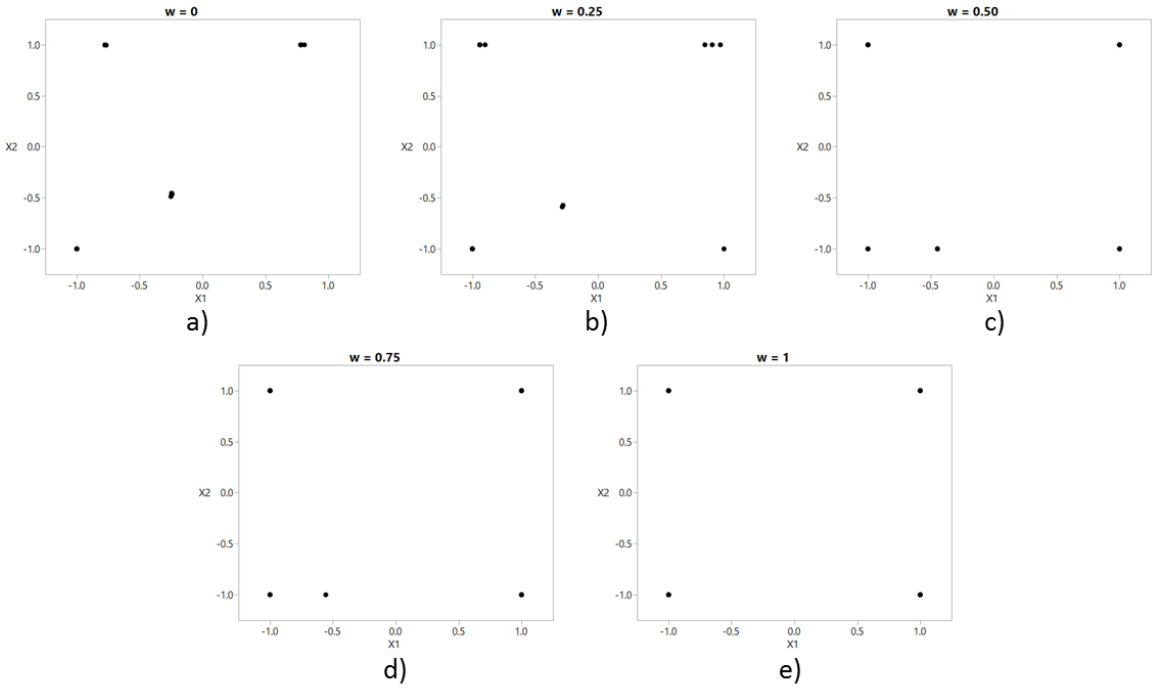

Figure 2.5. Weighted designs for ME2FI model with weights (a) $w_N = 0$ (binary optimal design), (b) $w_N = 0.25$, (c) $w_N = 0.50$, (d) $w_N = 0.75$, and (e) $w_N = 1$ (D-Optimal design)

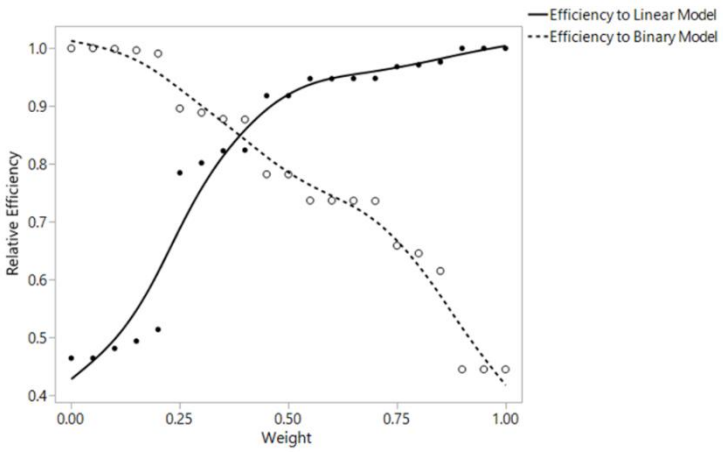

Figure 2.6. Relative D-Efficiencies for both responses (ME2FI model)

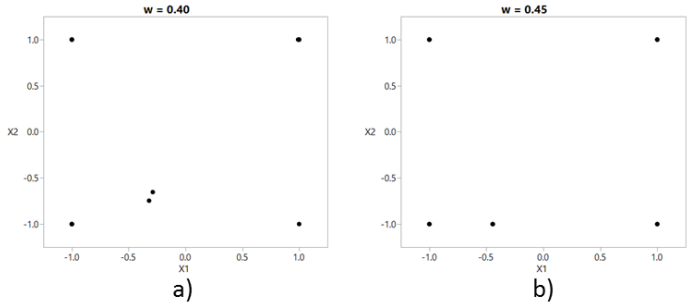

Figure 2.7. Designs for ME2FI model with weights (a) $w_N = 0.40$ and (b) $w_N = 0.45$

The FDS plots for the logistic regression model for designs based on several sets of weights are shown in Figure 2.8a. The prediction variance is notably poor for the logistic regression model when $w_N = 1$. Incorporating any value of $w_N < 1$ drastically improves the prediction variance for this binary model. We see similar behavior of the prediction variance for the linear response in Figure 8b. We can see the large tradeoffs in the prediction variance by choosing the design with $w_N = 0$. However, using a weight value $w_N > 0.5$, the prediction variance is nearly as small as for the D-optimal design.

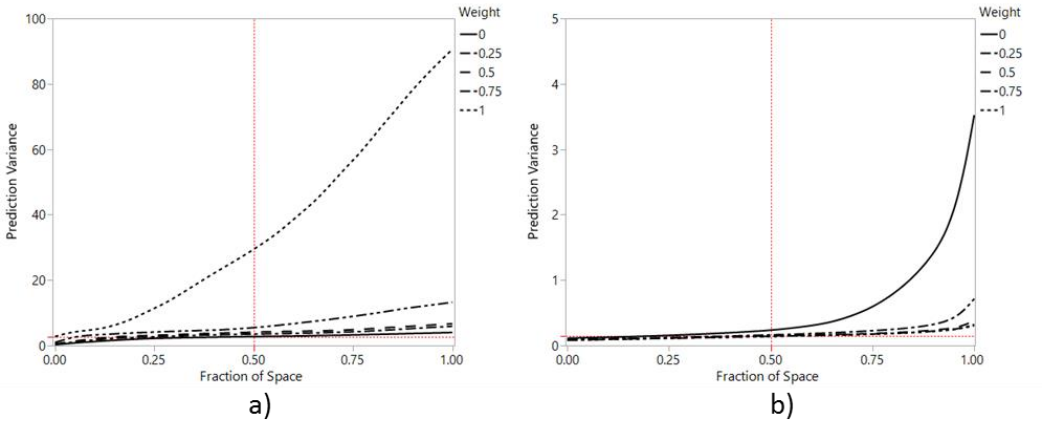

Figure 2.8. FDS plots for the ME2FI models for the (a) binary model and (b) normal theory model

2.4.3 Example: Full quadratic model

Finally, consider a two-factor study with sixteen runs for a full quadratic model. The prior information for the six parameters of the logistic regression model (assumed to be normally distributed) are:

$$1 \leq \beta_0 \leq 3$$

$$1.5 \leq \beta_1 \leq 4.5$$

$$-3 \leq \beta_2 \leq -1$$

$$1.5 \leq \beta_{11} \leq 4.5$$

$$-1.5 < \beta_{12} \leq -0.5$$

$$-6 \leq \beta_{22} \leq -2$$

A summary of these designs and their properties are displayed in Figures 2.10, 2.11, and 2.12. Including the quadratic terms in these models results in a large difference between these designs. As seen in Figure 2.10a, the binary optimal design does not have many points for values of $x_1 > 0.50$, emphasizing the difference in the estimated sizes of

effects of the model parameters for x_1 and those for x_2. Both the binary optimal design and D-optimal design have nine distinct design points. The designs with weights w_N between 0 and 1, have between 10 and 12 distinct design points. As the weight w_N increases, the distance between runs increases, so that more design points are pushed to the outer region of the design space.

Figure 2.10 shows the relative efficiency of the proposed designs compared to the design for the linear model and to the binary optimal design. Again, this efficiency plot allows us to better understand the effect of the value of the weight on the design efficiency. By selecting a weight value between 0.15 and 0.45, the resulting design is at least 80% efficient for both the linear model and the binary model.

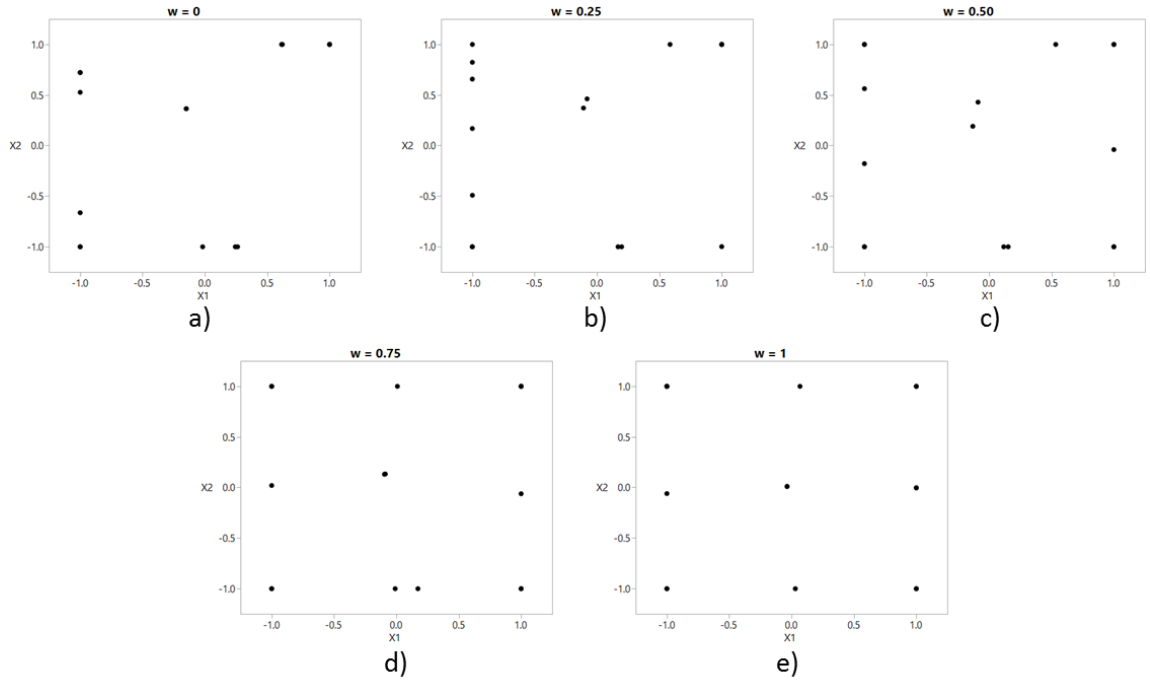

Figure 2.9. Weighted designs for quadratic model with weights (a) $w_N = 0$ (binary optimal design), (b) $w_N = 0.25$, (c) $w_N = 0.50$, (d) $w_N = 0.75$, and (e) $w_N = 1$ (D-Optimal design)

Figure 2.11a shows the FDS plots for the logistic regression model for the full quadratic model case. We see similar behavior in terms of the prediction variance performance that we saw for the main effect only model and main effects plus two-factor interactions model. For weights $w_N \leq 0.50$, the prediction variance is similar to the prediction variance for the binary optimal design. Similarly, Figure 2.11b displays the FDS plot for the linear model for this full quadratic model. For any weight $w_N > 0.5$, the prediction variance for the linear model is nearly the same as that for the D-optimal design.

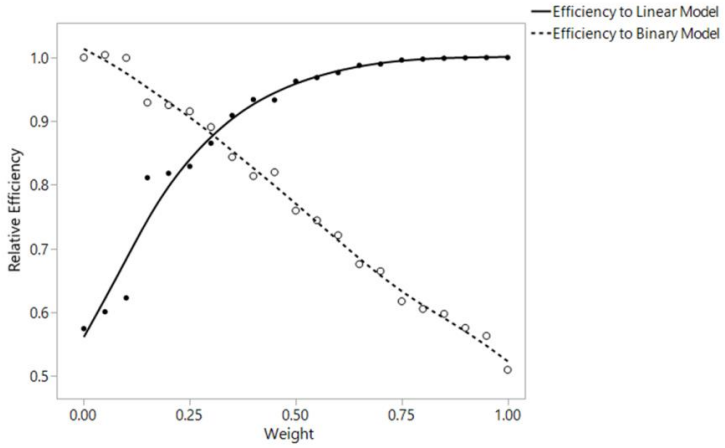

Figure 2.10. Relative D-efficiencies for both models (full quadratic model)

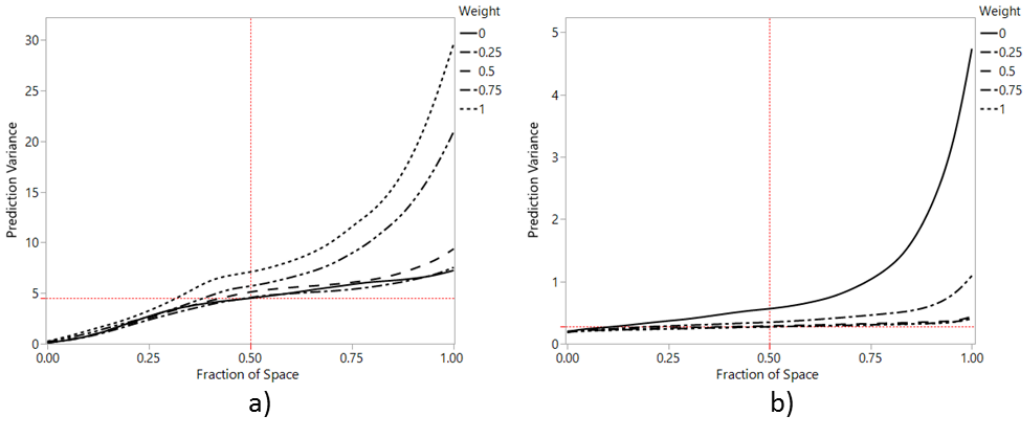

Figure 2.11. FDS plots (quadratic model) for (a) logistic regression model and (b) normal model

2.4.4 Comparison of Criteria

In section 2.3.3, two additional potential design criteria were considered to construct the optimal designs for dual response systems. The criterion in equation (2.11) scales each D-criterion and then combines the two scaled scores using an additive DF. The criterion in equation (2.12) uses an additive DF as well, but the only scaling is done by taking the log transformation of the D-criterion for the linear model. The designs, efficiency plots, and FDS plots for all three cases considered in section 2.4 are in Appendix A. Criterion (2.11) does not have as desirable an effect on the designs compared to the multiplicative DF as defined in equation (2.10). The designs using this criterion for the example considered here have large tradeoffs in design performance for the models. The criterion in equation (2.12) has a few advantages: 1) it does not require that the optimal designs for each model be found prior to generating the optimal design for a given weight w_N; and 2) it is an effective method to alter the designs over the range of the weight w_N. This criterion consistently provides designs across w where the design is simultaneously efficient for both models for the examples in this work. A disadvantage of this criterion, however, is the lack of consistent scaling of the two criteria comprising the design criterion. The ranges of the D-criterion and Bayesian D-criterion are very different which may allow one criterion to dominate the other simply because of the difference in ranges. The advantages of criterion (2.10) in section 2.3.3 solves the scaling concerns and is a natural choice for this application. The goal is to generate a design that is simultaneously efficient for both models. The multiplicative DF is a natural choice,

therefore, because this DF tends to penalize low criterion scores more than the additive DF.

2.5 Air-to-Ground Missile Simulation Example

Consider a larger example of this dual-response system for a scenario of an air-to-ground missile simulation. In this scenario, an aircraft is used to launch a missile at a desired target on the ground. A team needs to perform a screening experiment with five factors (a mix of continuous and categorical factors) to model two responses: the time to acquire the target, a continuous response, assumed to be normally distributed, and survival of the target, a binary response. The factors of interest are the aircraft type, the attack axis, the height above the target, maneuver, and the timing of the maneuver. Table 2.1 displays these five factors and their high and low levels. The team needs to determine which factors are active, whether there are any significant interactions between factors, and if either model exhibits curvature.

Table 2.1.

Factor descriptions and levels

Factor	Description	Low	High
X1	Aircraft Type	F-15E	F-16
X2	Attack Axis (Nose = 0 degrees)	0	180
X3	Height Above Target (HAT)	10K	18K
X4	Maneuver	45	135
X5	Timing of Maneuver	0	50

The team considers full quadratic models for both models. The model contains 20 parameters. Designs with $n = 28$ runs are constructed for weights w_N across the range of

0 to 1. Figure 2.12 shows the D-optimal design holding X1 (the categorical variable) and X5 constant at their respective levels. Figure 2.13 shows scatterplot matrices of the binary optimal design, design for weight $w_N = 0.1$, and the D-optimal design. The binary optimal design contains more points within the center of the design space, with several factors containing more than three levels. The weighted design maintains some of these properties, but also begins to approach the D-optimal design. The efficiency plot for this scenario in Figure 2.14 shows the tradeoff in design efficiency, most extreme for the logistic regression model. However, using a weight $w_N = 0.1$ improves the relative efficiency of the design for the binary model to 73% while maintaining 96% relative efficiency for the linear model.

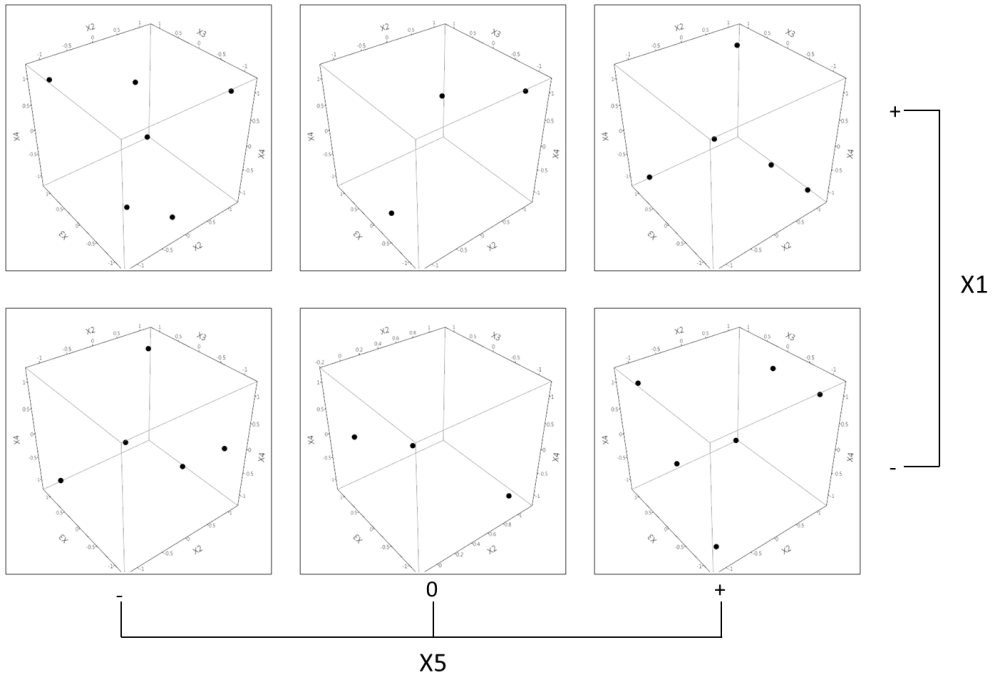

Figure 2.12. D-optimal design for 5-factor design with 28 runs (full quadratic model)

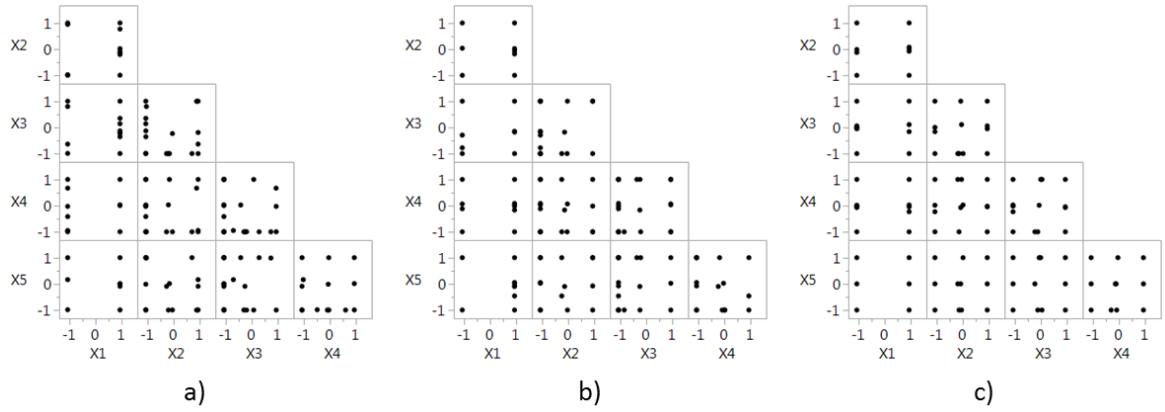

Figure 2.13 Scatterplot matrices of weighted designs for quadratic model with weights (a) $w_N = 0$ (binary optimal design), (b) $w_N = 0.1$, and (c) $w_N = 1$ (D-Optimal design)

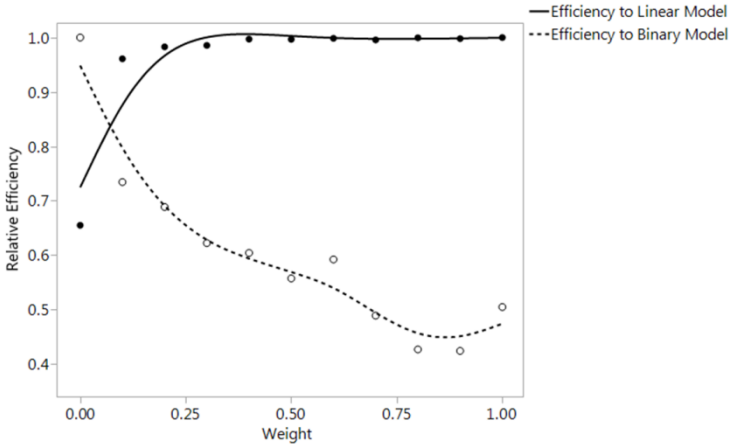

Figure 2.14. Efficiency plot of 5-factor 28-run design (quadratic model)

2.6 Separation

A potential by-product of small sample sizes or the choice of the priors for the logistic regression model parameters is that the obtained responses can lead to the inability to estimate the chosen model. In one case, all of the obtained data may fall into just one category. This scenario would not be unusual, for example, in a system with high reliability, where every trial results in a "pass." The second scenario that occurs is complete separation or quasicomplete separation for logistic regression models, situations where the maximum likelihood estimates of the model do not exist (Albert & Anderson,

1984). Complete separation occurs when there is a hyperplane that passes through the design space such that on one side of the hyperplane, all of the observations have a response value $y = 1$ and on the other side, all of the observations have a response value $y = 0$ (Agresti, 2013; Albert & Anderson, 1984). Mathematically, complete separation occurs if there exists any vector $\boldsymbol{a}$ such that:

$$\begin{gathered} \boldsymbol{a}'\boldsymbol{x}_i > 0 \text{ when } y_i = 1 \text{ and} \\ \boldsymbol{a}'\boldsymbol{x}_i < 0 \text{ when } y_i = 0 \end{gathered} \tag{4}$$

As Albert and Anderson (1984) show, if there is complete separation, then the maximum likelihood estimates of the logistic regression model do not exist. Figure 2.15a shows an example of complete separation for one factor. Because of the split in response values at approximately $x_1 = 0.5$, we can perfectly predict the response variable y based on the input factor x_1. Quasicomplete separation occurs if there exists any vector $\boldsymbol{a}$ such that:

$$\begin{gathered} \boldsymbol{a}'\boldsymbol{x}_i \geq 0 \text{ when } y_i = 1 \text{ and} \\ \boldsymbol{a}'\boldsymbol{x}_i < 0 \text{ when } y_i = 0 \end{gathered} \tag{5}$$

This type of separation occurs, in other words, when the hyperplane splitting the observations in the design space contains observations that have responses y equal to both 0 and 1. For example, the logistic regression model for the data shown in Figure 2.15b (where there is now an additional point at $x_1 = 0.5$ with a response value of $y = 0$) has quasicomplete separation. Albert and Anderson (1984) show that the maximum likelihood estimates also do not exist when quasicomplete separation occurs.

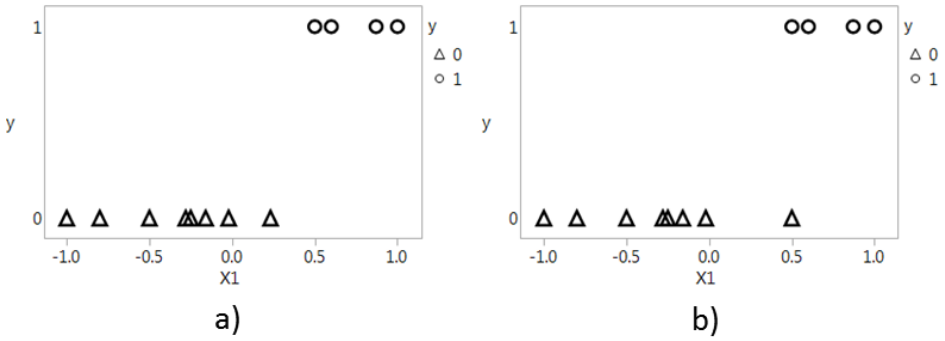

Figure 2.15. Example of (a) complete separation and (b) quasicomplete separation for a one-factor logistic regression model

If there is no complete or quasicomplete separation, then the maximum likelihood estimators exist and are unique (Albert & Anderson, 1984). Complete or quasicomplete separation in logistic regression models can occur frequently under several conditions, for example, when the run size is small or when many of the predictors are categorical and only take on two levels (Agresti, 2013). Different from the first scenario outlined previously, separation is more likely to occur when the range of expected probabilities of a success extends across a large portion of the range between 0 and 1. The likelihood of this occurring is compounded by small sample sizes.

The designs considered in this work are screening designs that will be used to identify important factors in a process. Therefore, being able to estimate the model parameters in both the linear model and the logistic regression model is essential. However, it is often difficult to detect separation when fitting the model. A MonteCarlo simulation was used to investigate the probability that the model was either inestimable (the first scenario) or had complete or quasicomplete separation (the second scenario) of these designs. 5000 datasets of responses were simulated for each run in the designs from section 2.4. In each simulation, the model parameters are randomly generated from the

given prior distributions. Responses for each run are then randomly sampled from a Bernoulli process with the probability of success equal to the expected value of y using the randomly generated model parameters. Separation was determined to occur if the estimated variance of any parameter estimate was greater than 5000, as suggested by Heinze and Schemper (2002). One weighted design of each model type in section 2.4 was used for this investigation. The results of this simulation are shown in Table 2.2. For all three model types, separation occurred at least 67% of the time for the weighted designs. To determine if this separation issue is unique to these weighted designs, we consider the issue of separation in the binary optimal designs for the same priors. Table 2.2 shows that separation was also a likely event for even the binary optimal designs.

Table 2.2.

Results of Monte Carlo Separation Simulation

Model	% Inestimable	% Separation
ME model (w = 0.65)	0.04	67.6
ME model (w = 0)	0.08	41.6
ME2FI model (w = 0.40)	0.08	91.7
ME2FI model (w = 0)	0.22	79.5
Quadratic model (w = 0.30)	0.08	76.5
Quadratic model (w = 0)	0.36	71.3

The response surfaces of the logistic regression models from the scenarios in sections 2.4.2 and 2.4.3 using the mean value of the model parameters are shown in Figure 2.16a and b, respectively. For the priors used in these test cases, the expected probability of success extends across the range 0 and 1. As comparison, consider a main effects and two factor interaction model and full quadratic model with a smaller range of

the expected probability of success (Figure 2.17a and b). However, separation occurred in 85% and 65% of the cases for these designs as well.

