Understanding Scaled Prediction Variance Using Graphical Methods for Model Robustness, Measurement Error and Generalized Linear Models for Response Surface Designs

by

Ayca Ozol-Godfrey

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in
Statistics

Christine M. Anderson-Cook, Chair
Eric P. Smith
G. Geoffrey Vining
William H. Woodall
Keying Ye

December 15, 2004
Blacksburg, Virginia

Keywords: Design Optimality, FDS Plots, Generalized Linear Models,
Model Robustness, Mixture Experiments, Measurement Error

Understanding Scaled Prediction Variance Using Graphical Methods for Model Robustness, Measurement Error and Generalized Linear Models for Response Surface Designs

Ayca Ozol-Godfrey

(ABSTRACT)

Graphical summaries are becoming important tools for evaluating designs. The need to compare designs in term of their prediction variance properties advanced this development. A recent graphical tool, the Fraction of Design Space plot, is useful to calculate the fraction of the design space where the scaled prediction variance (SPV) is less than or equal to a given value. In this dissertation we adapt FDS plots, to study three specific design problems: robustness to model assumptions, robustness to measurement error and design properties for generalized linear models (GLM). This dissertation presents a graphical method for examining design robustness related to the SPV values using FDS plots by comparing designs across a number of potential models in a pre-specified model space. Scaling the FDS curves by the G-optimal bounds of each model helps compare designs on the same model scale. FDS plots are also adapted for comparing designs under the GLM framework. Since parameter estimates need to be specified, robustness to parameter misspecification is incorporated into the plots. Binomial and Poisson examples are used to study several scenarios. The third section involves a special type of response surface designs, mixture experiments, and deals with adapting FDS plots for two types of measurement error which can appear due to inaccurate measurements of the individual mixture component amounts. The last part of the dissertation covers mixture experiments for the GLM case and examines prediction properties of mixture designs using the adapted FDS plots.

Dedicated

To John & My Family

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor, Dr. Christine M. Anderson-Cook for all her support and guidance throughout this research. She has always been there for me, not just during my dissertation but throughout my whole graduate study. She is a great friend and a mentor.

I would also like to express my appreciation to my committee members Dr. Eric P. Smith, Dr. G. Geoffrey Vining, Dr. William Woodall and Dr. Keying Ye for their valuable assistance and suggestions. I would also like to acknowledge Dr. Raymond H. Myers for his interest and support in my research.

I would also like to thank everyone in my department for supporting me in every way all these years and introducing me to the wonderful world of statistics.

Finally, I would like to thank my husband, John, for always believing in me and putting up with me during all this time. I would specially like to thank my parents, my brother and my aunty for always being there for me and supporting me all my life. I would also like to give my special thanks to my parents-in-laws for embracing me as one of their own.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION

In this thesis, new adaptations of the Fraction of Design Space plots are proposed. The Fraction of Design Space (FDS) plots, introduced by Zahran, Anderson-Cook and Myers, (2003), are new tools of examining a design's prediction variance properties through a graphical approach. They display the fraction of the design space where the scaled prediction variance (SPV) is less than or equal to a specific value for all observed values. The researcher can study the SPV distribution of a design by examining its FDS plot, and also compare designs.

In this dissertation, FDS plots will be adapted to study model robustness of various designs, to study standard designs (including mixture designs) for the Generalized Linear Models (GLMs) case and to study mixture designs for robustness to measurement errors. Usually, when the researcher is choosing a best design, he/she assumes that the particular model is correct. If after the data are collected, the research may find that some terms are not significant, then it may be that the chosen design is not that ideal for the reduced model. FDS plots are used to examine design robustness to the choice of various models for different designs. Little has been done in the literature to examine design prediction properties graphically for the GLM case. Khuri and Lee (1998) introduced a graphical approach for evaluating designs for nonlinear models using the scaled mean squared error of prediction. Robinson and Khuri (2003) adapted the Quantile Dispersion Graphs (QDGs) to compare designs for logistic regression models. Zahran (2002) has shown some examples of contour plots of both binomial and Poisson regression models for the 2^2 factorial design. FDS plots are adapted to study response surface designs for the GLM case. Since the prediction variance distribution for a design under the GLM framework depends on the parameter estimates, the robustness to parameter misspecification is also incorporated to the FDS plots. In addition, the robustness of mixture experiments to measurement error is studied. Measurement error may change the prediction properties of a mixture experiment. Therefore, we adapted FDS plots to study mixture designs'

robustness to measurement error. Last of all, mixture experiments for GLMs are discussed. The model form and the prediction distribution are studied using a binomial example.

The second chapter presents a review of the literature. Three important design optimality criteria are considered: D-, G- and V-optimality. Some response surface designs are defined which are compared for several properties in the following sections. The graphical tools to compare designs' prediction properties are studied: the Variance Dispersion Graphs (VDGs), the Quantile Dispersion Graphs (QDGs), and the Fraction of Design Space (FDS) plots. Finally, generalized linear models (GLMs) are defined.

In the third chapter we discuss robustness to model misspecification for any response surface design. An adaptation of the FDS plot is a design evaluation tool to illustrate this idea. The efficiency of a response surface design is model dependent, i.e., the design is typically chosen in accordance with an assumed model. If the assumed model is no longer appropriate, namely contains some unnecessary terms determined after the experiment is conducted, the prediction properties of the design may change due to the new model. The FDS plot is used to examine several types of models to consider the design's robustness properties to model misspecification. In addition to the 3- and 4-factor response surface designs studied, several mixture designs are examined for robustness to model misspecification. For any second order design, three types of models can be compared: the second order model, the first order model with interaction, and the first order model. An example with several oddly reduced models is given as well. FDS plots are scaled by the appropriate G-optimal values to help the researcher calibrate the robustness and desirability of the design more accurately. The FDS plots can also be used to calculate the G-efficiency and an approximate V-average for any model to help the researcher choose the best possible design.

In the fourth chapter, FDS plots are adapted to examine response surface designs for the GLM case. The form of the SPV for the GLM case is considered. A new term called penalized prediction variance (PPV) is introduced, which is important for evaluating

actual prediction. A comparison of designs including space-filling designs involving a response from a binomial distribution is given as an example to illustrate these methods. Both SPV and PPV are dependent upon the parameters in the model for the GLM case. Therefore, the robustness to initial parameter estimates should be taken into consideration. Several types of misspecifications are studied using the adapted FDS plots. The robustness to parameter misspecification is further investigated by considering a factorial design for GLM with a second order linear predictor. An example of a design involving a response from a Poisson distribution is also studied. The differences between choosing a model using a GLM or a variance stabilizing data transformation for a design are studied for the Poisson response case.

In the fifth chapter, the FDS plots are adapted to examine mixture experiments with measurement error. Measurement errors may arise when the measured mixture proportion is different than the desired mixture proportion. Steiner and Hamada (1997) studied the effect of measurement errors on the mixture proportions. The authors provided techniques for determining mixture proportions that are robust to measurement errors. Hamada, Martz and Steiner (2002) showed a Bayesian approach to account for mixing measurements errors. Two types of measurement errors in the component amounts are considered in this dissertation: relative errors and absolute errors. Both types of errors are randomly generated from two types of distributions: the normal distribution and the uniform distribution. Mixture designs are compared for robustness to measurement errors using FDS plots.

In the sixth chapter, mixture experiments under the GLM case are studied using FDS plots and contour plots. The model form for mixture experiments with the GLM is considered. Prediction variance properties of a mixture design with the GLM are explained using FDS plots and contour plots. The robustness of mixture experiments to misspecifications of the predicted means is studied using a binomial example as well as how mixture experiments may be improved in terms of their estimation properties using FDS plots.

Chapter II

LITERATURE REVIEW

II.1 Optimality Criteria

Design optimality criteria are often called the alphabetical optimality criteria because they are named by some of the letters of the alphabet. Kiefer and Wolfowitz (1959) were among the first authors who developed these optimality criteria. These are single number criteria where each one intends to capture a different aspect of the 'goodness' of a design. However a best design is typically more complicated than can be summarized by single numbers. These are just simplifications of the whole process of rating a design. Box and Hunter (1957), Box and Draper (1959, 1963, 1975), and Myers and Montgomery (2002, page 304) discussed some of the properties for comparing designs.

II.1.1 D-optimality

When considered historically, D-optimality (Kiefer, 1958) was the first alphabetical optimality criterion developed. It is also still among the most popular because of its simple computation, and the many available algorithms. The focus of D-optimality is on estimation of model parameters through good attributes of the moment matrix, which is defined as $M = \frac{X'X}{N}$, where $X'X$ is the information matrix, and N, the total number of runs, is used as a penalty for larger designs. D-optimality requires one to maximize the determinant of the moment matrix, i.e., a D-optimal design is the design, D^*, in the design space Ω such that the determinant of M, $|M(D^*)| = \max_{D \in \Omega} |M(D)|$. Under the standard normality assumptions, $|X'X|$ is inversely proportional to the square of the volume of the confidence region for the regression coefficients. Hence the larger the determinant of $X'X$, the better the estimation of the model parameters. Quite often for second order models, there is no finite D-optimal design, however one can still compare the results for a particular design to the theoretical values. D-efficiency of a design D

defined as D-efficiency = $\left(\frac{|M(D)|}{\max_{D\in\Omega} |m(D)|} \right)^{1/p}$, where p is the number of parameters in the model, allows the comparison between any design and the best design.

II.1.2 G-optimality

It may be that the aim of the practitioner is to have good prediction at a particular location in the design space. To attain this, Box and Hunter (1957) defined a variance function, i.e., the scaled prediction variance (SPV). The SPV provides a measure of the precision of the estimated response at any point in the design space. It is desirable that the distribution of the scaled prediction variance throughout the design space should be reasonably stable. Box and Hunter also defined the concept of rotatability which is attained when a design has the same value of scaled prediction variance for any two points which are the same distance from the center. The SPV is defined as

$$n(\boldsymbol{x}_o) = \frac{N \operatorname{var}(\hat{y}(\boldsymbol{x}_o))}{s^2} = N\, \boldsymbol{x}_o' (X'X)^{-1} \boldsymbol{x}_o \qquad \text{(II.1)}$$

where $\boldsymbol{x}_o$ corresponds to the location in the design space which is also a function of the model used, N is the total sample size penalizing the larger designs, $\operatorname{var}(\hat{y}(\boldsymbol{x}_o))$ is the variance of the estimated response at $\boldsymbol{x}_o$, X is the design matrix, and s^2 is the variance. For example, for a first order model with interaction, $\boldsymbol{x}_o' = (1, x_{10}, x_{20}, x_{10}x_{20})$. For a 2^2 factorial design with 1 center run (N=5) using a first order model with interaction,

$$n(\boldsymbol{x}_o) = N(1, x_{10}, x_{20}, x_{10}x_{20}) \begin{bmatrix} 1/5 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 1/4 \end{bmatrix} \begin{bmatrix} 1 \\ x_{10} \\ x_{20} \\ x_{10}x_{20} \end{bmatrix} = 1 + \frac{5}{4}(x_{10}^2 + x_{20}^2 + x_{10}^2 x_{20}^2). \qquad \text{(II.2)}$$

G-optimality makes use of the scaled prediction variance directly. The goal of G-optimality is to minimize the maximum SPV throughout the region of the design, i.e., a design which satisfies $\min [\max_{x\in R} n(\boldsymbol{x})]$ is a G-optimal design. Intuitively, G-optimality protects the experimenter against the worst case scenario being too undesirable. An interesting and an important result is that the lower bound for the maximum SPV is equal

to *p*, the number of parameters in the model (Myers, and Montgomery, 2002). Therefore, we define G-efficiency= $\frac{p}{\max_{x\in R} n(x)}$. Hence if a design with 3 factors for a second order model (this model has 10 parameters) is approximately 90% G-efficient, i.e., $0.90 = \frac{p}{\max_{x\in R} n(x)} = \frac{10}{\max_{x\in R} n(x)}$, then $\max_{x\in R} n(x) = \frac{10}{0.9} \approx 11.11$, meaning the maximum SPV for this design is not that much different from its optimum value.

II.1.3 V-optimality

V-optimality, also called Q-optimality and IV-optimality in the literature, is based on properties of the scaled prediction variance, SPV, as well. Instead of finding the maximum SPV in the region of interest, it makes use of the average of the SPV (throughout the region of interest). Hence this gives a measure of the overall distribution of the SPV. Box and Draper (1959, 1963) discussed the concept of average prediction variance. The aim of V-optimality is to minimize the average of the SPV throughout the whole region of interest. Even though, it is computationally more challenging, V-optimality is frequently most sensible in terms of measuring an important characteristic of the design. There is no known optimal bound for V-optimality. A V-optimal design is given by

$$V(D^*) = \min_{D\in\Omega} \frac{1}{K}\int_R n(x)dx \qquad \text{(II.3)}$$

where *R* is the region of interest and *K* is the volume of the region of interest. For example, for a 2^2 factorial design, a square region coded between -1 and 1, $K= 2^2 = 4$.

II.2 Some Response Surface Designs

In this section, some popular response surface designs are reviewed. Among the models considered are the first order model, $y = b_0 + \sum_{i=1}^{k} b_i x_i + e$, the first order model with interaction, $y = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i<j}\sum^{k} b_{ij} x_i x_j + e$, and the second order model,

$y = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i=1}^{k} b_{ii} x_i^2 + \sum_{i<j}^{k} \sum b_{ij} x_i x_j + e$, where y is the response, b's correspond to the parameters, x_i's are the design variables, and e is the random error with mean 0 and variance s^2. For more information on the basic models, refer to Myers and Montgomery (2002).

II.2.1 2^k Full Factorial Designs and 2^{k-s} Fractional Factorial Designs

Factorial designs are important in response surface methodology. Among their many applications are in screening experiments to help identify the important variables in an experiment to fit a first order model or a first order model with interaction, and as building blocks for other response surface designs, including the central composite designs which are discussed in the next section. The 2^k full factorial designs have only 2 levels for each of the k factors and these are usually coded as -1 for the low level and 1 for the high level.

Figure II.1 shows the design matrix and the geometry of the design for a 2^2 full factorial design.

$$\begin{matrix} x_1 & x_2 \end{matrix}$$
$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{pmatrix}$$

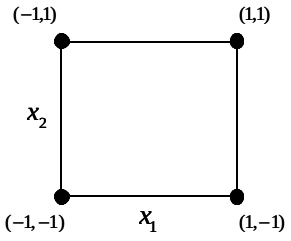

Figure II.1: a) Design Matrix for a 2^2 Factorial Design b) Geometry of a 2^2 Factorial Design

As the number of variables increases rapidly in an experiment, the number of runs needed to perform the full factorial design may become unrealistically large. For example, an experimenter might be interested in testing for 8 factors, meaning that he/she would need to perform $2^7 = 128$ runs which in most cases would be prohibitively expensive and time consuming. Hence the 2^{k-s} fractional factorial designs were introduced. The

2^{-s} corresponds to the fraction part of the design, i.e., only $\frac{1}{2^s}$ fraction of the total factorial runs is used. The optimal fractions are typically selected according to the resolution criteria. A fractional factorial design is said to be of resolution r if no effect involving i factors is aliased with effects involving less than r-i factors. The most commonly used resolutions are resolution III, IV, and V. The factorial fractional designs are based on three assumptions, the sparsity of effect principle, the projection property, and the sequential experimentation. Refer to Myers and Montgomery (2002) for more information.

II.2.2 Central Composite Designs (CCD)

The central composite design (CCD) introduced by Box and Wilson (1951) is one of the most popular response surface designs. It is designed to fit a second order model. It exists for spherical or cubical regions for $k \geq 2$, where k is the number of factors. Generally, this design has a small to moderate sample size with excellent estimation properties. The CCD has three components: the factorial points, the $2k$ axial points, and the n_c center runs. The factorial points used are either a 2^k factorial with levels at ± 1 or a 2^{k-s} resolution V fractional factorial, in which neither a main effect nor a 2-factor interaction is aliased with another main effect or a 2-factor interaction. These points are used primarily for the estimation of the linear terms, and the 2-way interactions. The axial points also called the star points, are at a distance of a from the design center. They mainly contribute to the estimation of the quadratic terms. The center runs are located in the center of the design space. They help estimate the quadratic terms as well as the pure error. For $k=3$, the design matrix and the geometry of the design are shown in Figure II.2.

The choice of the number of center runs provides flexibility to get a better estimate of the pure error and better power for tests. Moreover, the choice of the number of center runs affects the distribution of the SPV (Myers, and Montgomery, 2002). The choice of a is based on the region of interest. Depending on the value of a, the design may achieve rotatability. Rotatability is attained when a design has the same value of scaled prediction variance for any two points on the same distance from the center (Myers and

Montgomery, 2002). This can be achieved for $a = \sqrt[4]{F}$. For spherical regions, we commonly use $a = \sqrt{k}$ and n_c=3 to 5 helps to keep the distribution of the scaled prediction variance stable. For cuboidal regions, $a = 1$ is used, and no further improvement is achieved for n_c>2 (Myers and Montgomery, 2002).

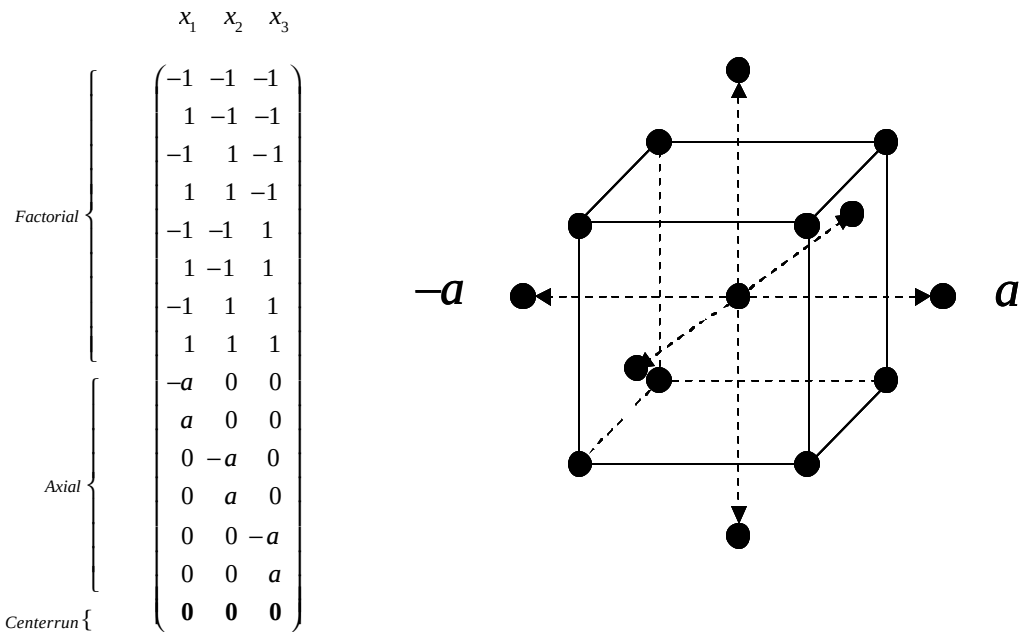

Figure II.2: a) Design Matrix for *k=3* CCD b) Geometry of *k=3* CCD

II.2.3 Small Composite Designs (SCD)

The small composite design was developed by Hartley (1959). This is a small economical design which exists for $k \geq 3$. The only modification to the CCD is that its factorial component is a fractional factorial design of resolution III. Therefore it suffers from poor estimation of some of the model parameters, specifically the main effects, and the 2-factor interactions. This design is mainly used when cost of each experimental run is high. Figure II.3 shows the design matrix and the geometry of the design for *k=3*.

$$
\begin{array}{l}
 \\
\textit{Factorial}\left\{\right. \\
 \\
 \\
\textit{Axial}\left\{\right. \\
 \\
 \\
\textit{Centerrun}\left\{\right.
\end{array}
\begin{array}{c}
\begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\
\left(\begin{array}{ccc}
-1 & -1 & -1 \\
1 & 1 & -1 \\
1 & -1 & 1 \\
-1 & 1 & 1 \\
\pm a & 0 & 0 \\
0 & \pm a & 0 \\
0 & 0 & \pm a \\
\mathbf{0} & \mathbf{0} & \mathbf{0}
\end{array}\right)
\end{array}
$$

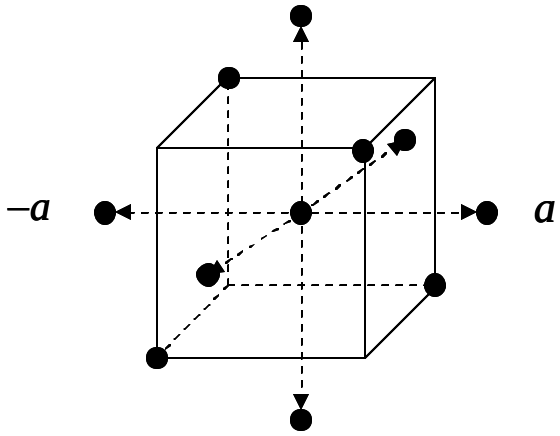

Figure II.3: a) Design Matrix for *k=3* SCD b) Geometry of *k=3* SCD

II.2.4 Plackett-Burman Composite Designs (PBCD)

The Plackett-Burman composite design is introduced by Draper (1985). This is a design for a second order model, and it exists for $k \geq 3$. Just like the CCD, it has three components: the factorial portion, the axial points, and the center runs. It uses *k* columns of a Plackett-Burman (1946) design for its factorial portion. The Plackett-Burman designs introduced in 1946 are 2-level fractional factorial designs of resolution III for studying up to *k=N-1* factors in *N* runs. Generally, the size of the PBCD is between a SCD and a CCD having the same number of factors. For example, for *k=4*, the SCD has 16 plus center runs, the regular CCD has 24 plus center runs, and the PBCD has 20 plus center runs. The design matrix for *k=5* is shown in Figure II.4.

$$
\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}'
=
\underbrace{\begin{pmatrix}
1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 & -1 & \pm a & 0 & 0 & 0 & 0 & \mathbf{0} \\
-1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 & 0 & \pm a & 0 & 0 & 0 & \mathbf{0} \\
1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 0 & 0 & \pm a & 0 & 0 & \mathbf{0} \\
-1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 & \pm a & 0 & \mathbf{0} \\
-1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & \pm a & \mathbf{0}
\end{pmatrix}}_{\textit{Plackett}-\textit{Burman}}{}'
$$

$\underbrace{\qquad\qquad}_{\textit{Axial}}$ $\underbrace{\quad}_{\textit{Centerrun}}$

Figure II.4: Design Matrix for *k=5* PBCD

II.2.5 Hoke Designs

Another class of small sized economical designs intended for a second order model for a cuboidal region is the Hoke (1974) designs. These designs exist for $k \geq 3$. They are based on a subset of the 2^k factorial and a subset of points containing one or more zeros. The Hoke designs have seven different types, D1 through D7 for any k. These smallest trace design types, D1 through D3 were chosen through the conjoining of sets involving assemblies of permutations of (-1, 0, 1). D4 through D7 were constructed through the augmentation of additional subsets to D1 and D3. D1 thorough D3 have fewer runs than D4 through D7. These designs may be appropriate if there is restriction preventing one corner from being considered. The D2 design matrix and the geometry of the design for *k=3* are shown in Figure II.5.

$$\begin{matrix} x_1 & x_2 & x_3 \end{matrix}$$
$$\begin{pmatrix} -1 & -1 & -1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

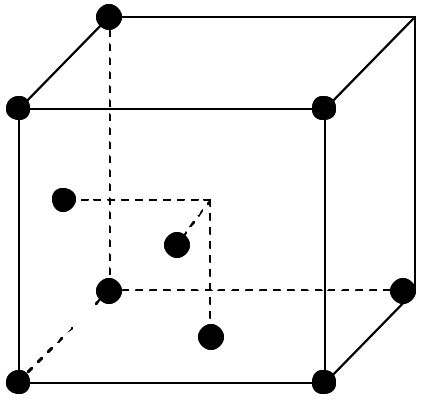

Figure II.5: a) Design Matrix for *k=3* Hoke D2 b) Geometry of *k=3* Hoke D2

II.2.6 Notz Designs

Another design for a second order model is the Notz (1982) design. This design has a cuboidal region and it exists for $k \geq 2$. The Notz design may be useful when one is seeking to avoid a design point in one of the corners but also wants a design with as few observations that still estimates all the parameters. It is constructed using a subset of the 2^k factorial, and a subset of points containing one or more zeros. Figure II.6 shows the design matrix and the geometry of the design for *k=3*.

$$\begin{array}{ccc} x_1 & x_2 & x_3 \end{array}$$

$$\begin{pmatrix} -1 & -1 & -1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

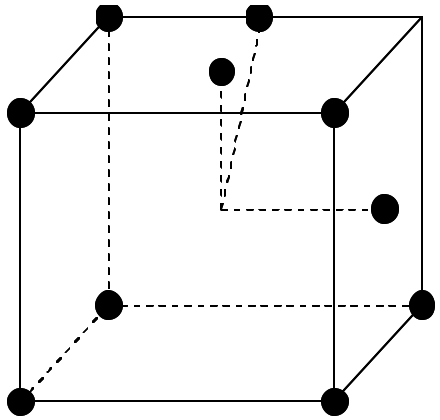

Figure II.6: a) Design Matrix for *k=3* Notz Design b) Geometry of *k=3* Notz Design

II.2.7 Box and Draper Designs

The Box and Draper design (1974) is a minimal point design for a cuboidal region with *N= ½ (k+1) (k+2).* This design was provided for a second order model and it exists for *k=2, 3, 4,* and *5*. The design matrix consists of a subset of a 2^k factorial plus $\frac{1}{2}k(k-1)$ points, of whose two coordinates are l, and $k-2$ coordinates -1, and k points, of whose coordinate is m, and $k-1$ coordinates -1. The values of l and m are chosen so as to maximize the determinant of $X'X$, to achieve a minimum generalized variance for the coefficient estimates. Figure II.7 shows the design matrix and the geometry of the design for *k=3*.

$$\begin{array}{ccc} x_1 & x_2 & x_3 \end{array}$$

$$\begin{pmatrix} -1 & -1 & -1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ -1 & 0.1925 & 0.1925 \\ 0.1925 & -1 & 0.1925 \\ 0.1925 & 0.1925 & -1 \\ -0.2912 & 1 & 1 \\ 1 & -0.2912 & 1 \\ 1 & 1 & -0.2912 \end{pmatrix}$$

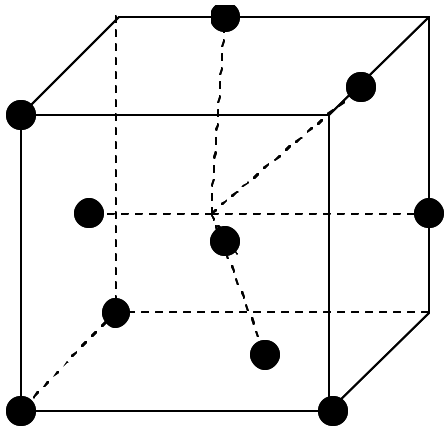

Figure II.7: a) Design Matrix for *k=3* Box and Draper b) Geometry of *k=3* Box and Draper

II.2.8 Mixture Experiments

Mixture experiments are a special type of response surface designs. Experiments where the factors are the ingredients or the components of a mixture are called mixture experiments. In mixture experiments, the choice of component levels is not independent. The response is assumed to depend only on the relative proportions of the components rather than the total amount in the mixture. For example, consider making a plain cake. The flavor of the cake depends on the relative proportions of its ingredients: baking powder, shortening, flour, sugar and water.