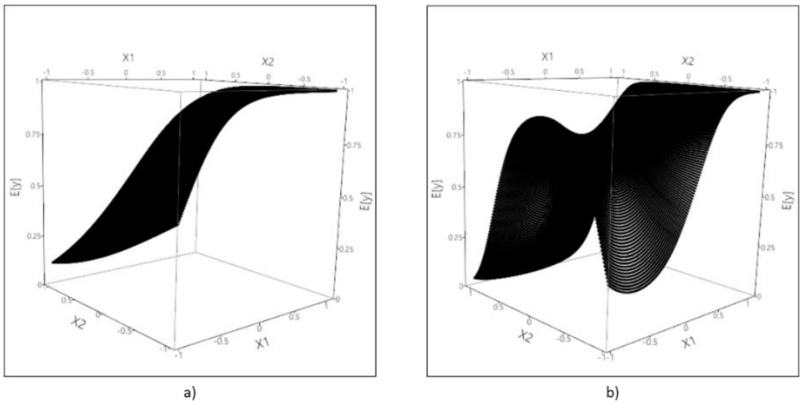

Figure 2.16. Response surface of logistic models for the models in section (a) 2.4.2 and (b) 2.4.3

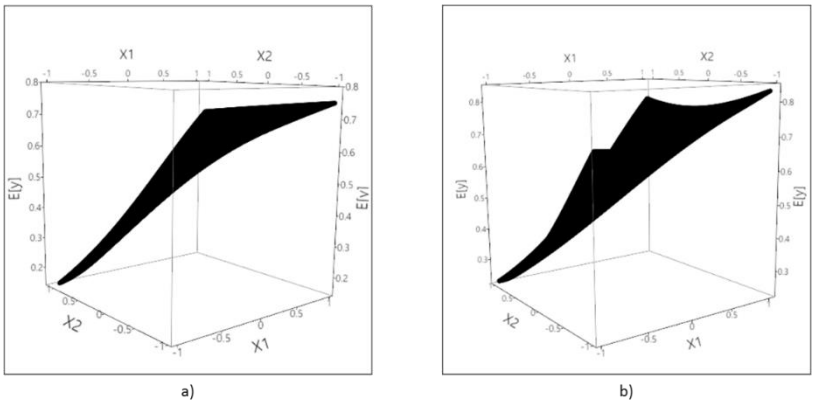

Figure 2.17. Response surface of logistic regression models for (a) ME2FI model and (b) quadratic model

Separation is an important issue that must be addressed when designing an experiment for binary responses, whether as part of a dual response system or not. The choice of the priors on the model parameters and the run size contribute to the chance of separation occurring. When the expected probability of success ranges from 0 to 1 and the design size is small, separation can occur often. There are some alternative analysis methods available that can "rescue" an analysis when maximum likelihood fails; for

example, a penalized likelihood function such as Firth's method (Firth, 1993). However, when fitting the logistic regression model, creating a design that provides a high probability of avoiding separation in the first place is preferred. The issue of separation for these designs (and for designs for logistic regression models in general) is left for future work.

2.7 Conclusions

A new weighted optimality criterion to create optimal designs for experiments when two responses are of interest has been introduced. One of these responses is assumed to be normally distributed while the other is binary. The described algorithm is available as a script in JMP (Version 12). Code can be provided upon request. The weighted criterion allows the user to specify a desired preference of one response over the other. For several test cases, we have shown how the design points change as the value of the weight changes.

Several graphic tools that may be of use to researchers were presented. The efficiency plots serve two purposes. One, they highlight the core issue of designing experiments for multi-response systems. A design created with only one model in mind will not be an efficient design for other responses, particularly when the responses follow different distributions. Second, these plots can help experimenters select the value of the weight w in the criterion by displaying the tradeoffs in design efficiency between the two models. The FDS plots also add a comparison of the prediction variance of the mean response for different values of w, assuming the given model is correct. This work

establishes the framework for many future research studies. One drawback of this design criterion is that there is an implicit assumption in independence of the response variables. Additional research needs to be done to investigate the consequences of this assumption.

CHAPTER 3

OPTIMAL DESIGN OF EXPERIMENTS FOR DUAL-RESPONSE SYSTEMS: PART II

3.1. Introduction

In chapter 2, we introduced new criteria to create computer-generated optimal designs for dual response systems. Our initial motivating example involved an air-to-ground missile simulation where several responses were of interest to the experimenters, including the time to acquire the target, the miss distance from the target, and survival of the target (a binary response). We focused on the case where the first response y_1 is assumed to have a normal distribution and the second response y_2 has a binomial distribution. In many of these multi-response systems, however, it is not unusual to find count data, which can be well modeled by a third response y_3 that has a Poisson distribution. In this work, we extend the methods of chapter 2 by incorporating a Poisson response into the methodology. We still consider dual-response systems; however, we introduce two new scenarios: 1) a dual response system where one response has a normal distribution while the second response has a Poisson distribution; and 2) a binomial-Poisson dual-response system.

While using priors can help overcome the uncertainty of model parameters for these nonlinear models, it also can drastically affect the validity of these designs when the supplied priors are very different from the truth. Sensitivity studies can provide

valuable information about the flexibility of these designs based on the supplied priors. In section 3.4, we discuss the results of the sensitivity study of the effects of the priors on the resulting design from designs in chapter 2 and in section 3.3.

3.2 Methods

3.2.1 Poisson Regression Model

A Poisson regression model, now included in the dual-response system, can be modeled using a generalized linear model (GLM) and is also nonlinear. The canonical link function for a Poisson regression model is the log link, defined as:

$$\eta = \ln(\boldsymbol{x}'\boldsymbol{\beta}). \tag{3.1}$$

Using the log link, the expected value of the mean is:

$$E[y_3] = e^{\boldsymbol{x}'\boldsymbol{\beta}}. \tag{3.2}$$

The log link is a popular choice because it ensures that all predicted values of y are nonnegative, a property of the Poisson distribution.

As discussed in chapter 2, the design criterion of a nonlinear model for a binary optimal design (defined in equation 2.7) is dependent on the parameters of the model. Recall that the asymptotic information matrix for a nonlinear model is defined as $\mathbf{I} = \mathbf{X}'\mathbf{V}\mathbf{X}$. The matrix $\mathbf{V}$ is a diagonal matrix that is a function of the mean and is dependent on the distribution of the GLM. For the Poisson distribution, the mean and variance are equal; therefore, the weight matrix $\mathbf{V}_{\mathrm{P}}$ is:

$$\mathbf{V}_{\mathrm{P}_{ii}} = e^{\boldsymbol{x}_i'\boldsymbol{\beta}}. \tag{3.3}$$

3.2.2 Design Criterion

The Bayesian D-criterion for the Poisson model using the asymptotic information matrix is:

$$\varphi_{\mathrm{P}} = \log \int |\mathbf{X}'\mathbf{V}_{\mathrm{P}}\mathbf{X}|\, f(\boldsymbol{\beta}_{\mathrm{P}}) d\boldsymbol{\beta}_{\mathrm{P}}. \tag{3.4}$$

For the remainder of this paper, we say Poisson optimal design to refer to the Bayesian D-optimal design for the Poisson response model. As in chapter 2, the Bayesian D-criterion for the Poisson model is scaled to a range between 0 and 1 so that the criteria can be more fairly combined in the proposed design criterion. The scaled criterion for the Poisson model $\tilde{\varphi}_P$ for a given value of the weight is divided by the value of the Poisson optimal design criterion (i.e., the design criterion when $w_P = 1$). This scaling shown below is analogous to the scaling done in chapter 2 for the linear model:

$$\tilde{\varphi}_{\mathrm{P}} = \frac{\varphi_{P,w}}{\varphi_{P,w_P=1}}. \tag{3.5}$$

The design criterion for the dual response system depends on the distributions of the two responses. In chapter 2, the criterion in equation (2.10) combines the scaled criteria for the linear model and binary model using the multiplicative DF. In section 3.3.1, results for dual-response systems with a normal and Poisson response are presented with the design criterion φ_{NP} is defined as:

$$\varphi_{NP} = \left(\tilde{\varphi}_N^{w_N}\right)\left(\tilde{\varphi}_P^{(1-w_N)}\right), \tag{3.6}$$

where w_N in [0,1] is the weight on the design criterion of the normally distributed response. Section 3.3.2 presents results for dual-response systems with a binary response and Poisson response. The design criterion φ_{BP} is defined as:

$$\varphi_{BP} = \left(\tilde{\varphi}_B^{w_B}\right)\left(\tilde{\varphi}_P^{(1-w_B)}\right), \tag{3.7}$$

where w_B is the weight of the design criterion of the binary response. The value of w_B must be between 0 and 1. Note that for this scenario, the model parameters for the binary model and Poisson model are not the same and are defined as $\boldsymbol{\beta}_B$ and $\boldsymbol{\beta}_P$, respectively. As in chapter 2, the linear predictor is assumed to have the same form for both models under consideration.

Given the number of runs n, the number of factors k, the weight of one of the model criteria (either w_N or w_B), the form of the model (main effects only, main effects plus two-factor interactions, or full quadratic), and the priors for the nonlinear model(s), we use the coordinate exchange algorithm as described in 2.3.4 which is adapted from Gotwalt et al. (2009) and Meyer and Nachtsheim (1995).

3.3 Results

3.3.1. Normal and Poisson Dual Response System

First consider a dual response system where one response follows a normal distribution while the other follows a Poisson distribution. We consider main effects models (Section 3.1.1), main effects model plus the two-factor interaction (3.1.2), and full quadratic models (3.1.3). The coordinate exchange algorithm searches for optimal designs that maximize the criterion defined in equation (3.6).

3.3.1.1 Main Effects Model

Consider a two-factor design with twelve runs where both models are assumed to have main effects only. The priors for the model parameters of the Poisson model are assumed to have a normal distribution with the following $\pm 2\sigma$ range.

$$1 \leq \beta_0 \leq 3$$

$$0.25 \leq \beta_1 \leq 0.75$$

$$-0.3 \leq \beta_2 \leq -0.1$$

The support points of the D-optimal design for the linear model and the Poisson optimal design are the same, a 2^2 factorial design. Figure 3.1 shows the designs for these two models with the number of replicates noted in the graph at each design point. While the support points for the D-optimal design are each replicated three times, the Poisson optimal design is unbalanced. Increasing the weight on the linear model from 0 to 0.05 causes the replicates to shift slightly (Figure 3.1b). Any weight w_N greater than 0.05 results in the D-optimal design. Figure 3.2 shows the efficiency plot, displaying the relative D-efficiency of the designs across the various weights to the Poisson optimal design (dashed line) and to the D-optimal design (solid line). The resulting designs and the efficiency plot both indicate that there is very little tradeoff in design efficiency for the normal and Poisson model as the relative efficiency of the designs is greater than 90% for both models. The resulting design with $w_N = 0.05$ provides relative efficiency greater than 98% for both models.

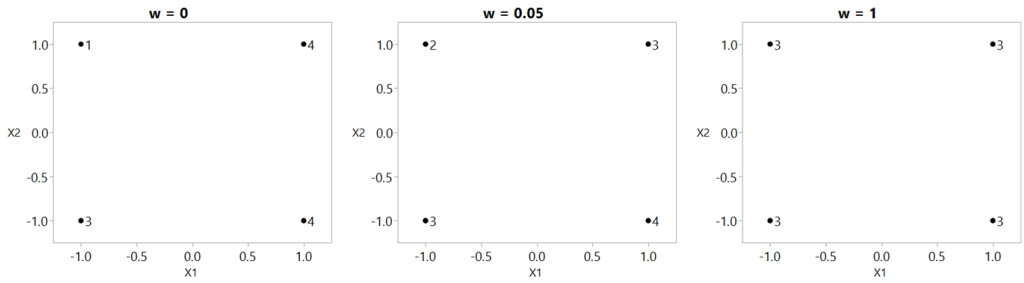

Figure 3.1. Designs for main effect only models with weights (a) $w_N = 0$ (Poisson optimal design), (b) $w_N = 0.05$, (c) $w_N = 1$ (D-optimal design)

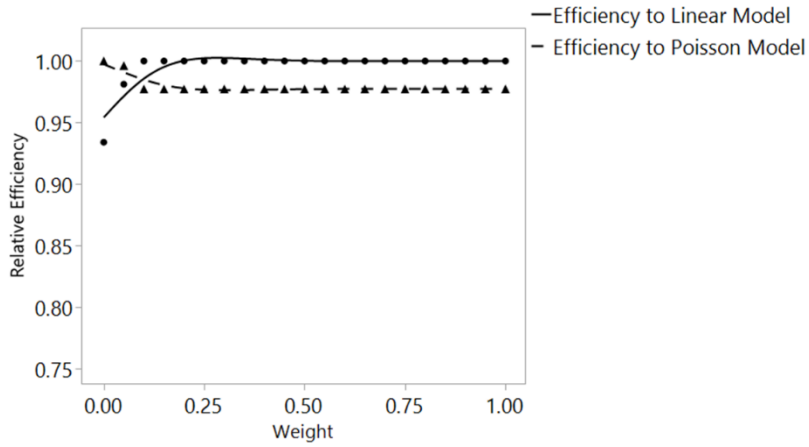

Figure 3.2. Relative D-efficiency of designs for normal and Poisson models (main effects model)

3.3.1.2 Main Effects Plus Two Factor Interaction Model

Consider a two-factor design with twelve runs where both models are assumed to have main effects plus the two-factor interaction. The priors for the model parameters of the Poisson model are assumed to have a normal distribution with the following $\pm 2\sigma$ range.

$$1 \leq \beta_0 \leq 3$$

$$0.25 \leq \beta_1 \leq 0.75$$

$$-0.3 \leq \beta_2 \leq -0.1$$

$$-1.5 \leq \beta_{12} \leq -0.5$$

Figure 3.5 shows the optimal design for the Poisson model ($w_N = 0$) and the optimal design for the linear model ($w_N = 1$). The D-optimal design is a replicated 2^2 factorial design. Including the interaction term in the model, however, has affected the support points of the design for the Poisson model. While the Bayesian D-optimal is balanced, only two corner points of the 2^2 factorial design are present in this design.

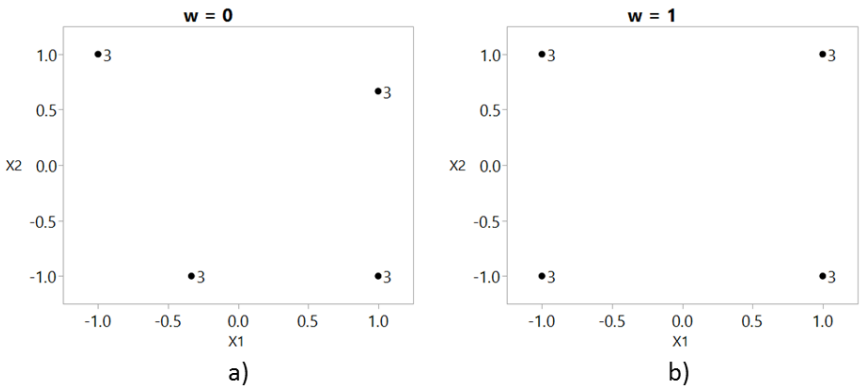

Figure 3.3. Designs for main effects plus two-factor interaction models with weights (a) $w_N = 0$ (Poisson optimal design) and (b) $w_N = 1$ (D-optimal design)

The efficiency plot (Figure 3.6) shows that there is a larger tradeoff in terms of relative efficiency for the linear model with the Poisson model. The Poisson optimal design is approximately 75% efficient as the replicated factorial design for the linear model. However, the factorial design is approximately 95% as efficient compared to the Poisson optimal design. Increasing the weight w_N to any value greater than 0 for this scenario results in the replicated 2^2 factorial design.

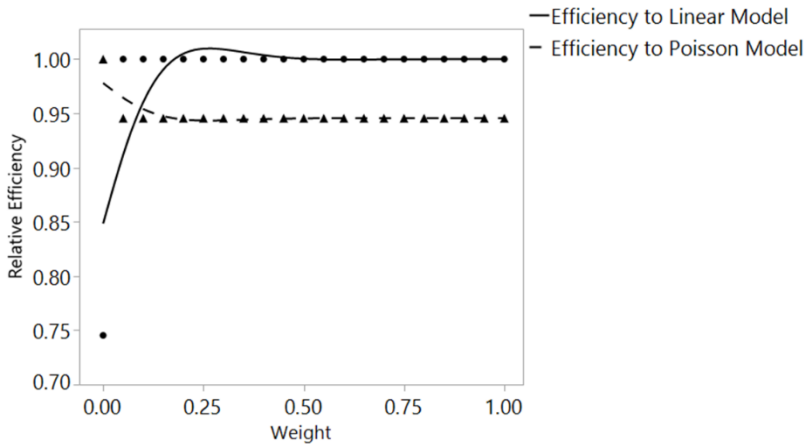

Figure 3.4. Relative D-efficiency of designs for normal and Poisson models (main effects plus two-factor interaction model)

3.3.1.3 Quadratic Model

Consider a two-factor design with sixteen runs where both models are assumed to have full quadratic models. The priors for the model parameters of the Poisson model are assumed to have a normal distribution with the following $\pm 2\sigma$ range.

$$1 \leq \beta_0 \leq 3$$

$$0.25 \leq \beta_1 \leq 0.75$$

$$-0.3 \leq \beta_2 \leq -0.1$$

$$-1.5 \leq \beta_{12} \leq -0.5$$

$$0.45 \leq \beta_{11} \leq 1.35$$

$$-1.2 \leq \beta_{22} \leq -0.8$$

The optimal designs for the Poisson model and linear model are shown in Figure 3.5a and b, respectively. The optimal design for the linear model is a face centered cube with six support points replicated. The Poisson optimal design is similar in structure to

the face centered cube. The corner points are all present in this design; however, the third level of factor x_1 is approximately equal to 0.5 as opposed to 0. In addition, the third level of x_2 is less than 0.

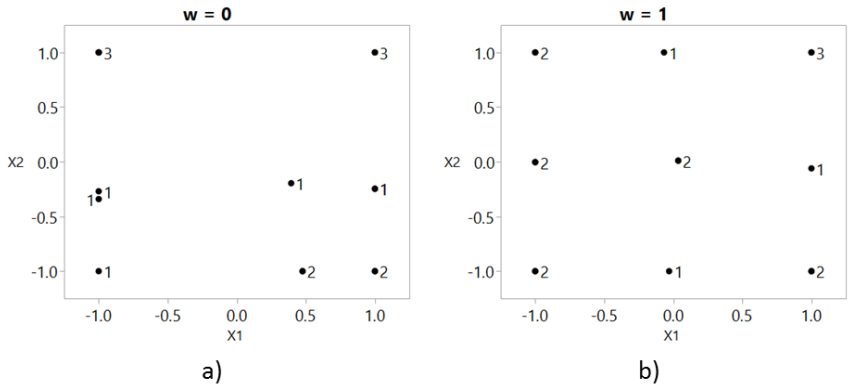

Figure 3.5. Designs for full quadratic models with weights (a) $w_N = 0$ (Poisson optimal design) and (b) $w_N = 1$ (D-optimal design)

The efficiency plot in Figure 3.6 shows that the relative efficiency of the optimal design for the Poisson regression model is approximately 80% as efficient as the face centered cube to model the linear model. On the other hand, the face centered cube is between 80% and 90% as efficient as the Poisson optimal design. The fluctuation in the relative efficiency for the Poisson model is due to changes in which support points of the face-centered cube are replicated.

From the scenarios considered in 3.3.1, the designs for the normal and Poisson models are very similar. Using the optimal design for the Poisson model tends to result in a larger tradeoff when modeling the linear model. The D-optimal design generally performs well compared to the Poisson model. The examples presented here represent only one set of priors on the nonlinear model parameters. In section 3.4, we investigate the effect of the priors of the nonlinear model on the resulting designs.

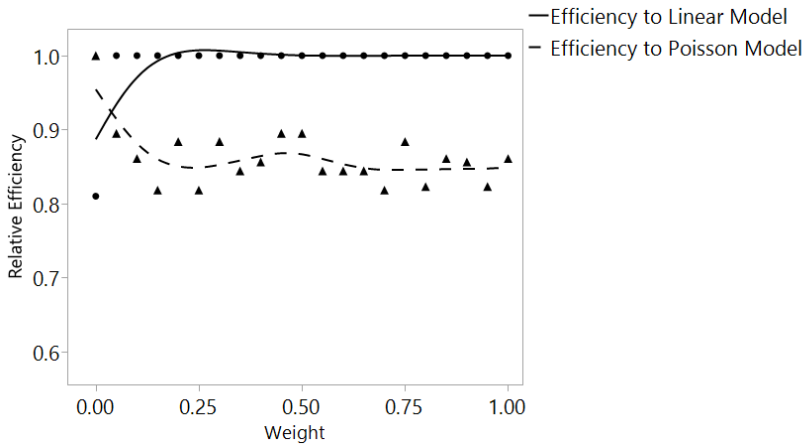

Figure 3.6. Relative D-efficiency of designs for normal and Poisson models (full quadratic model)

3.3.2 Binomial and Poisson Dual Response System

Now consider a dual response system where one response follows a binomial distribution while the other follows a Poisson distribution. We consider main effects models (Section 3.2.1), main effects model plus the two-factor interaction (3.2.2), and full quadratic models (3.2.3). The coordinate exchange algorithm searches for optimal designs that maximize the criterion defined in equation (3.7).

3.3.2.1 Main Effects Model

Consider a two-factor design with twelve runs where both models are assumed to have main effects only. The priors for both $\boldsymbol{\beta}_{\mathrm{B}}$ and $\boldsymbol{\beta}_{\mathrm{P}}$, the model parameters of the logistic regression model and Poisson model, respectively are shown in Table 3.1. These priors are all assumed to have a normal distribution with the specified $\pm 2\sigma$ range.

Table 3.1.

Range of Model Parameters for the Binary and Poisson Model

Parameter	Binary Model Priors		Poisson Model Priors	
	$\mu - 2\sigma$	$\mu + 2\sigma$	$\mu - 2\sigma$	$\mu + 2\sigma$
β_0	1	3	1	3
β_1	1.5	4.5	0.25	0.75
β_2	-3	-1	-0.3	-0.1

The optimal designs for the Poisson model and binary model for these sets of priors were seen in section 3.3.1.1 and 2.4.1, respectively, and are shown again in Figure 3.7a and c for comparison. Both of these designs have four distinct support points. While the optimal design for the Poisson model is an (unbalanced) factorial design, the optimal design for the binary model does not have any design points with $x_1 > 0.5$. Increasing the weight w_B to 0.5 results in a design with six distinct design points, three of which are corner points of the design region. The remaining design points indicate a compromise of the two optimal designs for the Poisson model and binary model. Note also that the $[-1, -1]$ run is replicated four times while the other design points occur only one or two times.

The efficiency plot in Figure 3.8 shows the extreme tradeoffs in using the optimal design of the binary response to model the Poisson response (dotted line) and vice versa (dashed line). The optimal design of the Poisson model is 42% as efficient as the optimal design of the binary model. In contrast, the optimal design of the binary model is 37% as efficient as the optimal design of the Poisson model. Using a weight $w_B > 0.05$ results in the optimal design for the binary model. However, when the weight $w_B = 0.05$, the

resulting design (Figure 3.7b) provides favorable behavior for both the binary and Poisson model (80% and 86% relative efficiency, respectively).

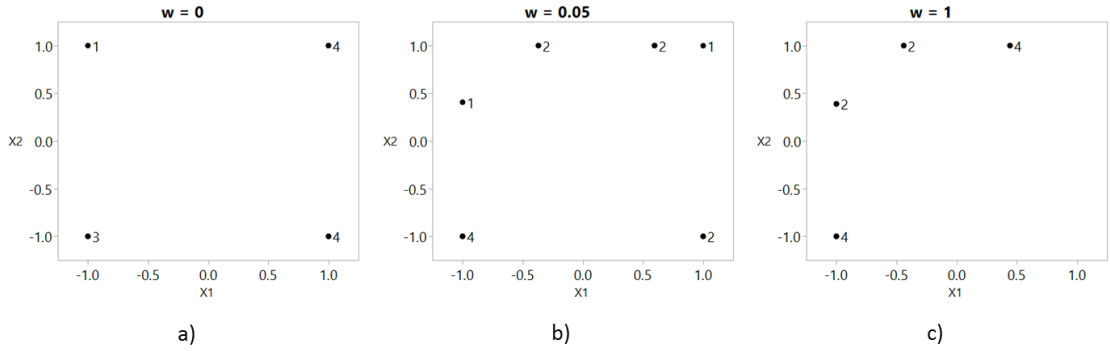

Figure 3.7. Designs for main effect only models with weights (a) $w_B = 0$ (Poisson optimal design), (b) $w_B = 0.05$, (c) $w_B = 1$ (binary optimal design)

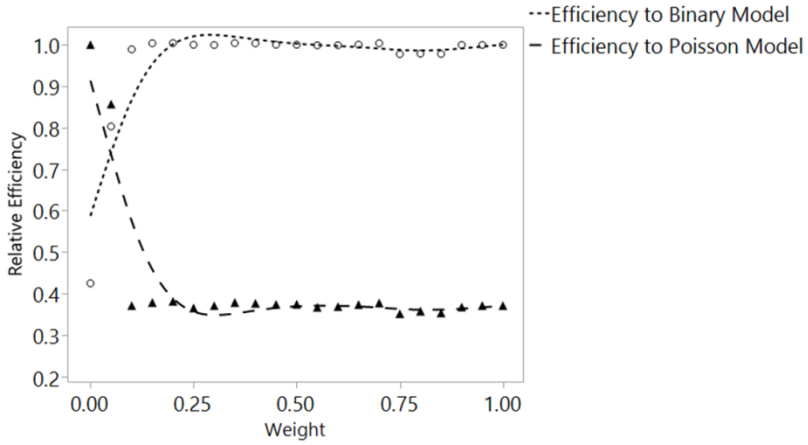

Figure 3.8. Relative D-efficiency of designs for binary and Poisson models (main effects only model)

3.3.2.2 Main Effect Plus Two-Factor Interaction Model

Consider a two-factor design with twelve runs where both models are assumed to have a main effects plus the two-factor interaction model. The priors for both $\boldsymbol{\beta}_{\mathrm{B}}$ and $\boldsymbol{\beta}_{\mathrm{P}}$, the model parameters of the logistic regression model and Poisson model, respectively

are shown in Table 3.2. These priors are all assumed to have a normal distribution with the specified $\pm 2\sigma$ range.

Table 3.2

Range of Model Parameters for the Binary and Poisson Model

Parameter	Binary Model Priors		Poisson Model Priors	
	$\mu - 2\sigma$	$\mu + 2\sigma$	$\mu - 2\sigma$	$\mu + 2\sigma$
β_0	1	3	1	3
β_1	1.5	4.5	0.25	0.75
β_2	-3	-1	-0.3	-0.1
β_{12}	-1.5	-0.5	-1.5	-0.5

The optimal designs for the Poisson model and binary model were individually seen in section 3.3.1.2 and 2.4.2, respectively, and are shown again in Figure 3.9a and d for comparison. Both designs essentially have four distinct design plots, with each replicated three times. The majority of the design points in Figure 3.9a are more concentrated in the lower triangular region of the design space. In contrast, the design points in Figure 3.9d are more concentrated in the upper triangular region of the design space. As the weight w_B increases, the design has an additional design point (Figure 3.9b and c). Comparing the design when $w_B = 0.1$ versus when $w_B = 0.25$, we see how the design slowly converges to the optimal design of the binary model.

Figure 3.10 shows the efficiency plot for this scenario. There are very large tradeoffs in relative efficiency for the two responses. The efficiency plot, however, shows a region of weights where there is favorable properties for both models ($0.05 \leq w_B \leq 0.35$. In particular, when $w_B = 0.1$, the relative efficiency of the design is greater than 80% for both models. When the weight w_B becomes greater than 0.35, the design remains the optimal design for the binary model.

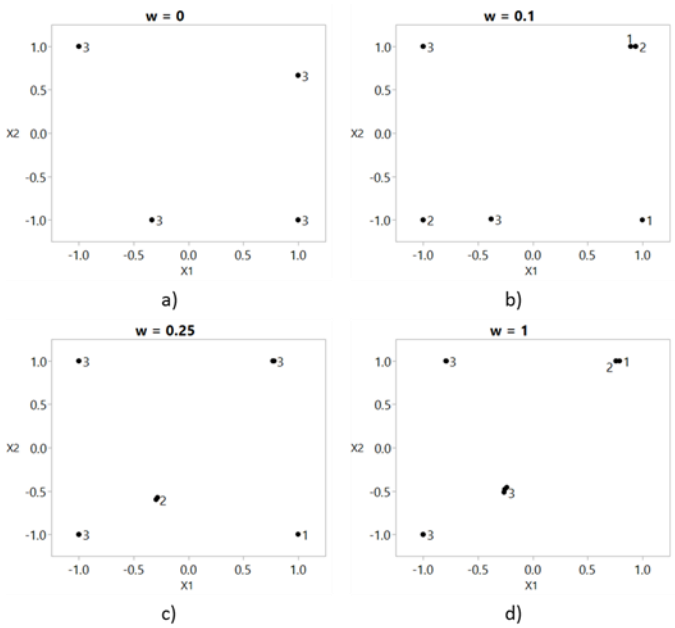

Figure 3.9. Designs for main effects plus two-factor interaction models with weights (a) $w_B = 0$ (Poisson optimal design), (b) $w_B = 0.1$, (c) $w_B = 0.25$, (d) $w_B = 1$ (binary optimal design)

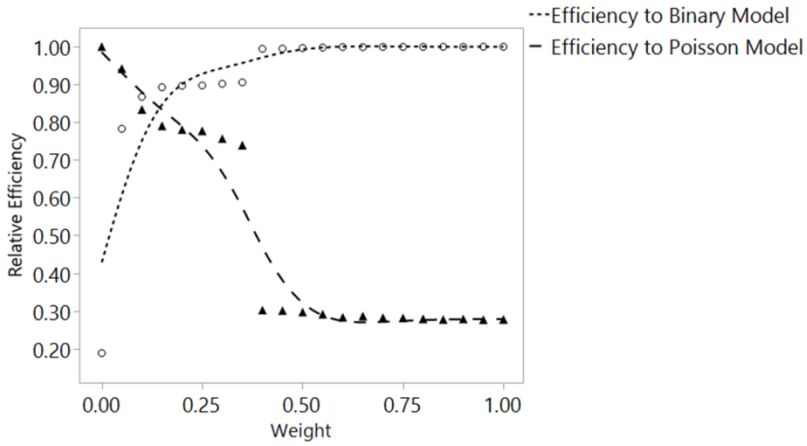

Figure 3.10. Relative D-efficiency of designs for binary and Poisson models (main effects plus two-factor interaction model)

3.3.2.3 Quadratic Model

Finally, consider a two-factor design with sixteen runs where both models are assumed to have a full quadratic model. The priors for both $\boldsymbol{\beta}_{\mathrm{B}}$ and $\boldsymbol{\beta}_{\mathrm{P}}$, the model parameters of the logistic regression model and Poisson model, respectively, are shown in Table 3.3. These priors are all assumed to have a normal distribution with the specified $\pm 2\sigma$ range.

Table 3.3

Range of Model Parameters for the Binary and Poisson Model

Parameter	Binary Model Priors		Poisson Model Priors	
	$\mu - 2\sigma$	$\mu + 2\sigma$	$\mu - 2\sigma$	$\mu + 2\sigma$
β_0	1	3	1	3
β_1	1.5	4.5	0.25	0.75
β_2	-3	-1	-0.3	-0.1
β_{12}	-1.5	-0.5	-1.5	-0.5
β_{11}	1.5	4.5	0.45	1.35
β_{22}	-6	-2	-0.6	-0.2

The optimal designs for the Poisson and binary model were first seen in sections 3.3.1.3 and 2.4.3, respectively, and are shown again in Figure 3.11a and d for comparison. The optimal design for the Poisson model has 9 distinct design points while the optimal design for the binary model has 11 distinct design points. The optimal design for the Poisson model has a cluster of points in the lower right corner of the design region. This design also has two design points in the upper corners replicated three times. The optimal design for the binary model does not have design points in the lower right corner. As the weight w_B increases, the design points spread out more so that there are

more levels of x_2 when x_1 is equal to -1. There is also an additional point in the middle of the design region.

The efficiency plot for this scenario (Figure 3.12) displays the efficiency of the design across the range of weights relative to the optimal design for each model. This figure again shows the large tradeoffs between the two designs in terms of design efficiency. When the weight w_B exceeds 0.35, the resulting design is the optimal design for the binary response. However, when $w_B = 0.1$, the relative efficiency of the design is 84% compared for both models.

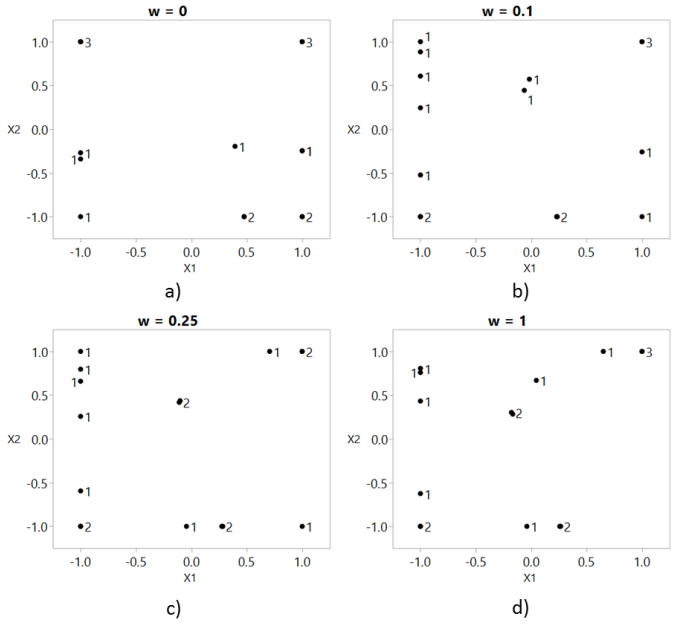

Figure 3.11. Designs for quadratic models with weights (a) $w_B = 0$ (Poisson optimal design), (b) $w_B = 0.1$, (c) $w_B = 0.25$, and (d) $w_B = 1$ (binary optimal design)

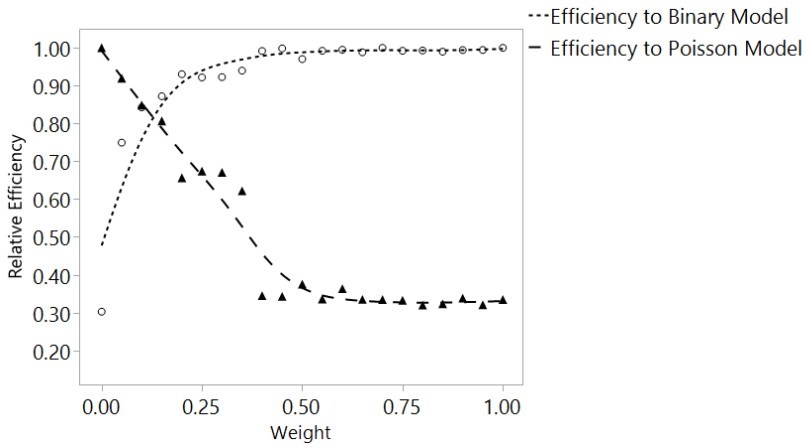

Figure 3.12. Relative D-efficiency of designs for binary and Poisson models (full quadratic model)

From these scenarios considered in 3.3.2, the designs for the binary and Poisson models are very different and result in large tradeoffs. This result highlights the importance of identifying a design that can balance the design properties of both designs. The examples presented here represent only one set of priors on the nonlinear model parameters. In section 3.4, we investigate the effect of changing the priors of the Poisson model on the resulting designs.

3.4 Sensitivity of Designs to Different Priors

The design for a nonlinear model can be affected by the prior distributions specified for the model parameters (Chaloner and Verdinelli, 1995). In 3.4.1, the effect of the priors of the Poisson model on the designs shown in the section 3.3 is explored. The Poisson distribution can be well modeled by the normal distribution when the expected value of the Poisson response is greater than 5 (Montgomery and Runger, 2003 p. 121), which may affect the designs for certain priors of the parameters of the Poisson model. In

section 3.4.2, we consider the sensitivity of the priors of the designs for the dual response system with normal and binomial responses shown in 2.4.1.

3.4.1. Sensitivity of Designs to Different Poisson Priors

Table 3.4 shows alternate priors used for the Poisson model with the priors used in section 3.3 listed as case 1. Because of the relationship between the normal and Poisson distribution, we investigate two alternative cases where the expected values of the response are much smaller or much greater than 10 throughout the design region. Case 2 represents a scenario where the expected value of the Poisson response is less than ten throughout the design region when the model parameters are equal to the specified mean values of each parameter (and for the three model types, main effects, main effects plus two factor interactions, and full quadratic). Case 3 represents the scenario where the expected value of the Poisson response is much greater than ten throughout the design region when the model parameters are equal to the mean values. We consider the same models as in section 3.3. For the dual response systems with the binomial response, the priors for the logistic regression model remained the same as in section 3.3.2.

The designs using the priors in case 2 (Table 3.4) for all three model types (main effects, main effects plus two-factor interaction, and quadratic) were very similar to those designs using the priors in case 1, for both dual-response systems with either the normal or binomial response. For the main effect plus two-factor interaction model, the optimal design for the Poisson model is the same as the D-optimal design for the linear model, a balanced 2^2 factorial design.

Table 3.4.

Range of Poisson Model Parameters for 3 Scenarios

Parameters	Case 1		Case 2		Case 3	
	$\mu - 2\sigma$	$\mu + 2\sigma$	$\mu - 2\sigma$	$\mu + 2\sigma$	$\mu - 2\sigma$	$\mu + 2\sigma$
β_0	1	3	0	2	2	4
β_1	0.25	0.75	0.15	0.65	0.75	1.25
β_2	-0.3	-0.1	-0.35	-0.15	-0.85	-0.65
β_{12}	-1.5	-0.5	0	1	0.5	1.5
β_{11}	0.45	1.35	$-.35$	1.20	0.45	1.35
β_{22}	-0.6	-0.2	-0.5	-0.05	-0.6	-0.2

The designs using the priors in case 3, however, differ more compared to the designs found in section 3.3 (see Appendix B for the designs and efficiency plots for this scenario). For the normal and Poisson dual-response system, there is a larger tradeoff in using the Poisson optimal design or D-optimal design to model both responses, particularly for the first order models. For the binomial and Poisson dual-response system, while the designs differ using these priors for the Poisson model, there are regions of the weight w_B where favorable design efficiency occurs for both models. This range of weights is also similar to the range of w_B in section 3.3.2 where the design efficiency to both models was favorable.

For the priors considered in these examples, while the designs may change as a result of the priors, the behavior of the DF across the range of weights is similar for both the normal and Poisson dual-response system as well as the binomial and Poisson dual-response system. The priors must be chosen carefully and investigated to fully understand their effect on the resulting designs. Particularly in the binomial-Poisson dual-response system, two sets of model parameters must be specified.

3.4.2 Sensitivity of Performance of Designs for Misspecifed Priors for Normal Binomial Example

In the design criterion specified in equation (2.10), the mean and standard deviation of the parameters of the logistic regression model must be specified by the user. Eliciting estimates of the mean of an unknown model parameter from subject matter experts is not an easy task; eliciting estimates of the variance is more challenging. Because of this challenge, it is important to evaluate the sensitivity due to the priors on the resulting designs. We consider the case of the two-factor main effects model for the normal-binomial dual-response system introduced in section 2.4.1. To investigate the sensitivity of the priors, we consider the effect of the designs if the mean or standard deviation of the model parameters have been misspecified by at most 20%. Table 3.5 shows the range of each prior parameter with "0" representing the true model parameters. The designs for this set of priors was discussed in section 2.4.1.

Because both the mean and the variance for the prior of each model parameter must be specified, there are six parameters that can be changed for the main effects model. We consider several combinations of high and low values of the three means and three standard deviations. Assuming that these values indicate misspecification of the priors, we compare the relative efficiency of the resulting designs for certain weights w_N to the designs using the true priors found in section 2.4.1. Specifically,

$$RE\ Linear\ Model = \left(\frac{|\mathbf{X}_w'\mathbf{X}_w|}{|\mathbf{X}_{true,w}'\mathbf{X}_{true,w}|}\right)100\% \tag{3.8}$$

$$RE\ Binary\ Model = \left(\frac{|\ \mathbf{X}_w'\mathbf{V}_B\mathbf{X}_w|}{|\mathbf{X}_{true,w}'\mathbf{V}_B\mathbf{X}_{true,w}|}\right)100\%. \tag{3.9}$$

Table 3.5

Range of Priors for Main Effects Model Normal-Binomial Dual-Response System

Prior		Low	Midpoint	High
		$-$	0	$+$
β_0	μ_0	1.6	2	2.4
	σ_0	0.4	0.5	0.6
β_1	μ_1	2.4	3	3.6
	σ_1	0.6	0.75	0.9
β_2	μ_2	-1.6	-2	-2.4
	σ_2	0.4	0.5	0.6

Several of the designs for these scenarios with misspecified priors are shown in Appendix B, along with the efficiency plots comparing the designs to the D-optimal design and binary optimal designs under the true priors. For the original priors, the binary optimal design has 52% relative efficiency for the linear model. The efficiency plot in section 2.4.1 indicates that a weight w_N provided reasonable behavior for both the binary model and the linear model. At the weight $w_N = 0.65$, the optimal design has 93% relative efficiency for the linear model and 72% as efficient as the binary design. Table 3.6 shows the relative efficiency of the misspecified designs for both the linear model and binary model compared to the designs using the true priors. The table shows the relative efficiency for weight $w_n = 0$ and $w_N = 0.65$.

Table 3.6

Sensitivity of Prior Parameters

Run	Combination	β_0		β_1		β_2		w_N	RE Linear Model (%)	RE Binary Model (%)
		μ_0	σ_0	μ_1	σ_1	μ_2	σ_2			
0	0 0 0 0 0 0	2	0.5	3	0.75	-2	0.5	0	100	100
1	- + + - - +	1.6	0.6	3.6	0.6	-1.6	0.6	0	80.8	71.3
	- + + - - +	1.6	0.6	3.6	0.6	-1.6	0.6	0.65	93.7	90.7
2	- - + + + -	1.6	0.4	3.6	0.9	2.4	0.4	0	70.3	78.6
	- - + + + -	1.6	0.4	3.6	0.9	2.4	0.4	0.65	86.4	114.9
3	+ - - - + +	2.4	0.4	2.4	0.6	2.4	0.6	0	102	67.4
	+ - - - + +	2.4	0.4	2.4	0.6	2.4	0.6	0.65	118	56.9
4	+ - + - - +	2.4	0.4	3.6	0.6	-1.6	0.4	0	80.0	82.4
	+ - + - - +	2.4	0.4	3.6	0.6	-1.6	0.4	0.65	98.1	96.3
5	+ + + + + +	2.4	0.6	3.6	0.9	2.4	0.6	0	69.4	94.7
	+ + + + + +	2.4	0.6	3.6	0.9	2.4	0.6	0.65	109.1	82.6

Table 3.6 shows that in some cases, the design performs well compared to the design with the true priors. Run 1, for example, with $w_N = 0.65$ is 93.7% and 90.7% as efficient as the original design for the linear model and binary model, respectively. Run 3 on the other hand, at weight 0.65 performs poorly for the binary model but better for the linear model compared to the original design. A value above 100 indicates that the design with the misspecfied priors has a higher design criterion for that weight compared to the design with the original priors. In run 3, the design at weight $w_N = 0.65$ is an unbalanced 2^2 factorial design (Figure B.x), a very different design compared to the corresponding design using the true priors (Figure 2.3). The efficiency plot shown in Figure B.18c shows that with this set of priors, the design at $w_N = 0.50$ provides more balanced behavior between the relative efficiencies of the two models. The other runs had fewer tradeoffs in design efficiencies when $w_N = 0.60$ or 0.65. Overall, when $w_N = 0$, i.e. the

binary optimal design, performs worse for the binary model and linear model compared to the design with the true priors.

3.5 Conclusions

In this chapter, dual-response systems which could include the normal, binomial, or Poisson distribution were introduced. For normal-Poisson systems, the optimal design for the normal model and the Poisson optimal design were very similar. There were fewer tradeoffs between these two designs. Changing the priors of the Poisson model parameters affected the design efficiency to some extent. However, the designs still have high relative efficiency. The tradeoffs between the binomial and the Poisson distribution are much greater. Using the weighted criterion allows for better balance between the extremes of choosing the optimal design for one response over the other.

The Bayesian D-criterion can be sensitive to the choice of the parameters, as shown in section 3.4.1. It is important to examine the effect the priors can have on the weighted designs. A sensitivity study is recommended to analyze how the designs (and thus their ability to fit the appropriate models) change if the priors are not correct. In section 3.4.2, we showed one example of this for the main effects model explored in Chapter 2 for the normal-binomial dual-response system.

CHAPTER 4

DESIGN OF EXPERIMENT SELECTION USING LAYERED PARETO FRONTS

4.1 Introduction

In chapters 2 and 3, several designs across ranges of weights on the design criteria were identified as potential choices for use in an experiment for a dual-response system. Only one design can be chosen to run in the experiment, so the question now shifts to how to choose the final design. In this chapter, we present a two-phase decision-making process using layered Pareto Fronts to help guide the user to a final decision. This method can be applied to the set of designs identified in chapters 2 and 3, for example.

4.1.1 Motivation

In a decision-making process, relying on only one objective or criterion can lead to oversimplified sub-optimal decisions which ignore important considerations. Incorporating multiple, and likely competing, objectives is critical during the decision-making process in order to balance the tradeoffs of all potential solutions. Provided one can quantify the objectives, there are methods to identify the best or optimal solution for a specified set of priorities. Lu et al. (2011) demonstrated a method using a two-stage Pareto Front approach that objectively eliminates non-contending solutions and then allows the decision-makers to make the final decision based on their priorities. However, there are many situations where we may be interested in several optimal solutions, not

just one. These types of decision processes may fall into one of two scenarios: 1) decision-makers want to identify the best N solutions to accomplish a goal or specific task, or 2) a decision is evaluated based on several primary objectives we can formally quantify along with secondary, qualitative priorities. In this second scenario, we identify several contending solutions (and eliminate non-contenders) using the primary, quantitative objectives. We then use the secondary, qualitative objectives to make the final decision.

4.1.2 Design of Experiments Application

In DOE, classical designs (such as fractional factorial and central composite designs) are known to have good overall performance for several design properties: Does the design provide good estimation and prediction for the fitted model, are we able to test model assumptions, is the design robust if things go wrong, is the design flexible, is it cost effective, etc (see Myers et al. (2016) p. 370 for a more comprehensive list of design criteria). However, these classical designs have some restrictions: they are limited to certain sample sizes and assume the design region is cuboidal or spherical. Computer-generated optimal designs (e.g. D-optimal, I-optimal, or alias-optimal designs) have greater flexibility to be tailored to match the sample size or design region required for a specific scenario. However, these designs are optimal with respect to only one criterion. Therefore, some of the favorable design properties exhibited in classical designs are typically not present in many computer-generated optimal designs. There has been increased interest in leveraging the advantages of both classical and computer-generated

optimal designs, creating or selecting a design optimal on multiple criteria (Lu et al., 2011). These designs balance tradeoffs in various measures of design performance and thus perform well for multiple criteria. These designs allow greater flexibility and are a better potential match to the experimenter's specified priorities.

We consider an application of a top N search in DOE that falls into the second scenario of formally considering several quantitative criteria, while also incorporating other qualitative measures to finalize the decision. When running a designed experiment, we typically only have time and resources for one experiment. When budgets are constrained and there are several factors under consideration, the design constraints often dictate a small number of runs relative to the number of factors. The choice of design is now critical as we must balance good design performance under an assumed model. Johnson and Jones (2010) outline all non-isomorphic orthogonal regular and non-regular 16-run designs for six, seven, and eight factors. In this paper, we consider the 8-factor case with 80 non-isomorphic designs from which to choose. We first evaluate the designs on two quantitative metrics to reduce the number of designs under consideration to a more manageable size. We then select a final design based on several qualitative measures including: preference for regular or non-regular designs, presence of replicates, and design geometry.

Since these designs are screening designs, we consider a first order plus two-factor interaction model and identify all active effects. The model has the form:

$$y = \beta_0 + \sum_{i=1}^{8} \beta_i x_i + \sum_{j>i} \sum_{i=1}^{8} \beta_{ij} x_i x_j + \varepsilon. \quad (4.1)$$

Note that this model has up to 37 terms (1 intercept, 8 main effects, and 28 two-factor interactions). Since the designs under consideration have only 16 runs, these designs are supersaturated (Satterthwaite, 1954). Although it is not possible to simultaneously estimate all 37 terms, we would like to be able to evaluate and identify all active main effects and two-factor interactions. In many cases, we can assume effect sparsity, where only a fraction of effects are active; however, we do not know before the experiment which factors are active and which are inactive. To help evaluate these designs and eliminate poor designs from further consideration, we use two common quantitative metrics: $\mathrm{E}[s^2]$ and $tr(\mathbf{AA}')$.

$\mathrm{E}[s^2]$ is a common metric for supersaturated designs and is defined as:

$$\mathrm{E}[s^2] = \frac{2}{f(f-1)} \sum_{i<j} s_{ij}^2, \tag{4.2}$$

where s_{ij} is the (i,j)th element in the $\mathbf{X}'\mathbf{X}$ matrix (where $\mathbf{X}$ is the model matrix excluding the intercept) and f is the number of terms in the model minus 1. In supersaturated designs, we would like the off-diagonal elements to be as small as possible so that the covariance between model terms is as small as possible (Myers et al., 2016 p. 201).

For the second criterion, a common response surface-based approach for supersaturated designs is to prioritize some terms and look at how their estimated parameters are impacted by other secondary terms. Here, for a screening design the approach is to fit the model with main effects only and choose a design that protects against aliasing if some of the two-factor interactions are active. The model then takes the form:

$$\boldsymbol{y} = \mathbf{X_1}\boldsymbol{\beta_1} + \mathbf{X_2}\boldsymbol{\beta_2} + \boldsymbol{\varepsilon}, \tag{4.3}$$

where the model matrix $\mathbf{X}$ is partitioned into $\mathbf{X_1}$ which contains the intercept and main effects only and $\mathbf{X_2}$ which contains the two-factor interactions. By considering the bias of $\boldsymbol{y}$ when only fitting the reduced model ($\boldsymbol{y} = \mathbf{X_1}\boldsymbol{\beta_1} + \boldsymbol{\varepsilon}$), we can use the alias matrix $\mathbf{A} = (\mathbf{X_1'X_1})^{-1}\mathbf{X_1'X_2}$ to quantify the impact of active secondary terms on estimating the primary terms. Minimizing the trace of $\mathbf{AA'}$ is a common metric to protect against model misspecification (Myers et al., 2016). We use the alias structure as defined here throughout the rest of this paper, including the example discussed in section 4.4.

In addition to these quantitative metrics, we also wish to incorporate several qualitative criteria into the final decision: regular or non-regular designs, presence of replicates, and design geometry. For example, we may have a preference for either a regular design (where terms are either completely aliased or orthogonal) versus a non-regular design (where there are likely fewer completely aliased model parameters, but there is non-zero correlation between some terms). To better understand the alias structure of these designs, we can also evaluate these designs using color maps of the correlation structure of the model parameters. We will return to this example in section 4.4 with a detailed case study.

4.2 Literature Review

Multi-objective optimization is not a new problem in the statistics literature, particularly in response surface methods. A common method for combining the responses into one overall score is through DF, made popular by Derringer and Suich (1980). Because the criteria likely have different units and are on different measuring scales, this

approach allows us to scale each criterion to values in the range [0,1] and then more fairly combine them. How you scale the criteria to this range requires careful consideration to ensure fair comparisons between these scaled criteria. For example, the range of one criterion may be much smaller (say, ranging from best to moderate) compared to the range of another criterion (which may range from best to terrible). We discuss four methods for scaling in section 4.3.3. Once the desirability score for each criterion has been calculated, the scaled criteria can be combined using one of several different functional forms. Two of the more common choices are the additive and multiplicative DFs. We investigate differences between these two DF in section 4.5. Additional information on multi-objective optimization methods can be found in Carlyle et al. (2000), Myers et al. (2016), and section 2.2.

4.3 Methods

4.3.1 Pareto Front Optimization

Another approach to multi-objective optimization is Pareto optimization - a technique used to optimize multiple criteria simultaneously while incorporating the experimenter's priorities. While the methods described above focus on algorithms to identify a single best solution, the Pareto Front (PF) approach seeks to identify a collection of competitive solutions that can be examined before making a final choice. The method is divided into two phases of analysis: 1) an objective phase where potential rational solutions are identified, eliminating non-contending choices from further consideration and 2) a subjective stage, where the remaining solutions are evaluated for

different prioritizations of the criteria. A final decision is made by evaluating the results from phase two of the analysis through the use of several graphical tools (Lu et al., 2011, Lu and Anderson-Cook (2013), and Lu et al., 2014). Anderson-Cook and Lu (2015) developed a five-step process called DMRCS (define, measure, reduce, combine, and select) to approach this general problem of evaluating competing objectives. PF optimization is an ideal method used in the reduce step of this process.

A solution in the criterion space is called a dominated solution if there exists another solution that is strictly better for at least one objective and at least as good for the remaining objectives. The PF is formed from the non-dominated points in the criterion region. The goal of the first phase of Pareto optimization is to eliminate the dominated, inferior points from further consideration since there is always at least one superior solution on the PF. A point on the PF cannot improve one objective without making at least one other criterion worse.

Figure 4.1 shows a representation of a PF for a scenario with two criteria, both of which are optimized by maximizing. Ideally, we would like a solution at the utopia point, the point that has the best possible values for all criteria simultaneously. The utopia point is represented by the triangle in Figure 4.1. However, this ideal solution rarely exists in practice since there are generally tradeoffs in the criteria. The points connected by the line in the figure make up the PF: the set of non-dominated points in the criterion region. The remaining points are dominated by at least one solution on the PF. Note that we would never choose these dominated non-PF points as the optimal solution as there is always at least one point on the PF that outperforms it.

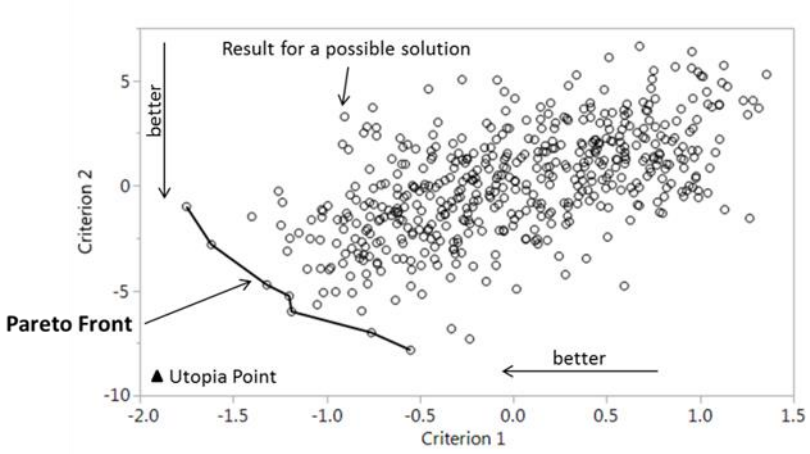

Figure 4.1. Graphical representation of a Pareto Front

Methods for constructing Pareto Fronts fit into two categories. The first identifies a subset of rational choices from an enumerated list of solutions. In this scenario, a decision-maker has evaluated a set of solutions on multiple criteria and wishes to choose the "best" choice based on the evaluated criteria. In the second category, finding the PF is incorporated into a search algorithm of candidate solutions. For example, a PF search was incorporated using the Pareto Aggregating Point Exchange (PAPE) algorithm to find designed experiments optimal for multiple design objectives (Lu et al. (2011), Lu and Anderson-Cook (2014), and Lu et al. (2014).

PF optimization provides several advantages compared to the multi-objective optimization methods described in section 4.2. Rather than provide users with a single optimal solution, PF optimization gives a set of choices from which to choose. While the method does not provide one (or *the*) solution, it does provide a set of rational promising choices. The first phase of the PF optimization algorithm eliminates non-contending solutions from the start, simplifying the problem and making the final decision more manageable. This first phase of the algorithm is also independent of any choice of scaling

on the criteria because the algorithm simply identifies the dominant points based on their evaluated criteria values. This method also separates the objective and subjective phases, and allows decision-makers to bring their own priorities into the second phase of the algorithm by examining the optimal solutions for different prioritizations (or weightings) of the criteria. Because of this flexibility, the method can accommodate several competing sets of priorities, particularly when multiple decision-makers are involved. The second phase of the algorithm also lets the user evaluate the sensitivity of a particular solution based on different weights on the criteria. The PF provides a range of reasonable values for each criterion and the second stage allows the user to compare a given solution to alternatives at any given weight combination of the criteria. Finally, the algorithm can be adapted to accommodate any number of criteria. However, we note that the computation time increases exponentially as the number of criteria increases. It is therefore recommended that practitioners be selective on the number of criteria used for optimization. Including too many criteria in the PF optimization not only reduces the computational efficiency, but also can result in mediocre solutions due to large trade-offs associated with many criteria.

4.3.2 Layered Pareto Front

Pareto optimization helps identify an optimal solution based on user priorities to balance multiple objectives or criteria. In this research, we do not want to identify just a single optimal solution, but the top N solutions from an enumerated list of candidates. To accommodate multiple solutions, we have adapted the two-phase PF approach (Lu et al.,

2011) to include a layered PF approach for the top N solutions. While the traditional PF looks for just the non-dominated solutions, layered PFs consider potential solutions that lie just behind the PF as potential candidates. The decision-maker may have one of two objectives for the layered PF approach: 1) given multiple quantitative criteria, they need to identify the top N solutions to accomplish a task; or 2) they must make a decision that is evaluated based on several primary quantitative criteria as well as secondary, qualitative priorities. In this second scenario (which we demonstrate in section 2.4), because secondary criteria will subsequently be considered, it can be helpful to have a larger pool of potential solutions from which to draw candidates. While a solution on the second or third layer is not the top choice regardless of how the criteria are prioritized, this solution could be a highly competitive choice when looking for the set of top N solutions. In the first objective phase of analysis, layered PFs are identified, eliminating the dominated choices from further consideration. In the second phase, the identified points from phase 1 are ranked from 1 to N for different weight combinations on the criteria to identify the top N solutions under different prioritizations. This Top N Pareto Front Search (TopN-PFS) algorithm was programmed in the *JMP* Scripting Language (V. 12) and is available as an Add-In in JMP.