Suppose we have a mixture experiment with q components, where x_i is the proportion of the i^{th} component. Then it must satisfy the following constraints:

$$0 \le x_i \le 1 \qquad \forall i = 1,2,...,q \tag{II.4}$$

$$\sum_{i=1}^{q} x_i = 1 \tag{II.5}$$

The first constraint keeps each mixture component proportion between 0% and 100%, and the second constraint makes sure that at any point in the mixture space, the total sum of the proportions of all the components adds up to unity.

Under these constraints, the experimental region for a mixture experiment is a regular (*q-1*) dimensional simplex. For $q = 2$, the factor space is a straight line, for three components, it is an equilateral triangle, and for four components, the region is a tetrahedron. The coordinate system for the values of $x_i = 1,2,....,q$ is a simplex coordinate system. For example, with three components, the vertices represent simple component mixtures denoted by $x_i = 1$, $x_j = 0$ for $i, j = 1, 2,$ 3 and $i \neq j$. The points on the sides represent mixtures where only two of the components are present, and the interior points represent mixtures where all three components are present. The simplex regions for $q = 2$ and $q = 3$ are shown in Figure II.8.

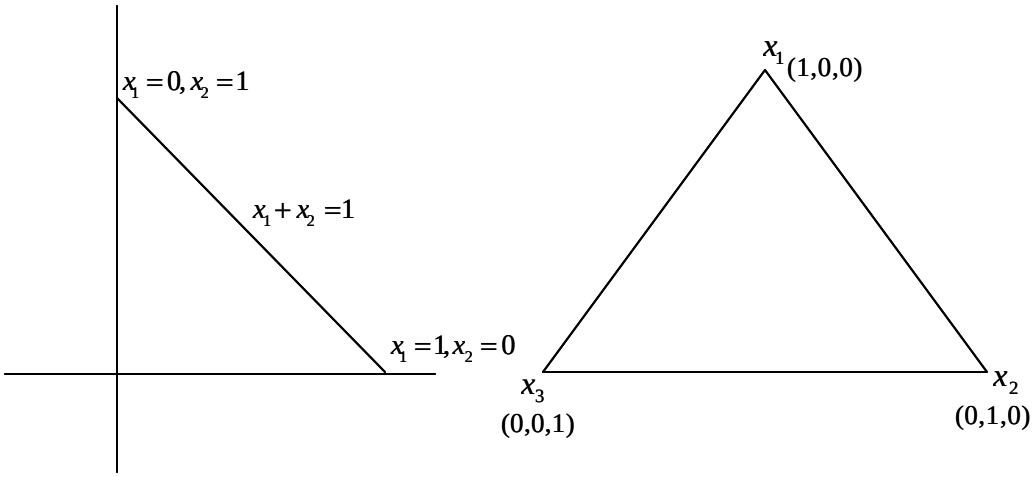

Figure II.8: Two- and Three-Component Simplex Regions for a Mixture Experiment

In addition to the two constraints, there may be additional constraints imposed on the component proportions. The constraints may be in the form of upper and lower bounds,

$$0 < L_i \le x_i \le U_i < 1, \qquad i = 1,.....,q, \tag{II.6}$$

or in the form of linear multicomponent constraints,

$$C_j \le A_{1j}x_j + A_{2j}x_j + ... + A_{qj}x_q \le D_j \qquad j = 1,2,.....,h \tag{II.7}$$

where the A_{ij}, C_j, and D_j are constants.

Due to the $\sum_{i=1}^{q} x_i = 1$ constraint, the form of the mixture model is different from the general polynomials used in the response surface methodology. The standard second order polynomial model for q factors is

$$h = b_o + \sum_{i=1}^{q} b_i x_i + \sum_{i=1}^{q} b_{ii} x_i^2 + \sum_{i<j} \sum b_{ij} x_i x_j . \tag{II.8}$$

When the constraints $\sum_{i=1}^{q} x_i = 1$ and hence $x_i^2 = x_i \left(1 - \sum_{\substack{j=1 \\ j \ne i}}^{q} x_j \right)$ are applied for the mixture designs, the model becomes

$$h = \sum_{i=1}^{q} b_i^* x_i + \sum_{i<j} \sum b_{ij}^* x_i x_j \tag{II.9}$$

where $b_i^* = b_o + b_i + b_{ii}$ and $b_{ij}^* = b_{ij} - b_{ii} - b_{jj}$ for $i, j = 1,2,....,q$, $i < j$. This model is known as the second order Scheffe′ (1958) model. The full cubic polynomial model is

$$h = \sum_{i=1}^{q} b_i^* x_i + \sum_{i<j} \sum^{q} b_{ij}^* x_i x_j + \sum_{i<j} \sum^{q} d_{ij} x_i x_j (x_i - x_j) + \sum_{i<j<k} \sum \sum^{q} b_{ijk}^* x_i x_j x_k \cdot \qquad \text{(II.10)}$$

A special case of the cubic polynomial model called the special cubic model is

$$h = \sum_{i=1}^{q} b_i^* x_i + \sum_{i<j} \sum^{q} b_{ij}^* x_i x_j + \sum_{i<j} \sum \sum^{q} b_{ijk}^* x_i x_j x_k \cdot \qquad \text{(II.11)}$$

There are other types of models used for conducting mixture experiments. Among them are the Cox model and the slack variable model. The Cox (1971) model is an alternative to the Scheffe′ model. The first order model is $h = b_o + \sum_{j=1}^{q} b_j x_j$, whereas the second order model is $h = b_o + \sum_{j=1}^{q} b_j x_j + \sum_{i=1}^{q} \sum_{j=1}^{q} b_{ij} x_i x_j$, and additional constraints are imposed to make the terms estimable. The Cox model parameters are given different interpretations than the Scheffe′ model parameters which are obtained by comparing the response values at points in the simplex to the response value obtained at a standard mixture point, ***s***. Smith and Beverly (1997) expressed the difference in the responses using a quadratic polynomial which makes it easier to understand several important properties of the Cox model.

Snee and Rayner (1982) proposed the slack-variable approach which leaves out the variable that usually has the largest proportion or is the most inactive compared to the other variables. The second order slack-variable model then involves one less component, and has the form $h = a_o + \sum_{i=1}^{q-1} a_i x_i + \sum_{i \le j} \sum^{q-1} a_{ij} x_i x_j$, where $a_o = b_q$, $a_i = b_i - b_q + b_{iq}$, $a_{ii} = -b_{iq}$, and $a_{ij} = b_{ij} - (b_{iq} + b_{jq})$.

Among the mixture designs that are commonly used are the simplex lattice designs (Scheffe′, 1958) and the simplex-centroid designs (Scheffe′, 1963). The simplex lattice

designs are also referred to as {q, m} simplex lattice designs. The q stands for the number of components and the proportions of $0, \frac{1}{m}, \frac{2}{m}, ..., 1$ are used for each component x_i. The simplex-centroid designs have $2^q - 1$ number of distinct points. They contain every non-empty subset of the q components where the components are present in equal proportions. These two designs will be revisited in the later chapters. See Cornell (2002) for more details on these two designs.

In addition to the mixture variables, there are variables called the process variables that do not form any portion of the mixture but whose levels when changed can affect the properties of the mixture components or the response. Going back to the plain cake example, baking time and baking temperature could be considered as process variables. If one changes the baking temperature from $350^{\circ}C$ to $375^{\circ}C$, this might affect how different combinations of ingredients influence the response. Scheffe′ (1963) was first to discuss the inclusion of process variables with mixture components.

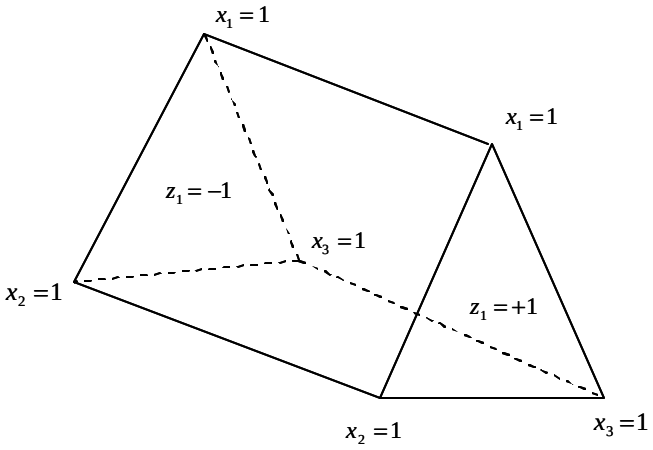

Figure II.9: The Combined Region of Interest for a 3-component Mixture Experiment with One Process Variable

With process variables in the experiment, the region of interest includes both the mixture components and the process variables. Let us assume that each process variable is coded between $z_i = +1$ and $z_i = -1$ for the low and high ranges being considered. Then the

region of interest for the process variables becomes an n dimensional hypercube. Therefore, the combined region of interest has dimension $(q-1)+n$. Figure II.9 shows the design space for a mixture experiment with three components, and a single process variable at two levels.

Cornell and Gorman (1984) presented mixture designs with process variables where only a fraction of the total number of possible design points is used. They used several scenarios of the fractions of the 2^n factorial for the process space. Cornell (1995) discussed some of the difficulties faced while building the model when a constrained-region mixture experiment with process variables or the total amount of the mixture is used.

When process variables are considered, complete randomization of the experimental runs may not be possible because of time or cost restrictions. In this type of a situation, split-plot type designs may be used. Cornell (1988) examined mixture experiments with process variables where the design is of the split-plot type. Kowalski, Cornell, and Vining (2002) proposed new split-plot designs using a new model form.

Piepel and Cornell (1994) used a five-component waste grout example to compare five approaches to analyze a mixture experiment: component proportions, mixture-amount, mixture-process variable, mathematically independent variables and the slack variable. Using a six-factor fish-patty example, Gorman and Cornell (1982) proposed a reparametrized model form which allows the process variables to be estimated distinctly from the effects of the mixture components. Montgomery and Voth (1994) discussed the impact of multicollinearity in mixture experiments. Cornell and Gorman (2003) introduced two model forms to encounter the problem of collinearity that arises when the mixture region is highly constrained.

Andere-Rendon, Montgomery and Rollier (1997) introduced a Bayesian modification to the dependency of D-optimal mixture designs on the assumed model. Lin, Myers and Ye

(2000) developed a Bayesian two-stage D-D optimal design for mixture experiments and mixture experiments with process variables.

Kowalski, Cornell and Vining (2000) proposed a new class of designs for mixture experiments with process variables which is based on central composite designs in the process variables. They presented a new model type to accommodate these new designs.

Chen, Li and Jackson (1996) presented a logistic regression model for analyzing data from a mixture experiment.

II.3 Variance Dispersion Graphs (VDGs)

A Variance Dispersion Graph (VDG) developed by Giovannitti-Jensen and Myers (1989) is a graphical tool to study a design's prediction properties. It displays the scaled prediction variance (SPV) throughout a multidimensional region on a single two dimensional graph. Assuming a spherical region, a VDG consists of three curves per design for standard linear models. First, the minimum of the scaled prediction variance, $n(\boldsymbol{x})$, displayed against r as

$$\min_{\boldsymbol{x}\in U_r}[n(\boldsymbol{x})]=\min_{\boldsymbol{x}\in U_r}\frac{N\,\mathrm{var}(\hat{y}(\boldsymbol{x}))}{\sigma^2}. \qquad \text{(II.12)}$$

Second, is the maximum of $n(\boldsymbol{x})$ displayed against r as

$$\max_{\boldsymbol{x}\in U_r}[v(\boldsymbol{x})]=\max_{\boldsymbol{x}\in U_r}\frac{N\,\mathrm{var}(\hat{y}(\boldsymbol{x}))}{\sigma^2}. \qquad \text{(II.13)}$$

Third is the average SPV plotted against a radius, r, from the center of the design space, is defined by

$$V^r=\gamma\int_{U_r} n(\boldsymbol{x}) \qquad \text{(II.14)}$$

where $U_r=\{x:\sum x_i^2=r^2\}$, denoting the integration over the surface of a sphere of radius, r, $\gamma=(\int_{U_r} d\boldsymbol{x})^{-1}$, and $n(\boldsymbol{x})$ is the scaled prediction variance.

In addition to its three curves, two horizontal curves, the 100% and 50% G-efficiencies are frequently displayed which correspond to $n(\boldsymbol{x}) = p$ and $n(\boldsymbol{x}) = 2p$, respectively, where p is the number of parameters in the model (Myers and Montgomery, 2002, page 396).

Figure II.10 shows the VDG of a 3-factor CCD for a second order model in a spherical region. The design is scaled to the unit sphere. It is rotatable, i.e., $a = \sqrt[4]{8}$. Therefore the minimum, the maximum, and the average SPV curves are identical. The number of center runs considered is one, three, and five. As the number of center runs increases, the SPV improves closer to the center of the design space.

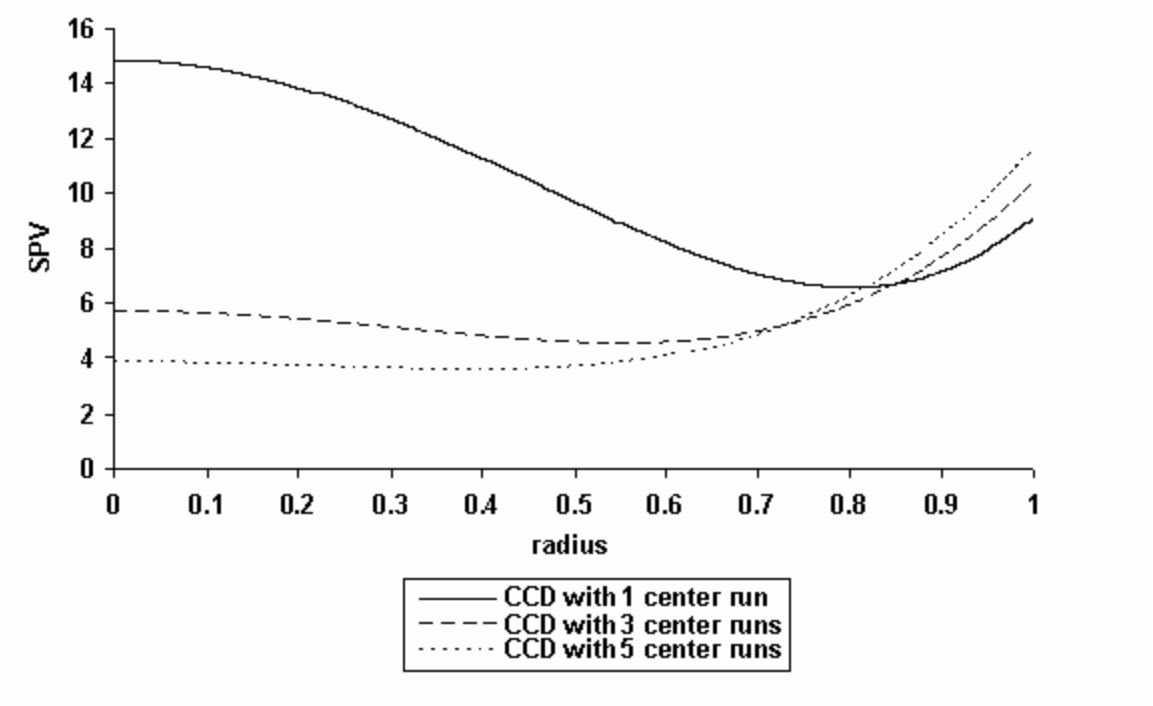

Figure II.10: VDG for Examining the Effect of Center Runs in CCDs

These graphs are informative and useful in terms of comparing designs on a fixed design space. Vining (1993) wrote a computer program in Fortran to generate the VDGs. Myers, Vining, Giovannitti-Jensen and Myers (1992), Borkowski (1995), and Block and Mee (2001) used the VDGs to examine Response Surface Designs on a spherical region.

Designs that have cuboidal regions were studied using the VDGs as well in Borkowski (1995), Rozum and Myers (1991), Block and Mee (2001), Myers and Montgomery (2002), Borror, Montgomery, and Myers (2002), and Park, Richardson, Borror, Ozol-Godfrey, Anderson-Cook, and Montgomery (submitted). In this case, instead of concentric circles, shrunken cubes are used. Moreover, Trinca and Gilmour (1998) used VDGs for comparing blocked response surface designs for both types of regions.

Mixture designs were also studied using the VDGs. Piepel and Anderson (1992), and Piepel, Anderson and Redgate (1993 a, b) developed VDGs for designs on polyhedral regions, specifically mixture experiments with constraints. Vining, Cornell and Myers (1993) used VDGs to plot the prediction variance along prediction rays (Cox rays) which are lines connecting the centroid of the design space to its vertices. Goldfarb, Montgomery, Borror and Anderson-Cook (2004) extended VDGs to three dimensions in order to study the prediction variance properties of mixture experiments with process variables. In the 3-D VDG for mixture-process designs, the process space shrinkage factor is placed along the x-axis, the mixture space shrinkage factor is along the y-axis, and the prediction variance values are plotted along the z-axis. The three dimensional VDG helps one to visualize the distribution of the prediction variance for both spaces separately. Moreover, by keeping one of the shrinkage factors of the two spaces at a constant value, one can examine the other space's prediction variance at that shrinkage level.

II.4 Quantile Dispersion Graphs (QDGs)

Quantile Dispersion Graph introduced by Khuri, Kim and Um (1996) described the distribution of the SPV using quantiles for a distance of r from the center of the design space. These plots display the quantiles of the SPV including the minimum and the maximum values. Later, Khuri, Harrison and Cornell (1999) used the QDGs to explore constrained regions, specifically mixture experiments. Robinson and Khuri (2003) considered generalized linear models and used the QDGs to display the scaled mean-squared error of prediction for comparing designs for logistic regression.

II.5 Fraction of Design Space (FDS) Plots

The FDS plot was first introduced by Zahran, Anderson-Cook and Myers (2003) to complement the Variance Dispersion Graph for both spherical and cuboidal design spaces. The plot also displays characteristics of SPV throughout a multidimensional region on a single 2 dimensional graph, this time with a single curve. The FDS plot shows the fraction of the design space at or below any SPV value. It is constructed by sampling a large number of values, say *n*, from throughout the design space and obtaining all of the corresponding SPV values which are then ordered and plotted against the quantiles (*1/n, 2/n,* …).

The difference between the FDS plot and the VDG is that the VDG hides the fact that different volumes are associated with each radii or shrinkage factor. Hence the viewer must mentally rescale the plots to focus more on the outer edges of the design, in order to gain an accurate perspective of the SPV of the region.

The FDS plot provides the experimenter information about the distribution of the SPV throughout the design space, including the minimum and the maximum SPVs. Similar to the VDG, the researcher can determine the approximate 50% and 100% G-efficiencies for a design by looking at an FDS plot. The idea is the more the fraction of the design space for a SPV which is close to the minimum, the better the design is. Moreover, the flatter the line, the more stable the SPV distribution for that design is.

Goldfarb, Anderson-Cook, Borror and Montgomery (2004) considered the FDS plots for non-regular regions, such as mixture experiments and mixture experiments with process variables. Park, Richardson, Borror, Ozol-Godfrey, Anderson-Cook, and Montgomery (submitted) studied cuboidal designs using FDS plots. They plotted the SPV values for the mixture space by holding the process space at constant shrinkage factors. This way, the experimenter is able to see which of the two spaces contributes more to changes in the SPV.

II.6 Generalized Linear Models (GLMs)

Generalized Linear Models are a general case of the linear models where the error term belongs to an exponential family type distribution. They were introduced by Nelder and Wedderburn (1972). See Wedderburn (1974) for more information. The density function for the exponential family type distribution is

$$f(y;q,f)=\exp\left\{\frac{yq-b(q)}{a(f)}+c(y,f)\right\}, \qquad \text{(II.15)}$$

where $a(\cdot), b(\cdot), \text{and } c(\cdot)$ are specific functions, q is the natural location parameter, and f is the dispersion parameter. Among the members of the exponential family of distributions are the binomial distribution, the Poisson distribution, the normal distribution, and many others. The observations, $y_1,.....,y_n$, are independently distributed each with mean $E(y_i)=m_i$, respectively. Each y_i has a distribution which is a member of the exponential family. $h_i = \boldsymbol{x}_i'b$ is the linear predictor of some function of $E(y_i)=m_i$. It involves the regressors, $x_1,.....,x_k$. The link function, g, is used to define the model and is given by $h_i=g(m_i)$ for $i=1,.....,n$. In this dissertation, we assume the canonical link $h_i=q_i$, where q_i is the parameter to be estimated.

For the GLM case, the information matrix becomes $I(X;\text{b})=\frac{(X'VX)}{[a(f)]^2}$ where $V=diag\{s_i^2\}$, and s_i^2 is a function of m_i. Hence the asymptotic variance-covariance matrix of $\boldsymbol{b}$, the parameter estimates, becomes $\text{var}(\boldsymbol{b})=I^{-1}(\boldsymbol{b})=(X'VX)^{-1}[a(f)]^2$.

In this dissertation, we concentrate on two particular cases of GLMs: the logistic regression, and the Poisson regression. For both of these cases, $a(f)=1$. Hence $\text{var}(\boldsymbol{b})=(X'VX)^{-1}$. For the logistic case, $s_i^2=n_i p_i(1-p_i)$ where n_i is the number of observations taken at $\boldsymbol{x}_i$, and p_i is the probability of success/failure. For the Poisson case, $s_i^2=e^{x_i'b}=m_i$ where m_i is the expected number of counts at $\boldsymbol{x}_i$. For more

information on the GLMs, see Myers, Montgomery and Vining (2002), and McCullagh and Nelder (1989).

Design optimality for GLMs is different than for the linear models because in the GLM case, the information matrix is a function of the unknown parameters. The V matrix above is a function of the unknown parameters that are estimated only after data have been collected. Hence design optimality becomes more complex. Factors like the reliability of the initial parameter estimates and the robustness of designs to these estimates should be considered.

There have been several approaches proposed in the literature to deal with the problem of initial estimates. The sequential designs approach, obtains the improved parameter estimates through a sequential stage procedure. Abdelbasit and Plackett (1983) explored a multi-stage procedure to come up with a D-optimal design for a one-variable logistic model. Minkin (1987) developed a two-stage procedure using the likelihood regions. Chaudhuri and Mykland (1993), Myers, Myers, Carter and White (1996) proposed a two stage design procedure for fitting the logistic model having independent variables with more than three levels.

The second procedure is a Bayesian approach which deals with assuming prior distributions on the parameters. Chaloner and Larntz (1989) derived a general theory for nonlinear models where there are no restrictions of equal spacing or equal weighting on the designs. Zacks (1977) and Tsutakawa (1980) used Bayesian methodology where expectations are taken to eliminate prior distributions to the parameters. Chaloner and Verdinelli (1995) gave a broad review of the Bayesian approach.

The third approach is the minimax procedure introduced by Sitter (1992). Robust designs to the poor initial parameter estimates are found by using more design points and a larger spread of the design space. This approach is straightforward and easy to implement through a computer algorithm.

Myers (1999) reviewed the current status of response surface methodology (RSM) and emphasized the importance of GLMs. Khuri gave an overview of the use of GLMs in the RSM (2001). Khuri (2003) reviewed the GLM issues in response surface methodology. Most recently, Myers, Montgomery, Vining, Borror and Kowalski (2004) discussed the new developments and applications in the RSM in the last 15 years, including the GLMs.

Chapter III

FRACTION OF DESIGN SPACE PLOTS FOR EXAMINING MODEL ROBUSTNESS

III.1 Abstract

When choosing between competing designs, it is typical to specify a design space and model on which to base the comparison. The prediction capabilities of the design, specifically G- and V-efficiency using scaled prediction variance (SPV), are based on this chosen model. After the data are collected, and individual effects are tested, some terms may not be significant in the model. In this case, the experimenter likely will decide to use a reduced model, which has only a portion of the terms included that were in the original model for which the design was chosen. This chapter presents a graphical method for examining design robustness related to the SPV values using Fraction of Design Space (FDS) plots by comparing designs across a number of potential models in a pre-specified model space. The FDS plots show the various distributions of the SPV throughout the design space for different models for a chosen design on the same graph. The methods are demonstrated on several examples for different models and on a variety of design spaces.

III.2 Introduction

When constructing an experiment, the aim of the researcher is to choose a design which allows for good estimation of the relationship between the explanatory factors and the response of interest. This relationship can be written as $y = h(x_1, x_2, ..., x_k) + e$ where y is the response, h is the true unknown function, $x_1, x_2, ..., x_k$ are the explanatory or the independent variables, and e is the error term that represents sources of variability not accounted for in h.

In response surface methodology (RSM), a standard approach to modeling the relationship is to approximate it with a low-order polynomial such as a second order model,

$$y = b_o + \sum_{i=1}^{k} b_i x_i + \sum_{i=1}^{k} b_{ii} x_i^2 + \sum_{i<j} \sum b_{ij} x_i x_j + e \ , \qquad \text{(III.1)}$$

where the b's correspond to the parameter coefficients to be estimated, $x_1, x_2, ..., x_k$ are the explanatory or the independent variables, and e is the error term. Before running the experiment, the researcher must decide on which type of design he/she wants to use to fit the approximate model. Choosing the design is dependent on the specific type of model. For example, the central composite design introduced by Box and Wilson (1951) is commonly used for second order polynomial models because its structure allows for good estimation of all model parameters. On the other hand, Plackett-Burman designs (1946) or factorial designs are most suitable for first order models.

Choosing between competing designs involves many trade-offs as this decision is based on many diverse goals for a good design. Some of these criteria are listed in Box and Draper (1987), and Montgomery (2001). One common approach to comparing designs is to consider the alphabetical optimality criteria, like D, A, G or V, which each attempt to summarize one important characteristic of the design. Typically, a design will have different efficiency values for different criteria. Other performance criteria are available when comparing designs based on a specific model such as the ability to measure pure error and lack of fit, the size of the design, good prediction variance properties, and many more. For more information, see Myers and Montgomery (2002, page 304).

The prediction properties of a design, specifically the scaled prediction variance (SPV), also depend on the chosen model. The scaled prediction variance, SPV, is defined as

$$n(\boldsymbol{x}_o) = \frac{N \ \mathrm{var}(\hat{y}(\boldsymbol{x}_o))}{s^2} = N \ \boldsymbol{x}_o' (X'X)^{-1} \boldsymbol{x}_o \ . \qquad \text{(III.2)}$$

Here, $\boldsymbol{x}_o$ corresponds to the location in the design space which is also a function of the model used, N is the total sample size penalizing the larger designs, $\mathrm{var}(\hat{y}(\boldsymbol{x}_o))$ is the variance of the estimated response at $\boldsymbol{x}_o$, X is the design matrix expanded to model form, and s^2 is the variance. For a 2-factor design with a second order model, X has dimension $N \times 6$ and correspondingly has $\boldsymbol{x}_o' = (1, x_{1o}, x_{2o}, x_{1o}x_{2o}, x_{1o}^2, x_{2o}^2)$. For a first order model with

interaction, $\boldsymbol{x}_o' = (1, x_{1o}, x_{2o}, x_{1o}x_{2o})$ and X has dimension $N \times 4$. A design having a desirable SPV distribution throughout the design space for a certain model would have a different distribution of SPV for another model. This frequently happens after the data are collected, and the experimenter finds that some effects may not be significant in the model. As a result, the experimenter will likely decide to choose a reduced model in subsequent analysis phases. When the reduced model is being used, the initial design might no longer have desirable prediction variance properties. The researcher needs a tool to assist him/her in choosing the best design regardless of the final model selected within a set of models considered likely in the possible model space.

Several methods to examine a design's prediction capabilities have been developed. Among these are the Variance Dispersion Graphs (VDGs) developed by Giovannitti-Jensen and Myers (1989), the three-dimensional VDGs by Goldfarb, Borror, Montgomery, and Anderson-Cook (2004), the Quantile Dispersion Graphs (QDGs) by Khuri, Kim, and Um (1996), the Fraction of Design Space (FDS) plots by Zahran, Anderson-Cook, and Myers (2003), and by Goldfarb, Anderson-Cook, Borror, and Montgomery (2004). In this chapter, we use the FDS plots to study the robustness of a design to model changes.

The FDS plot can be constructed in a number of different ways. Originally, the SPV values for the plots were calculated analytically using the software package Mathematica. While feasible for up to second order models in lower dimensions of up to 5 factors, in higher dimensions this approach became computationally too cumbersome and slow. Hence, other possibilities were considered. Goldfarb et. al. (2004) considered sampling on the perimeter of various shrinkage levels proportional to the relative volume associated with that portion of the total design space, as well as uniform sampling throughout the entire design space. Both of these approaches have been shown to give comparable results for a wide variety of applications at a fraction of the computational cost. For lower dimensional spaces, with 4 or fewer factors, an accurate FDS plot can be constructed with 2000-5000 points. For experiments involving 5 to 8 factors, simulations have shown that 10000 points is adequate. Borkowski (2003) has shown that evaluation

over a random set of points gives better estimates of the average prediction variance compared to the evaluation over a fixed set of points for cubodial designs for any polynomial model. To get more precise estimates of the maximum value, and hence the G-efficiency, in the design space, the uniform sampling can be supplemented with some of the corners of the design space.

The FDS plot shows the fraction of the design space at or below any SPV value. The plot is constructed by sampling a large number of values, say *n*, from throughout the design space and obtaining the corresponding SPV values. These n SPV values are then ordered and plotted against the quantiles (*1/n, 2/n*, …). The x-axis gives the quantiles of the design space ranging from 0 to 1, while the y-axis shows the SPV values. For a given point on the curve, we can extract what fraction of the total design space has SPV values less than or equal to the given value. By looking at the two ends of the line, we can see the minimum and maximum SPV values for the design. For example in Figure III.1, an FDS plot for a cuboidal region in three factors is shown for a central composite design (CCD) by Box and Wilson (1951). This design consists of three parts: the factorial points, the axial points, and the center runs. The F factorial points are a 2-level full factorial, with 2^3 combinations. The *2k* axial points are a distance a from the center of the design space. For the cuboidal region, we select $a = 1$. Two center runs $(n_c = 2)$ are used, giving a total sample size of 16 observations.

Both G- and V-efficiency can be summarized on the FDS plot. G-optimality seeks to minimize the maximum SPV throughout the region of the design. A design which satisfies $\min\left[\max_{x \in R} n(\boldsymbol{x})\right]$ is a G-optimal design where *R* is the chosen design region. The lower bound for the maximum SPV is equal to *p*, the number of parameters in the model (Myers and Montgomery, 2002, page 396). Hence a G-optimal design is one where the maximum SPV is equal to *p*. G-efficiency is defined as

$$\text{G-efficiency} = \frac{p}{\max_{x \in R} n(\boldsymbol{x})}. \qquad \text{(III.3)}$$

A 100% G-efficient design has its maximum SPV equal to p, and a 50% G-efficient design has maximum SPV equal to $2p$. These G-efficiency lines are displayed on an FDS plot. Hence, the researcher is able to obtain an approximate G-efficiency for a design.

V-optimality seeks to minimize the average SPV throughout the whole region of interest. A V-optimal design is one where

$$V(D^*) = \min_{D \in \mathrm{D}} \frac{1}{K} \int_R n(\boldsymbol{x}) d\boldsymbol{x} \qquad \text{(III.4)}$$

where R is the region of interest and K is the volume of the region of interest. The SPV is integrated over the design space to calculate the V-criterion. Accordingly, V-efficiency for a design D^* is defined as $\frac{\min_{D \in \mathrm{D}} V(D)}{V(D^*)}$. We propose an addition to the FDS plot that allows quantification relative to the V-efficiency to be obtained for a design. V-efficiency examines the size of the average SPV value over the region which we will define as V-average. The V-average value will be estimated by averaging the sampled SPV in the region of interest.

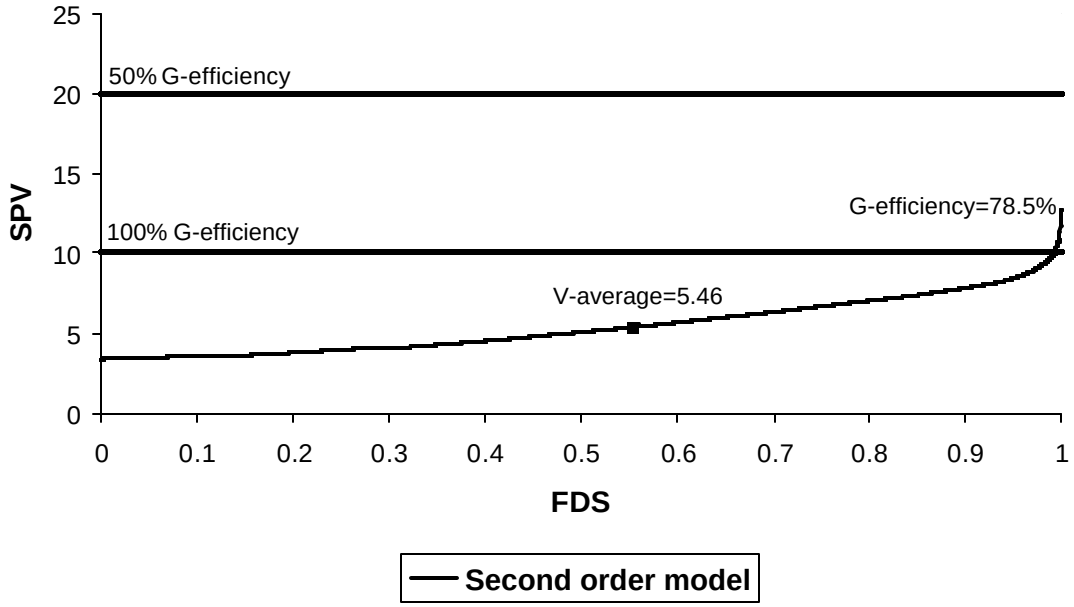

Figure III.1: FDS Plot for CCD $(n_c = 2)$ on a Cuboidal Region

In Figure III.1, the SPV values for the 3-factor CCD range between 3.35 and 12.73. The point at approximately (0.48, 5) can be interpreted as 48% of the total design space have a SPV value at or below 5. Note that a flatter curve implies the maximum and minimum SPVs are closer together giving a more stable distribution of the SPV. The number of parameters (G-optimal bound) of a second order model for $k=3$, is 10. Hence, the 100% G-efficiency = 10, and 50% G-efficiency = 20. From the above figure, the approximate G-efficiency is 78.5% (10/12.73). The approximate V-average for this design obtained by averaging all the sampled values, is 5.46. The corresponding FDS value for this point at approximately 0.56 indicates that the mean and median of the SPV distribution are not the same. This indicates that the distribution is positively skewed with a small number of large values influencing the average substantially.

FDS plots can be useful when comparing two or more designs (Zahran, Anderson-Cook, and Myers, 2003). They allow the researcher to compare the SPV distributions of designs as well as their G-efficiency and V-average values.

Unless the experimenter has prior evidence that the assumed model is the second order model, he/she should consider whether the chosen design will be appropriate for any reduced model. FDS plots allow the experimenter to study a design's prediction characteristics for any reduced model. The above FDS plots were obtained assuming a second order model. However, when the data are collected, the researcher may choose to use a lower order polynomial model, or an oddly reduced model. The reduced model may only have a fraction of the quadratic terms, or it may be a first order model with some of the interaction terms included. It may be such that the principal of hierarchy may not be appropriate. With this new chosen model, the design may no longer have its desired properties of the prediction variance. The distribution of the SPV may change drastically depending on which terms are excluded from the model. The experimenter wants to know how different if any, the SPV curves for each model for a certain design will be. The robustness of the design to model changes is determined by examining the behavior of the SPV curves.

Model robustness has been considered since Cox (1961) first introduced the concept. Hunter and Reiner (1965) narrowed the general concept to comparing two rival models, while Roth (1965) considered the case of more than two competing designs. Box and Hill (1967) improved the approach by taking into account the variances of the estimated responses. Andrews (1971) proposed a sequential design for screening out bad models using F tests. Atkinson and Federov (1975) developed sequential designs based on T-optimality for choosing between two potential regression models. Hill (1978) reviewed the procedures of experimental design for choosing between competing designs. Wu (1981) discussed model robustness for several randomization procedures such as the randomized complete block design and the Latin square design. Cook and Nachtsheim (1982) presented a generalized linear optimality criterion to cases where the model form was not known. This method was applicable to a situation in which the experimenter was unable to specify an exact model but knew that the true model was an unknown member of a finite set of known models. Vining and Myers (1991) proposed a graphical approach for evaluating designs in terms of their mean squared error of prediction aiding researchers in exploring design performance in the existence of model misspecification. DuMouchel and Jones (1994) proposed a Bayesian adjustment to D-optimality to generate designs more robust to model assumptions. Chipman (1996) applied a Bayesian approach to reduce the number of possible models through heredity properties. Using the hierarchical nature of different model terms, he developed prior relationships between predictors that were then incorporated into the stochastic search variable selection for any type of linear model. Andrea-Rendon, Montgomery and Rollier (1997) extended DuMouchel and Jones' approach to constrained and unconstrained mixture experiments. Li and Nachtsheim (2000) developed a class of model robust designs for estimating main effects and a combination of interactions. After obtaining an upper bound, g, on the number of possible two-way interactions from the experimenter, a model robust factorial design was conducted guaranteeing the estimability of any combination of the g interactions. Cheng and Wu (2001) made use of the projection properties of designs to combine the screening and the exploration phases in response surface methodology. Borkowski and Valeroso (2001) compared some response surface designs on the hypercube for reduced models using alphabetical optimality criteria. Bingham and Li

(2002) expanded Li and Nachtsheim's approach to the robust parameter design where the control-by-noise interactions were of interest. Heredia-Langner, Carlyle, Montgomery and Borror (2004) presented a technique to obtain model robust designs using genetic algorithms.

While much of the previous literature in model robustness has focused on how to find a best design, in this chapter we present a graphical approach for better visualizing model robustness. The FDS plot helps the researcher examine a design for any chosen linear model. Hence it gives an opportunity to compare a design's prediction variance properties based on different models simultaneously. This information is useful because the assumed model may change after the data are collected.

III.3 Examples Involving Second Order Response Surface Designs

We now consider some examples which compare a variety of designs. First, we consider possible designs for the cuboidal region in 3 factors. Then, we will consider some designs for the cuboidal region in 4 factors. The CCD design with $a = 1$ is considered. Another composite design, called the small composite design (SCD) by Hartley (1959), is similar to the CCD except that its factorial part is a fractional factorial design with resolution III. Also considered is the Plackett-Burman composite design (PBCD) by Draper (1985). It uses a Plackett-Burman design to replace the factorial portion of the design. Plackett-Burman designs (Plackett and Burman, 1946) are 2-level fractional factorial designs of resolution III for studying up to $k=n-1$ factors in n runs. Other designs considered are the Hoke (Hoke, 1974), Notz (Notz, 1982) and Box and Draper (Box and Draper, 1974) designs. The Hoke design has seven different types, D1-D7, for any number of factors, k. D1 through D3 have fewer runs than D4 through D7. Here we only consider D2, and D6 for $k=3$, and D2 and D5 for $k=4$. The Notz design is a subset of a 3^k factorial design, while the Box and Draper design is a minimal design. In Example 3, we will compare two three-factor mixture designs (Cornell, 2002).

III.3.1 Example 1: Some 3-factor Designs on a Cubodial Region

When we compare the six 3-factor designs using an FDS plot for a second order model: the CCD $(n_c = 2)$, the SCD $(n_c = 2)$, the Notz, the Hoke D2, the Hoke D6, and the Box and Draper design, we find some interesting results. Figure III.2 shows the FDS curves for these 6 designs on a single graph.

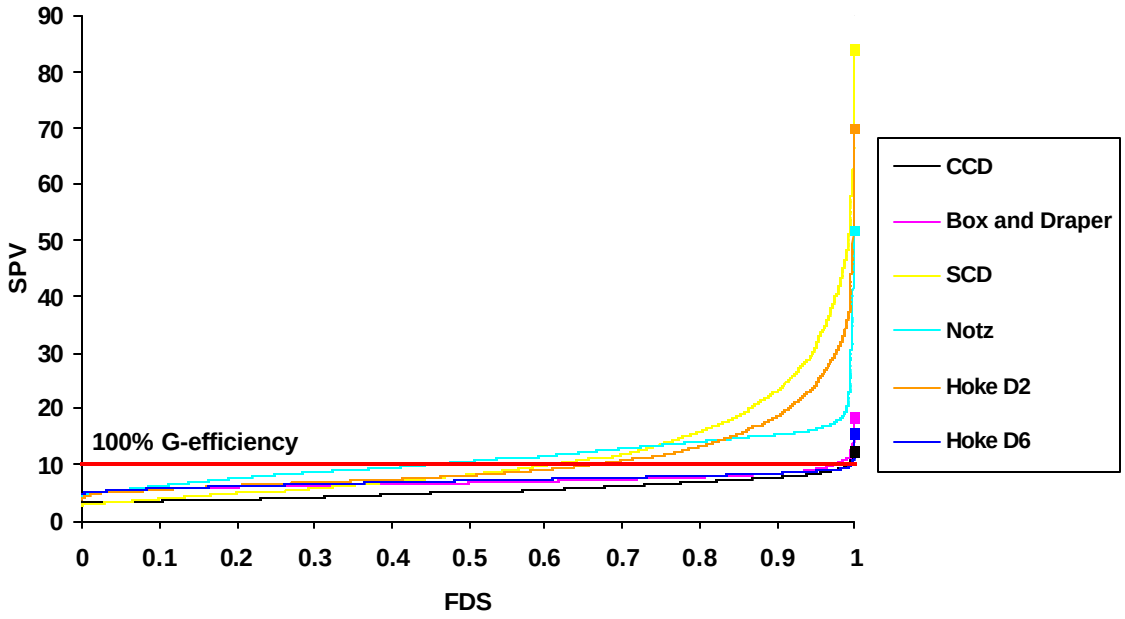

Figure III.2: Comparison of 3-factor Designs on a Cuboidal Design Space-I

Table III.1: Approximate G-efficiencies and V-averages for 3-factor Designs on a Cuboidal Region

Design	G-efficiency	V-average
CCD	78.5%	5.46
Hoke D6	65.1%	7.26
Box and Draper	54.3%	7.10
Notz	14.3%	10.99
Hoke D2	14.3%	10.59
SCD	11.2%	10.49

In Figure III.2, it is clear that three of the six designs have much flatter and lower FDS curves, indicating more desirable SPV distributions. The CCD, Box and Draper, and

Hoke D6 have the better prediction properties for the second order model. Table III.1 gives the approximate G-efficiencies and V-averages for all the designs. The CCD, Hoke D6, and the Box and Draper designs have higher G-efficiencies as well as low V-averages. Hence from now on, we consider only these designs.

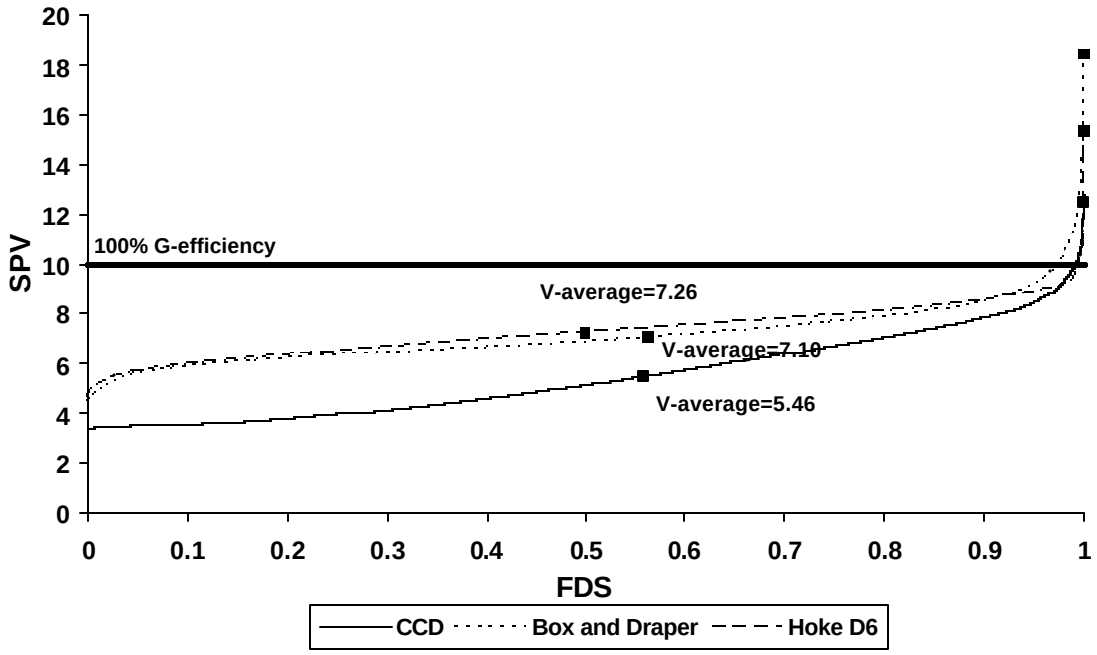

Figure III.3: Comparison of 3-factor Designs on a Cuboidal Design Space-II

Figure III.3 shows the rescaled FDS plot for these three designs. The CCD has lower SPV values for 97% of the total design space compared to the other two designs. The Box and Draper and the Hoke D6 designs have very similar SPV distributions. They are both uniform and flat except for close to the low SPV values and the large SPV values of the design space. Since CCD has the lowest SPV values for most of the design space, it has the lowest V-average value of 5.46. The V-average values for the Box and Draper and Hoke D6 designs are almost the same, which is expected given the similarity of their FDS plots. Examining the location of the V-averages, we find 58% of the total design space for the Box and Draper design has SPV less than or equal to 7.10, whereas 50% of the total design space for the Hoke D6 design has SPV less than or equal to 7.26. Even though these two designs have almost the same V-average values, the Box and Draper design has more design space at or below its average SPV. The maximum SPV for the

Box and Draper design is higher than the maximum SPV for the Hoke D6 design resulting in a lower G-efficiency of 54.3%. The CCD has the lowest maximum SPV resulting in the highest G-efficiency of 78.5%. However, the researcher should keep in mind the overall distribution of the SPV when comparing designs.

Now suppose that the second order model corresponds to the most complex model that the researcher feels might be needed to adequately model the data, but it is possible that there are some superfluous terms included in the model. Once the selected design has been run, individual terms can be tested and possibly removed if they are not enhancing the fit of the model. It should be noted that if a higher order model is actually required for fitting the underlying relationship, all of the models considered would likely not be able to even identify the lack of fit since they are designed to maximally estimate a second order model. Hence, for our purposes, model robustness is taken to mean that a design performs consistently well for all nested models within the largest one specified.

For the 3-factor second order model with 10 terms, there are a very large number of possible models nested within the full model. If we do not restrict ourselves to hierarchical models, then there are $2^9 = 512$ possible models (assuming that the intercept term is always included). If we only wish to consider hierarchical models, then there are still 95 possible models. As the number of factors considered in the experiment increases, the number of possible models grows at a dramatic rate. Clearly, any adaptation of the FDS plots to consider model robustness must streamline how to look at this aspect of the design.

Table III.2 gives a number of possible nested designs, which might be of interest when examining the 3-factor CCD. The corresponding FDS curves for these 7 models are shown in Figure III.4. Because of the symmetry of the CCD across all factors, these reduced models represent many of the hierarchical models that may be considered. For example, Model A with terms x_1, x_2, x_3, x_1x_2 would have the same FDS curve as x_1, x_2, x_3, x_1x_3 or x_1, x_2, x_3, x_2x_3, with the particular SPV values occurring in different locations in the design space. By comparing the curves we note an interesting feature of

the different designs: The first order model with interactions FDS curve has a very different slope from the remaining curves with similar numbers of terms. Hence, the addition of many interaction terms appears to have a much larger affect on the stability of the SPV values (as characterized by FDS curves that are closer to horizontal) than the addition of a similar number of quadratic terms.

Table III.2: Summary Information for the Oddly Reduced Models in Figure III.4 **(***k=3*)

Type of Model (*k=3*)	Terms in Model	# terms, *p* (G-optimal bound)	Maximum SPV
First Order Model	$x_1,\ x_2,\ x_3$	4	5.80
Model A	$x_1,\ x_2,\ x_3,\ x_1x_2$	5	7.80
First Order Model with Interactions	$x_1,\ x_2,\ x_3,\ x_1x_2,\ x_1x_3,\ x_2x_3$	7	11.80
Model B	$x_1,\ x_2,\ x_3,\ x_1x_2,\ x_1^2$	6	8.40
Model C	$x_1,\ x_2,\ x_3,\ x_1^2,\ x_3^2$	6	8.29
Model D	$x_1,\ x_2,\ x_3,\ x_1x_2,\ x_1^2,\ x_3^2$	7	10.29
Second Order Model	$x_1,\ x_2,\ x_3,\ x_1x_2,\ x_1x_3,\ x_2x_3,\ x_1^2,\ x_2^2,\ x_3^2$	10	12.73

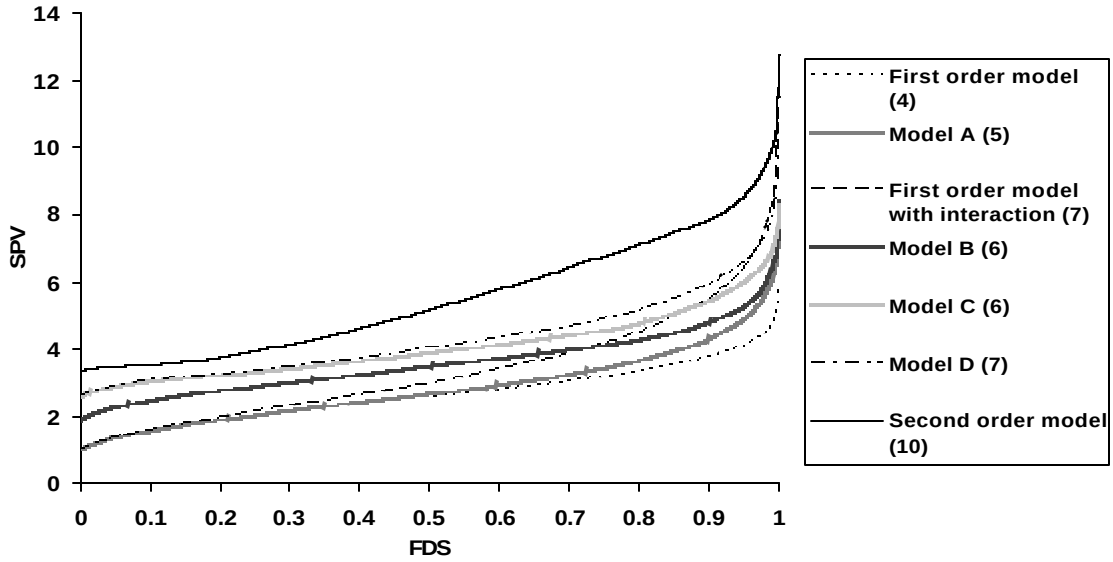

Figure III.4: FDS Plot for a 3-Factor CCD ($n_c = 2$) for Several Models (The legend gives the number of parameters for each model.)

Already just these 7 curves are somewhat overwhelming to process, so we note a number of general characteristics of the FDS curves that should allow us to further streamline the design comparisons. First, we note that as we reduce the model, the G-optimal bound is automatically reduced, since it is tied to the number of parameters in the model. Secondly, any model that is nested within another must have an FDS curve that is at or below the larger model for all fractions of the design space. This is a direct result of the form of the SPV, which means that adding terms to the model can only increase its SPV value. This is related to the classic bias-variance trade-off of model selection, where too small a model can suffer from bias problems by not being able to adequately describe the underlying relationship, while too large a model can suffer from inflated variance of prediction.

In some applications, there may be a natural subset of models to include in the FDS plot, in which case these "natural models" should provide the basis for the plot. However, in many applications, the researcher may have no idea about which of the many candidate models are likely to result. For these cases we suggest looking at a few common classes of models on a single FDS plot: the second order, the first order with interaction, and the first order models. By looking at Figure III.4, we see that these three models (the lines with symbols) form an "envelope" around the other models considered, and give a manageable summary of the prediction performance of the design across a broad class of models. By examining the plot in more details we note a few interesting characteristics: The maximum SPV values for the second order model and the first order model with interaction are 12.73, and 11.8 respectively. However, the G-efficiencies associated with them are far different, because of the number of associated parameters. The G-efficiencies of the design for these two models are 78.5% (10/12.73) and 59% (7/11.8), respectively.

Figure III.5 shows the comparison between the CCD and Hoke D6 designs, which were the top performers for the full second order model. As noted previously, the FDS curve for the second order model for the CCD is significantly lower than for the Hoke D6 design for most of the design space. For both the first order and first order with

interaction models, the Hoke D6 design performs better for most of the design space. Only the maximum SPV of the Hoke D6 is higher than the maximum SPV of the CCD for the first order model with interaction resulting in a lower G-efficiency. From Table III.3 which gives the G-efficiency and the V-average values for each design, it appears that if the best model turns out to be the second order model then the CCD would be the best choice. However, if many of the terms are not significant and a first order or first order model with interaction is suggested, the Hoke D6 may be more advantageous. Hence our recommendation of a preferred model may depend on how likely the experimenter thinks removing non-significant terms will be.