4.3.3 TopN-PFS Algorithm

The TopN-PFS algorithm is divided into two phases. The first is the objective stage where the points in the m layers of the PF are identified and m is a value specified by the user, and the remaining points are discarded from further consideration. To build the layers of the PF, we find the non-dominated points in the traditional PF, label them as

belonging to layer 1, and set them aside. We then build the next layer by examining the remaining points and finding the non-dominated points on the PF for this reduced set of points. These points are labeled as belonging to layer 2 and again set aside. This process continues until m layers have been identified. This phase of the algorithm corresponds to the reduce step in the DMRCS process outlined by Anderson-Cook and Lu (2015).

In the second subjective stage, DFs (Derringer and Suich, 1980) are used to rank the points from phase 1 for different combinations of weights on the criteria. The different weights allow users to compare changes in the solutions when changing the emphasis of the different criteria. However, the criteria, which measure different aspects of the potential choices, are likely measured on different scales. To allow for a fair comparison of each criterion, values x_{ij}, denoting the jth value in the dataset for criterion i, are scaled to values z_{ij} such that:

$$z_{ij} = \begin{cases} \dfrac{x_{ij} - worst_i}{best_i - worst_i}, \text{if criterion } i \text{ is maximized} \\ \dfrac{worst_i - x_{ij}}{worst_i - best_i}, \text{if criterion } i \text{ is minimized} \end{cases}, \tag{4.4}$$

where $worst_i$ and $best_i$ are defined as the worst and best values for the ith criterion, respectively. These scaled values are called "desirability scores." Using these scaling schemes, a value of 0 corresponds to the worst value and a value 1 corresponds to the best value. How $best_i$ and $worst_i$ are defined depends on the judgment of the decision-maker or the properties of the particular data set under consideration. Four options are considered here: 1) the best and worst values are calculated from the data only in the layered PFs; 2) The best and worst values are calculated from the entire dataset; 3) The best value is calculated from the data in the layered PFs and the user provides the worst

values for each criterion; 4) the user provides the best and worst values for each criterion. In the case of scaling options 1 or 2, for example, if criterion i is maximized, the worst value would be the minimum in the applicable dataset, while the best value would be the maximum. If criterion i is minimized in these scaling scenarios, the reverse is true. The third or fourth options for scaling may be appropriate when user-specified reference values are available or if the ranges of the criteria are vastly different. For example, if the range of criterion 1 covers excellent values to very poor values while the range for criterion 2 covers good to fair values, the user may want to supply values for $best_i$ and $worst_i$ in order to balance these discrepancies. These specifications can adjust the values for all the criteria in the range 0 to 1 on the desirability scale to be more comparable. These different options in scaling allow for more flexibility in phase two of the TopN-PFS algorithm; the user has more control over their priorities in this subjective stage of analysis. The choice of scaling can have a substantial impact on the results of the TopN-PFS algorithm. Hence, careful thought should be given to ensure that the choice of scaling matches with the decision-maker's intentions.

With the original data transformed to desirability scores z_{ij} (values between 0 and 1), we can fairly combine these values to compare them with different weighting schemes on the criteria. Two common DFs (the additive and the multiplicative DFs) are considered for k criteria:

$$Add\ DF_j = \sum_{i=1}^{k} w_i z_{ij} \tag{4.5}$$

$$Multi\ DF_j = \prod_{i=1}^{k} z_{ij}^{w_i}, \tag{4.6}$$

where the weights $w_i \geq 0$ satisfy $\sum_{i=1}^{k} w_i = 1$ and z_{ij} is the scaled value of the jth solution for the ith criterion. The choice of DF depends on the priorities of the user. The multiplicative DF penalizes low criterion values more than the additive DF. For example, if one choice performs very poorly for one of the criteria, it is very difficult for other criterion scores to compensate for this performance. The additive DF, on the other hand, is more forgiving of low criterion scores, since a high value for one criterion can trump a low value for another criterion. The choice of DF can impact which solutions are highlighted as best with this algorithm. We examine the effect of the DF under a variety of circumstances in a simulation study described in section 4.5.

DF scores are calculated for each solution identified in phase 1 of the algorithm and across different combinations of weights on the criteria that match the user's priorities. For each weight combination, the DF scores are ranked from highest to lowest to identify the top N solutions. This stage of the algorithm corresponds to the combine step in the DMRCS process. Several graphical summaries for comparing and selecting solutions are described in detail in section 4.4 (the case study).

4.3.4 Handling Ties

One interesting consideration, particularly in the DOE application, is how to handle ties. There are two types of ties that can occur in the TopN-PFS algorithm. The first type occurs in the first phase and results from two or more solutions that have the exact same values for all criteria. No matter how we scale the data, these solutions always have the same values of the criteria. Without taking tied solutions into consideration,

potential solutions could be lost in phase 1 of the algorithm. If there are two (or more) solutions with the same values for all criteria and one solution is on a layer in the layered PF, the matching solution would not be included because it is not strictly better on any one criterion. Therefore, for the general PF algorithm, the tied solution would not be added to the same layer in the PF as the point with which it is tied. Since we wish to include any potential solution that performs well for the criteria, we include all tied solutions as potential solutions. Therefore, the approach to handling ties is to choose a representative solution while building each layer in the PF. After the initial solutions in a layer in the PF have been identified, any remaining solutions that are tied with solutions in that layer are added to the same layer in the PF. The algorithm then continues to the next layer in the PF.

The second type of tie occurs in the second phase of the TopN-PFS algorithm. In the second phase, we calculate DF scores for each solution as described in the previous section. Although there may be solutions that are not tied in terms of their criteria values, more than one solution may have the same DF score for a particular weight combination of the criteria. Therefore, when we rank the potential solutions in terms of the DF score, there may be ties. If two or more solutions have the same DF score and are in the top N, these tied solutions are given the same ranking. If there are ties in the DF score, there could be more than N solutions identified for each weight combination of the criteria. Once at least N solutions have been identified, no additional solutions are included in the ranking; therefore, there may be fewer than N unique DF scores in phase two of the algorithm.

4.4 Case Study

We demonstrate the TopN-PFS algorithm for the DOE application described in section 4.1.2. We consider the 8-factor case with 80 non-isomorphic orthogonal regular and non-regular 16-run designs. The goal is to select a single design to run based on overall performance across multiple design characteristics which have been identified and measured in the first two steps of the DMRCS process. We first evaluate the designs on the two quantitative criteria $\mathrm{E}[s^2]$ and $tr(\mathbf{AA}')$ and use a layered Pareto Front in the third and fourth steps of the DMRCS process to reduce the number of choices to the top five choices. We then introduce several qualitative criteria (which include preference for regular or non-regular designs, presence of replicates, and design geometry) to select the final winner from the top five choices in the last step of the DMRCS process. Table 4.1 displays the values of the criteria $\mathrm{E}[s^2]$ and $tr(\mathbf{AA}')$ for these designs.

The first column lists the design numbers as listed in the catalog in Johnson and Jones (2010) and identifies the designs that are tied in terms of both criteria in parentheses. Using $m = 5$ PF layers, we first objectively eliminate non-contenders in phase 1 of the TopN-PFS algorithm. In phase 2, we use the additive DF to reduce the contenders to $N = 5$ choices. Several graphical tools are used to help choose between these remaining designs. Finally, we introduce our qualitative criteria to make a choice of which design to select. The steps of the algorithm are described in the following sections.

Table 4.1

Summary of Case Study Criteria

Design	$E[s^2]$	$tr(\mathbf{AA}')$	PF Layer
1	17.06667	21	>5
2 (10, 14, 19, 23, 44, 54, 59)	12.19048	15	>5
3 (8, 15, 24, 27, 35, 39, 49, 55)	12.19048	12	3
4 (17, 26, 42, 48, 77)	12.19048	9	1
5 (9, 16, 20, 25, 28, 30, 36, 50, 61, 63)	10.97143	12	1
6	17.06667	0	1
7 (32, 75)	12.19048	14.25	>5
11 (29, 34, 37, 43, 46, 51, 56, 57, 65, 69, 70, 74, 78)	11.58095	13.5	4
12 (40, 47, 58, 76)	11.58095	10.5	1
13	14.62857	15	>5
18	14.62857	6	1
21 (33, 52, 64, 71, 73)	11.58095	12.75	3
22	14.62857	18	>5
31 (45, 66, 79)	11.58095	12	>5
38	13.40952	12	4
41	13.40952	9	2
53	13.40952	15	>5
60	13.40952	16.5	>5
62 (80)	12.19048	13.5	5
67 (68)	12.8	10.5	2
72	12.8	15	>5

4.4.1 Phase 1 Analysis

We use $m = 5$ PF layers in phase 1 of the TopN-PFS algorithm. Note that $m \leq$ N in general. Figure 4.2 shows the criterion space for the 80 designs, with the designs in each Pareto Front layer identified by a unique symbol (and labeled with a representative design as in Table 4.1).

Table 4.1 also lists the layer that each design belongs to for this example, with ">5" indicating that a design is not on the top 5 layers and hence is not considered

further. Recall that the two criteria are both optimized by minimizing, so the bottom left hand corner of Figure 4.2 represents the ideal location for a design. Note that the designs in layer 1 are closest to the utopia point, followed by layer 2, etc. The non-labeled points in the figure are the non-contenders, which have been eliminated by the first phase of the algorithm.

Table 4.2 summarizes the composition of each PF layer, accounting for ties. There are 28 designs in the first layer; however, these 28 designs have only five unique sets of criteria values due to ties. Using five PF layers, we have reduced the number of designs under consideration to 67 designs, eliminating 13 from further consideration. In section 2.5, we consider the effects of using a smaller number of PF layers ($m < \mathrm{N}$) in the algorithm under several different scenarios.

Table 4.2

Summary of Layered Pareto Front

Pareto Front layer	Number of designs
1	28
2	7
3	15
4	15
5	2

With this reduced set of designs identified in phase 1, we move to phase 2 where user priorities are incorporated to make a decision on our top $\mathrm{N} = 5$ design choices. We demonstrate the algorithm using the additive DF. We scale the criteria values using the first option as discussed in section 4.3.3: the best and worst values for each criterion are calculated from the designs in the 5 layered PFs from phase 1. Several graphical tools to

help the user make a final decision during phase 2 of the selection process are described in the following sections.

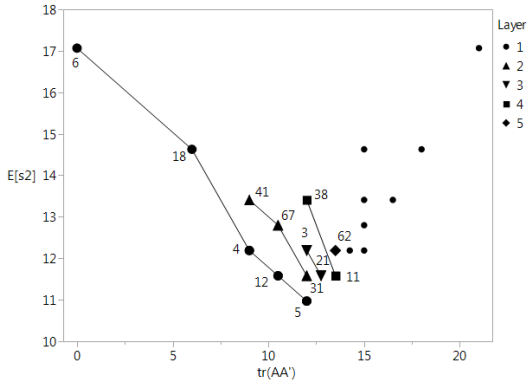

Figure 4.2. Scatterplot of E[s²] vs. tr(AA') for 16-run 8 factor design selection

4.4.2 Phase 2 Analysis

4.4.2.1 Mixture Plot

In the second, subjective phase of the TopN-PFS algorithm, the user selects the top N choices based on their priorities, which considers how the criteria are weighted in the DF. In many situations, one criterion has greater emphasis than another and thus the weight on that criterion is larger. In other situations, each criterion may be considered similarly important, and approximately equal weights are appropriate. Furthermore, in many decision-making problems, multiple decision-makers may have conflicting opinions on the importance (and therefore the weights) of each criterion. The TopN-PFS algorithm allows flexibility in the analysis of the phase 2 results with the top N solutions shown across all weight combinations.

The first graphical tool, the mixture plot (Figure 4.3), uses gray-scale to show the top N solutions across all evaluated weight combinations. The goal of this plot is to allow the decision-maker to identify best solutions for a given set of priorities when making a final decision. This plot shows how the top N solutions change as the emphasis on the criteria changes. Providing a range of options from which to select depending on how the criteria are weighted provides deeper understanding of the alternatives and allows maximum flexibility to the user. The horizontal axis of this plot is the weight placed on one criterion, in this example $\mathrm{E}[s^2]$. Because the criteria weights sum to one, the weight placed on $tr(\mathbf{AA}')$ is equal to one minus the weight on $\mathrm{E}[s^2]$. The vertical axis of the mixture plot shows the different options (in this example, designed experiments) under consideration. Black squares denote the top choice for a given weight combination and the lightest gray denotes the Nth choice. White space in this plot indicates that a particular option (e.g. design) was not a top N solution for that weight combination. Top solutions with tied scores have the same color shading, clearly identifying ties in DF scores.

In the design example, twenty designs are potential solutions, depending on how the criteria are weighted. To read the plot, select a weight combination and identify the solutions that have a square with any black or gray shading. For example, consider 0.25 weight on $\mathrm{E}[s^2]$ and 0.75 weight on $tr(\mathbf{AA}')$. At this weight combination, design 6 has the square with the darkest shading indicating it is the highest ranked solution with the largest DF score (equal to 0.75 and read from the plot by hovering the mouse over the given solution and weight within the JMP Add-In). Design 18 has the next highest score (0.52). Designs 4, 17, 26, 42, 48, and 77 are all tied for the 3rd highest DF score of 0.45.

In this example, since there were ties in DF scores, more than 5 design choices are available at this weight combination. Alternatively, we could consider a weight combination with 0.75 weight on E[s^2] and 0.25 weight on $tr(\mathbf{AA}')$. At this weight combination, 11 designs (5, 9, 16, 20, 25, 28, 30 36, 50, 61, and 63) are tied for the highest DF score of 0.78 (indicated in the mixture plot by the black squares for each of these designs). In general, these 11 designs perform favorably when more weight is placed on E[s^2] compared to $tr(\mathbf{AA}')$. This plot allows the experimenter to visually examine the tradeoffs as priorities vary. In addition, for a decision that has multiple decision-makers, the mixture plot facilitates a discussion when there are differing opinions on the emphasis for each criterion. The plot allows users to see how similar or different the top N solutions are for different weight combinations and shows how robust the top N solutions are to different weight combinations.

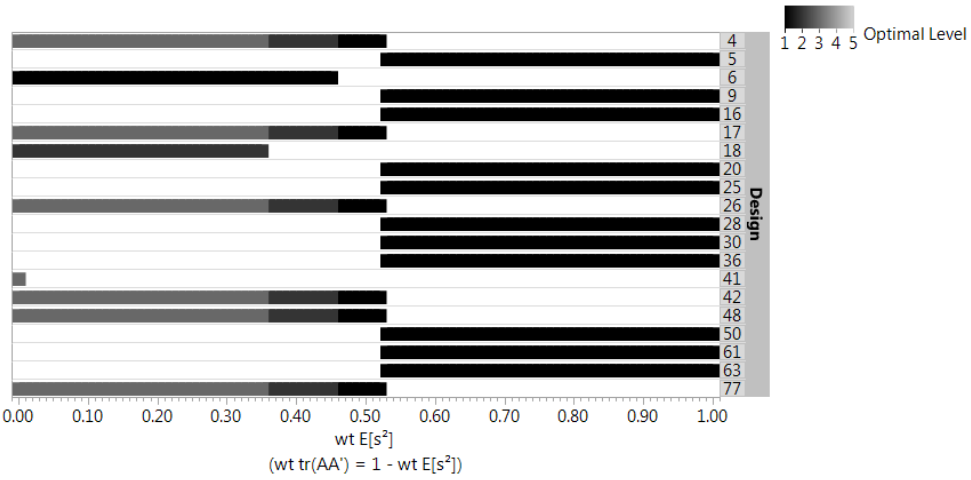

Figure 4.3. Mixture plot for 16-run 8 factor design selection

4.4.2.2 Proportion Plot

To better understand how frequently a choice was a top N solution, the proportion plot shown in Figure 4.4 can be used. This plot is a stacked bar chart that indicates how frequently each design is a top N solution across all of the evaluated weight combinations of the criteria. This plot indicates how robust a particular choice is across the range of criteria weightings. Only the choices that are a top N solution for at least one weight combination are displayed in this plot. The proportion plot uses the same gray-scale as the mixture plot to indicate the frequency a given solution was the top choice, second best choice, etc. The proportions displayed in this plot are determined from the entire set of weights that were evaluated in the mixture plot. If there is uncertainty on how the criteria should be weighted for a particular situation, this plot can provide insight on a choice's selection across the weight ranges. However, if it is known a priori that only a limited range of weights is of interest (as described in Lu et al., 2014), the plot can be adapted to summarize frequency for that subset of the weights.

The proportion plot can also highlight the importance of solutions on the second or third PF layer. While a choice on a lower layer cannot be the best solution for any particular weight, it can be highly ranked for many weight combinations. If we restricted our set of choices to only those in the first Pareto Front layer, we would miss out on these potentially important solutions which may have close to optimal performance based on the primary criteria and very good performance for the secondary qualitative criteria.
In the design example, of the twenty designs identified in the mixture plot of phase 2 of the TopN-PFS algorithm, 18 are a top 5 choice with similar frequency. Designs 4, 17, 26,

42, 48, and 77 are top 5 designs 52% of the time across all the weight combinations. However, these designs were the best choice for only a small fraction of the weights (7%). For many of the weight combinations, these designs were the third best design in terms of the DF score. Designs 5, 6, 9, 16, 20, 25, 28, 30, 36, 50, 61, and 63 were a top 5 choice 47.5% of the time across all the weight combinations. These designs were always ranked first for these weight combinations. In this example, because there were many tied designs (as seen in Table 4.1), the top 5 designs across all weight combinations were restricted to those designs only in the first PF layer.

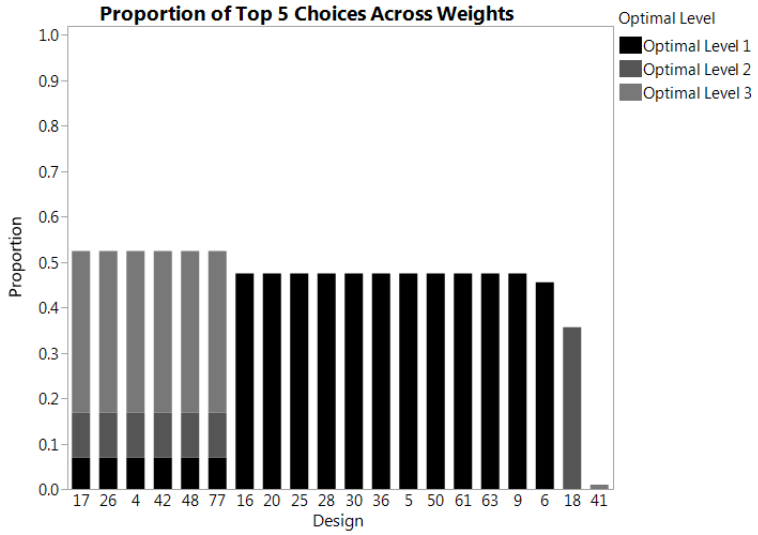

Figure 4.4 Proportion plot for 16-run 8 factor design selection

4.4.2.3 Parallel Plot

The previous two plots combine the criteria values into a single score. While these plots are useful in identifying important solutions in terms of their DF score, it is also important to understand how each choice compares in terms of the original criteria values. The parallel plot, introduced in Lu and Anderson-Cook (2014) and shown in

Figure 4.5 is a graphical tool to visually examine the tradeoffs of one solution over another in terms of the scaled criteria values. The best values of each criterion are on the top of the plot and the worst values are on the bottom. Each line in the plot corresponds to a particular solution (e.g. design). These lines are on a gray-scale where black indicates the highest scaled value for the first criterion and the lightest gray indicates the lowest scaled value for the same criterion. This plot is an ideal tool to visualize the tradeoffs in the criteria. If many lines cross, then there are more tradeoffs between the solutions.

In the design example, a black line indicates the highest scaled value for $\mathrm{E}[s^2]$ and light gray indicates its lowest scaled value. With several designs with tied values for the two criteria in this example, one line represents many designs. Eleven designs (5, 9, 11, 16, 20, 25, 28, 30, 36, 50, 61, and 63) have the best value of $\mathrm{E}[s^2]$, but a very poor value of $tr(\mathbf{AA}')$. In contrast, design 6 has a very poor value of $\mathrm{E}[s^2]$, indicated by the very light gray line, but has the best value of $tr(\mathbf{AA}')$. Relatively steep lines like these indicate extreme tradeoffs between the two criteria. Flatter lines indicate less tradeoff between the two criteria, but potentially no exceptional performance for either criterion. Design 18, for example, has less of a tradeoff between the two criteria, but neither criterion value is particularly good. In this example, there are substantial tradeoffs between the two criteria as no design simultaneously performs well for both criteria.

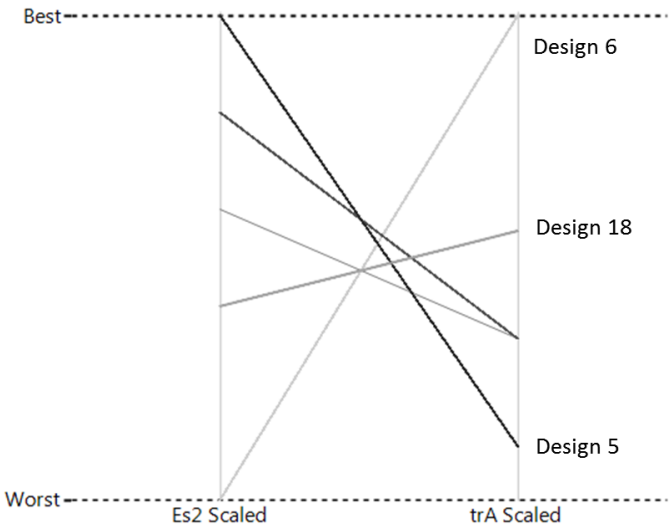

Figure 4.5 Parallel plot for 16-run 8 factor design selection

4.4.2.4 Synthesized Efficiency Plot

Using the previously described graphical tools together can help the experimenter make an informed decision on which solutions are best for their priorities. Another graphical tool to help explain the implications of each choice is the synthesized efficiency plot, first introduced by Lu and Anderson-Cook (2014). This plot provides the user with a visual tool that shows how each choice compares to the best available choice for any weight combination. This plot can facilitate a discussion when there is no consensus on which N choices to make in the final decision. For solutions that are not a top N solution for certain weight combinations, we can compare their DF scores to the best available at any particular weight combination.

The synthesized efficiency value at a given weight combination $\boldsymbol{w}$ is defined as:

$$Syn\ Eff_{\boldsymbol{w}} = \frac{DF\ Score_{\boldsymbol{w}}}{\max(DF\ Score_{\boldsymbol{w}})}. \tag{4.7}$$

This metric compares the efficiency of a solution at a particular weight combination relative to the best available solution at the same weight combination. Therefore, the top choice at a given weight combination always has a synthesized efficiency value of 1. Figure 4.6 displays the synthesized efficiency for the top N solutions across the weight combinations. The values of synthesized efficiency are color-coded on a blue to white scale, where dark blue indicates a high efficiency value and white indicates poor performance. A high value for synthesized efficiency means that a given solution has a similar DF score to the best available at that weight combination.

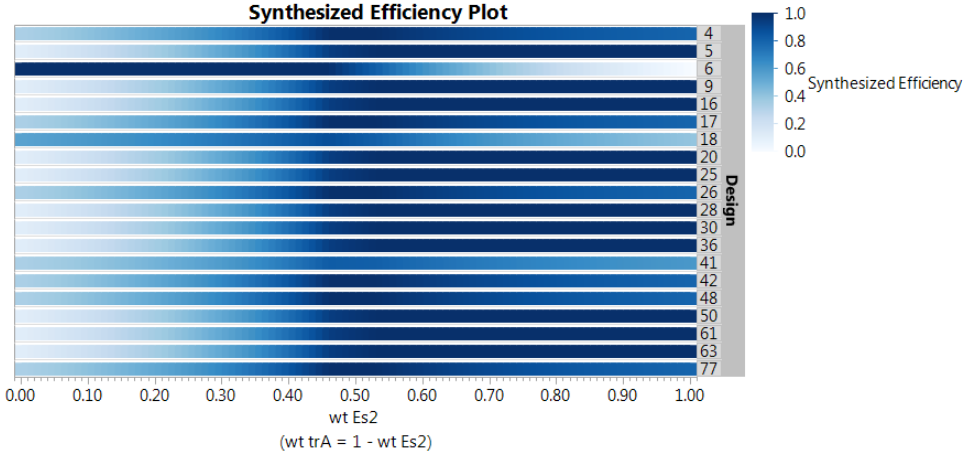

Figure 4.6 Synthesized Efficiency plot for 16-run 8 factor design selection

This plot helps explain the implications of choosing one solution over another in the top N set of solutions. In the design example, we see that design 6 (which is the top choice for a large portion of weights when $tr(\mathbf{AA}')$ is prioritized) has a large drop in synthesized efficiency as $\mathrm{E}[s^2]$ is weighted more heavily compared to $tr(\mathbf{AA}')$ (Figure 4.6). This is because design 6 has the best value for $tr(\mathbf{AA}')$, but the worst for $\mathrm{E}[s^2]$ which is also reflected in the parallel plot in Figure 4.5. Similar behavior, although not as extreme, is seen for the 11 designs that have high DF scores when $\mathrm{E}[s^2]$ is heavily

weighted. On the other hand, design 18 has more consistent DF scores and hence has more consistent medium-blue shading across the weight combinations.

4.4.2.5 N Comparison Plot

The TopN-PFS algorithm is dependent on the choice of N selected by the user. There may be some concern that the final decision may be strongly impacted by the value of N. A final graphical tool to help the user decide whether they should consider other values of N in the algorithm is the N Comparison plot shown in Figure 4.7. For each weight combination, we calculate the ratio of the DF scores for the (N+1)th best solution over that score for the Nth best solution:

$$NComp_{\boldsymbol{w}} = \frac{DFScore_{\boldsymbol{w},(N+1)}}{DFScore_{\boldsymbol{w},(N)}}. \tag{4.8}$$

This ratio is bounded between 0 and 1. Low values of this ratio indicate that the next best solution after the Nth choice is not very close to the Nth best solution, meaning it is not necessary to consider larger values of N in the algorithm. A high value of this ratio (close or equal to 1), indicates that the (N+1)th solution has a DF score very close to the Nth best solution. High values suggest that we may want to consider a larger value of N in our search. Note that a value of one indicates a tie in the N and (N+1)th solutions. By default, all tied solutions are automatically included as potential solutions in the TopN-PFS algorithm. This plot is particularly useful when ties are not a significant issue or when it is important to protect against a bad consequence of having nearly top N solutions excluded from consideration.

In the design example, there were many solutions with the same DF score for a given weight combination. Therefore, this plot (Figure 4.7) has a value of one for all weight combinations because of these tied solutions. For all of the weights, there are more than N solutions that have comparable DF scores and therefore performance. Because there were many ties, considering larger values of N is not necessary for this example.

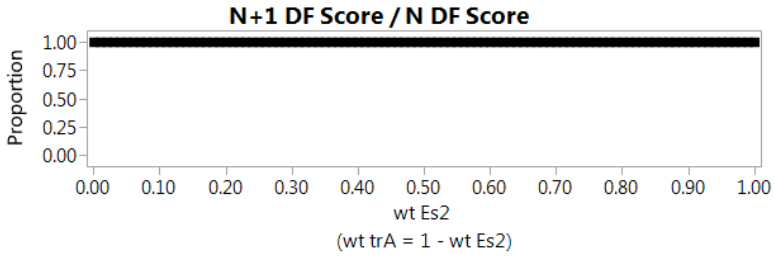

Figure 4.7 N Comparison plot for 16-run 8 factor design selection

4.4.3 Case Study Final Design Selection

Phase 1 of the TopN-PFS algorithm with $E[s^2]$ and $tr(\mathbf{AA}')$ as the two criteria of interest narrowed the design options from 80 designs to 67 designs. Using the additive DF in phase 2 then reduced the design options from 67 to 20. The experimenter now moves into the final stage of the DMRCS process by comparing the remaining solutions and making a final choice. If the experimenter is more concerned with model misspecification compared to the correlation of model parameters (e.g. $0.20 \leq wt\ \mathrm{E}[s^2] \leq 0.40$), the graphs suggest that designs 4, 6, 17, 26, 42, 48, and 77 are the strongest candidates for this emphasis since these designs are a top 3 choice across this range of weights. Note that designs 4, 17, 26, 42, 48, and 77 all have the same values for

the two criteria. Design 6 sacrifices performance in $\mathrm{E}[s^2]$, but has the minimum value of $tr(\mathbf{AA}')$. In addition, the synthesized efficiency plot in Figure 4.6 shows that these designs (with the exception of design 6) have competitive DF scores to the top 5 designs for weights when $\mathrm{E}[s^2]$ is emphasized more (say $0.6 \leq wt\ \mathrm{E}[s^2] \leq 0.9$).