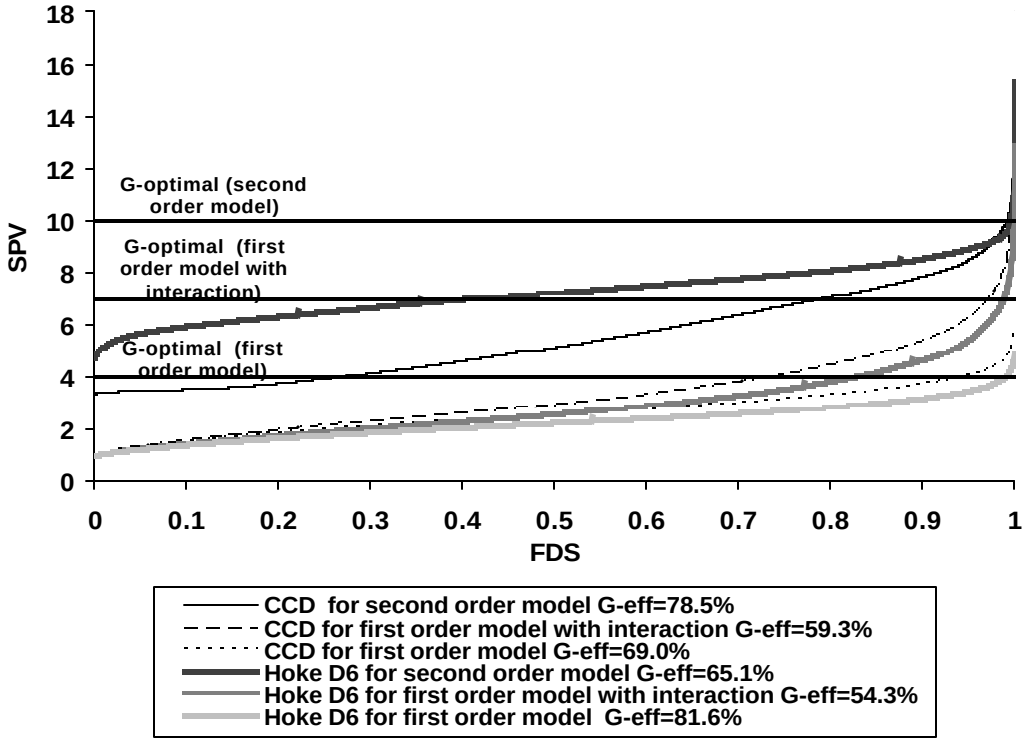

Figure III.5: FDS Plot for Comparing the CCD and the Hoke D6 for Several Model Types ($k=3$)

Table III.3: The Approximate G-efficiencies and the V-average Values for the CCD and the Hoke D6 for Several Models ($k=3$)

Design	Model Type	G-efficiency (Max SPV value)	V-average
CCD with 2 Center Runs	Second Order Model	78.50% (12.73)	5.46
	First Order Model with Interaction	59.30% (11.80)	3.26
	First Order Model	69.00% (5.80)	2.60
Hoke D6	Second Order Model	65.10% (15.36)	7.26
	First Order Model with Interaction	54.30% (12.90)	2.89
	First Order Model	81.60% (4.90)	2.30

Since, each of the models may be associated with a different number of terms, making comparisons relative to the optimal G-bound, p, for each model becomes difficult. Scaling each of the FDS curves by their corresponding G-optimal bound for each model will help the experimenter compare the models on a similar scale. Figure III.6 shows the G-scaled version of the FDS plot for the same CCD, where each curve graphs $\frac{SPV}{p}$.

Hence, a G-optimal design will have maximum value at FDS=1 of one. This approach allows us to consider model robustness most directly. If all of the curves associated with different models for a particular design are close together on the G-scaled FDS plot, then the design is robust to the models considered and the experimenter can expect similar prediction performance relative to the G-optimal bound.

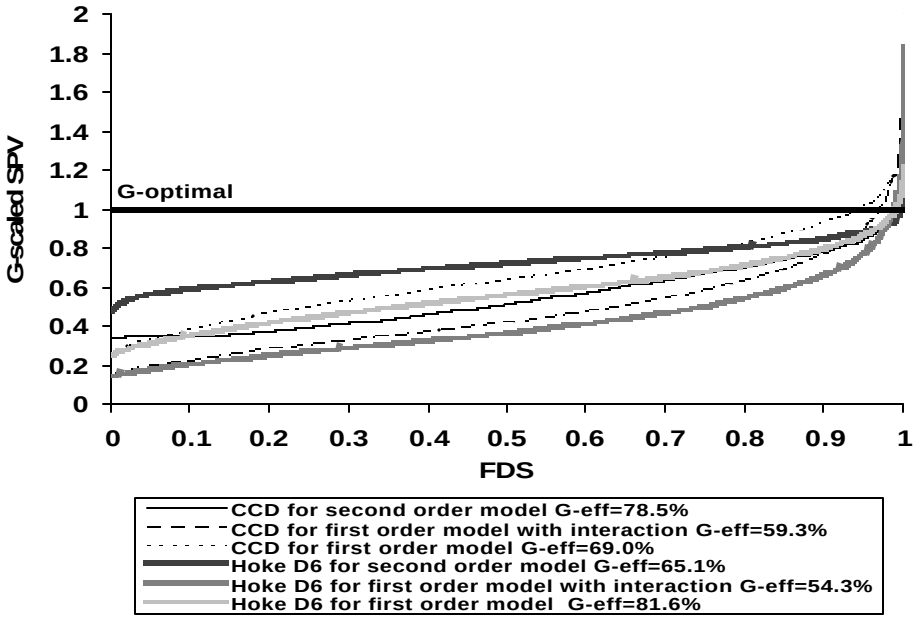

Figure III.6: G-scaled FDS Plot for Comparing the CCD and the Hoke D6 for Several Model Types ($k=3$)

For the CCD design in Figure III.6, we note that the new scaling has changed the order of the curves. The second order model no longer has the highest SPV value towards the end of fraction of the design space, because the SPV values for the second order model are all scaled by 10. In fact it has the most stable prediction throughout the region – a characteristic which is typically sought when choosing a design. For both designs, the first order model with interaction now has the lowest scaled-SPV values for most of the design space, but then has the largest extreme G-scaled SPV values. By examining the new adaptations of the FDS plots, we are able to gain more detailed insights into the relative performance of the various designs for a number of nested models

III.3.2 Example 2: Some 4 Factor Designs on a Cuboidal Region

Now, we compare some 4-factor second order designs on a cuboidal region: the CCD $(n_c = 4)$, the SCD $(n_c = 4)$, the PBCD $(n_c = 4)$, the Notz design, the Hoke D2 and the Hoke D5 designs, and the Box and Draper design. Table III.4 shows the G-efficiencies and V-averages for the various designs. In this case, selecting the top two or three designs is more difficult because of the differences in performance based on the G- and V-

criteria. The SCD design does not compete well based on either of the prediction criteria, while the CCD dominates both. Based on the V-averages, the next two best designs are the PBCD and Hoke D5, while for G-efficiency the better choices would be the Box and Draper, Notz and Hoke D2 designs. In this case, we assumed that the researcher in this case is more interested in the average SPV value throughout the design space, namely V-efficiency, and hence focus subsequently on the CCD, PBCD and Hoke D5 designs for model robustness.

Table III.4: The Approximate G-efficiencies and the V-average Values of the Designs in Figure III.7 for the Second Order Model ($k=4$)

Design	G-efficiency (Max SPV value)	N	V-average
CCD	69.80% (21.48)	28	8.64
PBCD	29.00% (51.76)	24	10.10
Notz	39.10% (38.33)	15	12.34
Hoke D2	39.10% (38.33)	15	12.34
Hoke D5	31.40% (47.69)	19	10.61
Box and Draper	53.30% (28.14)	15	16.39
SCD	10.80% (139.43)	20	19.69

We are able to compare the CCD, Hoke D5, and PBCD for model robustness in Figure III.7. The CCD performs the best for the second order model and the first order model with interaction. For the second order model, both the CCD and the PBCD perform similarly for small SPV values, but then the CCD maintains better stability for large values of SPV. For the first order model with interaction, the CCD and the Hoke D5 are quite similar in performance for the majority of the design space. The Hoke D5 has slightly lower SPV values than the CCD and the PBCD for the first order model, but for this model all three designs seem to perform quite similarly.

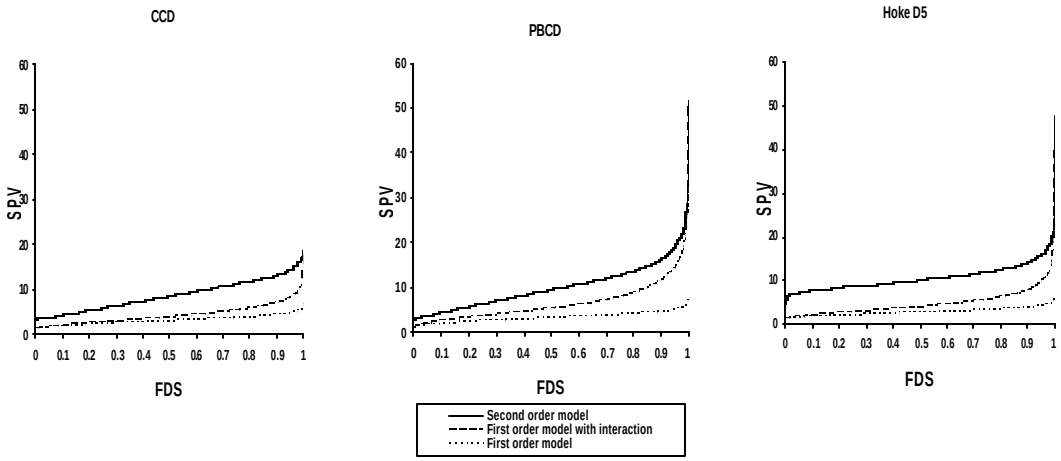

Figure III.7: FDS Plot for Comparing the CCD, the Hoke D5 and the PBCD for 3 Types of Models (*k=4)*

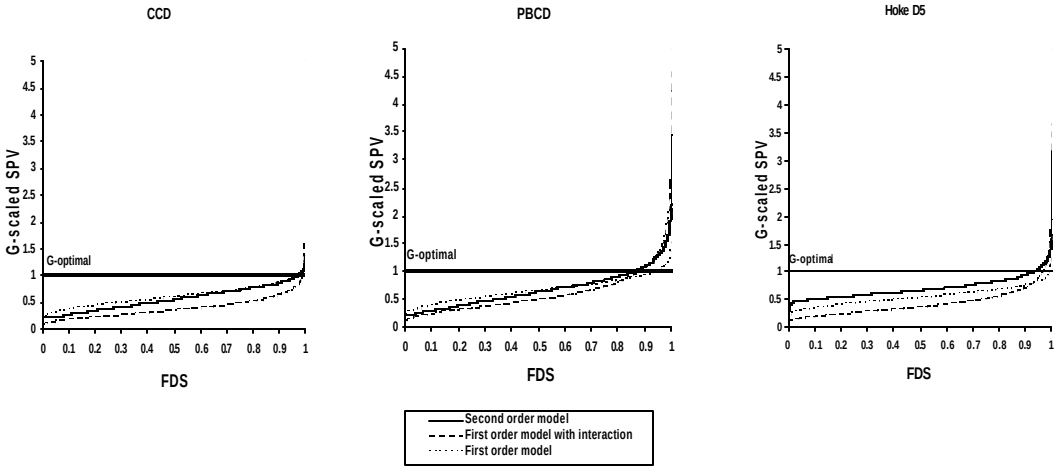

Figure III.8: G-scaled FDS Plot for Comparing the CCD, the Hoke D5 and the PBCD for Three Types of Models (*k=4*)

For the G-scaled FDS plot in Figure III.8, we note that for all three designs the first order model with interaction has lowest values for most of the design space but has dramatically worse maximal values for the PBCD and Hoke D5. For the CCD, the second order model has almost the same G-scaled SPV distribution as the first order model. For the majority of the design space, all three designs appear to have good robustness properties, while only the CCD maintains that stability for all three models at the largest

SPV values. In this case, the CCD appears to be a clear winner for the preferred design, since it has the best G- and V-efficiencies and has best or close to best SPV values for the nested models within the second order model.

If the experimenter cannot afford to use as many runs as the above three designs require and he/she does not mind the increase in the SPV, he/she may want to consider the Notz design and the Hoke D2 as well. Both of these two designs are small and saturated. Figure III.9 shows the FDS curves for the Notz design and the Hoke D2 for the three types of models.

In Figure III.9, for any of the three models, the SPV curves are almost the same. Therefore, depending on which design points the experimenter is allowed to use in the experiment, he/she may be able to choose one design over the other.

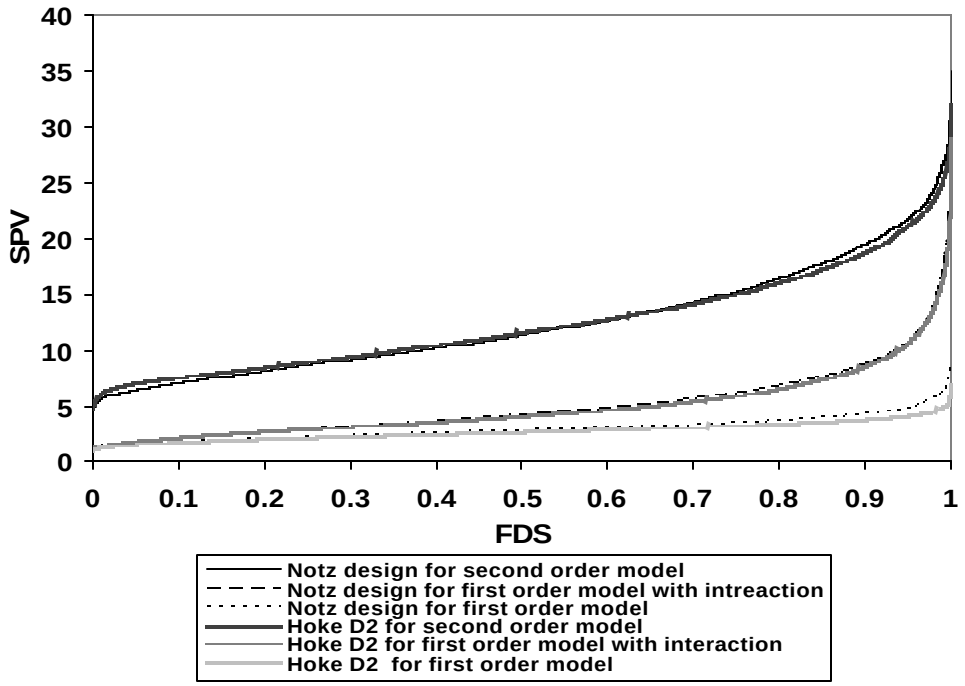

Figure III.9: FDS Plot for Comparing the Notz and Hoke D2 Designs for Three Types of Models ($k=4$)

III.3.3 Example 3: Comparing Mixture Designs

We now consider another type of response surface design, a mixture design (Cornell, 2002). A mixture experiment is an experiment where the factors are the ingredients or the components of a mixture, and the response depends on the relative proportions of the components. The levels of the components are dependent. A mixture experiment with q components, where x_i is the proportion of the i^{th} component, must satisfy the following constraints:

$$0 \le x_i \le 1 \qquad \forall i = 1,2,...,q \tag{III.5}$$

$$\sum_{i=1}^{q} x_i = 1 \tag{III.6}$$

The first constraint keeps each mixture component proportion between 0% and 100%, and the second constraint makes sure that for any point in the mixture space, the total sum of the proportions of all the components adds up to unity.

The form of the mixture model is different from the general polynomials used in the response surface methodology due to the $\sum_{i=1}^{q} x_i = 1$ constraint. The second order canonical mixture model is

$$h = \sum_{i=1}^{q} b_i x_i + \sum_{i<j} \sum b_{ij} x_i x_j \,, \tag{III.7}$$

where b_i corresponds to the expected response at the pure component i, $x_i = 1, x_j = 0, j \neq i$. The term $b_{ij} x_i x_j$ corresponds to the quadratic terms needed in addition to the basic linear model. This model is also known as the second order Scheffe′ (1958) model. The full cubic polynomial model is

$$h = \sum_{i=1}^{q} b_i^* x_i + \sum_{i<j} \sum^{q} b_{ij}^* x_i x_j + \sum_{i<j}^{q} \sum d_{ij} x_i x_j (x_i - x_j) + \sum_{i<j<k} \sum \sum^{q} b_{ijk}^* x_i x_j x_k \,. \tag{III.8}$$

A special case of the cubic polynomial called the special cubic is

$$h = \sum_{i=1}^{q} b_i^* x_i + \sum_{i<j} \sum^{q} b_{ij}^* x_i x_j + \sum_{i<j} \sum \sum^{q} b_{ijk}^* x_i x_j x_k \,. \tag{III.9}$$

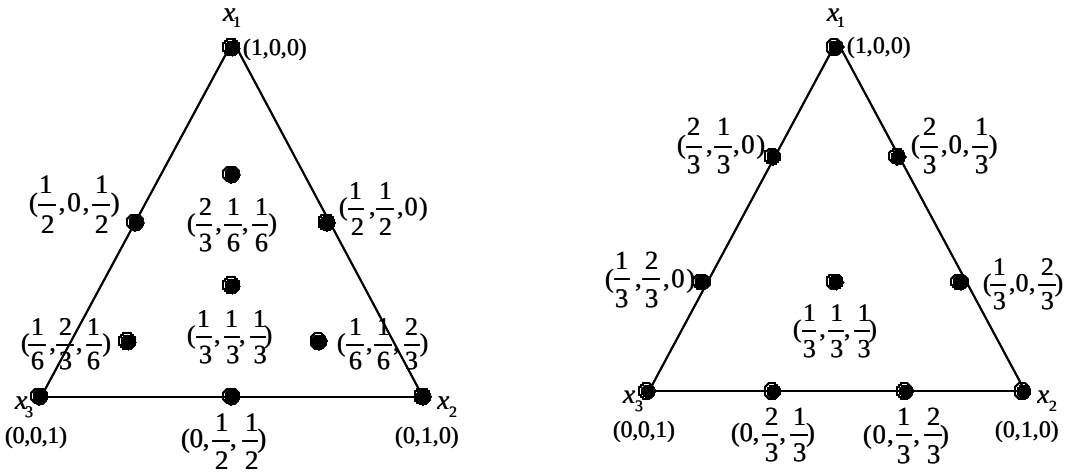

Figure III.10: a) Simplex-centroid Augmented with Interior Points ($k=3$) b) {3, 3} Simplex Lattice ($k=3$)

Figure III.10 shows two possible designs, each with 10 design points, for the design space of a 3-factor mixture experiment. The first in Figure III.10.a is a simplex-centroid design augmented with three interior points and maximally supports the fitting of a special quartic model, $h = \sum_{i=1}^{q} b_i^* x_i + \sum_{i<j}^{q} \sum b_{ij}^* x_i x_j + \sum_{i<j<k} \sum \sum \sum b_{iijk}^* x_i^2 x_j x_k$ with 9 parameters. A second possible design, shown in Figure III.10.b is the {3, 3} simplex lattice. See Cornell (2002, page 23) for more details. This design maximally supports the fitting of a general cubic model with 10 parameters. However, the highest order model estimable by both models is the special cubic. The lack of fit degrees of freedom available for the two models would estimate different types of model misspecification.

Goldfarb, Anderson-Cook, Borror and Montgomery (2004) introduced the FDS plots for mixture and mixture-process designs. Figure III.11 shows the FDS plot for each of the two designs for three types of models: the special cubic, the second order model, and the first order model. A large sample of locations over the mixture simplex was generated from a Dirichlet distribution to obtain the SPV values. For more information about the sampling method see Rubenstein (1982). The G-optimal bounds for the first, the second,

and the special cubic models are 3, 6 and 7, respectively. The simplex-centroid performs better than the simplex lattice for the entire design space for the special cubic model. Especially, for large SPV values, the simplex-centroid is superior to the simplex lattice. However, for the first order model, the simplex lattice has lower SPV values for the whole design space. For the second order model, the curves for the two designs intersect at the 43rd percentile. Below this level, the simplex-centroid performs better, and above this level, the simplex lattice has lower SPV values.

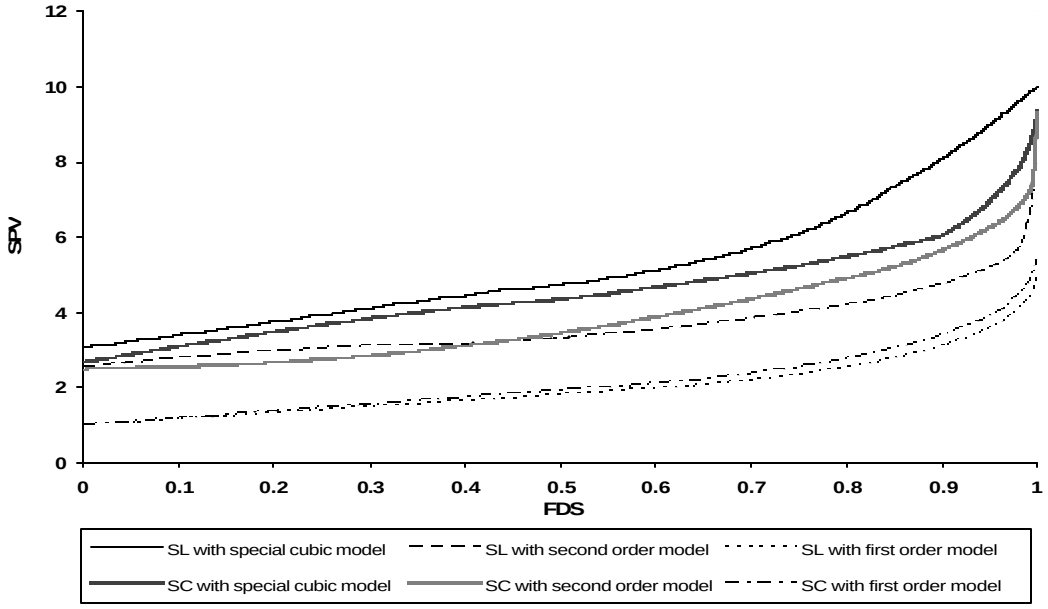

Figure III.11: FDS Plot for Comparing Two Mixture Designs ($k=3$): the {3, 3} Simplex Lattice (SL) and the Simplex-centroid Augmented with Interior Points (SC)

Table III.5: The Approximate G-efficiencies and the V-average Values for the Mixture Experiments for Several Models

Design type	Model type	G-efficiency (Max SPV value)	V-average
Simplex Lattice	Special Cubic	70.00% (10.00)	5.24
	Second Order	67.70% (8.86)	3.60
	First Order	60.00% (5.00)	1.99
Augmented Simplex-centroid	Special Cubic	74.90% (9.34)	4.56
	Second Order	64.50% (9.30)	3.82
	First Order	55.10% (5.44)	2.12

Table III.5 gives additional information about the two designs' G-efficiencies and their V-averages.

Cornell (1986) provided a detailed comparison of these two designs. The paper noted that both designs estimate the special cubic, quadratic and first order models well, with differing emphases of what the lack of fit terms are measuring for each design. The FDS plot results match with his findings in terms of the average variance of $\hat{y}(\boldsymbol{x}_o)$ (V-average). The overall conclusion of Cornell (1986) was that the augmented simplex-centroid is a better choice based on several diverse criteria. It is quite clear from Figure III.11, that the augmented simplex-centroid has a notable advantage over the simplex lattice based on the special cubic SPV values. However, for reduced models such as the second or first order models, that pattern is reversed. Hence, the FDS plots allow for a more detailed examination of the SPV values for the classes of possible models common to both designs and enhance our ability to assess two competitive designs.

Figure III.12 shows the G-scaled SPV values for both designs. For the simplex lattice, the second order model and the first order model intersect. The second order model performs better for most of the design space except for the initial fractions compared to the first order model. The special cubic model still has the highest G-scaled SPV values for most of the design space except for the last 0.02 fraction where the first order model has the worst scaled-SPV values. For the simplex-centroid, even though the special cubic model may have lower scaled-SPV values towards the latter fractions of the design space, eventually it reaches the high scaled-SPV values of the second order model. In this case, the simplex-centroid design appears to be the more robust design, since all three lines for the special cubic, second, and first order models are most similar. The simplex lattice design performs best for the second order model for large values of SPV, but the design performance depends more heavily on which final model is selected.

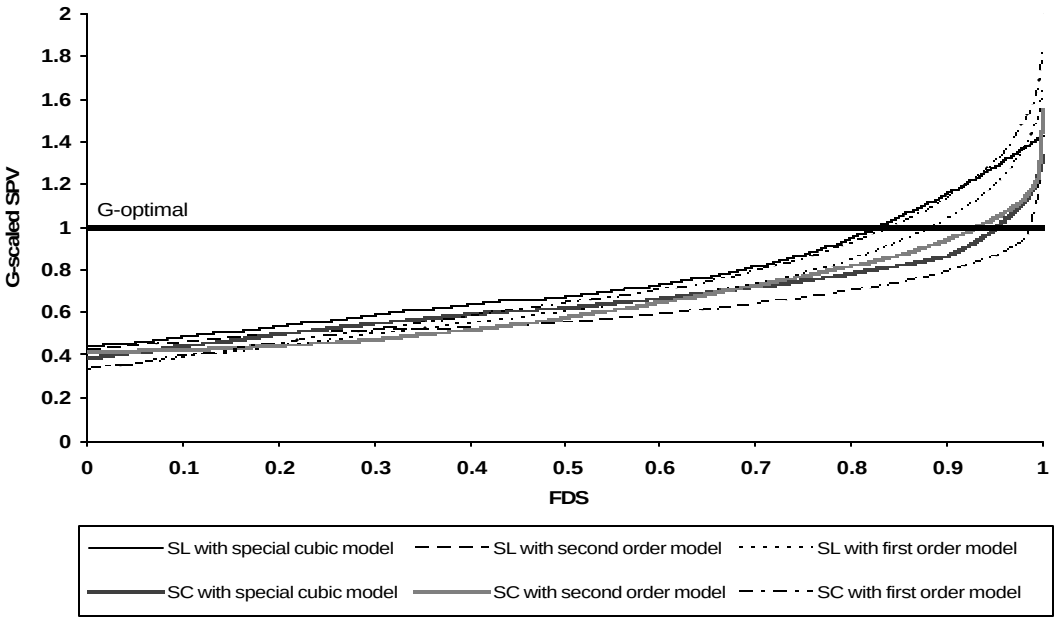

Figure III.12: G-scaled FDS Plot for the Two Mixture Designs ($k=3$): the {3, 3} Simplex Lattice (SL) and the Simplex-centroid Augmented with Interior Points (SC)

III.4 Conclusion

Since frequently the full model for which a design may have been chosen does not turn out to be the final model used in the analysis phase, considering the model robustness properties of a design can be an important consideration. Adaptations of the FDS plot can assist the researcher in choosing a design that will perform well for a broad range of nested models within the maximal model specified. In addition to looking at a curve's behavior, one can see the approximate G-efficiency and V-average value for that design with the specific model directly from the FDS plot.

By noting that nested models must have SPV ranges that are less than or equal to the larger model from which they came, looking at strategic subsets of models can make design assessment more manageable. In this chapter, we considered second order, first order with interaction and first order models as the primary subset models for standard factors in cuboidal regions, and special cubic, second order and first order models for the mixture experiment situation.

To examine model robustness most directly, we scaled the FDS plot values by the corresponding G-optimal value for each model. This approach helps the researcher to compare the models with different number of terms on a common scale and more accurately calibrate the desirability of the design.