The user is only able to run one experiment, so we now transition from a formal quantitative comparison and include some qualitative criteria into our final decision. None of these seven designs have replicated design points, so this qualitative criterion does not distinguish between alternatives. Of these seven designs, two designs (4 and 6) are regular designs while the remaining five are non-regular designs. For designs with a large number of factors, color maps of correlations can be effective to get a better understanding of effect confounding. These plots, introduced by Jones and Montgomery (2010) as a method to easily visualize the alias relationships in fractional factorial designs, show the absolute value of the correlations between any two potential model parameters. For the 8 factor design, this is a 36 x 36 matrix where each row and column represents the 8 main effects and 28 two-factor interactions. The main effects are listed first and two-way interactions second. The plot can be thought of as having 3 sections of interest: the top left corner (8 x 8) shows the pattern of correlation between main effects. The bottom right corner (28 x 28) shows the correlations between all pairs of two way interactions. The off-diagonal sections (top right (8 x 28) or bottom left (28 x 8)) show the correlation structure for main effect by two way interactions. Figure 4.8 shows the color map of correlations for design 17. Light gray squares indicate a correlation of 0, black squares indicate the two parameters are completely confounded, and darker gray squares indicate partial aliasing. In Figure 4.8, for example, we see that no main effects

are aliased with other main effects; however, main effects A, C, and D are each completely aliased with a two-factor interaction. All main effects except A and B are partially aliased with several two-factor interactions, 9 pairs of interactions are completely correlated with each other, and there are many pairs of interactions that are partially aliased. Ideally, we would like this plot to have mostly light gray as this indicates that the parameter estimates are not correlated with other parameters. The correlation maps are shown for the remaining six designs in Figure 4.9. Note that the design structure for designs 42, 48, and 77 (Figure 4.9.d, e, and f, respectively) are quite complicated compared to the other four designs.

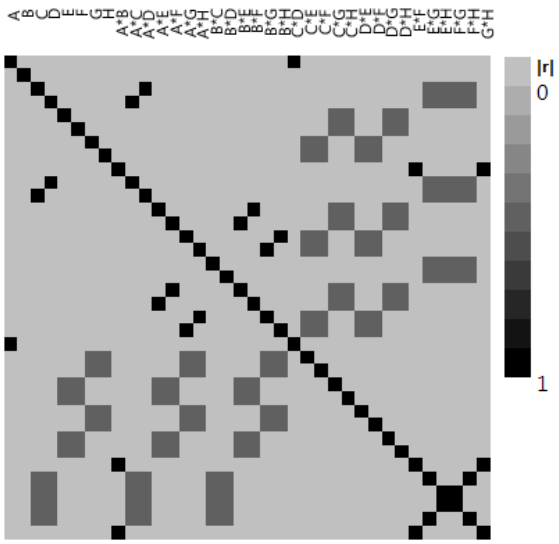

Figure 4.8 Color map of correlation for design 17

The final design decision depends on the preferences of the experimenter. If the experimenter prefers non-regular designs over regular designs, then we can eliminate designs 4 and 6 as options. Non-regular designs may be more advantageous in this setting because non-regular designs do not have complete aliasing between all of the model

parameter estimates. In particular, the regular design 4 has main effects completely aliased with two-factor interactions, a disadvantage in this scenario as the objective is to identify active main effects and two-factor interactions. This leaves a choice between non-regular designs 17, 26, 42, 48, and 77, which all have the same values for the two criteria. When considering the non-regular designs, we therefore would like to consider designs which have a minimum number of main effect by two-factor interaction aliasing. Design 17 has several main effects completely confounded with two-factor interactions, removing this design from further consideration.

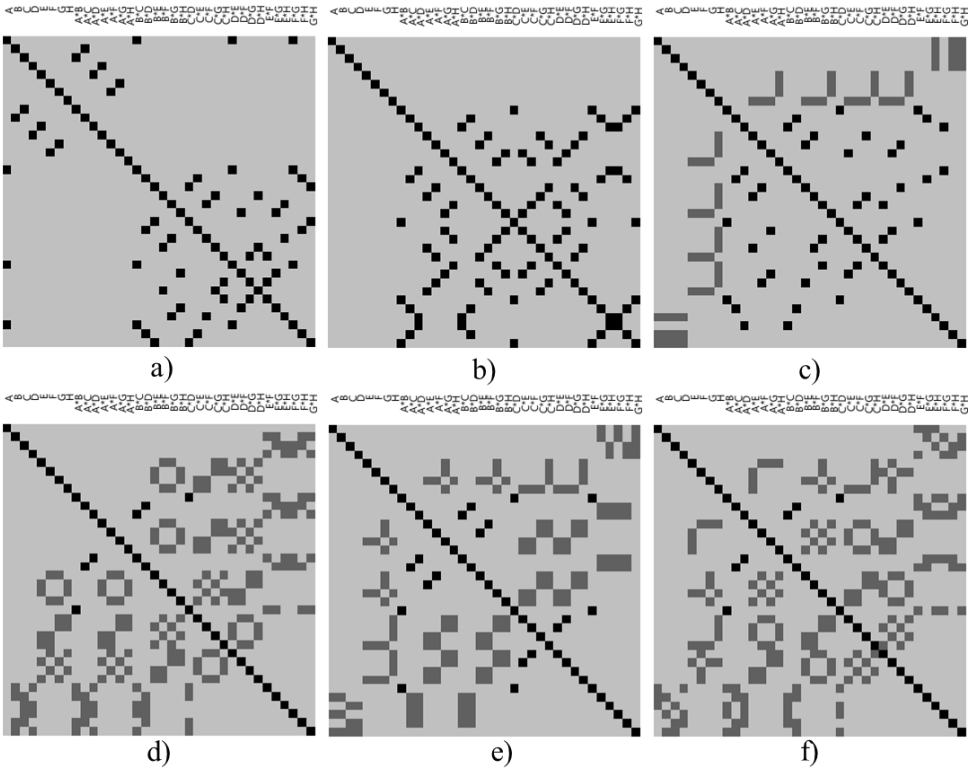

Figure 4.9 Color map of correlations for designs (a) 4, (b) 6, (c) 26, (d) 42, (e) 48, and (f) 77

To make the final decision, we can consider the geometry of the design by considering the defining equations, balance of factor levels, and design resolution. The starting design for all of the 8-factor 16-run designs in this catalog is a full factorial

design for the first four factors. The defining equations for these final four designs (Johnson and Jones, 2010) are given as:

Design 26: E=BCD,F=ABD,G=ACD,H=1/ 2[AD+AB+AC − ABCD]

Design 42: E=BCD,F=1/ 2[AC+ACD+ABC − ABCD], G=1/ 2[AC+AB+ABD − ACD], H=1/ 2[ABD+ABC+ABCD − AB]

Design 48: E=BCD,F=ACD,G=1/ 2[AD+AC+ABC − ABD], H=1/ 2[ABC+ABD+ABCD − AB]

Design 77: E=ABC,F=1/ 2[AD+ABD+ABCD − ACD], G=1/ 2[AD+ACD+BCD − BD], H=1/ 2[ABD+CD+BCD − AD].

From these defining equations (and Figure 4.9) we see that all of these designs are resolution III designs since each has partial aliasing between main effects and two-factor interactions. Design 26, in particular, has favorable behavior with main effects and two-factor interaction aliasing, but at the cost of more complete aliasing between pairs of interactions. Design 48 has the next highest number of completely aliased pairs of interactions while maintaining favorable partial aliasing of main effects and two-factor interactions. Designs 42 and 77 have the fewest number of completely aliased pairs of interactions, but each with considerably more partially aliased pairs of effects. The choice between these final four designs also depends on the goal and resources available for the experiment. If the primary goal is to identify main effects only, design 26 is favorable, but would potentially acquire additional runs if there are also active two-factor interactions. Designs 42 or 77 would be good choices if only one experiment is possible since there are fewer completely confounded pairs of interactions.

Jones and Montgomery (2010) provide a recommendation from the same catalog of designs detailed in Johnson and Jones (2010). The authors recommend design 67

which has $\mathrm{E}[s^2] = 12.80$ and $tr(\mathbf{AA}') = 10.5$. Table 4.1 shows that this design is on the second PF layer and is outperformed on both of these criteria by other designs in the catalog. However, the authors chose their recommended design primarily so that the number of completely confounded pairs of effects is zero and then sought to minimize $\mathrm{E}[s^2]$. The color map of correlations of effects for design 67 is shown in Figure 4.10. Note how there are only black and medium gray shadings in the plot, indicating no effect is completely confounded with another effect. While not on the first PF layer, this design highlights the importance of the second or third PF layers. This solution may be outperformed on the two quantitative criteria by other designs; however, it is as competitive a choice in terms of alternate qualitative criteria. The choice of quantitative criteria used in the TopN-PFS is a critical decision that must be carefully considered in the define stage of the DMRCS process. Focusing solely on select quantitative criteria without considering a more diverse set of quantitative objectives can lead to over-simplified solutions that miss potentially desirable choices. A wider set of quantitative criteria could identify additional and competitive solutions in the TopN-PFS process.

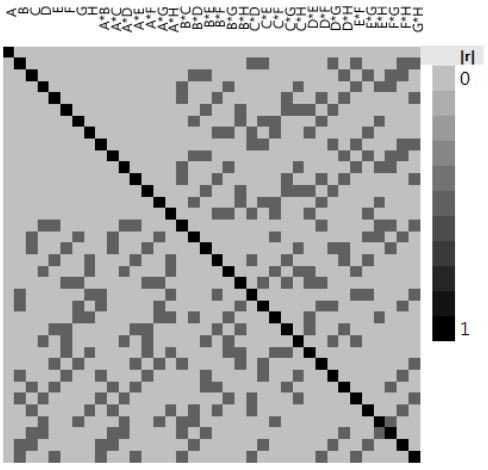

Figure 4.10 Color map on correlations for design 67

4.5 Simulation Study and Results

We conducted a simulation study to investigate the properties of the layered PFs under a variety of conditions. The study focuses on three questions: 1) What are the general properties of the PF layers; 2) How many layers are necessary to determine the top N solutions; and 3) How do the top N solutions change for different degrees of tradeoff between criteria and DFs.

4.5.1 Simulation Description

For this simulation, we used the TopN-PFS JMP add-in for a decision based on two criteria. Without loss of generality, we looked at maximizing the objectives. The values for these two criteria were randomly generated for three different data types: "normal," "uniform," and "convex." In the normal scenario, the two criteria values are uncorrelated and randomly generated from a Normal(0,1) distribution. This scenario typically produces a set of points that are a circular cloud of values with more values near the center of the bivariate distribution. In the uniform scenario, the two criteria values are also uncorrelated, but are each randomly generated from a Uniform(0,1) distribution. This scenario typically produces a collection of points with a more square shape and more even spread. In the convex scenario, we generate data so that the PF layers are anticipated to be convex in shape, representing strong trade-offs between the two criteria. To do this, we generate a large quantity of data (10,000 observations) from a Uniform(0,1) distribution for the two criteria C_1 and C_2. We then select the points in the

portion of the region that satisfy the expression $C_2 < 1 - \sqrt{C_1}$ and add a small amount of noise from a Normal(0,0.1) distribution to the C_2 coordinate. We use this subset of data from which to sample to get the intended sample size. Figure 4.11 shows representative samples for each of these scenarios for a sample size of $n = 50$.

For each of the three data types, we consider sample sizes $n = 50$, 500, and 1000. We also compare the additive and multiplicative DFs. These options (data type, sample size, and DF) give a total of 18 scenarios under investigation. We consider Top N values of N $= 1, 2, 3, 4,$ and 5 and number of PF layers $m = 1, \ldots,$ N. For each of the 18 scenarios, 5000 datasets were generated and evaluated.

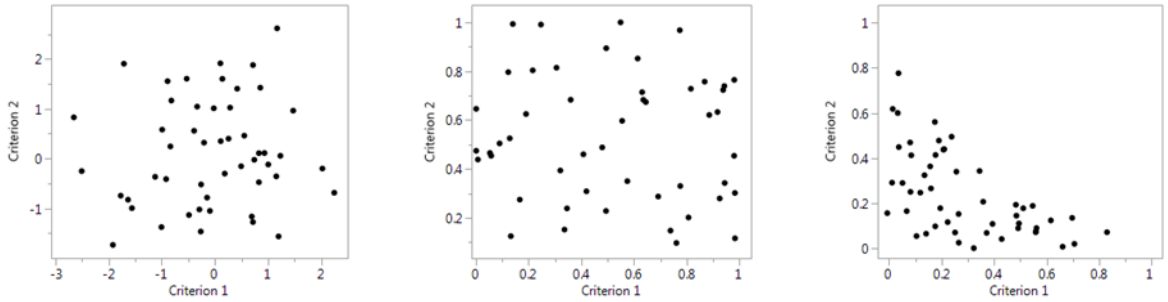

Figure 4.11 Representative data sets for each data type scenario: Normal, Uniform, and Convex (n = 50)

4.5.2 General Properties of PF Layers

We present key findings here with full results and analysis of the simulation study in Appendix B. In general, the number of points in each PF layer increases for higher PF layers, regardless of the data type. The number of points in each layer for the convex case is larger than those for the normal and uniform cases. In this scenario, there are more tradeoffs between the two criteria, causing fewer points to be dominated (and therefore,

fewer points are eliminated during Phase 1 of the TopN-PFS algorithm). The number of points on the PF layers for the normal and uniform data types was similar (Figure C.1).

We investigated the number of unique solutions (the total number of solutions ranked as a top N solution), for each scenario. The overall proportion of unique solutions for N = 5 was calculated as the number of unique potential solutions for a dataset divided by the number of points in the five PF layers of that dataset. The convex data has a smaller proportion of unique solutions across the sample sizes and DFs compared to the normal and uniform cases (Figure C.2). While the PF layers tend to be richer for the convex data, only a small proportion of these points are ever ranked solutions. A striking difference between the two DFs is that the proportion of unique solutions for the multiplicative DF tends to be higher across sample sizes and data types, likely because it is less forgiving of poor criterion values. These results highlight the importance of selecting a DF to match study goals since the final choice can depend greatly on the form of the DF.

4.5.3 Relationship Between N and m

A key question of interest is how many PF layers are necessary to identify the top N solutions. Using fewer layers eliminates more non-contending solutions in phase 1 of the algorithm and could identify the top N solutions more quickly. However, fewer layers can also miss potential solutions in phase 2. To understand the right balance between N and m, we identified the number of times an optimal level N solution (N = 1, …, 5) came from PF layer $m = 1, \ldots, \mathrm{N}$ for each dataset. We then determined the proportion of the

simulated datasets where m layers were needed to obtain all top N solutions. Figure 4.12 shows the proportion of times N-1 and N-2 layers were sufficient to identify all top N solutions for N = 2, …, 5. As N increases, N-1 PF layers capture all top N solutions more often. Using fewer PF layers for N = 2 or 3 leads to a high risk of missing potential solutions; therefore, it is recommended to use $N = m$ PF layers for these cases. For N = 4 and N = 5, there is some loss in information across all the weights by using N-1 or N-2 layers, but this loss in information is not as drastic as for smaller values of N. The convex-shaped data generally identified all top N solutions with fewer PF layers compared to the other two data types. Also the multiplicative DF required more PF layers to obtain all top N solutions.

We also considered what percent of top N solutions were missed across all weight combinations on the criteria if $m <$ N PF layers were not enough to identify all top N solutions. Figure 4.13 displays the mean proportion of missed solutions for each of the 18 scenarios for N = 2, …, 5 when using $N - a$ PF layers ($a = 1, 2, 3, 4$). As Figure 4.13 shows, using N-1 PF layers for N = 4 and N = 5 does not result in substantial loss in potential solutions. However, fewer layers would lead to compromised results. We also wanted to understand how often the higher-numbered PF layers contributed to the top N solutions. We may have a large number of points in layer 3, for example, that are never a top 5 solution. For smaller sample sizes, the layer 2 and points are more frequently identified as a top N solution (Figure C.5, Figure C.6). For the normal and uniform data types and when $n = 50$, points in layer 3 contribute to the top N solutions regularly (Figure C.7). However, as the sample size increases, their importance decreases. For the convex data, points in layers 3, 4, and 5 are rarely a top N solution.

4.5.4 Algorithm Runtime

Finally, using a computer with an Intel Core 2 2.40 GHz Processor and 4 GB RAM, we measured the run time of phase 1, phase 2, and creating the graphical output for the largest scenario ($n = 1000, \mathrm{N} = 5, \text{and } m = 5$) for all combinations of data types and DFs, each with 5000 generated datasets. The algorithm ran on average between 5 and 6 seconds for all cases except the convex data with the multiplicative DF (average run time = 10 seconds). With N PF layers, we are guaranteed to get all of the top N solutions with manageable runtime (Figure C.8, Figure C.9). Hence we recommend using $m = \mathrm{N}$ in practice.

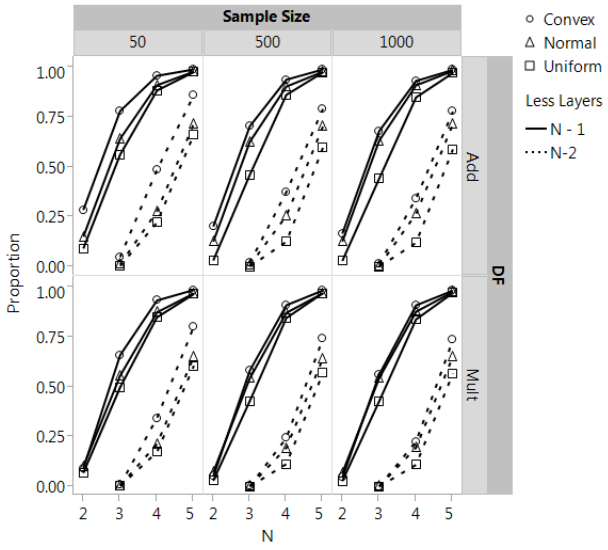

Figure 4.12 Proportion of simulations N-1, N-2 layers were enough to identify the top N solutions

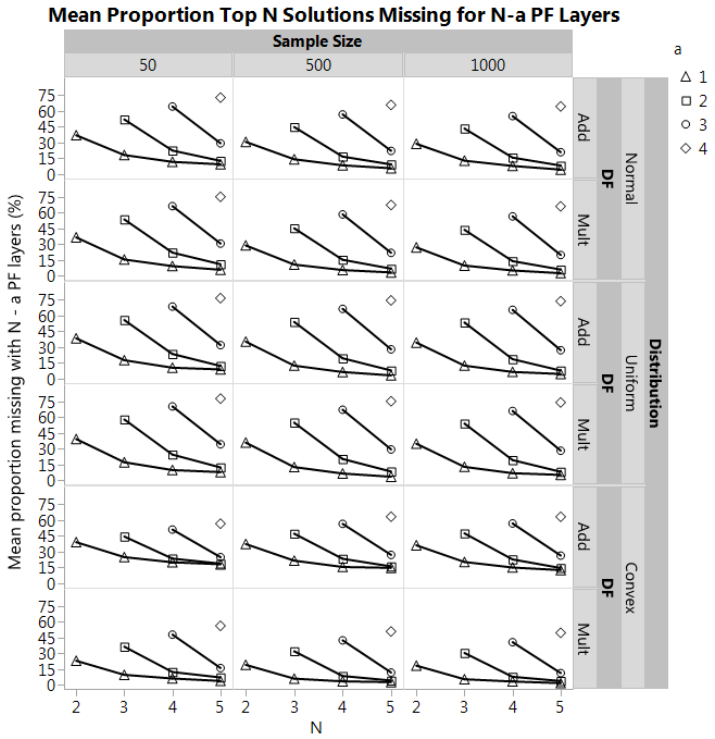

Figure 4.13 Mean Proportion the top N solutions were missed when using N-a PF layers

4.6 Conclusions

We presented an extension of a two-stage Pareto front optimization method for multiple objectives using layered Pareto fronts. Using multiple criteria when making a decision adds flexibility and realism compared to a single objective. Layered Pareto Fronts eliminate non-contenders and highlight options from which to choose when selecting the top N solutions in an enumerated list of objects. We presented several graphical tools to guide users to make their decision using their priorities. These graphical tools allow for a flexible decision and facilitate a discussion when there are differing opinions on the weighting of criteria. We also explored several properties of the

layered Pareto fronts under different conditions. Decision-makers must be thoughtful as they consider options in the TopN-PFS algorithm as the DF and scaling of the criteria values substantially impact results of the algorithm. We also showed that although for values of N = 4 and N = 5, fewer Pareto Front layers do not result in a large loss of potential solutions, the algorithm runs so quickly, we recommend using $m = \text{N}$. The TopN-PFS algorithm has been implemented in as a JMP Add-In and can be downloaded by contacting the author.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

Multi-response systems are prevalent in industrial settings; however, many designs are planned with only one type of response in mind. The choice of design for an experiment where multiple types of responses are of interest did not previously have a clear formal answer. In this work, a weighted optimality criterion was presented for dual-response systems where the distributions of the two responses are different.

In Chapter 2, the optimality criterion was introduced for a dual-response system where one response follows a normal distribution and the other response is binary. Several variations of this criterion were explored including changes in scaling the individual design criteria and the choice of DF. The multiplicative DF was chosen to emphasize the importance of balancing the design efficiency for both response models. This DF penalizes poor performance of one criterion more heavily than the additive DF. The value of the weight in the design criterion heavily affects the resulting design. It is recommended to explore the resulting designs across many weights to better understand how the designs change when the weight on one criterion changes.

As a result of the designs in Chapter 2, the issue of separation for fitting the logistic regression model was further explored in a simulation study. Designs with small runs may be more likely to have separation, so that the model parameters cannot be estimated. The simulation study showed that separation was likely to occur with the run size and specified priors in the designs from the examples in Chapter 2. However, the

simulation study also showed that separation is likely to occur for binary optimal designs, not just the weighted designs presented in this work. Furthermore, changing the priors so that the range of probabilities was not quite as extreme did not greatly decrease the chance of separation occurring. Using the distinct points of the designs constructed and then including additional runs at each location may help decrease this likelihood. More work needs to be done on constructing designs, including binary optimal designs in general, that decrease the chance of separation occurring.

In Chapter 3, we continued exploring designs for dual-response systems where the Poisson distribution was included as a potential response. Several examples were presented for different model types. The tradeoffs between the binomial and Poisson models is quite significant, indicating the need for a design that works reasonably well for both models simultaneously. In addition, the effect the specified priors have on the resulting designs was explored by changing the priors of the Poisson model. It is recommended for any response system where the priors must be specified to perform a sensitivity study to better understand the effects of the priors on the resulting designs.

Finally, Chapter 4 discussed a two-phase method to identify the top N items when evaluating each item across multiple objectives. The method utilizes layered Pareto fronts to identify not just the optimal solution, but several ordered optimal solutions. Including the additional layers of non-dominated solutions allows the decision-maker to evaluate additional solutions that may have been discarded had only the first Pareto front layer had been used. The method emphasizes the importance of the priorities of the decision-maker(s), and provides the user with a set of options to evaluate based on their own priorities. This can facilitate a discussion when multiple decision-makers are involved.

While the method was presented using an application in design of experiments, the approach is applicable in a variety of scenarios in practice.

The methodology developed in this work opens up many topics for future research. The algorithm described in Chapter 2 requires the user to supply the value of the weights in the design criterion. For the examples presented throughout Chapters 2 and 3, the changes to the design as the weight changed were shown in several plots. Incorporating a Pareto front search into the design algorithm would provide more information on the effect of the weights on the resulting designs. In addition, the Pareto front would then provide a set of reasonable designs so that the user can make an informed decision while balancing the criteria of both responses.

The distributions considered in this work were the normal, binomial, and Poisson distributions. Responses following other types of distributions could also be considered, for example, the gamma distribution. Including the gamma distribution as a potential distribution would allow a new method to model the mean and variance of a response simultaneously. Using the log link for the gamma response, the information matrix is only dependent on the coefficient of variation r (Myers et al., 2013). Other distributions in the exponential family could also be included.

The focus in this work is on dual-response systems; however, in many systems there are more than two responses of interest. A natural extension of this work is to include more than two types of responses in the design criterion. Understanding the effect the weights play on the resulting designs is critical as the number of responses increases. In addition, by including additional nonlinear models in the design criterion, the effects of the chosen priors for each model parameter likely has a substantial impact on the

design. In addition, the distribution of the priors can also affect the design; investigating the effect of a uniform, exponential, or lognormal distribution for the model parameter priors can also be explored.

Chapter 2 discussed the issue of complete and quasi-complete separation in logistic regression models. For both these weighted designs and binary optimal designs for a logistic regression model, separation can occur frequently for different priors on the model parameters and for different run sizes. While there are ways, such as Firth's method, that can rescue an analysis when separation has occurred, constructing designs where there is a smaller likelihood of separation occurring is preferred.

On a more general note, research can also be done to explore design augmentation, split plot studies, and blocking for multi-response systems.

REFERENCES

Agresti, A. *Categorical Data Analysis.* Hoboken, NJ: Wiley & Sons Inc., 2013.

Albert, A. & Anderson J.A. "On The Existence of Maximum Likelihood Estimates in Logistic Regression Models." *Biometrika* 71, 1 (1984): 1-10.

Ames, A. E., Mattucci, N., MacDonald, S., Szonyi, G., & Hawkins, D. M. "Quality Loss Functions for Optimization Across Multiple Response Surfaces." *Journal of Quality Technology* 29, 3 (1997): 339-346.

Anderson-Cook, C. M., & Lu, L. "Much Needed Structure: A New 5-Step Decision-Making Process Helps You Evaluate, Balance Competing Objectives." *Quality Progress* 48, 10 (2015): 42-50.

Carlyle, W. M., Montgomery, D. C., & Runger, G. C. "Optimization Problems and Methods in Quality Control and Improvement." *Journal of Quality Technology* 32, 1 (2000): 1-17.

Chaloner, K., & Larntz, K. "Optimal Bayesian Design Applied to Logistic Regression Experiments." *Journal of Statistical Planning and Inference* 21, 2 (1989): 191-208.

Chaloner, K., & Verdinelli, I. Bayesian Experimental Design: A Review. *Statistical Science* 10, 3 (1995): 273-304.

Chang, F., Huang, M. L., Lin, D. K. J., & Yang, H. "Optimal Designs for Dual Response Polynomial Regression Models." *Journal of Statistical Planning and Inference* 93 (2001): 309-322.

Chernoff, H. "Locally Optimal Designs for Estimating Parameters." *The Annals of Mathematical Statistics* 24, 4 (1953): 586-602.

Costa, N. R. P. "Simultaneous Optimization of Mean and Standard Deviation." *Quality Engineering* 22, 3 (2010): 140-149.

Del Castillo, E., & Montgomery, D. "A Nonlinear Programming Solution to The Dual Response Problem." *Journal of Quality Technology* 25, 3 (1993): 199-204.

Del Castillo, E., Montgomery, D. C., & McCarville, D. R. “Modified Desirability Functions for Multiple Response Optimization.” *Journal of Quality Technology* 28 (1996): 337-345.

Del Castillo, E., Shu-Kai, F., & Semple, J. “The Computation of Global Optima in Dual Response Systems.” *Journal of Quality Technology* 29, 3 (1997): 347-353.

Derringer, G., & Suich, R. “Simultaneous Optimization of Several Response Variables.” *Journal of Quality Technology* 12 (1980): 214-219.

Draper, N. R., & Hunter, W. G. “The Use of Prior Distributions in The Design of Experiments for Parameter Estimation in Non-Linear Situations: Multiresponse Case.” *Biometrika* 54, 3 (1967): 662-665.

Fedorov, V., Wu, Y., & Zhang, R. “Optimal Dose-Finding Designs with Correlated Continuous and Discrete Responses.” *Statistics in Medicine* 31, 3 (2012): 217-234.

Firth, D. “Bias Reduction of Maximum Likelihood Estimates.” *Biometrika* 80, 1 (1993): 27-38.

Gotwalt, C. M., Jones, B. A., & Steinberg, D. M. “Fast Computation of Designs Robust to Parameter Uncertainty for Nonlinear Settings.” *Technometrics* 51, 1 (2009): 88-95.