When comparing three factor designs in a cuboidal region, the CCD design appears best for the full second order model based on both the G- and V-criteria, while if a reduced model is likely the Hoke D6 may be preferred. For the four-factor design in a cuboidal region, the CCD design appears best for both G- and V-efficiency for all the models considered.

When comparing three-component mixture designs (the augmented simplex-centroid and the simplex lattice), the simplex-centroid design appears to be more robust, since all three lines for the special cubic, second, and first order models are most similar. The simplex lattice design performs best for the second order model for large values of SPV, but design performance depends more heavily on which final model is selected.

In addition to reducing the order of the model, another option to study model robustness of a design is to consider reducing the number of factors. This would correspond to one or more of the factors in the experiment having no influence on the response either through linear, quadratic or interaction terms.

In this dissertation, only designs on a cuboidal region or mixture designs in unconstrained regions are considered. This approach would work for any design space and any set of nested models. The software for creating the FDS plots in Excel using Visual Basic is available by request from the author. See Appendix-A for examples of different outputs of the computer programs and the source codes.

Chapter IV

FRACTION OF DESIGN SPACE PLOTS FOR GENERALIZED LINEAR MODELS

IV.1 Abstract

Several methods have been developed for comparing the prediction variance properties of designs graphically such as the Variance Dispersion Graphs (VDGs), the Quantile Dispersion Graphs (QDGs), and the Fraction of Design Space (FDS) plots. Initially it was assumed that the form of the model was linear with respect to the parameters. In this chapter, FDS plots are adapted to evaluate designs for generalized linear models (GLMs). In such a situation, initial parameter estimates of the model need to be provided by the experimenter in order to estimate the prediction variance. Since if the parameters were already precisely known, there would be no need to conduct the experiment, the question of robustness to misspecification of the initial parameter estimates is an important one. FDS plots provide a graphical way of assessing the relative merits of different designs under a variety of types of parameter misspecification. These new plots are studied for the logistic regression and the Poisson regression models with canonical links. Examples are given in which several response surface designs are compared using the FDS plots. In addition, a new prediction quantity called the penalized prediction variance (PPV) is introduced. While for linear models the SPV has both information about actual prediction and also the theoretical optimum of designs, two separate functions of the prediction variance, PPV and SPV, respectively, are needed for the GLM case.

IV.2 Introduction

In this chapter, we suggest a new adaptation of the graphical tool, Fraction of Design Space (FDS) plots, by Zahran, Anderson-Cook and Myers (2003) for generalized linear models (GLMs). A new quantity, the penalized prediction variance (PPV) is defined which gives insights about actual prediction. This complements the SPV which gives information about the G- and V-optimality of the designs. We plot the SPV and the PPV values using the FDS plots to compare response surface designs for the GLM case.

Several graphical methods have been developed in the literature to study a design's prediction variance properties. Among these are the Variance Dispersion Graphs (VDGs) by Giovanitti-Jensen and Myers (1989), the three-dimensional VDGs by Goldfarb, Borror, Montgomery and Anderson-Cook (2004), the Quantile Dispersion Graphs (QDGs) introduced by Khuri, Kim and Um (1996), and the Fraction of Design Space Plots introduced by Zahran, Anderson-Cook and Myers (2003), and the FDS Plots for mixture and mixture-process designs by Goldfarb, Anderson-Cook, Borror and Montgomery (2004). Among the papers that used the VDGs to compare designs on a spherical region are Myers, Vining, Giovannitti-Jensen and Myers (1992), Borkowski (1995), and Block and Mee (2001). Designs on a cuboidal region were compared using the VDGs as well. See Borkowski (1995), Rozum and Myers (1991), Block and Mee (2001), Myers and Montgomery (2002), Borror, Montgomery and Myers (2002), and Park, Richardson, Borror, Ozol-Godfrey, Anderson-Cook, and Montgomery (submitted) for more information. Piepel and Anderson (1992), and Piepel, Anderson and Redgate (1993 a, b) developed VDGs for mixture experiments.

For the standard linear model of the form $\boldsymbol{y} = X\boldsymbol{b} + \boldsymbol{e}$, the SPV is defined as

$$n(\boldsymbol{x}_o) = \frac{N \operatorname{var}(\hat{y}(\boldsymbol{x}_o))}{\boldsymbol{s}^2} = N\, \boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o \,, \qquad \text{(IV.1)}$$

where $\boldsymbol{x}_o$ corresponds to a point in the region of interest, $\operatorname{var}(\hat{y}(\boldsymbol{x}_o))$ is the variance of the estimated response at $\boldsymbol{x}_o$, X is the design matrix expanded to model form, and *N* is the total number of runs. SPV is only a function of the number of runs, the model, the design space, the particular location in the design space, and the design. It corresponds to the prediction variance, multiplied by *N* to allow for comparisons between designs of different sizes, and divided by a constant, $\boldsymbol{s}^2$, since this is typically unknown until the data are collected.

FDS plot shows the fraction of the design space at or below any SPV value. The plot is constructed by randomly sampling a large number of values, say *n*, from throughout the design space and obtaining the corresponding SPV values. These *n* SPV values are then ordered and plotted against the quantiles (*1/n, 2/n, …*). The x-axis gives the quantiles of

the design space ranging from 0 to 1, while the y-axis shows the SPV values. For a given point on the curve, we can extract what fraction of the total design space has SPV values less than or equal to the given value. By looking at the two ends of the line, we can see the minimum and maximum SPV values for the design. See Zahran, Anderson-Cook and Myers (2003) for more details.

IV.3 Generalized Linear Models (GLMs)

When the response of interest has a distribution other than the normal distribution but belongs to the exponential family, generalized linear models (GLMs) are frequently used. The data might represent the number of defects in an assembly line, or it might be the number of injuries that occur in a coal field during a certain period (Myers, Montgomery and Vining, 2002, Hamada and Nelder, 1989). Two members of the exponential family have received significant attention in the literature: the binomial and the Poisson distributions. Suppose there is a regression structure with regressors, $x_1,.....,x_k$, and the response is binary (0 or 1). Observations of this nature arise, for example, in an industrial environment, where at the end of each independent experimental run, the item is deemed defective or not. Examples of Poisson data arise when the interest is in the number of counts of events that follow independent Poisson distributions. For example, the number of incidents involving damage to ships of a specified type over a given period of time (Hamada and Nelder, 1989). In the case of industrial experiments, quite often the experimenter has a notion of the region of operability. He/she has a prior knowledge about the restrictions that should be used for the x_i's, to yield information of interest.

The members of the exponential family of distributions have probability density functions of the following general form:

$$f(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\} \qquad \text{(IV.2)}$$

where $a(\cdot), b(\cdot), \text{and } c(\cdot)$ are specific functions. θ is called the natural location parameter and ϕ is the dispersion parameter. $a(\phi)$ is a function of dispersion, $b(\theta)$ is a function of the location parameter and $c(\cdot)$ is a function of ϕ and y. The probability density

function for the Poisson distribution is given by $\exp[yq - e^q - \ln(y!)]$ where $a(f) = 1$, $b(q) = e^q$ and $c(y,f) = -\ln(y!)$. The probability density function for the binomial distribution is given by $\exp[yq - n\ln(1+e^q) + \ln\binom{n}{y}]$ where $a(f) = 1$, $b(q) = n\ln(1+e^q)$ and $c(y,f) = \ln\binom{n}{y}$. See Myers, Montgomery and Vining (2002) for additional information. For a generalized linear model, the observations, $y_1,.....,y_n$, are assumed to be independent with means $m_1,.....,m_n$, respectively. Some function of the mean, $h_i = \boldsymbol{x}_i'\boldsymbol{b}$ is the linear predictor of $E(y_i) = m_i$. The link function, g, which is a monotonic differentiable function, is used to define the model, i.e., $h_i = g(m_i)$. If the link function is equal to the parameter, q_i, it is called the canonical link function, which we assume throughout the remainder of this chapter. For the binomial distribution, the canonical link function is the logistic link defined as $h_i = \ln\left(\frac{P_i}{1-P_i}\right)$. For the Poisson case, it is the log link defined as $h_i = \ln(m_i)$. We focus on the canonical link because it is an important class of link functions and is commonly used.

IV.4 Scaled Prediction Variance (SPV) and Penalized Prediction Variance (PPV)

When we consider how to study prediction variance for the GLM case, two natural measures of the quality of prediction emerge. First, we consider the SPV and how to adapt it for generalized linear models. The SPV for the linear case is given by equation IV.1. Since the variance, s^2, is constant for any observation, it can be divided out of the equation leaving only the $\boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o$ term. In this case, the shape of the prediction is not influenced by removing the s^2 term, and means that the quantity can be evaluated before data are collected.

When the variance for different observations throughout the design space is not a constant, the form of the SPV becomes different. For the GLM case with canonical link, the prediction variance of the estimated response at location $\boldsymbol{x}_o$ is defined as

$$\text{var}(\hat{y}(\boldsymbol{x}_o)) = \text{var}(y(\boldsymbol{x}_o))^2 \boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o, \qquad \text{(IV.3)}$$

where $\text{var}(y(\boldsymbol{x}_o))$ corresponds to the variance of the observation at $\boldsymbol{x}_o$. See Myers, Montgomery and Vining (2002, page 111). $X'VX$ is a function of the design chosen, and V is defined as V = diag(s_i^2). s_i^2 is the variance at a design point $\boldsymbol{x}_i$ for $i=1,\ldots,s$ where s is the number of design points. The variance, s_i^2, is typically a function of the number of observations at that design point. When the variance of the estimated mean response is scaled by the specific variance of any observation and multiplied by N, the SPV for a GLM becomes

$$n(\boldsymbol{x}_o) = N \text{ var}(y(\boldsymbol{x}_o))\boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o \ . \qquad \text{(IV.4)}$$

$n(\boldsymbol{x}_o)$ is a function of the number of runs, the location, the design and now also the parameters in the model. Note that this scaling now involves dividing by different values for each location. For the GLM case, the G- and the V-optimal values are still in the correct scale as the SPV with the G-optimal value being p, the number of parameters in the model. This can be justified by using a transformation of the X matrix to return to the standard linear models case. See Appendix-B for more details. By considering the characteristics of the SPV, the researcher is still able to determine the G-efficiency and find the V-average value for a design.

Secondly, we introduce the penalized prediction variance (PPV) appropriate for the case of a GLM. The PPV for the standard linear model situation is defined as

$$r(\boldsymbol{x}_o) = N \text{ var}(\hat{y}(\boldsymbol{x}_o)) = N s^2 \boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o, \qquad \text{(IV.5)}$$

which is the prediction variance of the estimated mean response, $\text{var}(\hat{y}(\boldsymbol{x}_o))$, multiplied by the number of runs in the design. The distinction between a PPV and a SPV for a linear model is that the PPV is *not* scaled by dividing by s^2. Assuming s^2 is known, the researcher is able to study the actual prediction variance for any design point in the design space while still compensating for larger designs. For the linear model case, the PPV is just a fixed multiple of the SPV, so the shape and stability of both of these for a given design will be similarly conveyed by either quantity. Since s^2 is generally

unknown and would need to be estimated for the linear models case, SPV is usually used to obtain information about a design's prediction properties.

The PPV for the GLM case is important because the researcher is able to study the effects of initial parameter estimates on the actual prediction variance before the data are collected.

The PPV for GLM is defined as

$$r(\boldsymbol{x}_o) = N \text{ var}(\hat{y}(\boldsymbol{x}_o)) = N \text{ var}(y(\boldsymbol{x}_o))^2 \boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o, \qquad \text{(IV.6)}$$

where $\text{var}(y(\boldsymbol{x}_o))$ is the variance of an observation at location $\boldsymbol{x}_o$. V is defined as before and both V and $\text{var}(y(\boldsymbol{x}_o))$ are dependent on the initial parameter estimates. Therefore PPV is dependent on the parameter estimates.

We consider two specific cases of GLMs: logistic and Poisson regression models. In the logistic regression case, the observations are assumed to be from a binomial distribution, and the two possible outcomes can be thought to represent success and failure. When we use the canonical link, the model is written as $p(\boldsymbol{x}_i) = \frac{1}{1+e^{-\boldsymbol{x}_i'\boldsymbol{\beta}}}$ where $\boldsymbol{x}_i'\boldsymbol{\beta}$ is the linear predictor, and $p_i = p(\boldsymbol{x}_i)$ is the probability of success or failure at location $\boldsymbol{x}_i$.

The SPV becomes

$$n(\boldsymbol{x}_o) = N[p_o(1-p_o)]\boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o, \qquad \text{(IV.7)}$$

where V = diag(s_i^2) = diag($n_i p_i(1-p_i)$) for $i=1,\ldots s$. Here the SPV is a function of the unknown parameters, b 's, as $p(\boldsymbol{x}_i) = \frac{1}{1+e^{-\boldsymbol{x}_i'\boldsymbol{\beta}}}$.

The PPV is

$$r(\boldsymbol{x}_o) = N[p_o(1-p_o)]^2 \boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o. \qquad \text{(IV.8)}$$

The difference between the SPV and the PPV is that the variance portion is squared for PPV.

Another common GLM case is the Poisson regression model. The data are modeled with the canonical log link as $\ln(m_i) = \boldsymbol{x}_i'\boldsymbol{ß}$, where $m_i = e^{x_i'ß}$ is the expected count at location $\boldsymbol{x}_i$. The SPV becomes

$$n(\boldsymbol{x}_o) = N\boldsymbol{\mu}_o\boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o \qquad \text{(IV.9)}$$

where V=diag(s_i^2)=diag(m_i) making it a function of the parameters. The PPV is

$$r(\boldsymbol{x}_o) = N\boldsymbol{\mu}_o^2\boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o. \qquad \text{(IV.10)}$$

The SPV and the PPV for the GLM case are both dependent on the unknown parameters, b's, through the var($y(\boldsymbol{x}_o)$) and the V terms in the formulas.

IV.5 Obtaining the FDS Plot for the GLM Case

To obtain the FDS plot for a design with a GLM, the researcher first needs to specify the model of interest, the design space, the number of runs, and the initial parameter estimates. Using this information, he/she can calculate the SPV throughout the design space to produce the FDS plot. For example, consider a manufacturing company which produces plastic molding. They are interested in constructing an experiment with an appropriate design for a model to calculate the probability of a non-defect. As a function of two factors x_1 (temperature), and x_2 (time), and their interaction. They consider a 2^2 factorial design in a square region coded so that each variable is in the range [-1, 1]. A first order model with interaction in the linear predictor with the logistic link has the form

$$\ln\left(\frac{P_i}{1-P_i}\right) = b_o + b_1x_{1\,i} + b_2\,x_{2\,i} + b_{12}x_{1}x_{2i}, \qquad \text{(IV.11)}$$

which gives $p_i = \dfrac{1}{1+\mathrm{e}^{-(b_o + b_1x_{1\,i} + b_2\,x_{2\,i} + b_{12}x_{1}x_{2i})}}$ for the probability of a non-defect at a given location. The practitioner is asked to estimate the probability of non-defects at each of the corners and gives (0.93, 0.60, 0.95, 0.88) for {(-1, -1), (-1, 1), (1, -1), (1, 1)}, respectively. Suppose the design with equal number of observations at each design location is chosen. For simplicity of discussion assume there are 5 observations at each corner. Then the design space with observations is shown in Figure IV.1.

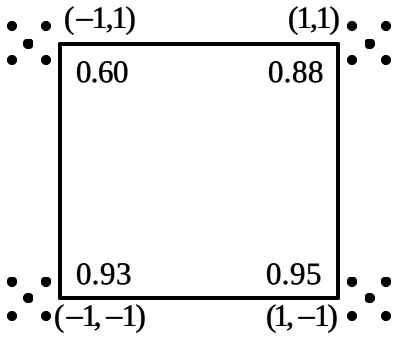

Figure IV.1: A 2^2 Factorial Design with 5 Observations at Each Design Point

From the practitioner's estimate above, we can calculate the initial parameter estimates as $\hat{b}' = (\hat{b}_0, \hat{b}_1, \hat{b}_2, \hat{b}_{12})' = (2, 0.5, -0.8, 0.3)'$. Figure IV.2 shows the FDS plot using SPV for this design in the square region. This factorial design is a G-optimal design since the obtained maximum SPV (shown at FDS value equal to 1) is equal to the lower bound for the maximum SPV of 4, equal to the number of parameters in the model. The SPV values for this design range between 1.29 and 4. The approximate V-average for this design is obtained by averaging all the sampled values, and it is equal to 2.09. The corresponding FDS value for this average is 0.58 indicating that the distribution is positively skewed.

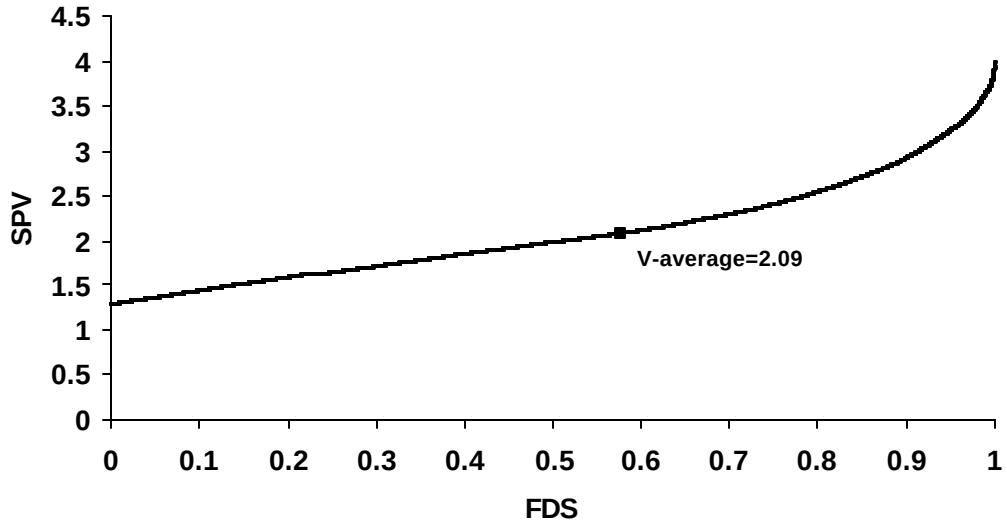

Figure IV.2: FDS Plot (Using SPV) for 2^2 Factorial Design for a Logistic Regression Model with Initial Parameter Estimate, $\hat{b}' = (2, 0.5, -0.8, 0.3)'$

Figure IV.3 shows the FDS plot using PPV for this design. The minimum and the maximum PPV values are 0.11 and 0.96, respectively. For most of the fraction of the design space, the slope of this curve is flatter than the curve of the FDS plot using SPV values. However, the larger PPV values increases at a much faster rate than the SPV.

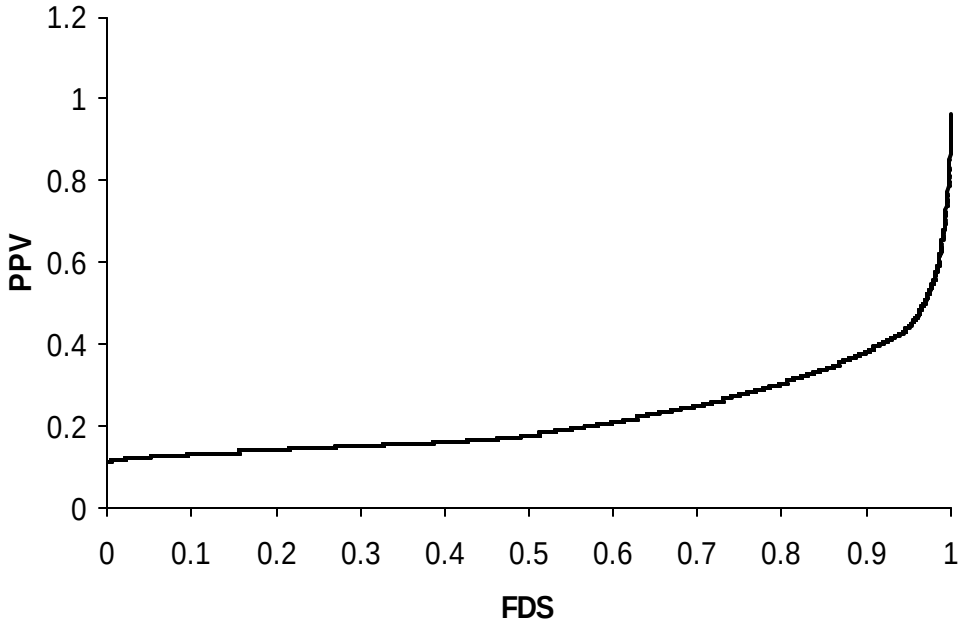

Figure IV.3: FDS Plot (Using PPV) for 2^2 Factorial Design for a Logistic Regression Model with Initial Parameter Estimate, $\hat{b}' = (2, 0.5, -0.8, 0.3)'$

When comparing two or more designs, the procedure remains the same as for the standard linear models case, except one still needs to estimate the initial parameter estimates for the GLM case. Figure IV.4 shows the FDS plot for comparing two designs. The first design, D1, is the same as in Figure IV.1. The second design, D2, has the same total size but a different allocation of points, with equal observations at each of the factorial locations and center runs (say 4 observations at each location). D2 has a slightly lower minimum SPV value than D1 for the given initial parameter estimates. This is due to inclusion of center runs for D2 which provide an improved estimate at the center. However, for most of the design space, D1 performs better than D2.

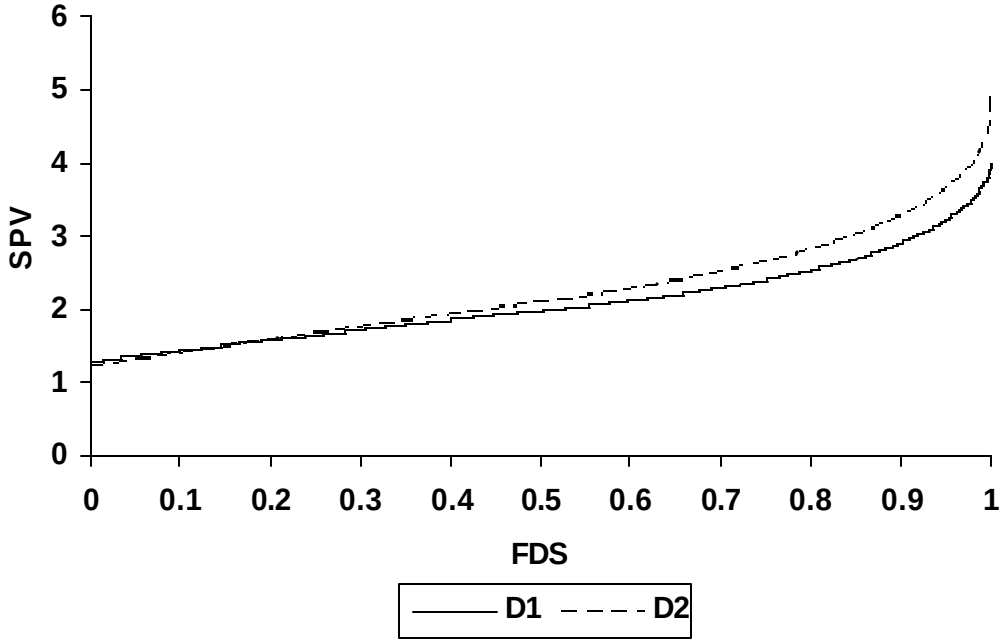

Figure IV.4: FDS Plot (Using SPV) for Comparing Two Designs for a Logistic Regression Model with Initial Parameter Estimate, $\hat{b}' = (2, 0.5, -0.8, 0.3)'$

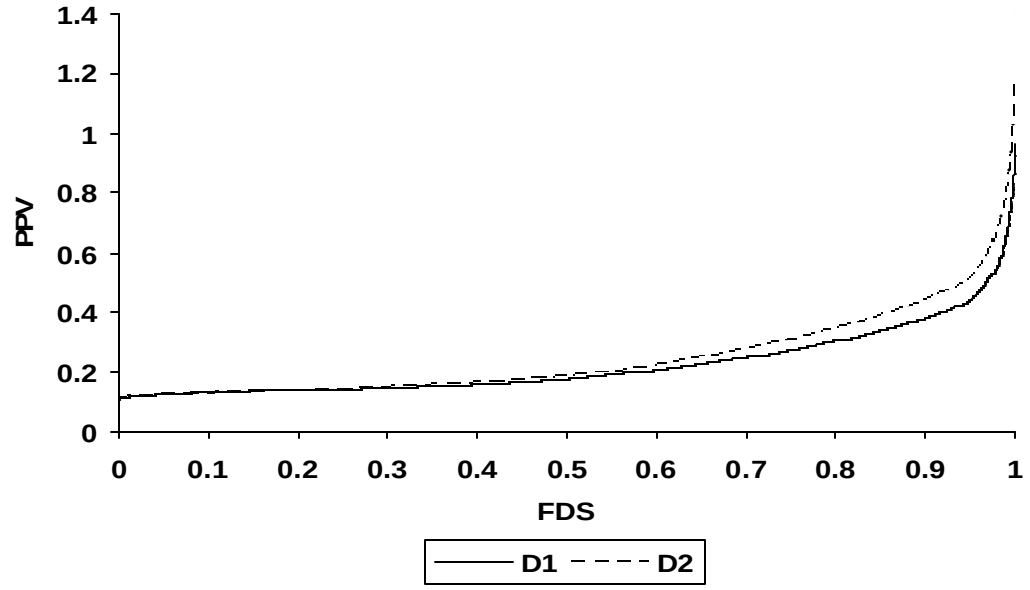

Figure IV.5: FDS Plot (Using PPV) for Comparing Two Designs for a Logistic Regression Model with Initial Parameter Estimate, $\hat{b}' = (2, 0.5, -0.8, 0.3)'$

Figure IV.5 shows the PPV values for both designs. D2 has approximately equal penalized prediction values compared to D1 for the 22% of the design space where prediction is best, but D1 is better for the remaining 78% of the total design space. The additional $\text{var}(y(\boldsymbol{x}_o))$ in the PPV changes the shape of the curves for both designs considerably showing that the ratio of best to worst prediction has been dramatically accentuated with the PPV form. This is important for the practitioner interested in actual prediction since the performance of D2 is considerably worse at the maximum PPV (=1.18) compared to the D1 value of 0.96.

IV.6 Comparison of Space Filling Designs to Standard Factorial Designs for the GLM Case Using FDS Plots

The GLM-adapted FDS plots can be used for comparing any type of chosen design. In this section, we compare some space filling designs to the standard response surface designs under the GLM framework. Space filling or exploratory designs are designs where the design points are chosen to fill the design space. In some experiments, space fillings designs may be preferred against the standard response surface designs. This might occur in higher dimensions or when the researcher is interested in fitting a very complex model to the data. In this dissertation, we compare some space filling designs to the standard factorial designs for the GLM case in terms of their prediction properties using FDS plots.

Space filling designs are generally preferred when there are restrictions on the number of runs that can be run and when the number of levels for each factor needs to be greater than two. These designs are commonly used for computer experiments with many factors and an assumed complex model.