Gueorguieva, I., Aarons, L., Ogungbenro, K., Jorga, K. M., Rodgers, T., & Rowland, M. “Optimal Design for Multivariate Response Pharmacokinetic Models.” *Journal of Pharmacokinetic and Pharmacodynamics* 33, 2 (2006): 97-124.

Heinze, G. and Schemper, M. “A Solution to the Problem of Separation in Logistic Regression.” *Statistics in Medicine* 21 (2002): 2409-2419.

Heise, M. A., & Myers, R. H. “Optimal Designs for Bivariate Logistic Regression.” *Biometrics* 52, 2 (1996): 613-624.

JMP (1987-2015). (12th ed.) SAS Institute Inc.

Johnson, M. E., & Jones, B. “Classical Design Structure of Orthogonal Designs with Six to Eight Factors and Sixteen Runs.” *Quality and Reliability Engineering International* 27, 1 (2010): 61-70.

Johnson, R. T., & Montgomery, D. C. “Choice of Second-Order Response Surface Designs for Logistic and Poisson Regression Models.” *International Journal of Experimental Design and Process Optimisation* 1, 1 (2009): 2-23.

Jones, B., & Montgomery, D. C. “Alternatives to Resolution IV Screening Designs in 16 Runs”, *International Journal of Experimental Design and Process Optimisation* 1, 4 (2010): 285-295.

Khuri, A., & Conlon, M. “Simultaneous Optimization of Multiple Responses Represented by Polynomial Regression Functions.” *Technometrics* 23, 4 (1981): 363-375.

Kim, K., & Lin, D. K. “Dual Response Surface Optimization: A Fuzzy Modeling Approach.” *Journal of Quality Technology* 30, 1 (1998): 1-10.

Lin, C.F.D., Anderson-Cook, C.M., Hamada, M.S., Moore, L.M., Sitter, R.R. “Using Genetic Algorithms to Design Experiments: A Review.” *Quality and Reliability Engineering International* 31 (2015): 155-167.

Lind, E., Goldin, J., & Hickman, J. “Fitting Yield and Cost Response Surfaces.” *Chemical Engineering Progress* 56, 11 (1960): 62-68.

Lu, L., & Anderson-Cook, C. M. “Balancing Multiple Criteria Incorporating Cost Using Pareto Front Optimization for Split-Plot Designed Experiments.” *Quality and Reliability Engineering International* 30, 1 (2014): 37-55.

Lu, L. and Anderson-Cook, C.M. “Rethinking the Optimal Response Surface Design for a First-Order Model with Two-Factor Interactions, when Protecting Against Curvature.” *Quality Engineering* 24, 3 (2012): 404–421.

Lu, L., & Anderson-Cook, C. M. “Adapting the Hypervolume Quality Indicator to Quantify Trade-Offs and Search Efficiency For Multiple Criteria Decision Making Using Pareto Fronts.” *Quality and Reliability Engineering International* 29, 8 (2013) 1117-1133.

Lu, L., Anderson-Cook, C. M., & Lin, D. K. “Optimal Designed Experiments Using a Pareto Front Search for Focused Preference of Multiple Objectives.” *Computational Statistics & Data Analysis* 71 (2014): 1178-1192.

Lu, L., Anderson-Cook, C. M., & Robinson, T. J. “Optimization of Designed Experiments Based on Multiple Criteria Utilizing a Pareto Frontier.” *Technometrics* 53, 4 (2011): 353-365.

Lu, L., Chapman, J. L., Anderson-Cook, C. M. “A Case Study on Selecting a Best Allocation of New Data for Improving the Estimation Precision of System and Sub-System Reliability Using Pareto Fronts.” *Technometrics* 55, 4 (2013) pp. 473 - 487.

Meyer, R. K., & Nachtsheim, C. J. “The Coordinate-Exchange Algorithm for Constructing Exact Optimal Experimental Designs.” *Technometrics* 37, 1 (1995): 60-69.

Montgomery, D. C. *Design and Analysis of Experiments.* Hoboken, NJ: Wiley & Sons Inc., 2013.

Montgomery, D. C. and Runger, G.C. *Applied Statistics and Probability for Engineers*. Hoboken, NJ: Wiley & Sons Inc., 2003.

Myers, R. H., & Carter, W. H., Jr. “Response Surface Techniques for Dual Response Systems.” *Technometrics* 15, 2 (1973): 301-317.

Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. *Response Surface Methodology.* New Jersey: Wiley, 2016.

Myers, R. H., Montgomery, D. C., Vining, G. G., & Robinson, T. J. *Generalized Linear Models*. John Wiley & Sons. 2012.

Ortiz, F. J., Simpson, J. R., Pignatiello, J. J., & Heredia-Langer, A. “A Genetic Algorithm Approach to Multiple-Response Optimization.” *Journal of Quality Technology* 36, 4 (2004): 432-450.

Ozol-Godfrey, A., Anderson-Cook, C., Robinson, T.J. “Fraction of Design Space Plots for Generalized Linear Models.” *Journal of Statistical Planning and Inference* 138 (2008): 203-219.

Pignatiello Jr, J. J. “Strategies for Robust Multiresponse Quality Engineering.” *IIE Transactions* 25, 3 (1993): 5-15.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. *Numerical Recipes: The Art of Scientific Computing.* Cambridge, MA: Cambridge U. Press, (1986).

Satterthwaite, F. E. "Random Balance Experimentation." *Technometrics*, 1, 2 (1959): 111-137.

Tang, L., & Xu, K. "A Unified Approach for Dual Response Surface Optimization." *Journal of Quality Technology* 34, 4 (2002): 437-447.

Vining, G. "A Compromise Approach to Multiresponse Optimization." *Journal of Quality Technology* 30 (1998): 309-313.

Zahran, A.R., Anderson-Cook, C.M., & Myers, R.H. "Fraction of Design Space to Assess Prediction Capability of Response Surface Designs." *Journal of Quality Technology* 35, 4 (2003): 377-386.

APPENDIX A

ANALYSIS OF DESIGNS OF CHAPTER 2 USING ALTERNATIVE CRITERIA

A.1 Main Effects Model Only

Figures A.1, A.2, and A.3 display the designs, efficiency plot, and FDS plots for a main effects model using the optimality criterion as defined in (2.11). The priors for the logistic regression model are the same as those described in section 2.4.1.

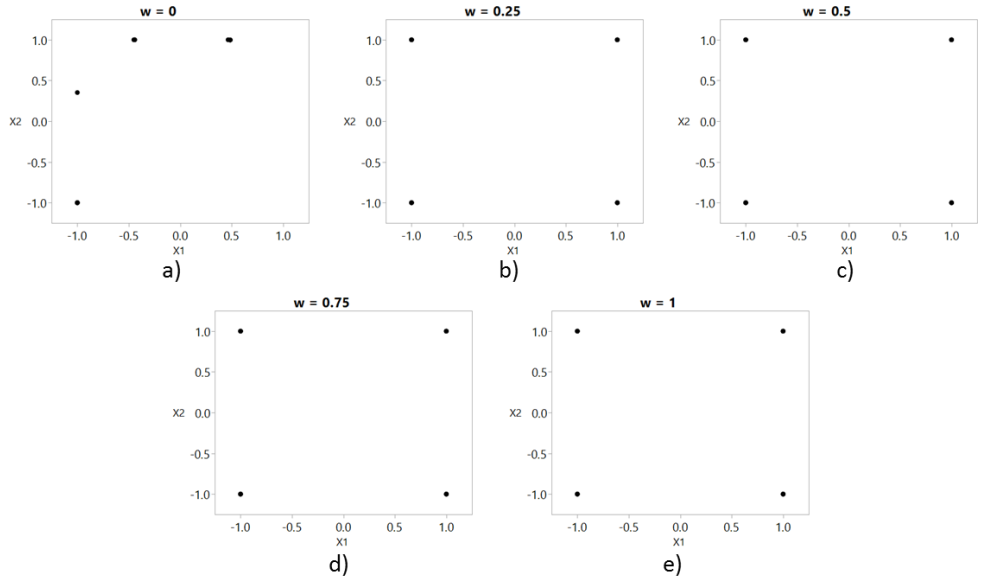

Figure A.1. Designs for main effect only models generated using criterion (2.11) with weights (a) $w_N = 0$ (binary optimal design), (b) $w_N = 0.25$, (c) $w_N = 0.50$, (d) $w_N = 0.75$, and (e) $w_N = 1$ (D-optimal design)

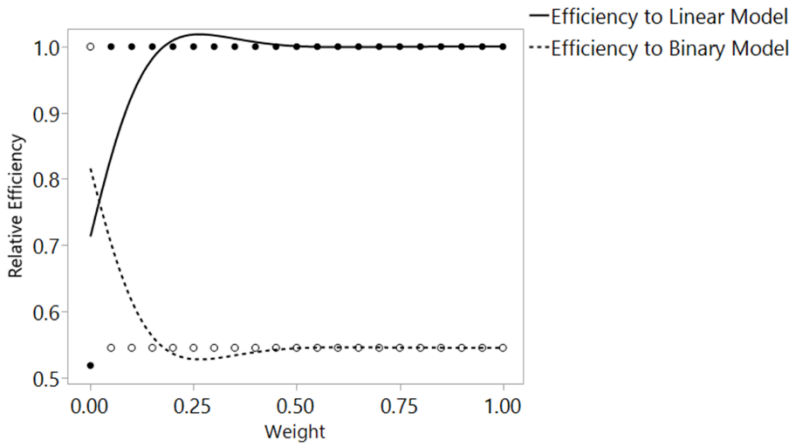

Figure A.2. Relative D-efficiency of designs generated using criterion (2.11) for the linear model (solid line) and binary model (dashed line) (main effects model)

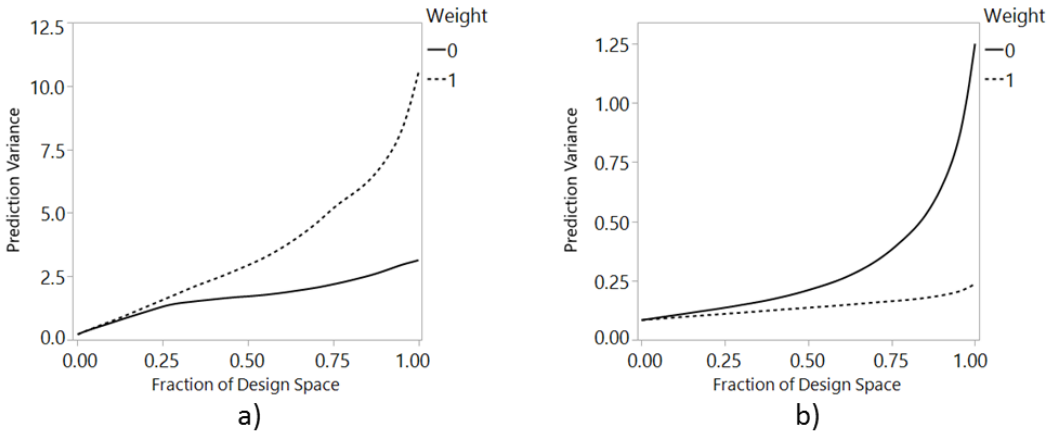

Figure A.3. FDS plots of designs generated using criterion (2.11) for main effects models for (a) the logistic regression model and (b) the linear model

Figures A.4, A.5, and A.6 display the designs, efficiency plot, and FDS plots for a main effects model using the optimality criterion as defined in (2.12). The priors for the logistic regression model are the same as those described in section 2.4.1.

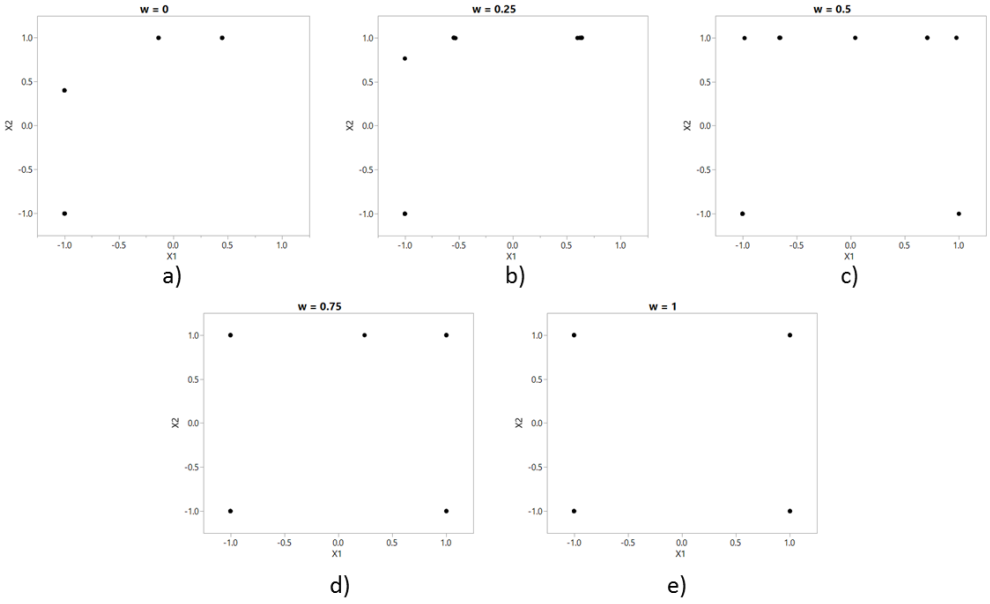

Figure A.4. Designs for main effect only models generated using criterion (2.12) with weights (a) $w_N = 0$ (binary optimal design), (b) $w_N = 0.25$, (c) $w_N = 0.50$, (d) $w_N = 0.75$, and (e) $w_N = 1$ (D-optimal design)

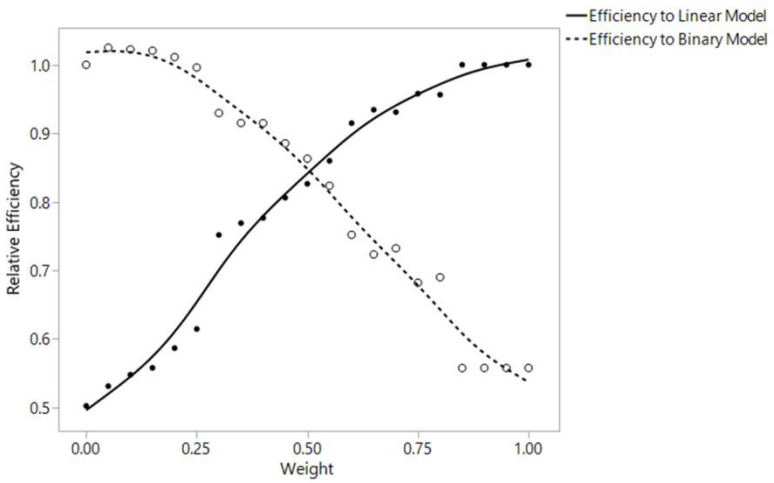

Figure A.5. Relative D-efficiency of designs generated using criterion (2.12) for the linear model (solid line) and binary model (dashed line) (main effects model)

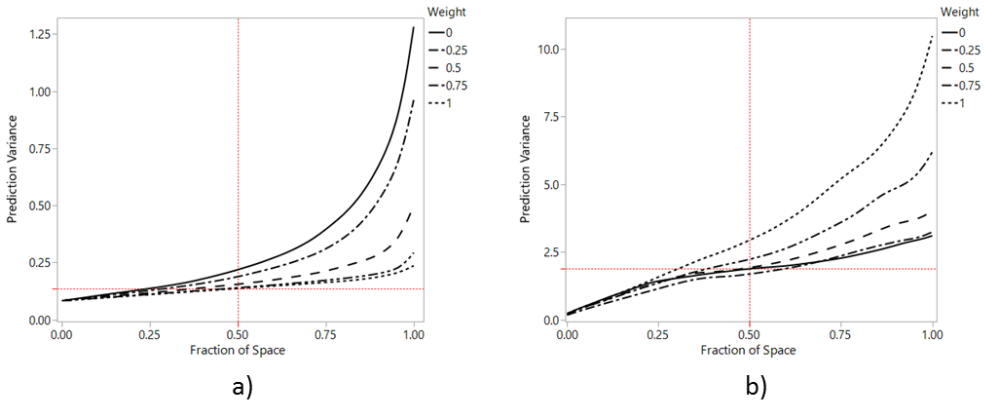

Figure A.6. FDS plots of designs generated using criterion (2.12) for main effects models for (a) the logistic regression model and (b) the linear model

A.2 Main Effects Plus 2 Factor Interaction Model

Figures A.7, A.8, and A.9 display the designs, efficiency plot, and FDS plots for a main effects plus two factor interaction model using the optimality criterion as defined in (2.11). The priors for the logistic regression model are the same as those described in section 2.4.2.

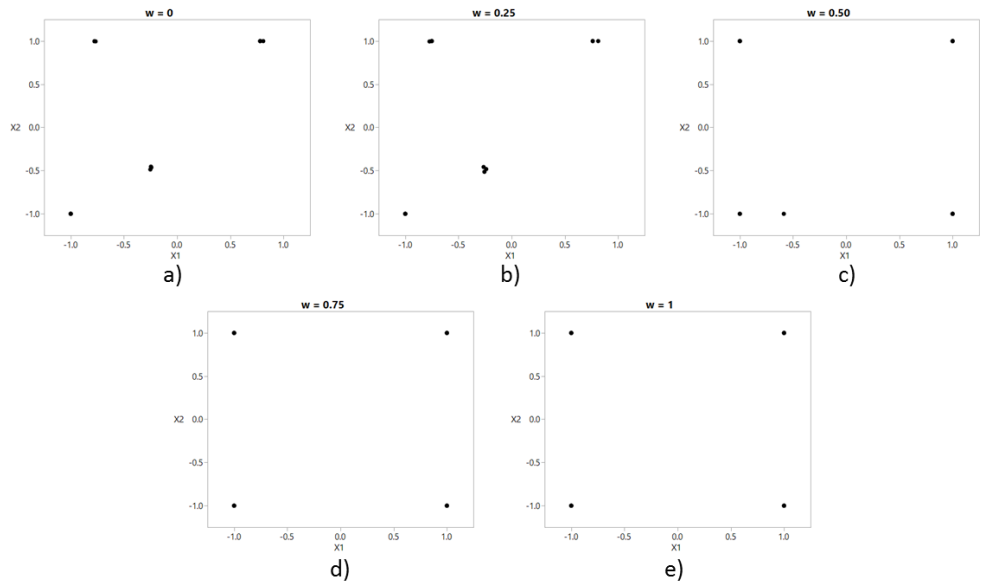

Figure A.7. Designs for ME2FI models generated using criterion (2.11) with weights (a) $w_N = 0$ (binary optimal design), (b) $w_N = 0.25$, (c) $w_N = 0.50$, (d) $w_N = 0.75$, and (e) $w_N = 1$ (D-optimal design)

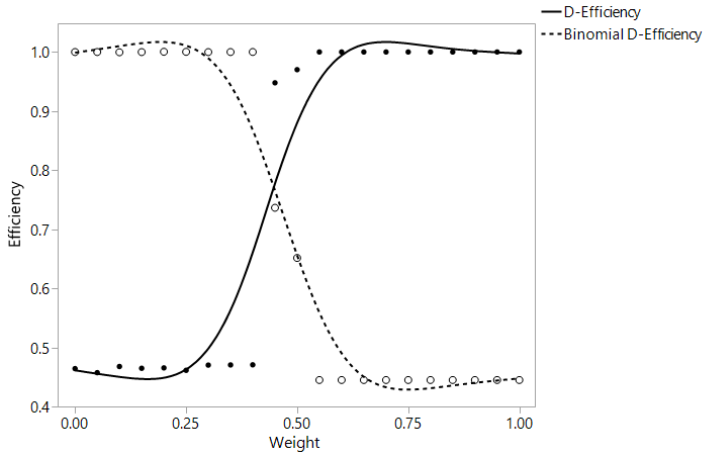

Figure A.8. Relative D-efficiency of designs generated using criterion (2.11) for the linear model (solid line) and binary model (dashed line) (ME2FI model)

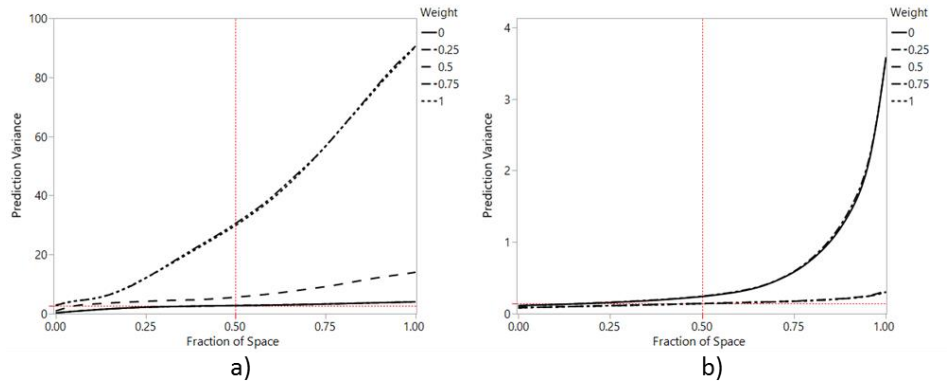

Figure A.9. FDS plots of designs generated using criterion (2.11) for ME2FI models for (a) the logistic regression model and (b) the linear model

Figures A.10, A.11, and A.12 display the designs, efficiency plot, and FDS plots for a main effects plus two factor interactions model using the optimality criterion as defined in (2.12). The priors for the logistic regression model are the same as those described in section 2.4.2.

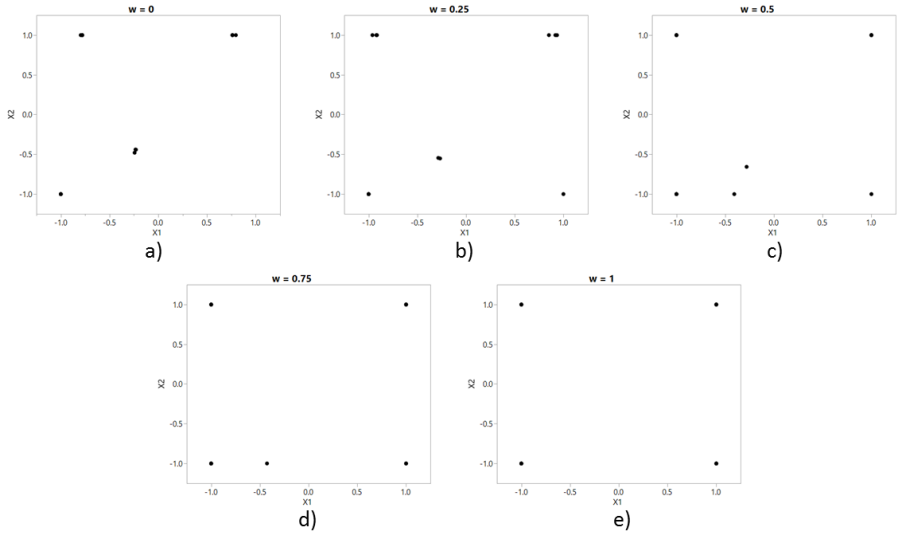

Figure A.10. Designs for ME2FI only models generated with criterion (2.12) with weights (a) $w_N = 0$ (binary optimal design), (b) $w_N = 0.25$, (c) $w_N = 0.50$, (d) $w_N = 0.75$, and (e) $w_N = 1$ (D-optimal design)

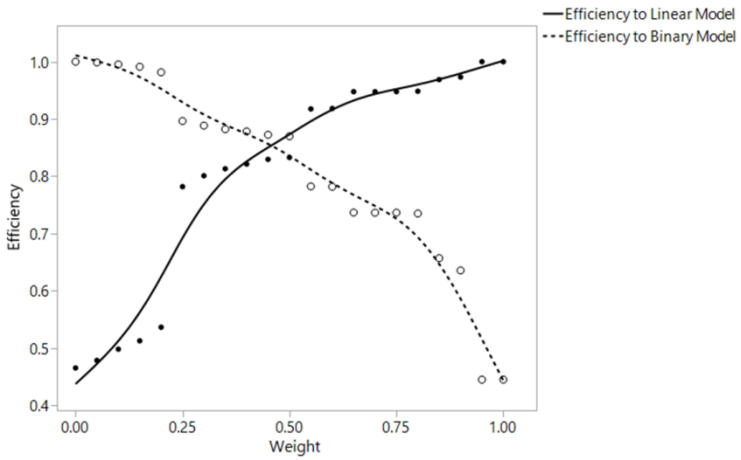

Figure A.11. Relative D-efficiency of designs generated using criterion (2.12) for the linear model (solid line) and binary model (dashed line) (ME2FI model)

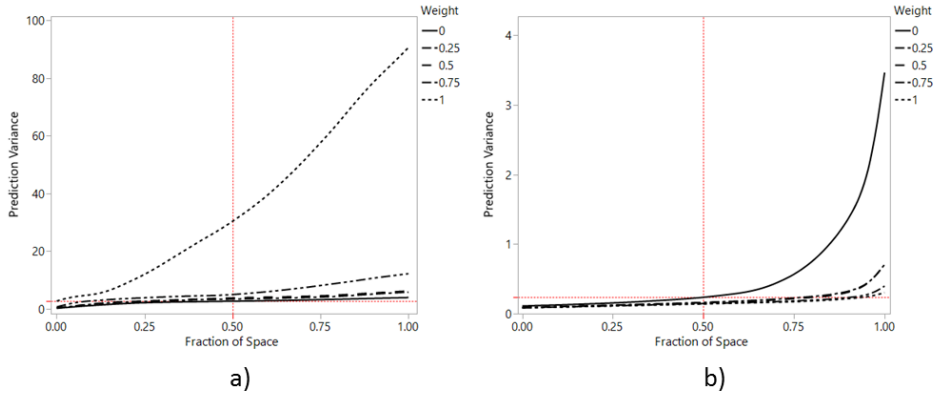

Figure A.12. FDS plots of designs generated using criterion (2.12) for ME2FI models for (a) the logistic regression model and (b) the linear model

A.3 Full Quadratic Model

Figures A.13, A.14, and A.15 display the designs, efficiency plot, and FDS plots for a full quadratic model using the optimality criterion as defined in (2.11). The priors for the logistic regression model are the same as those described in section 2.4.3.

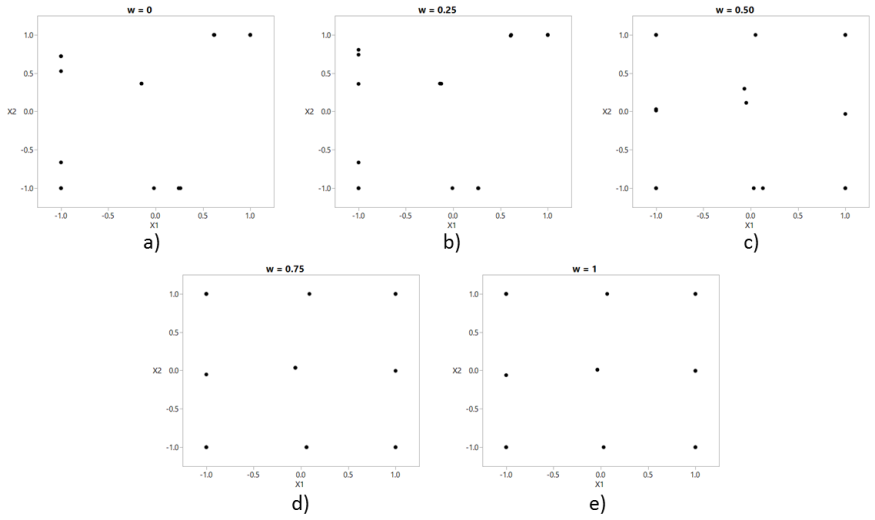

Figure A.13. Designs for full quadratic models generated using criterion (2.11) with weights (a) $w_N = 0$ (binary optimal design), (b) $w_N = 0.25$, (c) $w_N = 0.50$, (d) $w_N = 0.75$, and (e) $w_N = 1$ (D-optimal design)

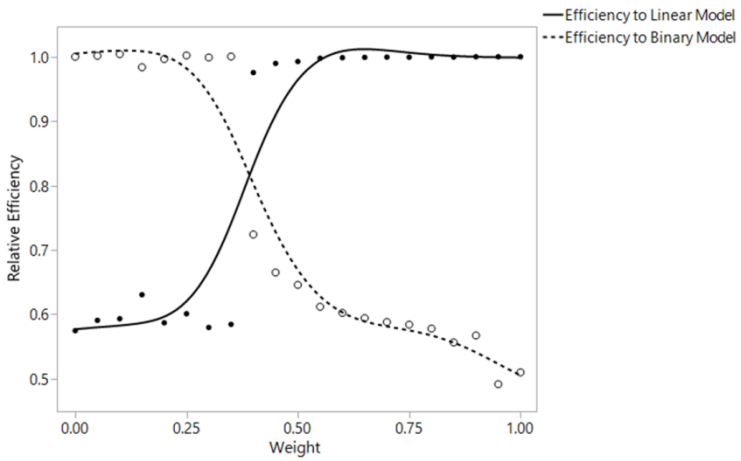

Figure A.14. Relative D-efficiency of designs generated using criterion (2.11) for the linear model (solid line) and binary model (dashed line) (quadratic model)

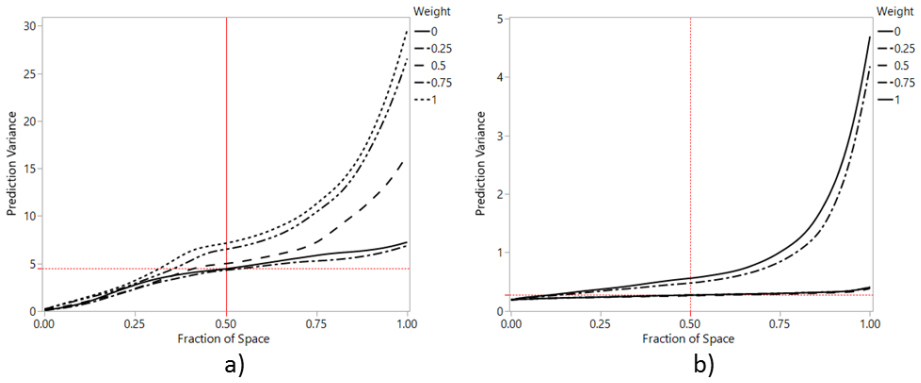

Figure A.15. FDS plots of designs generated using criterion (2.11) for quadratic models for (a) the logistic regression model and (b) the linear model

Figures A.16, A.17, and A.18 display the designs, efficiency plot, and FDS plots for a full quadratic model using the optimality criterion as defined in (2.12). The priors for the logistic regression model are the same as those described in section 2.4.3.

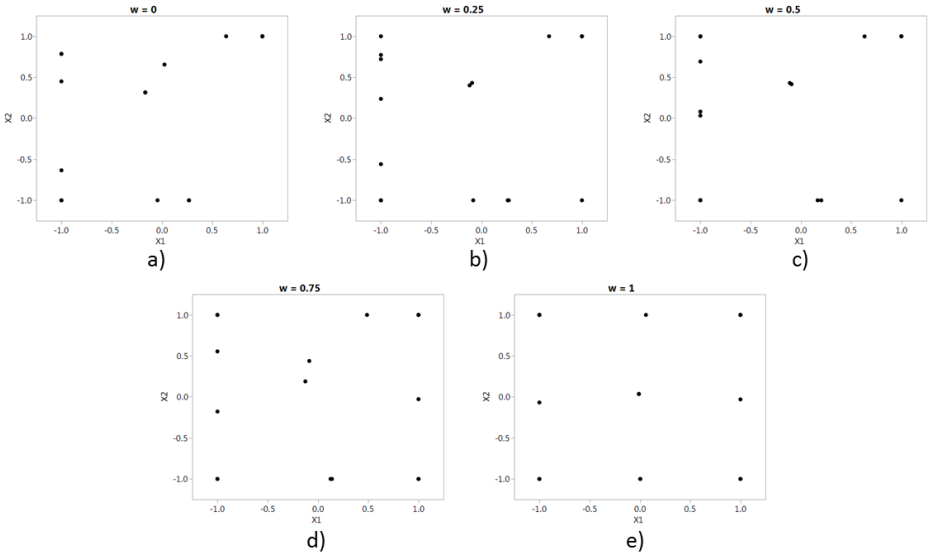

Figure A.16. Designs for full quadratic models generated with criterion (2.12) with weights (a) $w_N = 0$ (binary optimal design), (b) $w_N = 0.25$, (c) $w_N = 0.50$, (d) $w_N = 0.75$, and (e) $w_N = 1$ (D-optimal design)

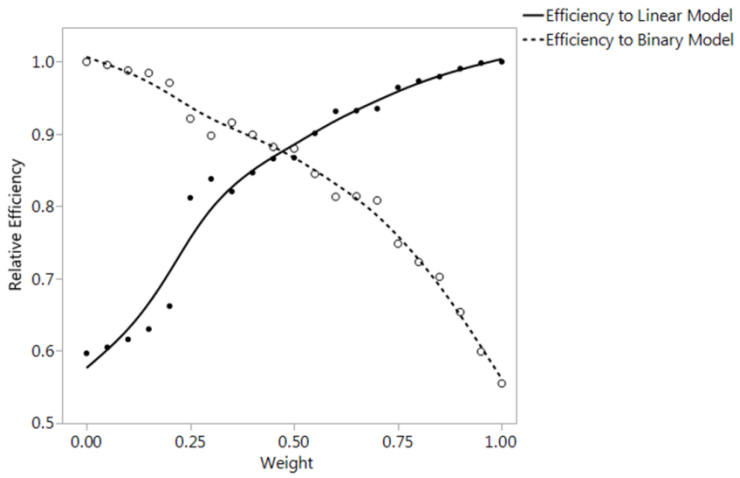

Figure A.17. Relative D-efficiency of designs generated using criterion (2.12) for the linear model (solid line) and binary model (dashed line) (quadratic model)

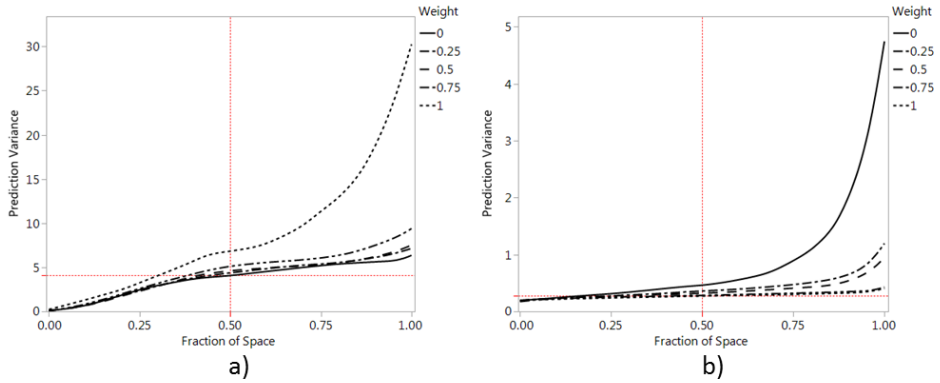

Figure A.18. FDS plots of designs generated using criterion (2.12) for quadratic models for (a) the logistic regression model and (b) the linear model

APPENDIX B

DESIGNS AND EFFICIENCY PLOTS OF DESIGNS IN SECTION 3.4

B.1. Results of Sensitivity (3.4.1)

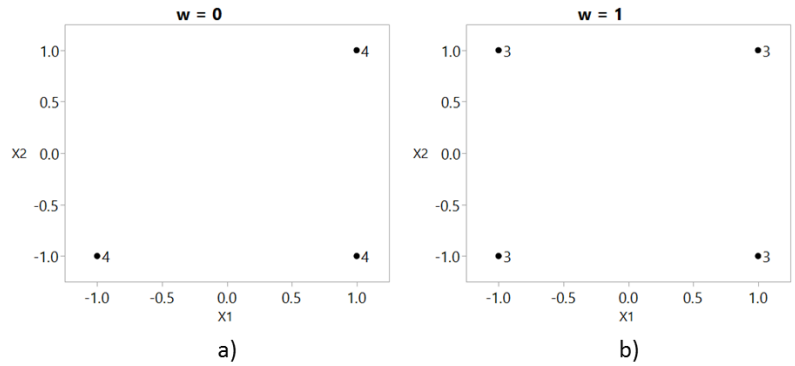

Figure B.1. Designs for main effect only models with weights (a) $w_N = 0$ (Poisson optimal design), (b) $w_N = 1$ (D-optimal design) for Case 3 priors of Poisson model

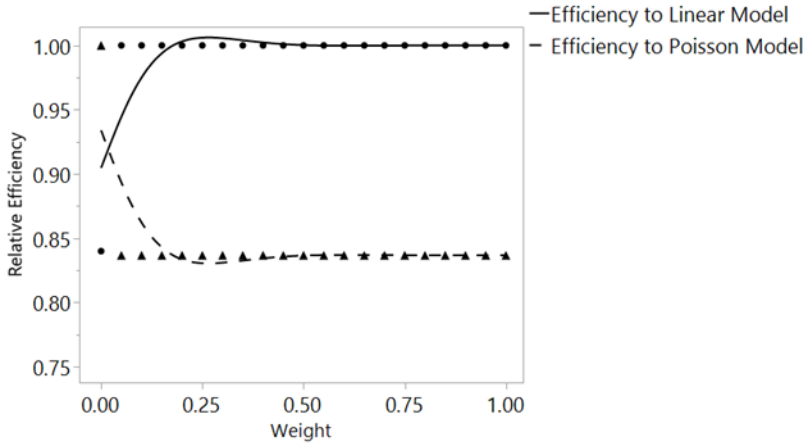

Figure B.2. Efficiency plot for normal, Poisson dual-response system with Case 3 priors for nonlinear model (main effects model)

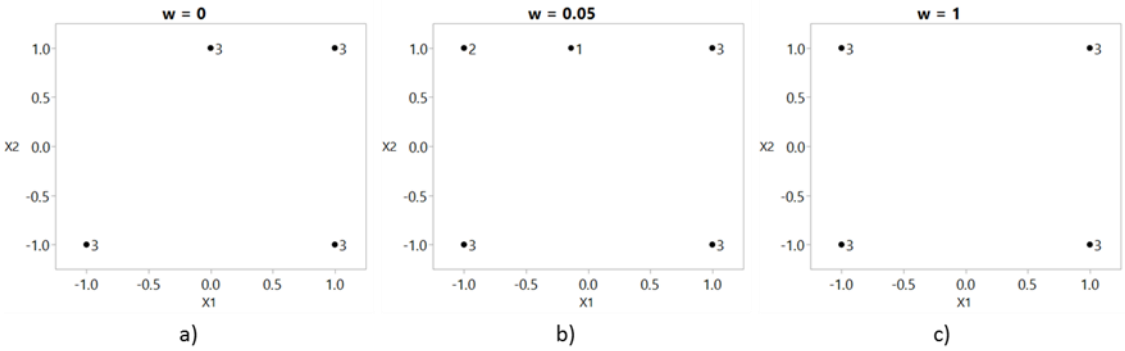

Figure B.3. Designs for main effect plus two-factor interaction model with weights (a) $w_N = 0$ (Poisson optimal design), (b) $w_N = 0.05$, and (c) $w_N = 1$ (D-optimal design) for Case 3 priors of Poisson model

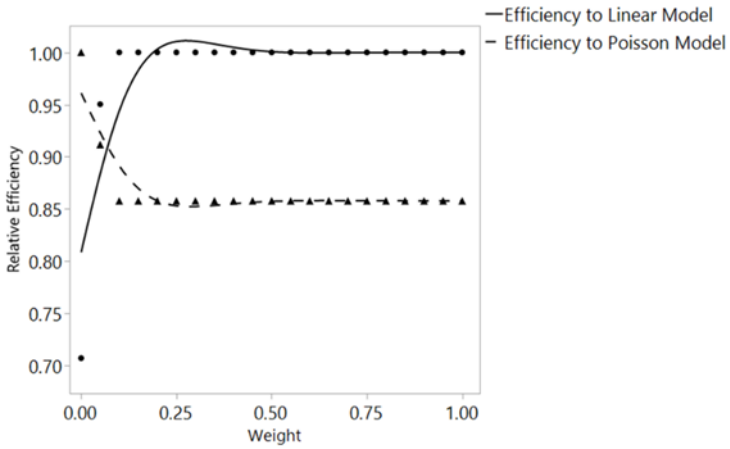

Figure B.4. Efficiency plot for normal, Poisson dual-response system with Case 3 priors for nonlinear model (main effects plus two-factor interaction model)

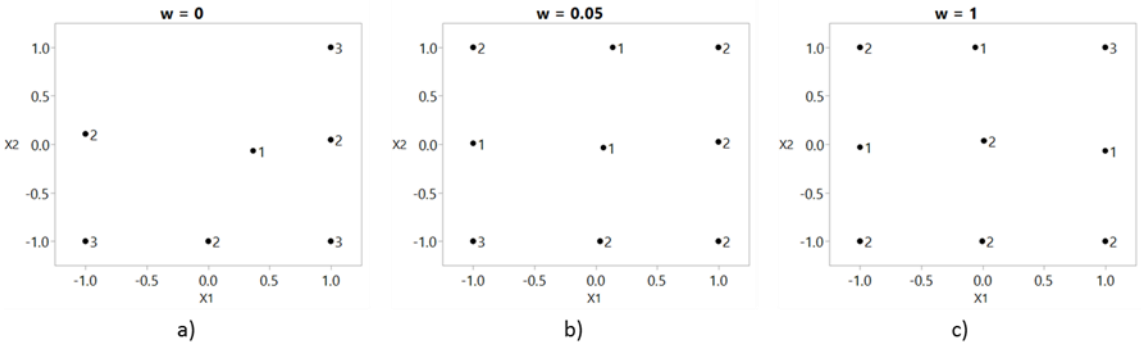

Figure B.5. Designs for quadratic model with weights (a) $w_N = 0$ (Poisson optimal design), (b) $w_N = 0.05$, and (c) $w_N = 1$ (D-optimal design) for Case 3 priors of Poisson model

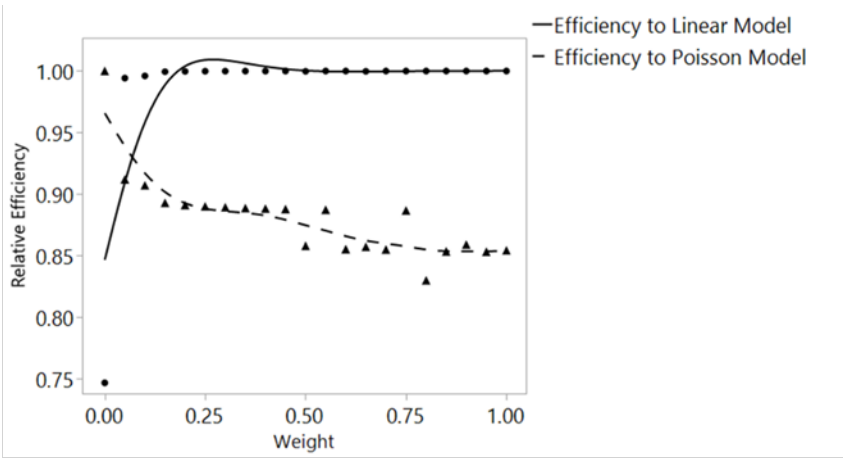

Figure B.6. Efficiency plot for normal, Poisson dual-response system with Case 3 priors for nonlinear model (quadratic model)

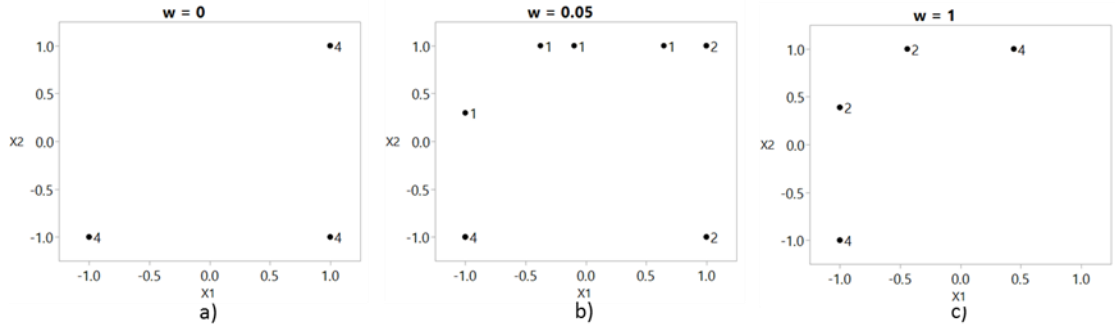

Figure B.7. Designs for main effects model with weights (a) $w_B = 0$ (Poisson optimal design), (b) $w_B = 0.05$, and (c) $w_B = 1$ (binary optimal design) for Case 3 priors of Poisson model

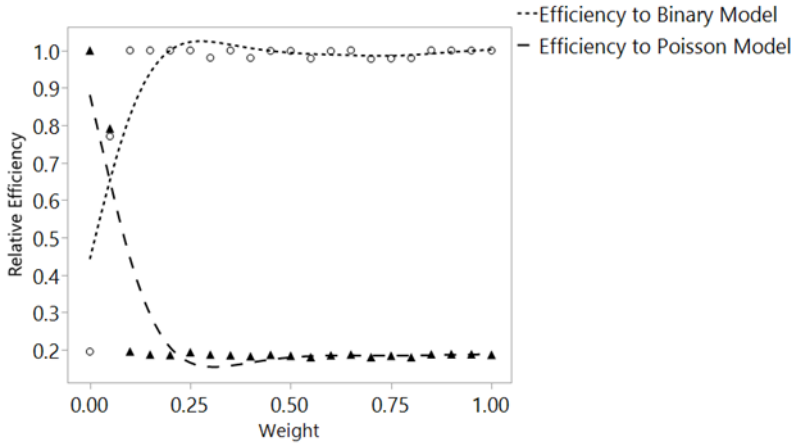

Figure B.8. Efficiency plot for binomial, Poisson dual-response system with Case 3 priors for nonlinear model (main effects model)

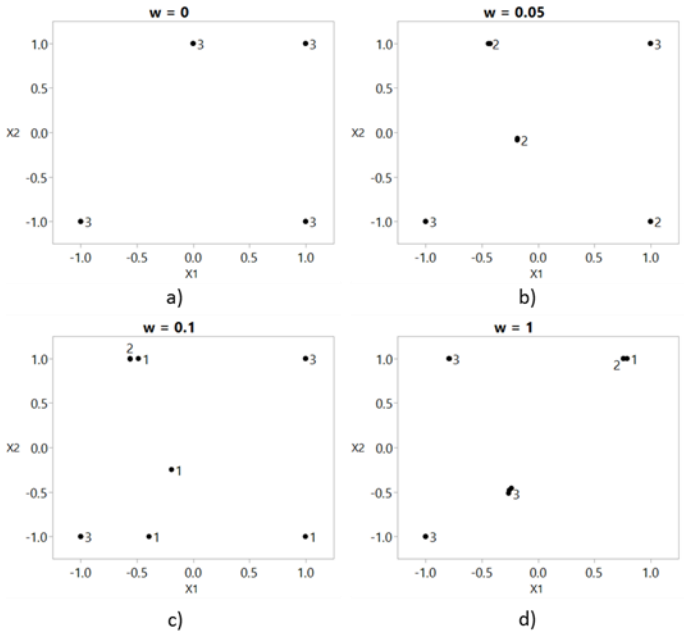

Figure B.9. Designs for main effects plus two-factor interaction model with weights (a) $w_B = 0$ (Poisson optimal design), (b) $w_B = 0.05$, (c) $w_B = 0.1$, and (d) $w_B = 1$ (binary optimal design) for Case 3 priors of Poisson model

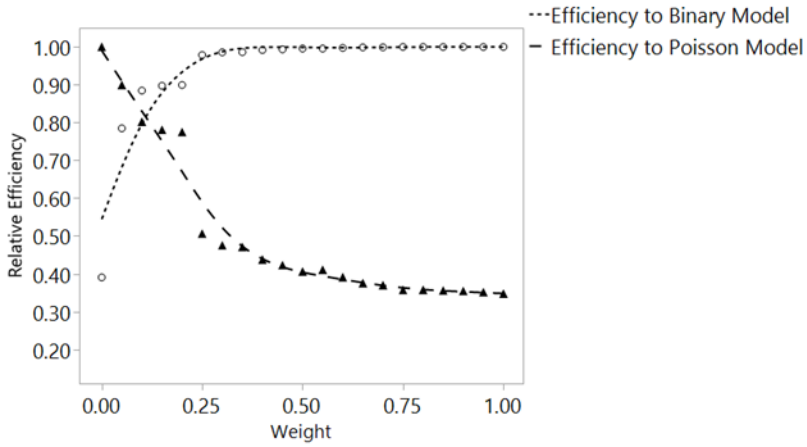

Figure B.10. Efficiency plot for binomial, Poisson dual-response system with Case 3 priors for nonlinear model (main effects plus two-factor interaction model)

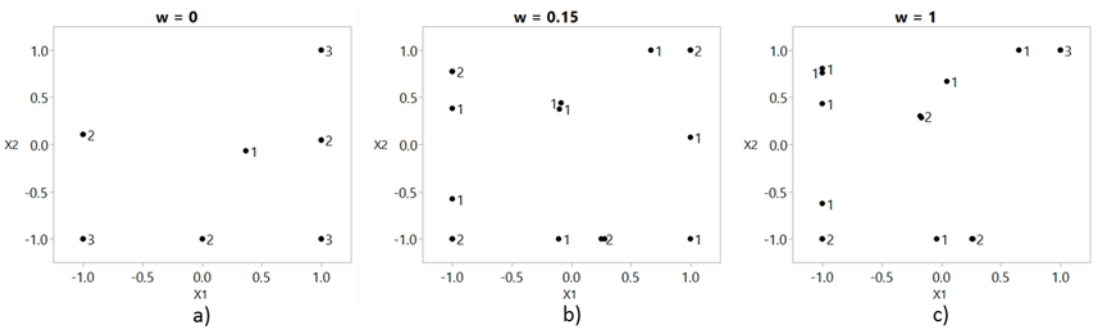

Figure B.11. Designs for quadratic model with weights (a) $w_B = 0$ (Poisson optimal design), (b) $w_B = 0.15$, and (c) $w_B = 1$ (binary optimal design) for Case 3 priors of Poisson model

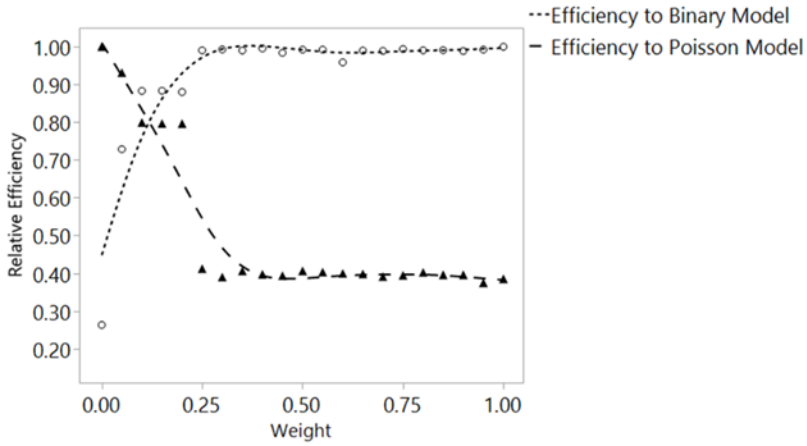

Figure B.12. Efficiency plot for binomial, Poisson dual-response system with Case 3 priors for nonlinear model (quadratic model)

B.2. Results of Sensitivity (3.4.2)

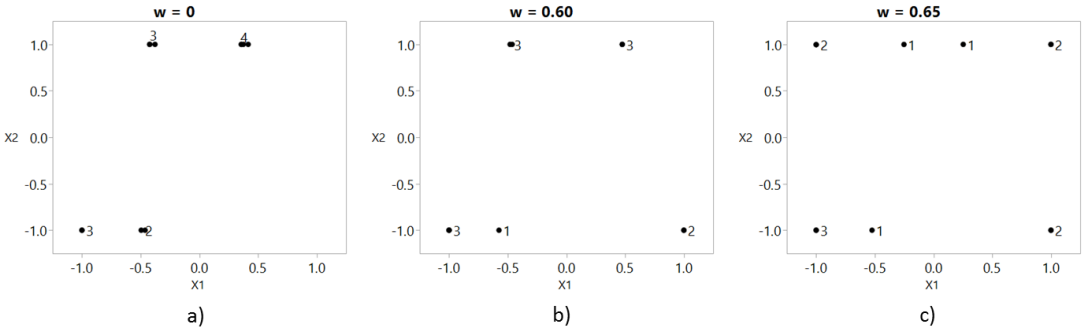

Figure B.13. Selected designs for run 1 of misspecified priors for normal-binomial dual-response system. (a) $w_N = 0$, (b) $w_N = 0.60$, and (c) $w_N = 0.65$

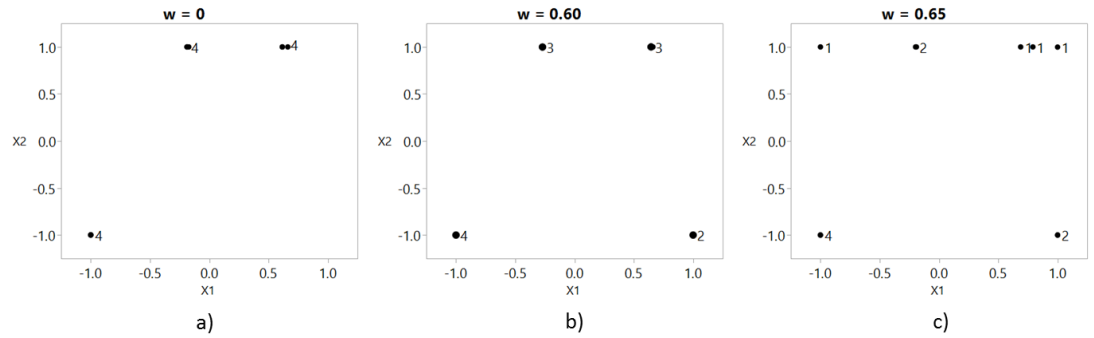

Figure B.14. Selected designs for run 2 of misspecified priors for normal-binomial dual-response system. (a) $w_N = 0$, (b) $w_N = 0.60$, and (c) $w_N = 0.65$

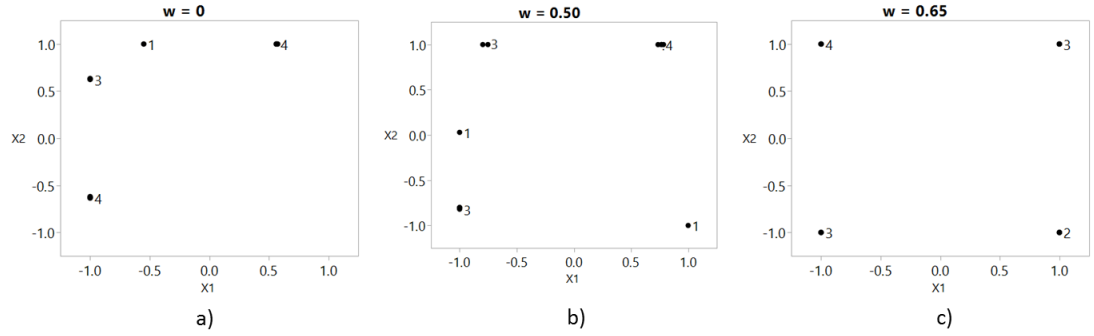

Figure B.15. Selected designs for run 3 of misspecified priors for normal-binomial dual-response system. (a) $w_N = 0$, (b) $w_N = 0.50$, and (c) $w_N = 0.65$

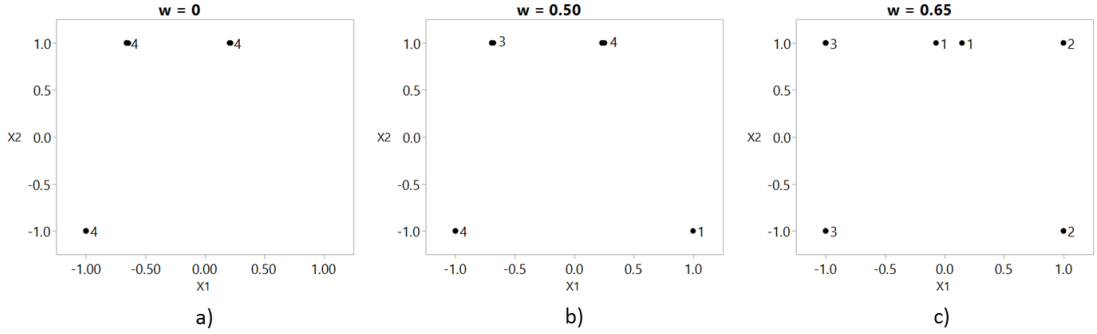

Figure B.16. Selected designs for run 4 of misspecified priors for normal-binomial dual-response system. (a) $w_N = 0$, (b) $w_N = 0.50$, and (c) $w_N = 0.65$