There are two common types of space filling designs: the uniform design and the Latin hypercube design. Popular since 1980 (Fang, 1980), a uniform design is a space filling design obtained by comparing the distribution of its design points to the uniform distribution. The Latin hypercube design introduced by McKay, Beckman and Conover (1979) is a space filling design where each randomly permuted column in the design

matrix is randomly matched to form the design. Even though under a finite experimental domain, Latin hypercube and uniform designs may be similar, they can be quite different for higher dimensions. This is because Latin hypercube designs are balanced in only one-dimension for all levels of each factor. However, the uniform designs require uniformity and balance in all dimensions.

Both designs have been shown to be robust to model assumptions. Since both design types were intended to be used where the model form is very complex, we would not necessarily expect them to do well when the assumed model is only a first order polynomial.

In addition to the standard two factor designs D1 and D2 compared in Section IV.5, we consider four space filling designs: D3, D4, D5 and D6. All six designs were chosen to have 20 observations. D3 is a uniform design with 4 levels of each factor with replicates at different design locations. D4 is a Latin hypercube design with 5 replicates at each design location and 4 levels of each factor. D5 is another Latin hypercube design with 10 levels for each factor and 2 replicates at each design location. Last of all, D6 is another uniform design having 5 levels for each factor with replicates at different design locations. The design points of D3 and D6 were obtained from the UD-web (http://www.math.hkbu.edu.hk/UniformDesign). Figure IV.6 shows the design points for designs D3, D4, D5 and D6. The design locations for D6 fall on a circle inside the design region with an additional location in the center. D3 is uniformly distributed all over the design space. Latin hypercube designs D4 and D5's design points are not unique. These designs were chosen from many possible designs with similar characteristics where some may be better than others. As before, D1 and D2 are the factorial designs having design points at the four corners and D2 has additional runs in the center.

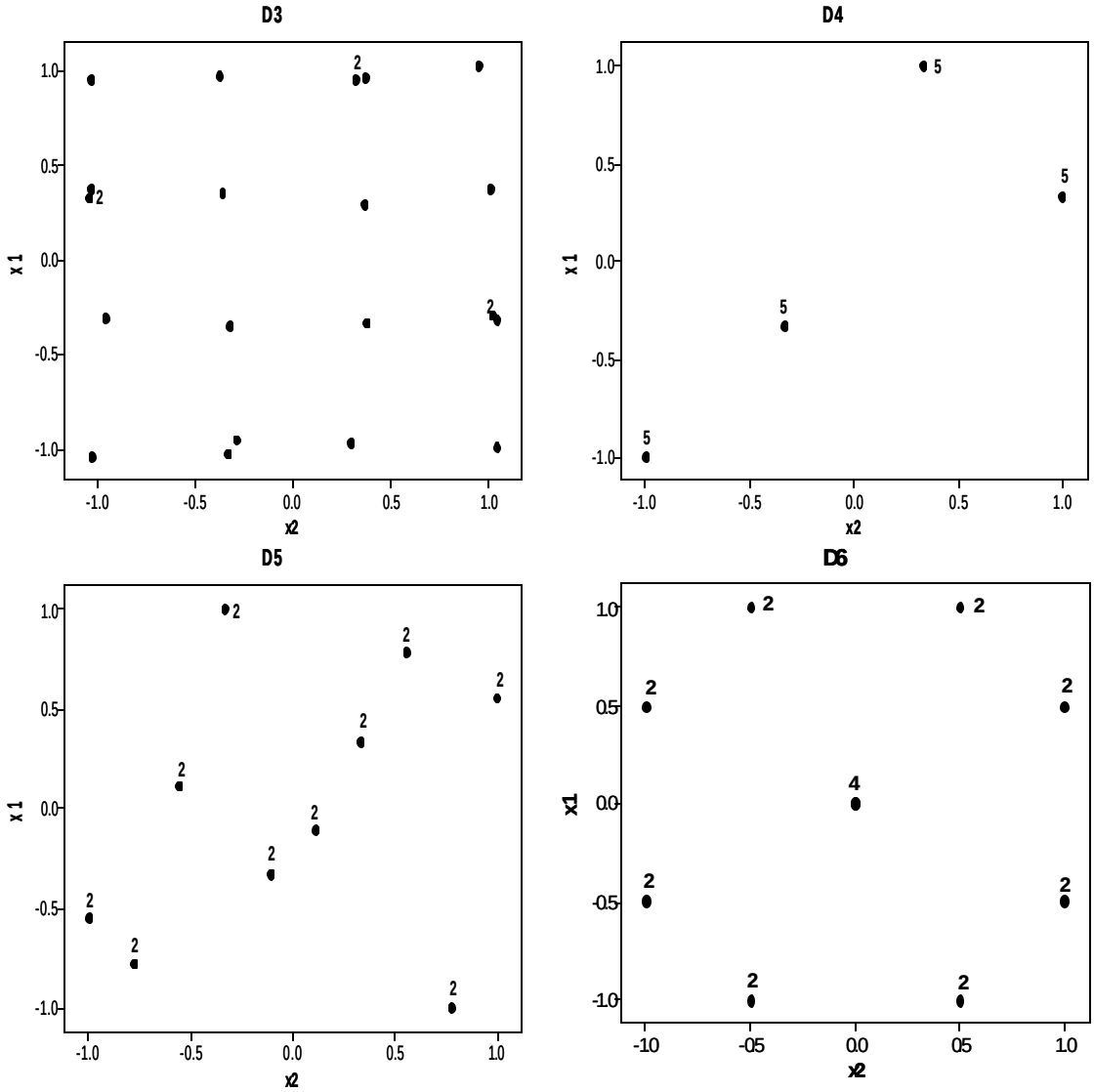

Figure IV.6: The Geometries of D3, D4, D5 and D6

Assuming the practitioner specified the same initial parameter estimates, $\hat{b}' = (2, 0.5, -0.8, 0.3)'$, Figure IV.7 shows the FDS plots using SPV values for five of the designs. The FDS curve for D4 has been omitted since D4 performed very poorly in terms of SPV values compared to the other designs. In addition, Table IV.1 shows the minimum and the maximum SPV and PPV values for each design. As seen from Figure IV.7, the standard response surface designs D1 and D2 performed considerably better compared to the other designs in terms of SPV values. D5, the second worst Latin hypercube design, has the highest SPV values for most of the design space. The uniform

designs D3 and D6 perform considerably better than the Latin hypercube designs but still have higher SPV values compared to the standard response surface designs. Looking at the distribution of the design points for each design, the FDS curves make sense. D4 has observations only at one corner and the rest of its design points are away from the corners resulting in large regions of the design space that are not close to any observations and lead to worse prediction at the corners. Similarly, the randomly sampled design points for D5 are away from the corners inflating the maximum SPV value of the design.

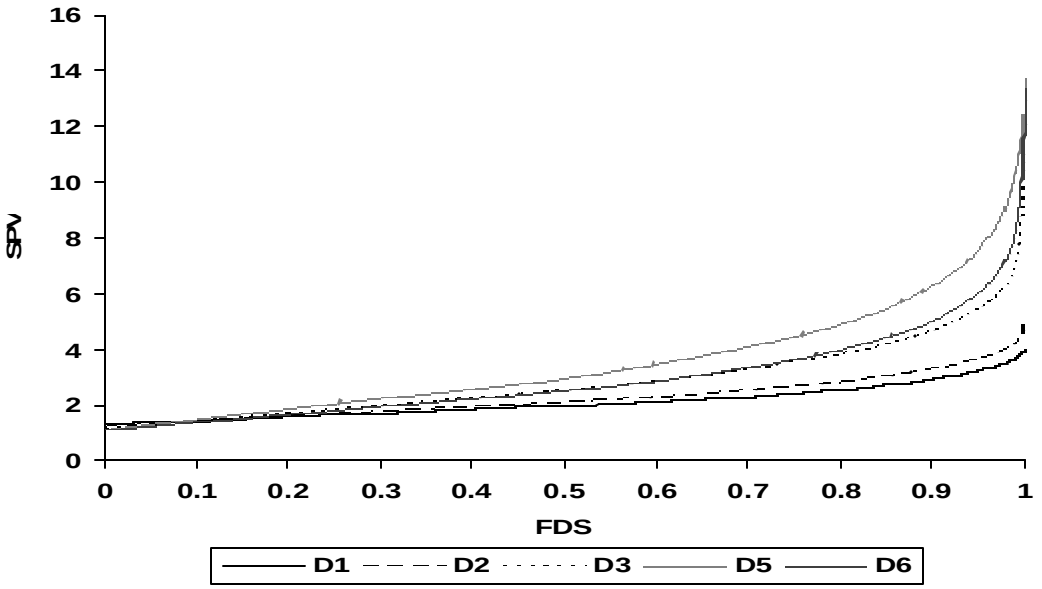

Figure IV.7: FDS Plot (Using SPV) for Comparing Standard Designs to Space Filling Designs for a Logistic Regression Model with Initial Parameter Estimate, $\hat{b}' = (2, 0.5, -0.8, 0.3)'$

Table IV.1: Minimum and Maximum SPV/PPV values for D1, D2, D3, D4, D5 and D6

	D1	D2	D3	D4	D5	D6
Min SPV	1.29	1.22	1.24	1.61	1.14	1.12
Max SPV	4	4.9	10.16	132	14	13
Min PPV	0.11	0.11	0.12	0.14	0.13	0.12
Max PPV	0.96	1.18	2.44	31.74	3.30	3.22

Figure IV.8 shows the FDS plots of the designs using PPV values. Similar to the SPV distribution, D4 performs significantly worse than the rest of the designs in terms of its PPV distribution. Therefore its FDS curve was omitted from Figure IV.8. The other Latin hypercube design D5, performs poorly compared to the rest of the designs as well. The uniform designs D3 and D6 perform very similarly to each other in terms of their PPV distributions, but they still perform worse compared to the standard designs D1 and D2. Even though these space filling designs performed poorly compared to the standard response surface designs in terms of prediction for the first order model with interaction, these designs may be appropriate to use when the particular model form is not known or very complex, when there is model uncertainty or when the researcher is interested in using a non-parametric fitting method.

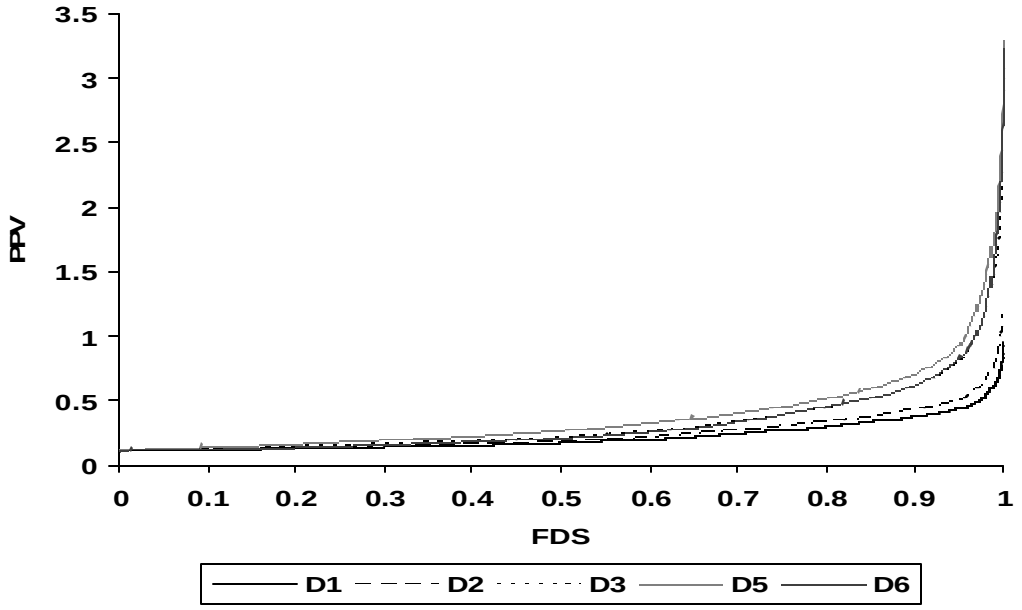

Figure IV.8: FDS Plot (Using PPV) for Comparing Standard Designs to Space Filling Designs for a Logistic Regression Model with Initial Parameter Estimate, $\hat{b}' = (2, 0.5, -0.8, 0.3)'$

As seen in this section, similar to the linear models case, FDS plots can be effectively used to compare designs for the GLM case.

IV.7 Studying Design Robustness for Binomial Data with a First Order Linear Predictor

Previously we have compared designs based on an assumed set of parameter values that were estimated by the practitioner. The problem is that the practitioner is unlikely to know the parameter values precisely. Frequently, practitioners may have some idea of what ranges of non-defects (or successes) might be expected, but they would expect not to get the initial estimates correct.

Since the SPV and PPV values depend on the unknown parameters in the model, the comparison on competing designs should take into account robustness to parameter misspecification. That is, a design that is more robust to the initial parameter misspecification while still having good prediction variance properties would be preferred. The robustness property can be examined by looking at the distributions of the SPV and PPV values for different assumed parameter estimates. Parameter misspecification will be studied in two ways: in terms of the $\hat{\beta}$'s in the model, and the predicted mean values, $\hat{\mu}$'s. The parameter misspecification in terms of the $\hat{\beta}$'s have drawbacks which will be developed subsequently. The parameter misspecification in terms of the $\hat{\mu}$'s is likely most natural for practitioners interested in prediction. There are several types of misspecification which may be of interest to be compared. These will be illustrated first using the binomial distribution.

Consider again the plastic molding example with binomial data for the logistic link with 2 factors and a first order model with interaction where the restricted design space is a square. Assume the set of initial parameter estimates are $\hat{\beta}' = (\hat{b}_0, \hat{b}_1, \hat{b}_2, \hat{b}_{12})'$, and the practitioner thinks that the estimates might be misspecified by as much as g %, then we may want to explore the SPV and PPV distributions for $\pm g$ % away from the original estimates. The upwards adjustments for say g =10%, would be $\hat{b}_i + 0.10 \times |\hat{b}_i|$ which means $\hat{b}_i + 0.10 \times |\hat{b}_i| = 1.1\hat{b}_i$, for $\hat{b}_i > 0$ and $\hat{b}_i + 0.10 \times |\hat{b}_i| = 0.9\hat{b}_i$ for $\hat{b}_i < 0$. Similarly for -10%, $\hat{b}_i - 0.10 \times |\hat{b}_i| = 0.9\hat{b}_i$ for $\hat{b}_i > 0$, and $\hat{b}_i - 0.10 \times |\hat{b}_i| = 1.1\hat{b}_i$ for

$\hat{b}_i$<0. To study the robustness to this form of misspecification, new FDS plots for either SPV or PPV values are constructed with the new $\hat{\boldsymbol{ß}}$. Our simulation results show that the FDS curves using PPV values corresponding to the $\pm g$ % misspecification surround the FDS curves of PPV values that correspond to all the possible $\hat{\boldsymbol{ß}}$ vectors that have misspecification of all the parameters of less than or equal to g % in either direction around the original estimates. However, this is not the case for the SPV values. Therefore, for the SPV values, the $\pm g$ % misspecification bands do not include all the possible misspecified $\hat{\boldsymbol{ß}}$ vectors that may occur within the desired range.

We have simulated a large number of $\hat{\boldsymbol{ß}}$ vectors within the fixed $\pm g$ % misspecification range and obtained the corresponding SPV/PPV values at each $\boldsymbol{x}_o$ in the design space to see if these SPV/PPV values fall within the SPV/PPV values corresponding to the $-g$% misspecification and the $+g$% misspecification at each $\boldsymbol{x}_o$. From our simulation results, we see that the interpretation of these fixed percentages is different depending on whether the researcher is interested in the PPV or the SPV distribution of the design. Considering the PPV distribution, we can conclude that while it is unlikely that all of the parameters will be misspecified by the same amount, the $\pm g$ % provides an idea of the effect on the PPV for over or underestimating the $\boldsymbol{ß}$'s. The $-g$% and the $+g$% provide bounds on what FDS curves can be obtained from the many possible misspecified $\hat{\boldsymbol{ß}}$ vectors within the range. As a result, the choice of g, say at 10, 20 or 30, can reflect the level of uncertainty of the practitioner's ability to estimate the $\boldsymbol{ß}$'s. (However, this is not the case when the researcher is interested in the SPV distribution of a design. The $\pm g$ % does not form a reasonable bound.)

Consider the binomial example where the practitioner's best guess for the parameter estimates is $\hat{\boldsymbol{ß}}' = (\hat{b}_0, \hat{b}_1, \hat{b}_2, \hat{b}_{12})' = (2, 0.5, -0.8, 0.3)'$. The practitioner is relatively uncertain about the parameters, and hence the FDS plot examines the change in design performance if $\hat{\boldsymbol{ß}}$ is misspecified by $\pm 20\%$. The misspecified values become

$\hat{ß}' = (2.4, 0.60, -0.64, 0.36)'$ for +20% and $\hat{ß}' = (1.6, 0.40, -0.96, 0.24)'$ for -20%. From our simulations, FDS curves for the PPVs corresponding to these two $\hat{ß}$ vectors form a band around the PPV values that could be obtained if each single $\hat{b}_i$ was misspecified by a different amount between $\pm 20\%$. Therefore, the $\pm 20\%$ bands can be used to provide an idea of the effect on the PPV for over/underestimating the $ß$'s. To show this, to each $\hat{b}_i$ corresponding to the $\hat{ß}$ vector we randomly assigned a misspecification percentage from a uniform distribution of (-0.20, 0.20). Each $\hat{b}_i$ was misspecified by a different percentage and the PPV distribution corresponding to the misspecified $\hat{ß}$ vector was calculated. This procedure was repeated a large number of times to obtain the PPV distribution corresponding to each $\hat{ß}$ vector. Among these PPV distributions, at each x_o, the number of PPV values that fall within the PPV range corresponding to the $\pm 20\%$ misspecification range was counted.

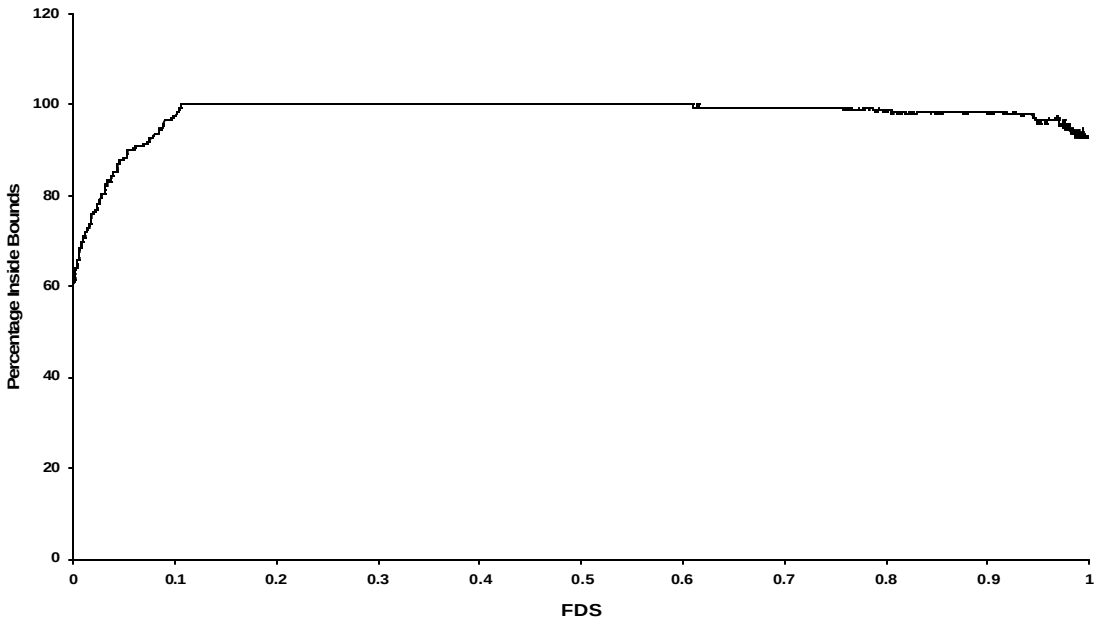

Figure IV.9: Plot of Percentage inside Bounds of $\pm 20\%$ Misspecification Range (Considering PPVs)

Figure IV.9 shows the plot of the percentage of the PPV values calculated using differently misspecified $\hat{\beta}$ vectors that fall inside the $\pm 20\%$ misspecification band at each FDS value. For example, at FDS=0.5, 100% of the PPVs defined by differently misspecified $\hat{\beta}$ vectors fall within the PPV values corresponding to the (-20%, 20%) range. Towards the ends of the fraction of design space, since the PPV values corresponding to the -20% and the 20% misspecifications become very close to each other, it is harder for the PPV values of the misspecified $\hat{\beta}$ vector to fall within the $\pm 20\%$ misspecification range. Therefore the proportion within the bounds is slightly less than 100%. The important conclusion is that a large portion of the proportions were equal to or close to 100%.

In addition to calculating the proportion, the minimum and the maximum PPV values among the simulated PPV distributions from misspecified $\hat{\beta}$ vectors were found at each FDS value. These were plotted along with the FDS curves corresponding to the initial $\hat{\beta}$ guess and the $\pm 20\%$ misspecification bands. Figure IV.10 shows the PPV values corresponding to the initial $\hat{\beta}$ vector, the $\pm 20\%$ misspecification bands and the minimum and the maximum PPV curves. The $\pm 20\%$ misspecification bands match up closely with the minimum and the maximum PPV curves, indicating that it is reasonable to use these $\pm 20\%$ misspecification bands to obtain information about the range of effects on the PPVs for over/underestimating the β's. The $\pm 10\%$, $\pm 30\%$ and the $\pm 50\%$ misspecification values were also calculated and similar results were obtained.

The three curves: D1, -20% and 20% misspecification curves can now be compared to obtain information about D1's robustness to parameter misspecification using PPV values. Towards the end of the fraction of design space, the three curves converge together and reach approximately the same PPV value. The PPV values at FDS=1 are 0.96, 1, and 0.86 for the initial parameter estimate, the -20% misspecification, and the 20% misspecification, respectively. The 20% misspecification curve has lower PPV values than both the initial estimate curve and the -20% misspecification curve. Even

though prediction precision is somewhat different for most of the design space for the three cases, it becomes similar as one gets closer to the end of the fraction of design space. Hence, for the practitioner who may be mostly interested in protecting against the worst possible prediction variance, this design has good robustness.

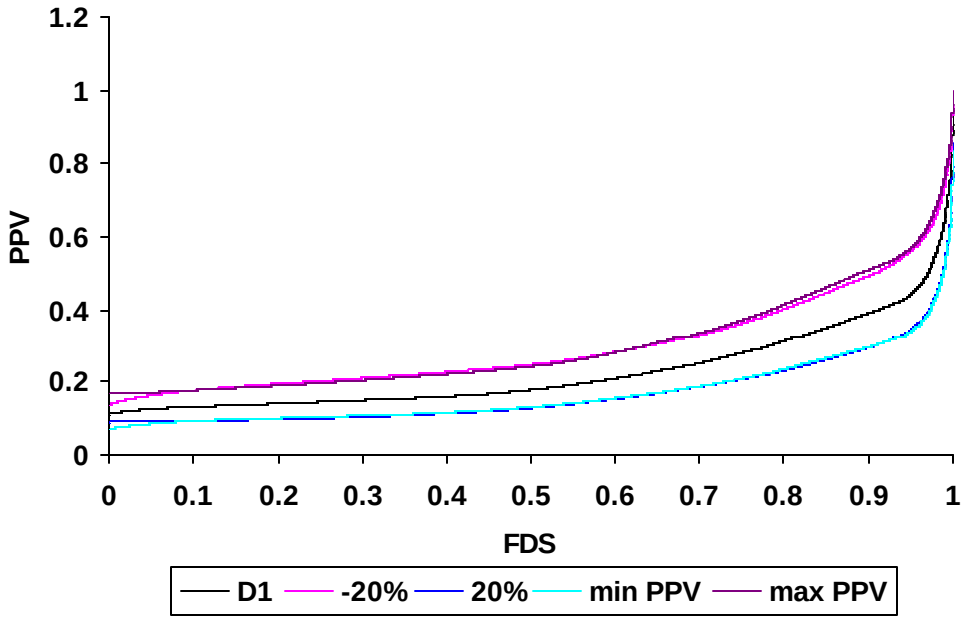

Figure IV.10: FDS Plot (Using PPV) of $\pm 20\%$ Misspecification Range and the Minimum/Maximum PPVs

When the researcher is interested in studying robustness for the SPV values, it performs very differently compared to the results from the PPV values. This is due to the difference between the definitions of SPV and PPV at a location in the design space. Recall the PPV and the SPV for the GLM at a location in the design space are defined as $N \operatorname{var}(\hat{y}(\boldsymbol{x}_o))$ and $N \dfrac{\operatorname{var}(\hat{y}(\boldsymbol{x}_o))}{s_o^2}$, respectively where both $\operatorname{var}(\hat{y}(\boldsymbol{x}_o))$ and s_o^2 depend on the $\hat{\boldsymbol{\beta}}$ vector. The PPV is directly related to the value of a predicted mean. For example, for a $\hat{p}_i$ value at a design location (corresponding to a certain $\hat{\boldsymbol{\beta}}$ vector) approaching 0 or 1, the corresponding PPV will be lower compared to a location with a $\hat{p}_i$ value closer to 0.5 since the variance of a binomial variable is higher when the probability is closer to

0.5 and it is lower when the probability is close to 0 or 1. However, for the SPV since each variance of the estimated mean response is divided by the variance of an observation at that location, the behavior is less predictable. The SPV will not change directly with the behavior of the predicted means (i.e., the $\hat{\beta}$ vector). That is why not all the SPV values corresponding to the simulated $\hat{\beta}$ vectors within the ±20% misspecification range fall within the range. Figure IV.11 shows the percentage of the SPV values calculated using differently misspecified $\hat{\beta}$ vectors that fall inside the ±20% misspecification bands at each FDS value. The percentage of all $\hat{\beta}$ estimates that lies between the ±20% bands is between 20% and 50% for the first third of the fraction of designs space. For the second third of the FDS range, it is around 50% and it decreases rapidly towards the end of the fraction of design space. At the four design locations (the corners of the square) with maximum SPV values, the ratio increases to nearly 100%.

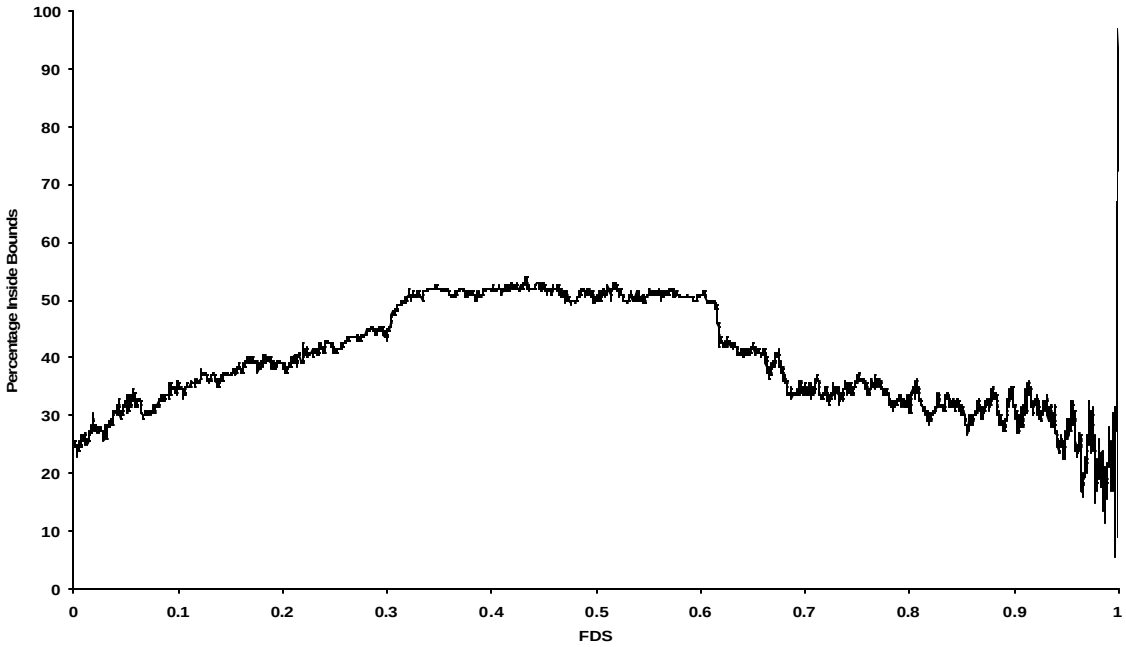

Figure IV.11: Plot of Percentage inside Bounds of ±20% Misspecification Range (Considering SPVs)

Figure IV.12 shows the SPV values corresponding to the initial $\hat{\beta}$ vector, the ±20% misspecification bands and the minimum/maximum SPV curves. In this case, the ±20%

misspecification bands do not form a bound around the minimum/maximum SPV curves indicating that it is not sensible to use these $\pm 20\%$ misspecification bands to obtain information about the range of effects on the SPVs for over/underestimating the β's. The separation between the $\pm g$ % misspecification curves and the minimum/maximum SPV curves becomes more evident as g increases.

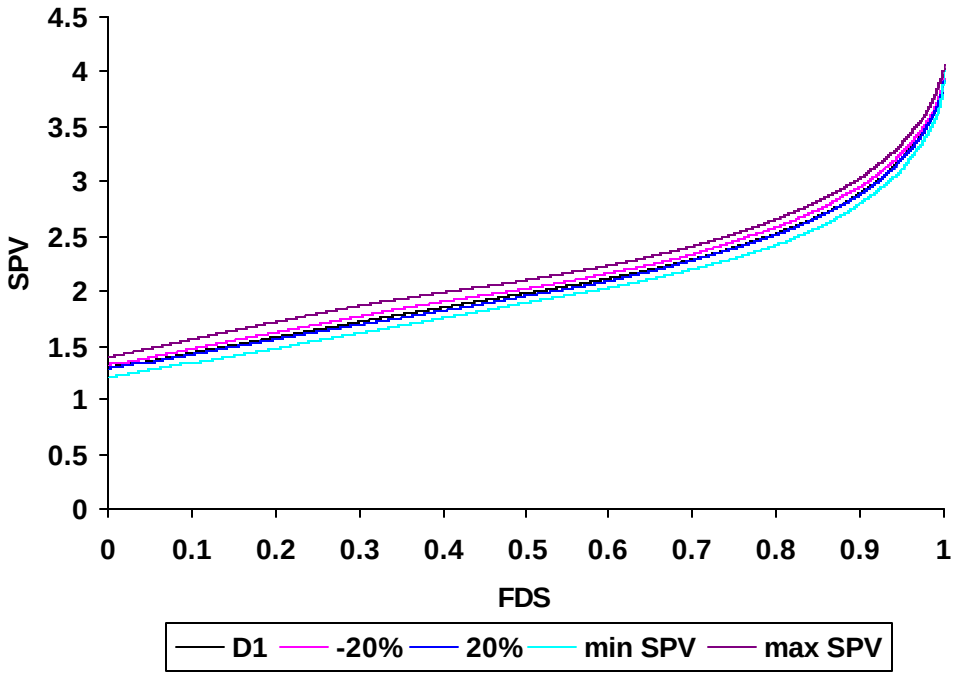

Figure IV.12: FDS Plot (Using SPV) of $\pm 20\%$ Misspecification Range and the Minimum/Maximum SPVs

In addition to the problems with bounding the effects of the misspecification, this approach to summarizing misspecification in terms of the $\hat{\beta}$'s may not be particularly realistic for many applications since rarely will the practitioner think of model prediction in terms of the $\hat{\beta}$ vectors. Recall how the estimates were obtained for the previously considered example: the practitioner estimated the probability of non-defects in the design region, and then the $\hat{\beta}$'s were calculated from this.

The second approach to considering design robustness to parameter misspecification examines the predicted mean values, $\hat{\boldsymbol{p}}$'s, directly. We consider three possible types of misspecifications in terms of the predicted mean values, $\hat{\boldsymbol{p}}$'s. The researcher can consider increasing or decreasing the spread of the range, shifting the range, or changing the location of the predicted mean values. More details on each of these are provided in the discussion of the example.

Recall for this example that the predicted mean values, $\hat{\boldsymbol{p}}$, are estimated to be $\hat{\boldsymbol{p}}' = (\hat{p}_{-1,-1},\ \hat{p}_{-1,1},\ \hat{p}_{1,-1},\ \hat{p}_{1,1})' = (0.93,\ 0.60,\ 0.95,\ 0.88)'$ for the 2^2 full factorial design with a first order model with interaction. The range for the predicted mean values is 0.35 since $(\min \hat{p}_i,\ \max \hat{p}_i) = (0.6, 0.95)$. The researcher may be interested in how the distribution of the SPV changes as the spread of the range of the predicted mean values is changed. Assume that the practitioner anticipates misspecifying the minimum and the maximum values in the following ways. If the minimum value is increased by 0.1, and the maximum value is decreased by 0.05, the range becomes 0.2 where $(\min \hat{p}_i,\ \max \hat{p}_i) = (0.7,\ 0.90)$. Another possibility is the researcher may increase the range to 0.38, and obtain $(\min \hat{p}_i,\ \max \hat{p}_i) = (0.5,\ 0.98)$. For calculation purposes, the predicted mean values for the other two locations are taken to be the average of the minimum and the maximum predicted mean values.

Figure IV.13 shows the FDS plot using SPV values for the three cases described. The increase in the spread of the range results in a large increase in the SPV values throughout the design space. When the spread of the range is decreased, the SPV values decrease. Notice as well that the shape of the FDS curve for the largest range of parameters has a very different shape than the other ranges. Table IV.2 displays the minimum and the maximum SPV and PPV values, the approximate G-efficiencies, and the V-averages for the three cases described. For this example, as the maximum predicted mean value approaches 1, the G-efficiency decreases. This confirms the result of Zahran and Myers (2003).

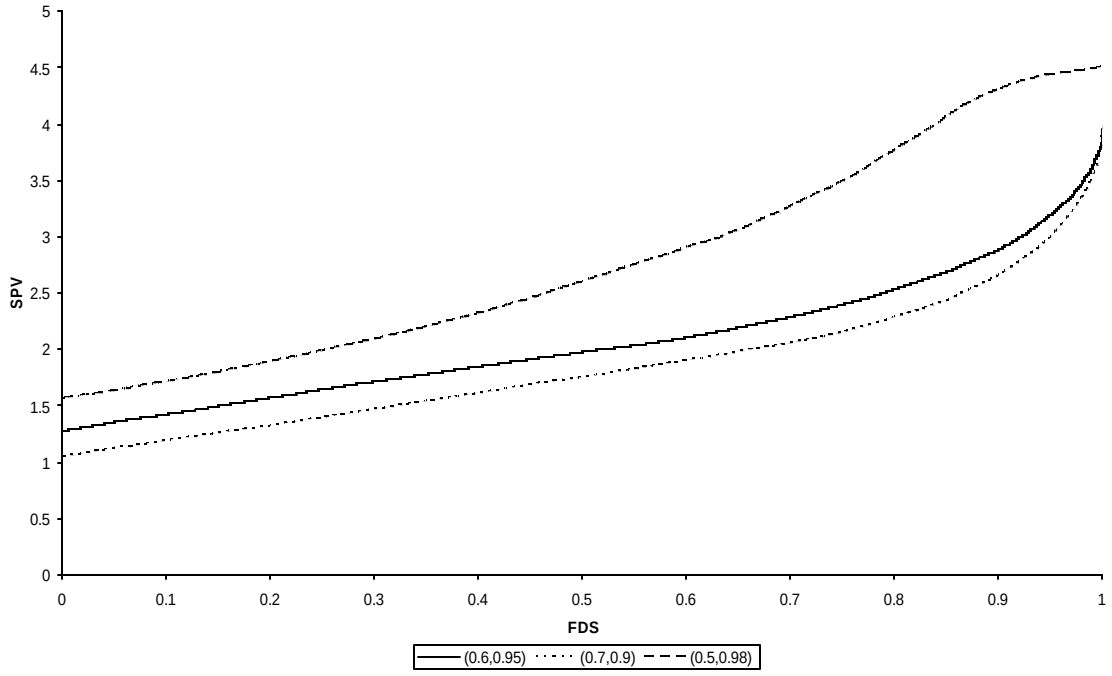

Figure IV.13: FDS Plot (Using SPV) for the Initial Predicted Mean Value, $\hat{\boldsymbol{p}}' = (0.93, 0.60, 0.95, 0.88)'$, and the Increased/Decreased Predicted Mean Value Range

Table IV.2 Summary of SPV and PPV values, Approximate G-efficiencies and V-averages for the Three Different Ranges Shown in Figure IV.13

Range	Min SPV	Max SPV	G-efficiency	V-average	Min PPV	Max PPV
(0.60, 0.95)	**1.3**	**4.0**	**100%**	**2.09**	**0.11**	**0.96**
(0.70, 0.90) (decreased)	1.1	4.0	100%	1.86	0.18	0.84
(0.50,0.98) (increased)	1.6	4.5	89%	2.80	0.08	1

Figure IV.14 shows the FDS plot using PPV values for ranges described in Table IV.2. The range (0.60, 0.95) has the smallest PPV for most of the fraction of design space whereas range (0.5, 0.98) has considerably the highest PPV values for most of the fraction of design space. The curves intersect towards the beginning and the end of the fraction of the design space. The range (0.5, 0.98) starts with the lowest PPV and ends with the highest PPV values. This can be explained by the fact that if $\hat{p}_i$ is chosen to be 0

or 1, then the PPV will become zero. Towards the end of the fraction of design space, all three curves possess steeper slopes, and they get very close to each other at the extremes. The reason why the PPV curves are considerably different from the SPV curves is due to the additional $\text{var}(y(\boldsymbol{x}_o))$ in the PPV formula. Notice that for both SPV and PPV values, the design is only moderately robust to the initial parameter estimates. Expanding or reducing the range of $\hat{\boldsymbol{p}}$ values makes considerably more difference to design performance than shifts in the $\hat{\boldsymbol{ß}}$'s.

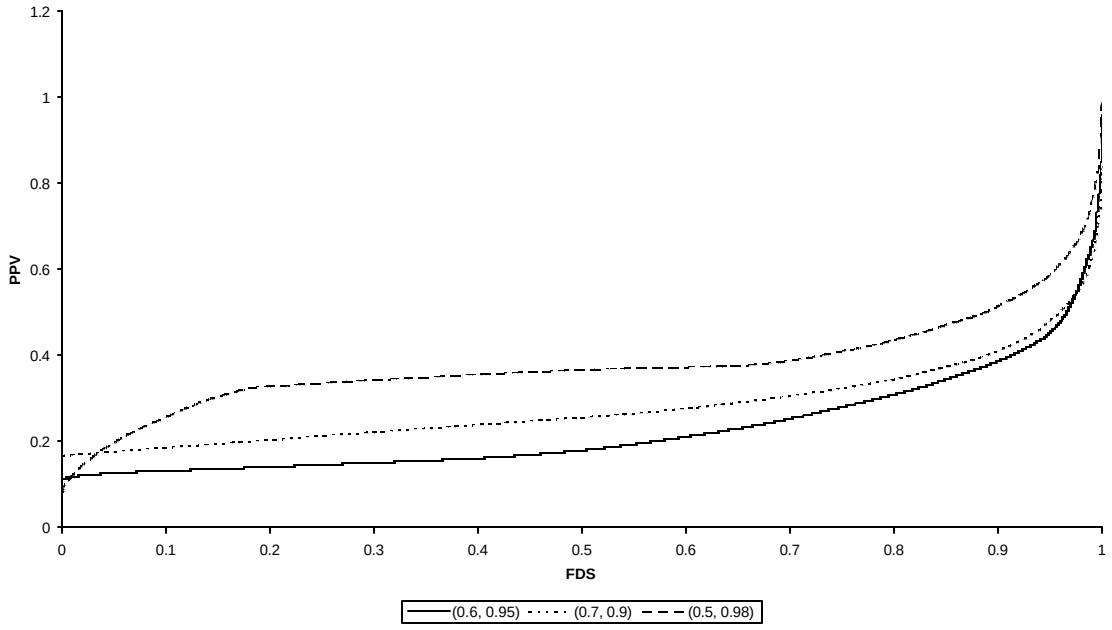

Figure IV.14: FDS Plot (Using PPV) for the Initial Predicted Mean Value, $\hat{\boldsymbol{p}}' = (0.93,\ 0.60,\ 0.95,\ 0.88)'$, and the Increased/Decreased Predicted Mean Value Range

Another type of misspecification that can be considered is to examine the effect on the SPV and PPV values when range of the predicted probabilities of success is shifted. Considering our example where the predicted mean values, $\hat{\boldsymbol{p}}$, are $\hat{\boldsymbol{p}}' = (\hat{p}_{-1,-1},\ \hat{p}_{-1,1},\ \hat{p}_{1,-1},\ \hat{p}_{1,1})' = (0.93,\ 0.60,\ 0.95, 0.88)'$, the range of the predicted means can be shifted either by increasing or decreasing it. When the minimum and the maximum values are decreased by a certain amount, 0.03 in this case, the range

becomes $(\min \hat{p}_i, \max \hat{p}_i) = (0.57, 0.92)$. When the range is increased by 0.03, it becomes $(\min \hat{p}_i, \max \hat{p}_i) = (0.63, 0.98)$. The predicted mean values for the other two locations are again assumed to take the average of the minimum and the maximum predicted mean values.

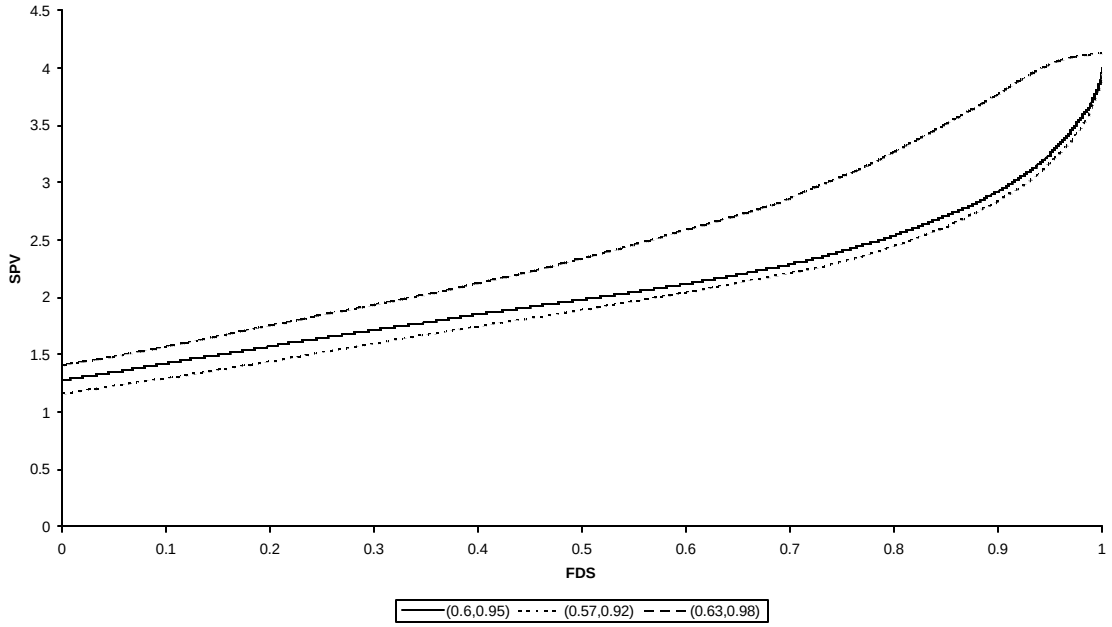

Figure IV.15: FDS Plot (Using SPV) for the Initial Predicted Mean Value, $\hat{\boldsymbol{p}}' = (0.93, 0.60, 0.95, 0.88)'$, and the Shifted Predicted Mean Value Range

Table IV.3 Summary of SPV and PPV values, Approximate G-efficiencies and V-averages for the Three Different Ranges Shown in Figure IV.15

Range	Min SPV	Max SPV	G-efficiency	V-average	Min PPV	Max PPV
(0.6, 0.95)	**1.3**	**4.0**	**100%**	**2.09**	**0.11**	**0.96**
(0.57, 0.92)	1.2	4.0	100%	1.99	0.22	0.98
(0.63,0.98)	1.4	4.1	98%	2.5	0.08	0.93

Figure IV.15 compares the three cases using SPV values. When the range is shifted down (0.57, 0.92), the SPV values decrease, and when it is shifted up (0.63, 0.98), the SPV values increase. The change in SPV values for the positive shift is larger compared to the negative shift. In this example, the comparison of both the shift and the increase in range of the estimated means yields similarly shaped curves. Table IV.3 shows the minimum

and the maximum SPV and PPV values, the approximate G-efficiencies, and the V-averages for the three cases described. Again, as the maximum predicted mean value approaches 1, the G-efficiency decreases.

Figure IV.16 shows the FDS plot using PPV values. Similar to the example where the spread of the range was changed, the curves for the range shifts obtained using PPV values are considerably different than when using SPV values. The range (0.57, 0.92) has the highest PPV values over 99.9% of the total design space. Even though the range (0.63, 0.98) has lower values at the beginning of the fraction of design space compared to the range (0.6, 0.95), its PPV values increase considerably as one moves towards the middle and the end. All three curves approach approximately the same value for maximum PPV values.

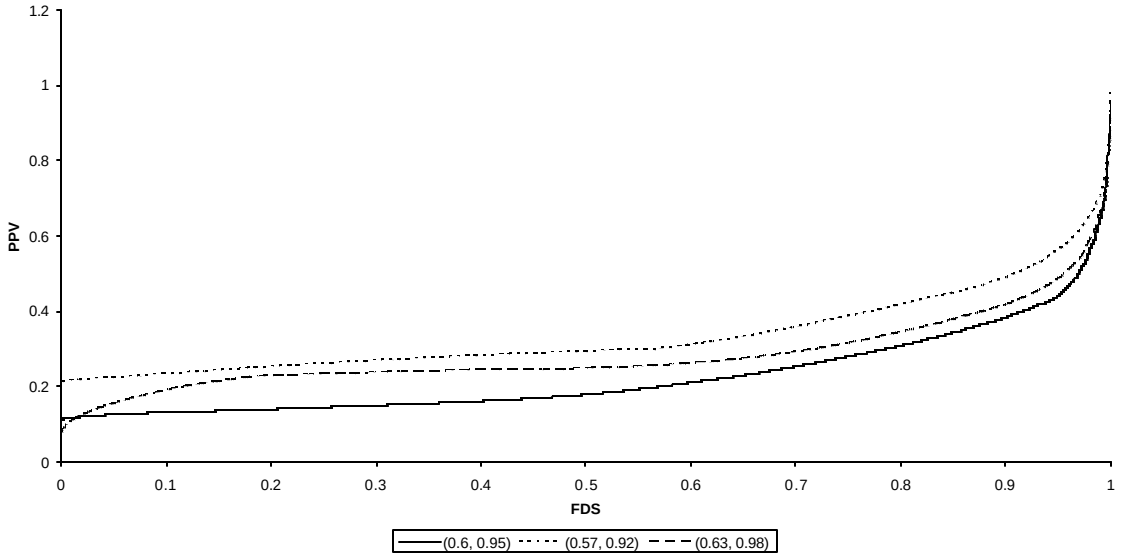

Figure IV.16: FDS Plot (Using PPV) for the Initial Predicted Mean Value, $\hat{\boldsymbol{p}}' = (0.93,\ 0.60,\ 0.95,\ 0.88)'$, and the Shifted Predicted Mean Value Range

Notice the conflicting information between the SPV and PPV values. When examining the G-efficiency using the SPV scaling, the range (0.57, 0.92) has the smallest values, while on the PPV scale this is reversed for most of the ranges of the variances. Hence, distinct from the linear model case where PPV and SPV are just fixed multiples of each

other, for the GLM we now need to prioritize between the theoretical properties of the SPV and the practical interpretability of the PPV.

Another type of misspecification in the predicted mean values which the practitioner may wish to consider occurs when the locations of the predicted mean values are changed. This type of misspecification only becomes important for non-symmetric designs with different number of observations at each design location. For a symmetric design, changing the locations of the predicted mean values will not change the performance of the design. This type of misspecification is important for the practitioner who is interested in the effect of misspecifying where particular predicted mean values occur.

Consider the design previously introduced, D1 (2^2 factorial design with 5 observations at each design location with a first order model with interaction in the linear predictor with the canonical logistic link). Consider a second design, D2, which assumes the same predicted mean values at the design locations as D1 but is non-symmetric in terms of its allocation of observations as shown in Table IV.4. In D2, the allocation of the observations is chosen in a way to attempt to stabilize the unequal variances, by allocating more observations to locations with larges variances (with $\hat{p}_i$ values near 0.5). Since it is unlikely that this will be accurately estimated before collecting data, the practitioner can consider the location of the predicted mean values being misjudged and occurring as in D2* to study how robust this design is to location changes. Therefore, the locations of the predicted mean values are changed in such a way that the predicted mean value with the smallest variance ($\hat{p}_i$ close to 0 or 1) is placed at the corner with the largest number of observations, to create a less than optimal allocation scenario.

The motivation for putting more observations at locations with larger variance locations as in D2 can be seen in Table IV.4 and Figure IV.17 where we see that this type of allocation gives the most stable PPV distribution among the three designs. D1 gives relatively good performance in terms of the shape of its PPV distribution and it does not need as detailed prior information to create the design whereas D2* corresponds to using an unbalanced design and guessing wrong on where the predicted probability values will

be. D2* has lower PPV values than D2 for the 62% smallest PPV values in the design space. For the rest of the design space, D2's PPV values are lower than D2*'s PPV values and the maximum PPV values are much smaller for D2 (only 37.5% of the size of D2*). Starting at FDS=0.77, D1 has higher PPV values than D2 as well.

Table IV.4: Summary of SPV and PPV values for the 2^2 Factorial D1, D2 and D2*

Design	$\hat{\boldsymbol{p}}' = (\hat{p}_{-1,-1}, \hat{p}_{-1,1}, \hat{p}_{1,-1}, \hat{p}_{1,1})$	$\underline{n}' = (n_{-1,-1}, n_{-1,1}, n_{1,-1}, n_{1,1})'$	Min SPV	Max SPV	Min PPV	Max PPV
D1	(0.93, 0.60, 0.95, 0.88)	(5, 5, 5, 5) symmetric allocation	1.29	4.0	0.11	0.96
D2	(0.93, 0.60, 0.95, 0.88)	(4, 8, 3, 5) improved allocation	1.29	6.67	0.17	0.6
D2*	(0.88, 0.95, 0.60, 0.93)	(4, 8, 3, 5) allocation for $\hat{\boldsymbol{p}}$ misjudged	1.21	6.67	0.08	1.6

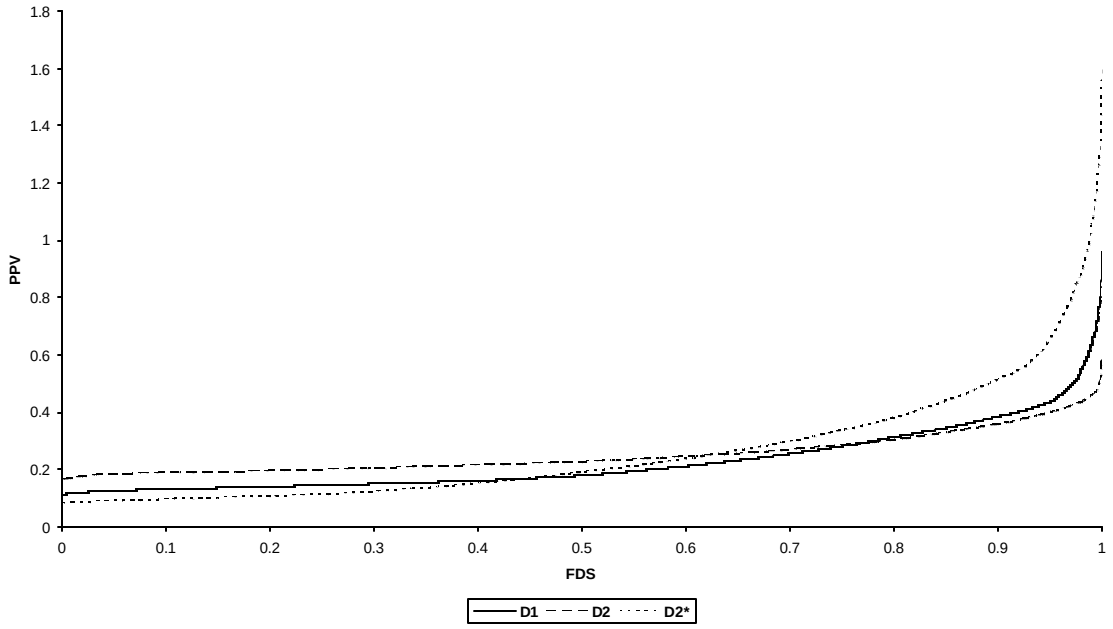

Figure IV.17: FDS Plot (using PPV) for D1, D2, and D2*

When we look at Figure IV.18 for studying the SPV values, we see that the shapes of the curves look vastly different from the PPV values in Figure IV.17. All three curves

increase gradually for small SPV values and then the slope increases towards the end of the fraction of design space. D2 has the highest SPV values among the three designs. D2* has the smallest SPV values for 60% of the total design space and D1 has the smallest SPV values for the largest 40% of the total design space. D1 and D2 have the same minimum SPV values. D2 and D2* have the same maximum SPV, therefore the same G-efficiency of 60%. Even though D1 may be chosen as the best design for theoretical purposes, the practitioner may prefer D2, to obtain a more overall stable prediction variance and a smaller worst case prediction variance. However, the superiority of D2 for PPV is highly dependent on correct parameter estimation, and hence equal allocation designs are recommended unless the relationship between the response and the factors is well understood.