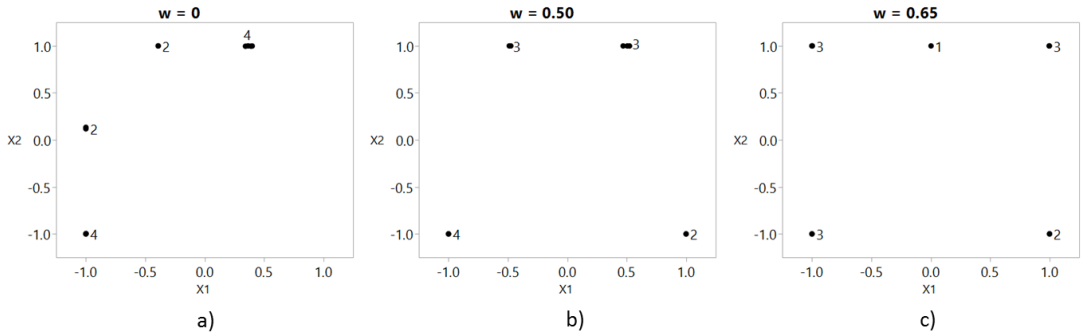

Figure B.17. Selected designs for run 4 of misspecified priors for normal-binomial dual-response system. (a) $w_N = 0$, (b) $w_N = 0.50$, and (c) $w_N = 0.65$

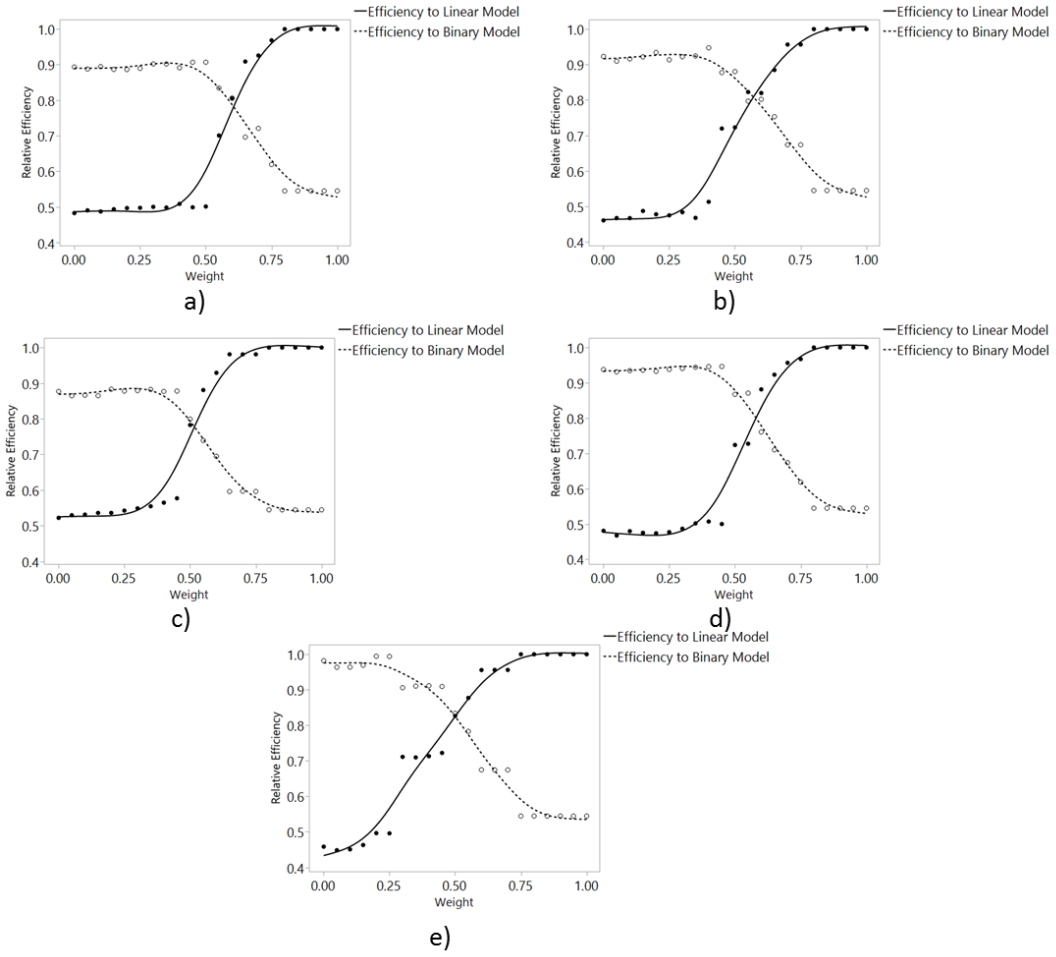

Figure B.18. Efficiency plots of designs with misspecified priors from section 2.4.2. (a) run 1, (b) run 2, (c) run 3, (d) run 4, and (e) run 5.

APPENDIX C

TOPN-PFS SIMULATION STUDY FULL RESULTS

We present a detailed analysis of the results from the simulation study described in Section 4.5. The simulation study focused on three questions: 1) What are the general properties of the Pareto Front (PF) layers; 2) How many layers are necessary to determine the top N solutions; and 3) How do the top N solutions change for different degrees of tradeoff between criteria and DFs. We randomly generated values for two criteria under three different scenarios: "normal, "uniform," and "convex." We considered sample sizes $n = 50, 500, \text{and } 1000$ and used the additive or multiplicative DF, leading to 18 scenarios.

C.1 Layered Pareto Front Characteristics

These scenarios allow us to consider the effect of the shape and size of the dataset on the properties of the PF layers. Note that the DF has no effect on the composition of the PF layers as forming these layers does not depend on the DF. Figure B.1 summarizes the number of points in each layer by the data type and sample size for this simulation study. For sample sizes $n = 500$ and $n = 1000$, the number of points in each layer typically increases for higher PF layers regardless of the data type. However, the number of points in each layer for the convex case is larger than those for the normal and uniform cases. In the convex scenario, since there are more tradeoffs between the two criteria, leading to fewer dominated points eliminated, we expect the PF layers to be richer for this correlated data structure. There is also more variability in the number of points for these larger sample sizes across all three data types. For $n = 50$, the number of points on each layer was similar for each of the data types; however, in the convex scenario, the number of points in each layer decreases slightly for the higher PF layers. This result is most likely an artifact of the richer PF layers for the convex scenario and the relatively low sample size. Regardless of the sample size, the number of points in each PF layer is similar for the normal and uniform scenarios. This finding is more surprising as we had expected the normal case to have more points per layer in general since the uniformly distributed data is more spread more evenly over the criterion region.

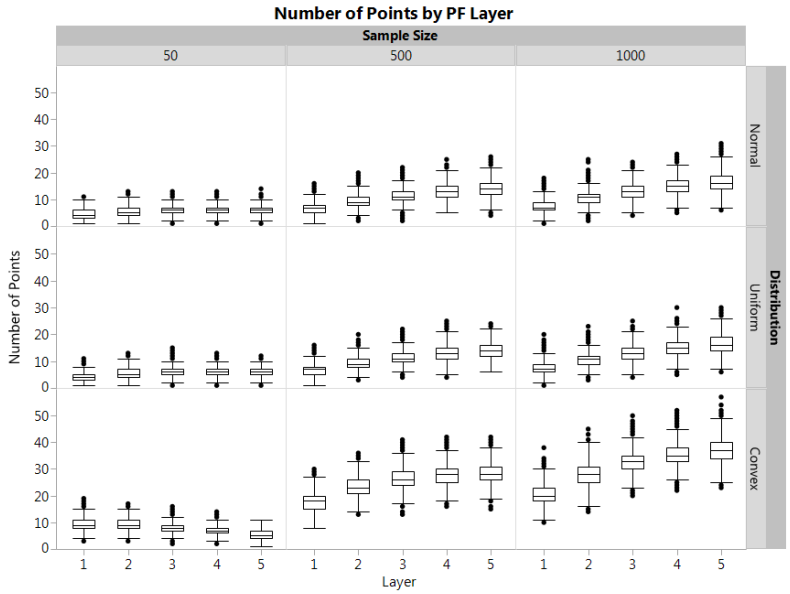

Figure C.1. Summary of the number of points in each PF layer by data type and sample size

C.2 Proportion Unique Top N Solutions

We were also interested in characterizing the number of unique solutions (the total number of solutions that were ever a top N solution for each scenario), across all possible weights of the two criteria. We consider the effect of the DF on these results as the choice of DF is critical in the final solution. We determined an overall proportion of unique solutions for N = 5 by finding the number of unique potential solutions for a dataset and dividing by the number of points in the five PF layers of that dataset. This proportion is summarized in Figure C.2 for each of the 18 scenarios. For $n = 50$, there is more variability in the proportion of unique solutions, where values typically range from 20% to 80%. The convex data has a smaller proportion of unique solutions across the sample sizes and DFs compared to the normal and uniform cases. While the PF layers tend to be richer for the convex data, only a small proportion of these points are ever potential solutions. As the sample size increases, a smaller proportion of points in the PF layers are a top 5 solution, and again we see very little difference between the normal and uniform data. We can interpret this result as the number of unique solutions rising more slowly than the number of points on the PF as sample size increases.

There is a striking difference between the additive and multiplicative DFs as the proportion of unique solutions for the multiplicative DF tends to be higher across all sample sizes or data types, likely because it is less forgiving of poor criterion values. This difference is particularly strong for the convex data. These results highlight the importance of selecting the DF to match the priorities of the study as the final solution may depend greatly on the DF choice.

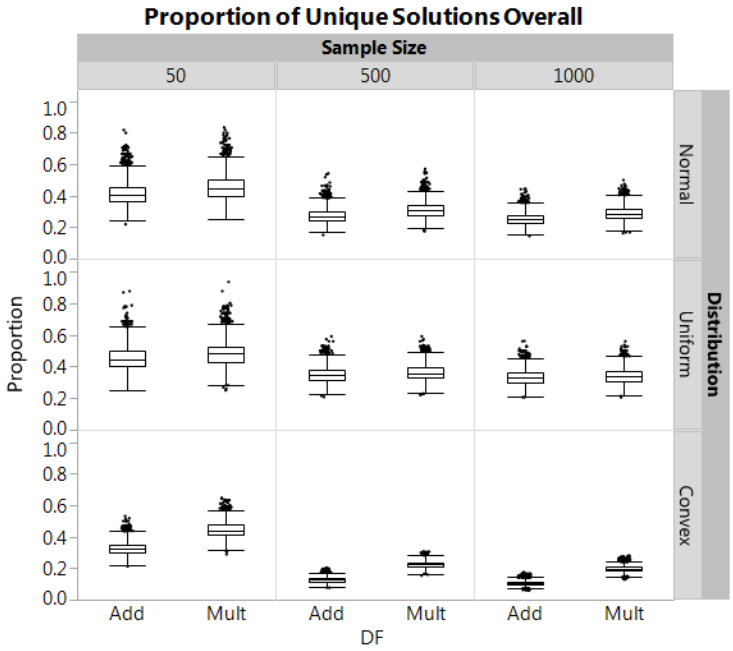

Figure C.2. Proportion of unique solutions for N = 5 and $m = 5$

C.3 Frequency Layer m Needed

One question of interest related to computational efficiency for populating the PFs was how often were $m = 1, \ldots, \mathrm{N} - 1$ layers sufficient to identify the top N solutions across all possible weightings of the criteria. We identified the number of times an optimal level N solution (N = 1, …, 5) came from PF layer $m = 1, \ldots, \mathrm{N}$ for each simulated dataset. We then determined the proportion of the 5000 simulation iterations where m layers were needed to obtain all of the top N solutions. Figure C.3 shows the proportion of times N-1 and N-2 layers were sufficient to identify all top N solutions for N = 2, …, 5. For $\mathrm{N} = 2$ for example, one layer was sufficient to identify the top 2 solutions at most 29% of the time for convex data, $n = 50$, and the additive DF. However, using one PF layer in all the other scenarios typically identified the top 2 solutions 15% or less of the time. The convex-shaped data typically identified all top N solutions with fewer PF layers compared to the other two data types. There was more variability in the number of layers required between the three data types for the additive DF. When using the multiplicative DF, more PF layers were needed more often to obtain all top N solutions. This pattern was consistent across all sample sizes.

As N increases, N-1 PF layers capture all top N solutions more often. For example, 4 PF layers identified all top 5 solutions in over 97% of the simulated datasets in each of the 18 scenarios. Three PF layers identified all top 4 solutions in over 84% of the simulations in each of the scenarios. As Figure C.3 shows, however, using N-1 layers for N = 2 and N = 3 failed to identify all top N solutions frequently. For N = 4 and N = 5,

using N-1 or N-2 layers typically does not result in substantial loss in potential solutions. Fewer layers than that would lead to compromised results.

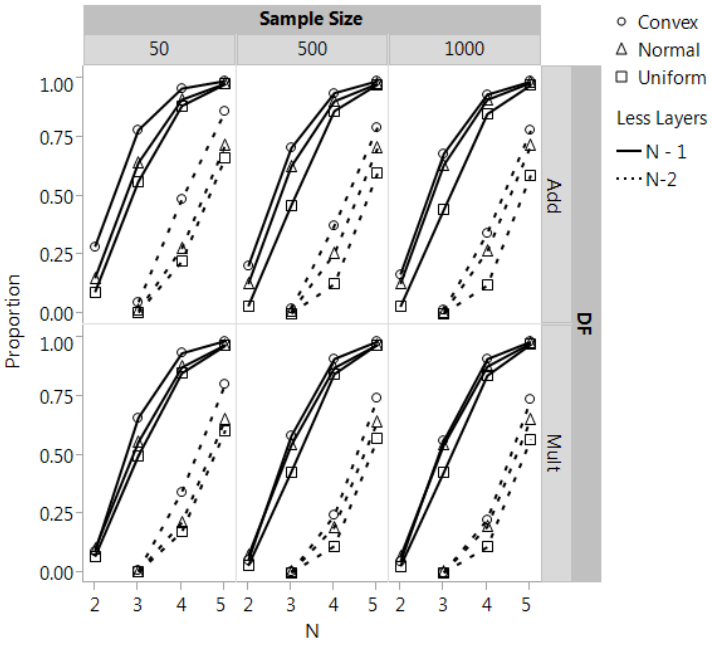

Figure C.3. Proportion of simulations N-1, N-2 layers were enough to identify the top N solutions

C.4 Solutions missed when using fewer layers

We also considered what percent of top N solutions were missed across all weight combinations of the criteria if $m <$ N PF layers were not enough to identify all top N solutions. Specifically, if m PF layers were not enough to identify all top N solutions, what percent of top N solutions were missed across all weight combinations on the criteria?

Figure C.4 displays the mean proportion of missed solutions for each of the 18 scenarios for N = 2, …,5 when using N $- a$ PF layers ($a = 1, 2, 3, 4$). For example, if we want the top 5 solutions, across all the possible weight combinations, how often do we miss solutions when using 1, 2, 3, or 4 PF layers? If we were to use $m <$ N PF layers, we would like to see a small proportion of missed solutions. Figure B.4 reinforces the previous conclusion about the large amount of information lost when using fewer PF layers when N = 2 or N = 3. Using one layer when N = 3 with an additive DF, for example, results in 43-55% of the solutions across all weight combinations missed for the normal and uniform data.

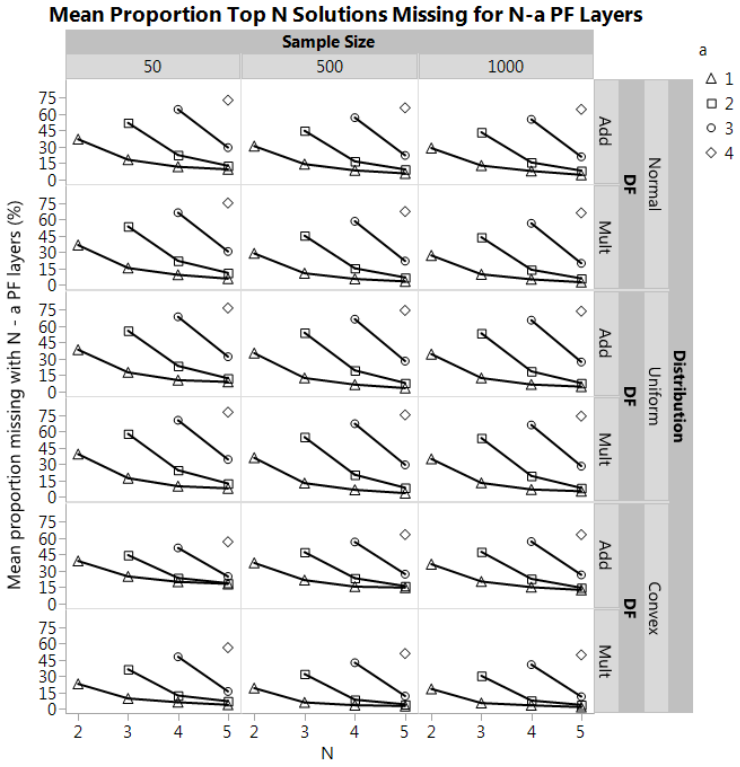

Figure C.4. Mean proportion the top N solutions were missed when using N-a PF layers

We previously saw a small loss in information when using fewer PF layers for N = 4 and N = 5. As Figure C.4 shows, using only one layer for N = 4 or one or two layers for N = 5 leads to too many potential solutions missed, regardless of the data type, sample size, or DF. There was very little difference between the proportion of missed solutions when using three or four PF layers for N = 5 across all 18 scenarios. For the convex data and multiplicative DF, there is minimal penalty of using N-1 or N-2 layers; however, there is a bigger loss for this data when using the additive DF. For the normal and uniform cases, the behavior was similar for all sample sizes and the two DFs. Using 3 layers for N = 4 resulted in 5-10% missed solutions; using 2 layers resulted in 12 – 22% missed solutions. Using 4 layers for N = 5 resulted in 2-5% missed solutions; using 3 layers resulted in 6 – 13% missed solutions. Therefore, N-1 PF layers for N = 4 and N = 5 may not result in too substantial a loss of potential solutions. However, fewer layers than that would not be appropriate.

C.5 Proportion layer m points contribute to Top N Solutions

To further understand the relationship between the number of PF layers and the top N solutions, we also consider how often a point in each PF layer is a top N choice. The goal was to understand how often the outer layers contribute to the top N solutions.

We may have a large number of points in layer 3, for example, that are never a top 5 solution no matter how we weight the criteria.

We consider each of the top 3 layers separately, and examine where solutions from each layer contribute to the top N. Figure C.5 shows the proportion of the time that a point from layer 1 was included as a top 5 solution for all 18 scenarios. Here a value of 0 denotes points that were never a top 5 solution. We see that most points in the convex data scenario are never a top N solution. This result is not too surprising as this type of data leads to a rich layered PF. There are many more points competing for a position in the top 5 compared to the normal and uniform cases. There is not a large difference between the results for the two DFs for either data type. The behavior for the normal and uniform data types is also similar.

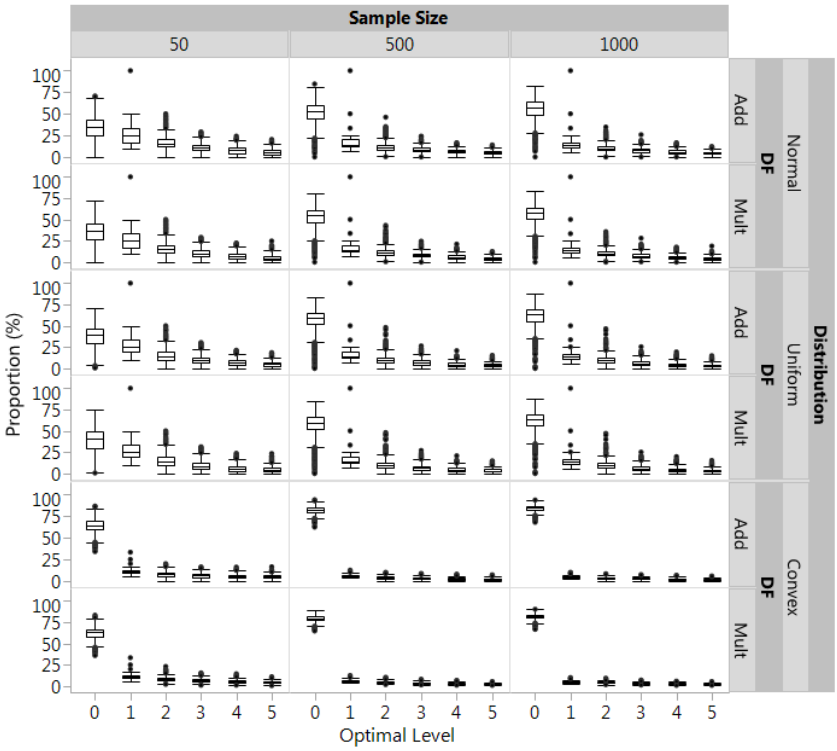

Figure C.5. Proportion Layer 1 points are a top N solution

Figure C.6 shows the proportion of time that points from layer 2 in the PF were a top 5 solution. Note that by the properties of the layered PF, a point in layer 2 can never be the best solution, as indicated by the proportion being necessarily equal to zero for all data sets in the figure. Similar behavior occurs for points in layer 2; layer 2 points are rarely a top 5 solution in the convex data scenario. We see how sample size plays a more important role here. For smaller sample sizes, the layer 2 points are more frequently identified as a top N solution. Similar behavior occurs for the layer 3 points (Figure C.7). For the normal and uniform data types and when $n = 50$, points in layer 3 contribute to the top N solutions regularly. However, as the sample size increases, their importance decreases. For the convex data, points in layers 3, 4, and 5 are rarely a top N solution.

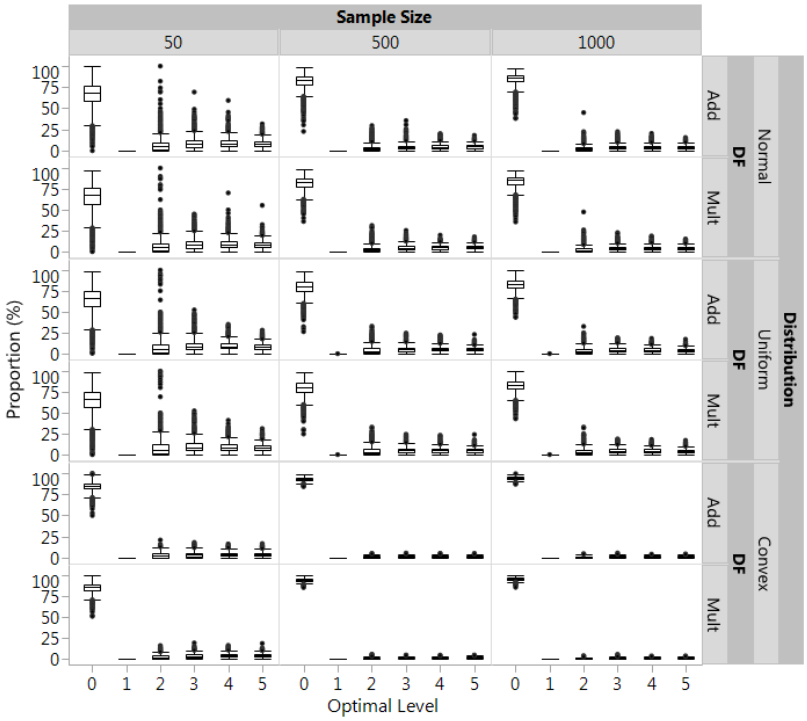

Figure C.6. Proportion layer 2 points are a top N solution

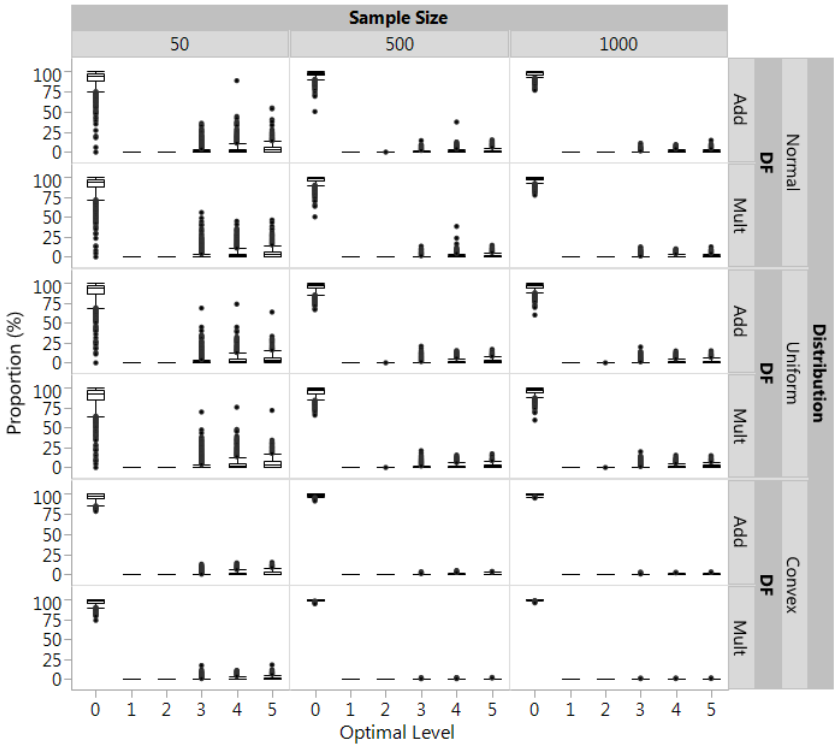

Figure C.7. Proportion Layer 3 points are a Top N solution

C.6 Time Simulation

The results from this simulation indicated that for $N \leq 3$, at least three PF layers are required to ensure there is no substantial loss in the top N solutions. However, for N = 4 and N = 5, the results indicated that while there would be some loss of information using fewer layers, this loss occurred infrequently and was not too severe. Using a computer with an Intel Core 2 2.40 GHz Processor and 4 GB RAM, we measured the run time of phase 1 (the objective phase where potential rational solutions are identified), phase 2 (the subjective stage where the remaining solutions are evaluated for different prioritizations of the criteria), and creating the graphical output for $n = 1000$, N = 5, and $m = 5$, the biggest scenario in the previous simulation. We compared the results for the three different data types and two DFs. For each of these 6 scenarios, 5000 datasets were generated with Figure C.8 showing the overall runtimes of the algorithm. The algorithm ran on average between 5 and 6 seconds for all cases except the convex data with the multiplicative DF (average run time = 10 seconds). Figure C.9 shows the contributions to the overall time for each of the three computational components examined in this simulation. The plots show a large portion of the runtime is due to the graphical output of the algorithm. Since N PF layers guarantees getting all of the top N solutions with a manageable runtime, we therefore recommend using $m =$ N in practice.

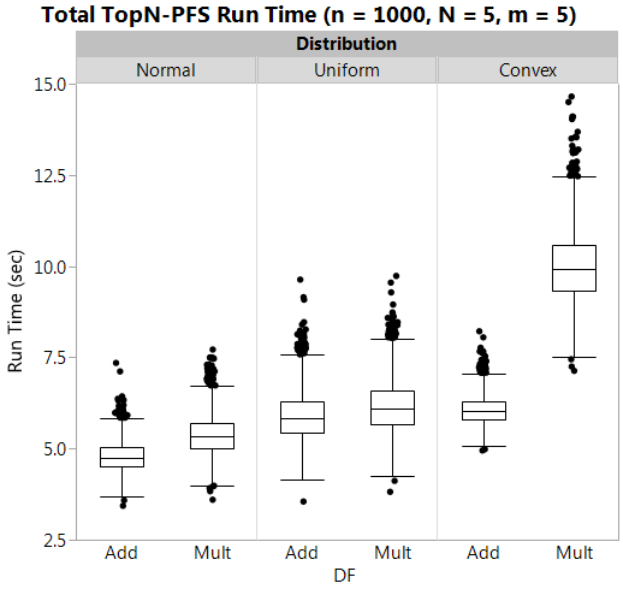

Figure C.8. Runtime simulation results

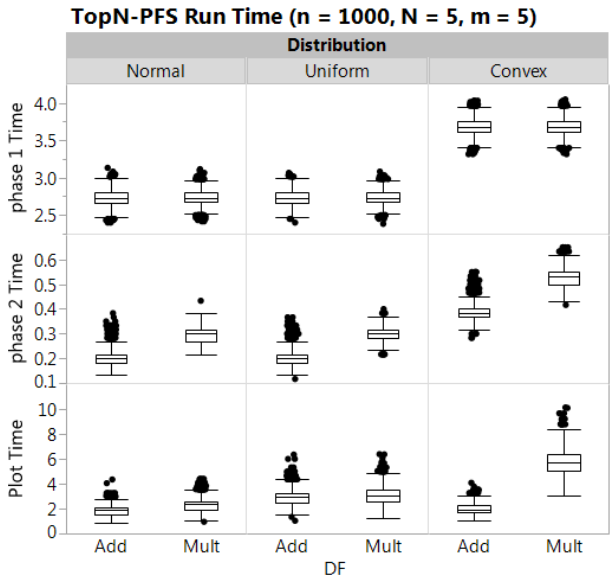

Figure C.9. Runtime simulation results partitioned by phase