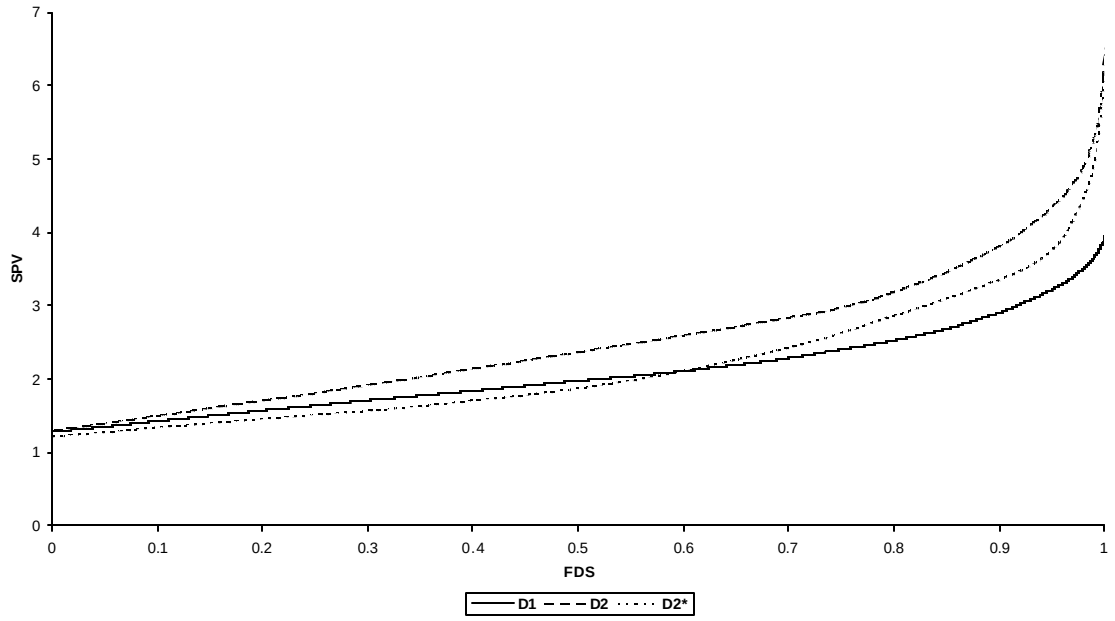

Figure IV.18: FDS Plot (using SPV) for D1, D2, and D2*

IV.8 Studying Design Robustness for Binomial Data with a Second Order Linear Predictor

The prediction variance performance of response surface designs for GLM is an important issue that needs to be better understood. The adapted FDS plots are good tools for obtaining information about the properties of the distribution of the prediction variance for a standard design under the GLM framework. In the previous section, we

used FDS plots to illustrate the prediction properties of a 2^2 factorial design for a logistic regression model with a first order linear predictor. In this section, we examine the case of the second order linear predictors. The procedure for obtaining an adapted FDS plot is the same for a design with a logistic regression model with a second order linear predictor. However, the prediction variance properties of a standard design for the GLM with a second order linear predictor are now more complex since the underlying surface of the predicted means is more complicated. However, utilizing the FDS plots can potentially help the researcher obtain a better understanding of the design's robustness properties.

Consider again the plastic molding example with binomial data (where a success is a non-defect) with 2 factors. An equally replicated 3^2 full factorial might be an appropriate design choice for a second order GLM model with an assumed square design space. If the linear predictor is a second order model for the canonical logistic link, the model looks like:

$$\ln\left(\frac{p_i}{1-p_i}\right) = b_o + b_1 x_{1i} + b_2 x_{2i} + b_{12} x_{1i} x_{2i} + b_{11} x_{1i}^2 + b_{22} x_{2i}^2 . \qquad \text{(IV.12)}$$

Equivalently, $p_i = \frac{1}{1+e^{-(b_o + b_1 x_{1i} + b_2 x_{2i} + b_{12} x_{1i} x_{2i} + b_{11} x_{1i}^2 + b_{22} x_{2i}^2)}}$ models the probability of a non-defect at any location, *i*.

Since the model has become more complicated, it is less likely that when the practitioner is asked to give initial estimates, he/she will be able to accurately estimate much more than the range of predicted means that might be obtained. Moreover, it is very unlikely that the particular locations of the predicted means will be precisely guessed. The practitioner might give an initial range of (0.55, 0.95) for the means, and a guess at the locations, called G1. Table IV.5 explores other location combinations in the same range to allow us to assess robustness.

Figure IV.19 shows the contour plots of the first two cases, G1 and G2. Even though the ranges of the predicted means are the same, the shape of the surfaces described is quite

different. The minimum predicted mean for G1 is near the (-1, 1) corner whereas for G2, it is along the $x_2 = -1$ edge. We wish to investigate how this affects the prediction properties of the design.

Table IV.5: Possible Predicted Probabilities for Range = (0.55, 0.95)

Design Location	(1,1)	(1,-1)	(-1,1)	(-1,-1)	(-1,0)	(1,0)	(0,-1)	(0,1)	(0,0)	Range
G1	**0.79**	**0.93**	**0.55**	**0.94**	**0.69**	**0.78**	**0.95**	**0.72**	**0.77**	**(0.55, 0.95)**
G2	0.80	0.71	0.95	0.67	0.88	0.78	0.55	0.83	0.73	(0.55, 0.95)
G3	0.55	0.56	0.57	0.95	0.9	0.92	0.66	0.7	0.72	(0.55, 0.95)
G4	0.55	0.6	0.57	0.59	0.7	0.8	0.83	0.95	0.65	(0.55, 0.95)
G5	0.92	0.93	0.95	0.55	0.7	0.89	0.8	0.85	0.65	(0.55, 0.95)
G6	0.55	0.95	0.75	0.75	0.75	0.75	0.75	0.75	0.75	(0.55, 0.95)
G7	0.9	0.55	0.95	0.89	0.75	0.92	0.94	0.88	0.85	(0.55, 0.95)
G8	0.85	0.87	0.9	0.55	0.95	0.7	0.71	0.92	0.93	(0.55, 0.95)
G9	0.8	0.71	0.95	0.75	0.88	0.9	0.55	0.93	0.91	(0.55, 0.95)
G10	0.58	0.75	0.95	0.75	0.75	0.55	0.75	0.56	0.75	(0.55, 0.95)
G11	0.95	0.93	0.87	0.85	0.77	0.55	0.92	0.9	0.94	(0.55, 0.95)
G12	0.63	0.75	0.73	0.65	0.57	0.58	0.55	0.7	0.95	(0.55, 0.95)

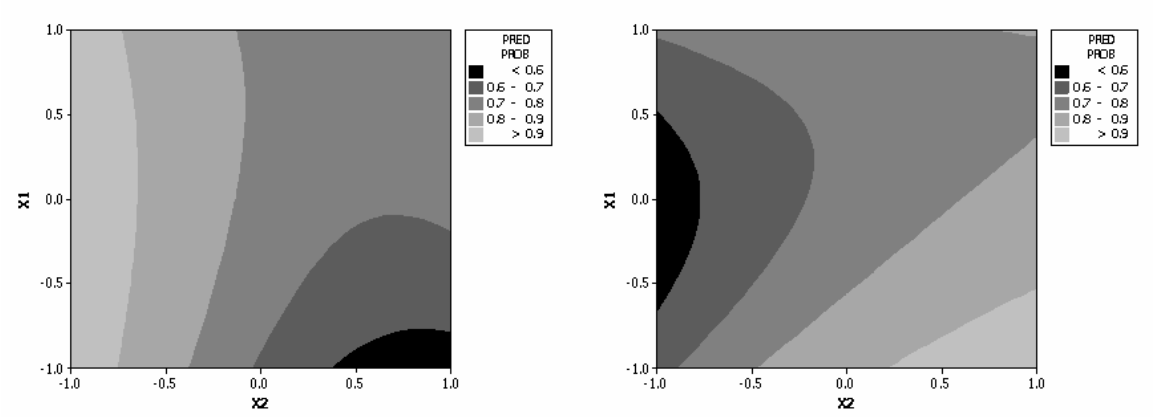

Figure IV.19: Contour Plots of the Predicted Probabilities for G1 and G2, respectively

Figures IV.20 shows the contour plots of the PPVs for G1 and G2, respectively. Recall for a binomial distribution, the predicted means closer to 0.5 will have higher variances and those are closer to 0 or 1 will have lower variances. Since the $x_2 = -1$ edge has the

highest predicted means for G1, it has the lowest PPV values and similarly the (-1, 1) corner has the highest PPVs because its predicted means are closest to 0.5. However, the distribution of the PPVs for G2 is very different. The highest PPVs occur at the (-1, -1) corner and lowest values are at (-1, 0.9). Even though there are noticeable differences in terms of where a specific PPV value occurs, this may or may not change the overall actual prediction performance of the design.

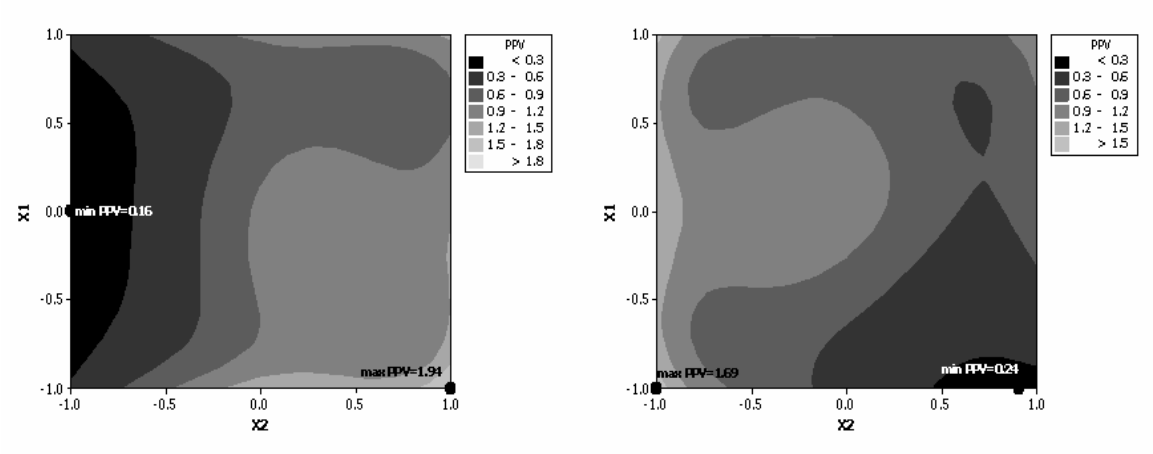

Figure IV.20: Contour Plots of PPVs for G1 and G2, respectively

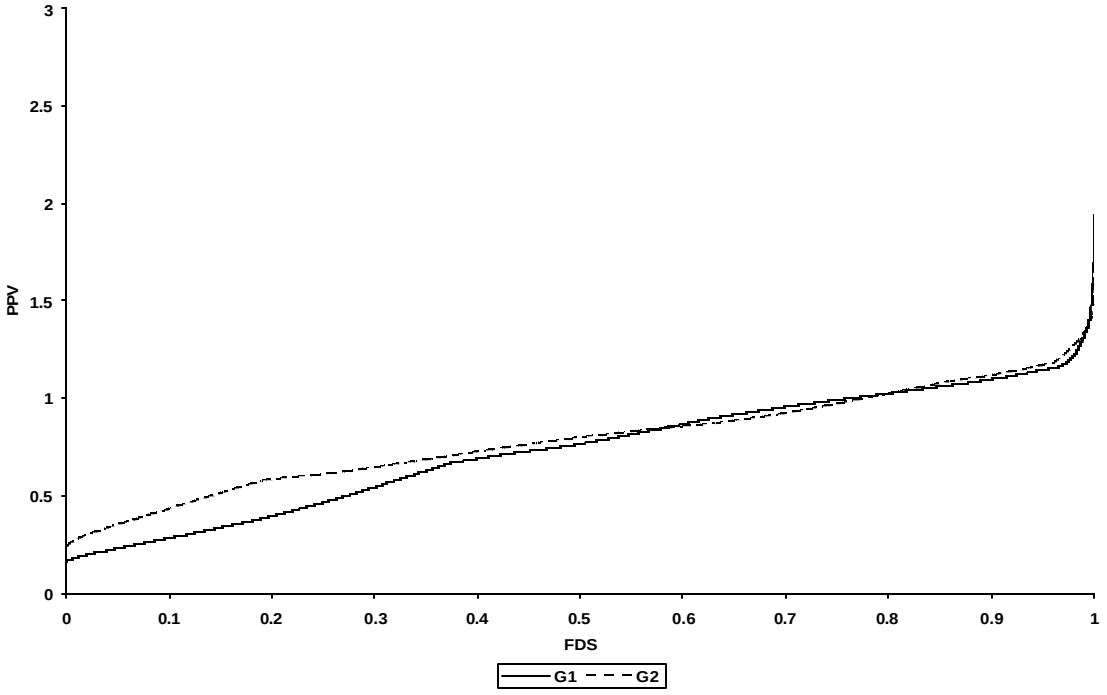

Figure IV.21: FDS Plots of G1 and G2 (using PPVs)

Figure IV.21 shows the FDS plots for G1 and G2 using PPVs, where as long as an appropriate range of the predicted probabilities is captured, the distributions of the PPVs for the two cases are similar.

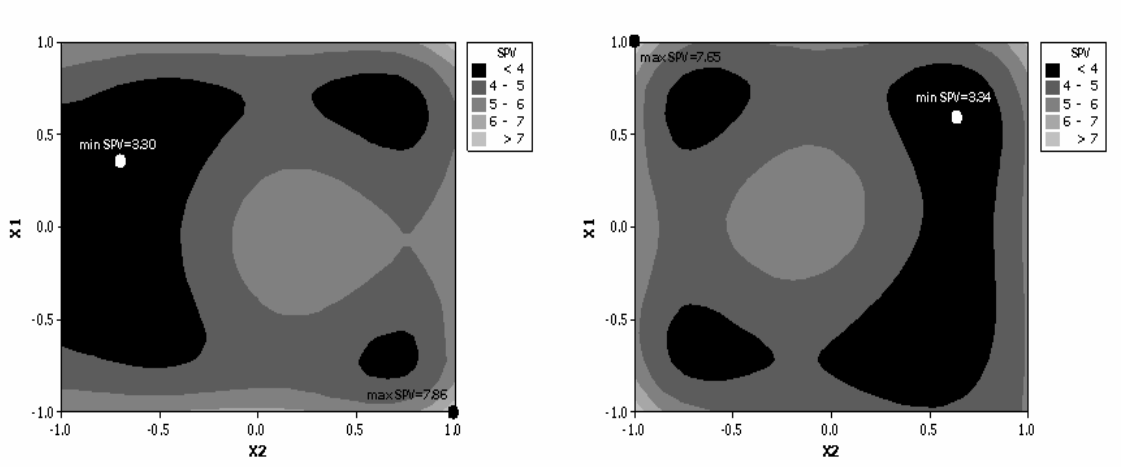

Figure IV.22: Contour Plots of SPVs for G1 and G2, respectively

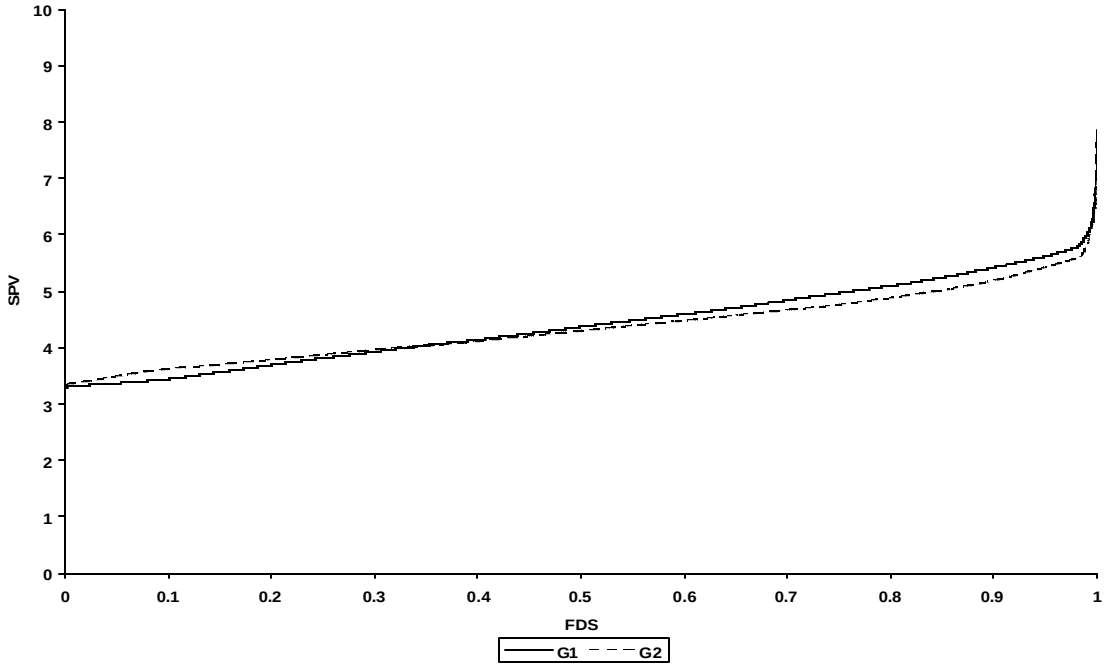

Figure IV.23: FDS Plots of G1 and G2 using (SPVs)

In Figure IV.22, the contours show that the distributions of the SPVs vary considerably for G1, the highest SPV (7.86) occurs at the (-1, 1) corner and the lowest SPV (3.30) occurs at the (0.35, -0.69) location. For G2, the lowest SPV (3.34) occurs at (0.59, 0.64) and the maximum SPV value of 7.65 is at (1, -1). Figure IV.23 shows the FDS plots for G1 and G2 using SPVs. Similar to the PPV distributions, as long as an appropriate range of the predicted probabilities is captured, the distribution of the SPVs of the two cases are similar to each other.

By looking at an FDS plot the researcher cannot obtain information about the exact location of a specific prediction variance. However, FDS plots are very useful in terms of providing information about the general profiles of the distributions of the SPV and the PPV values. These plots are helpful to describe the overall characteristics of the SPV/PPV distributions of a design unlike contour plots. As the number of factors increase, it becomes hard to view contour plots. On the other hand, FDS plots can be obtained for a design with any number of factors. FDS plots give the researcher information about the overall trends in the distribution of SPV and the PPV values.

To gain a better understanding of the possible difference in the predictions of a variety of underlying surface shapes, all with a similar range of observed means, we simulated a number of possible surfaces. We considered a number of different cases where some cases had more predicted means close to the bottom end of the (0.55, 0.95) range, some cases had more predicted means close to the top of the range and some had more predicted means close to the average of the range. Figure IV.24 shows the PPV plots of the same 3^2 factorial design with the same range of predicted probabilities but with a selection of different values and at different locations to study how robust the design is to the initial parameter estimates. As seen from Figure IV.24, there is considerable variability in the FDS curves as the location and the values of the predicted probabilities change. From Figure IV.25 a similar dispersion exists for the SPV values as well.

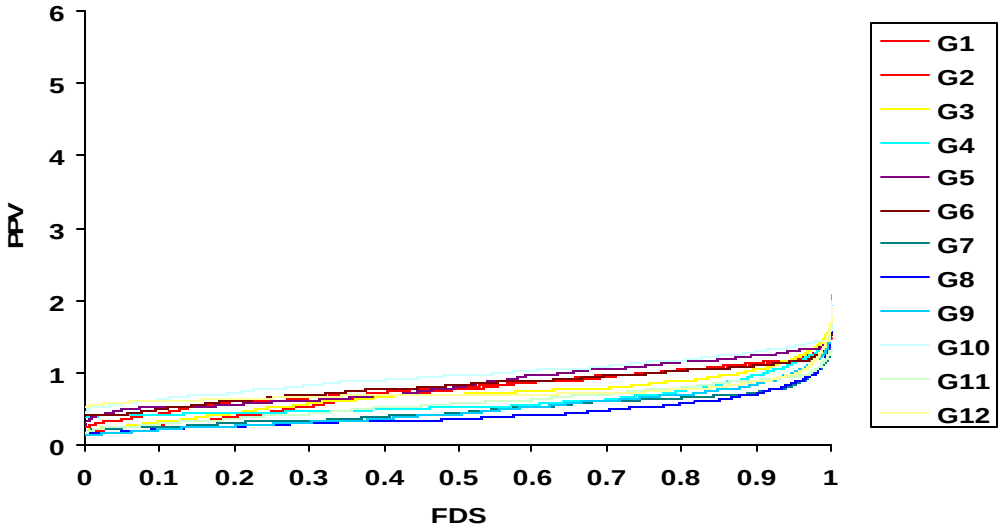

Figure IV.24: FDS Plot (Using PPV) for Cases: G1-G12 with Equal Allocation

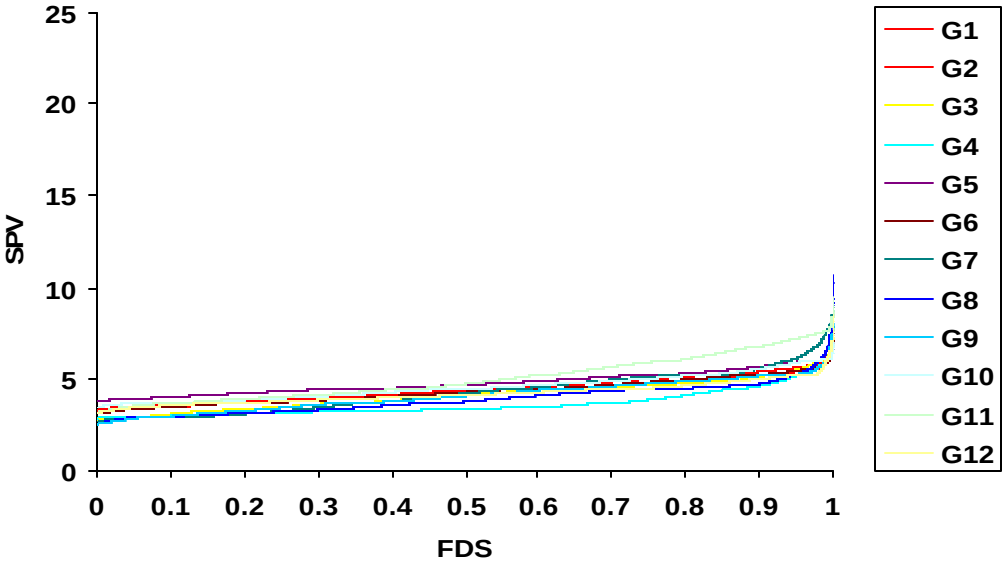

Figure IV.25: FDS Plot (Using SPV) for Cases: G1-G12 with Equal Allocation

In previous sections, we have discussed that the prediction variance can be made more stable by using unequal allocations at different design points. We now examine this for the second order model. There are two possibilities: The first assumes the ideal allocation of the observations chosen assuming good knowledge of what means will be observed

and to account for the larger variances known to occur for $\hat{p}_i$ values near 0.5. In the second case we consider how this allocation might suffer if location of the predicted mean values is being misjudged to study how robust the design is to location changes. The ideal unequal allocation is frequently not very realistic, but can allow explanation of the most stable prediction distribution. The second case is used to see how the design performs with a misjudged allocation. For all the three allocation cases, the y-axes of the FDS plots using either SPV or PPV values are on the same scale to distinguish the differences more clearly.

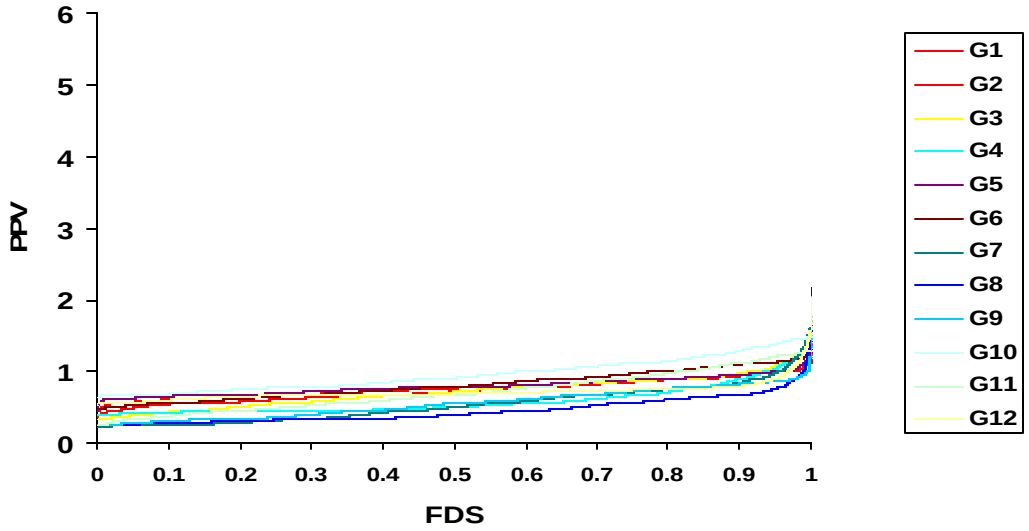

Figure IV.26: FDS Plot (Using PPV) for Cases: G1-G12 with Improved Allocation

Figure IV.26 shows the PPV distribution of the design with the different predicted probability cases assuming an improved allocation. Overall the curves look less variable and more horizontal than the allocation of Figure IV.24. This is to be expected since the allocation was chosen for just this purpose. The motivation for putting more observations at locations with larger variances is to be able to estimate better at those locations that are thought to have larger variances based on the estimated proportion of defects at each location. This can be seen in Table IV.6 for G1 and G2. Since the variance at each design point is $s_i^2 = p_i(1-p_i)/n_i$ where n_i is the number of observations at that point, the improved allocation selects sample sizes at each location to make s_i^2 as close to constant as possible. The misjudged allocation corresponds to studying how robust the design is to

location changes. The effect of a misjudged allocation can also be seen in Table IV.6 for G1 and G2.

Table IV.6: Improved/Misjudged Allocations for G1 and G2

Design	$\underline{\boldsymbol{n}}' = (n_{-1,-1}, n_{-1,1}, n_{1,-1}, n_{1,1})'$	Min PPV	Max PPV
G1	(112, 38, 166, 38, 144, 115, 32, 136, 119) improved allocation	0.39	1.24
G2	(96, 124, 29, 133, 63, 103, 149, 85, 118) improved allocation	0.45	1.33
G1	(119, 136, 32, 144, 38, 115, 166, 38, 112) misjudged allocation	0.13	5.47
G2	(118, 85, 149, 63, 133, 103, 29, 124, 96) misjudged allocation	0.18	2.79

To be able to obtain a total percentage gain/loss compared to the improved/misjudged allocations, the average of the SPV/PPV distributions of 20 different predicted probability cases were used. Compared to the equal allocation case in Figure IV.24 there is an average 7% decrease in the maximum PPV. This is due to the fact that the FDS curves using PPVs become more stable with an improved allocation case. However, there was a 1% increase in the average PPV and a 34% increase in the minimum PPV.

Figure IV.27 shows the PPV distribution of the design with the different predicted probability cases assuming a misjudged allocation. There is a considerably wider spread of the curves, especially for large PPV values. Compared to the equal allocation case in Figure IV.24 there is an 89% increase in the maximum PPV of the misjudged allocation case, a 19% increase in the average PPV and a 9% decrease in the minimum PPV. This is due to the fact that the prediction variance can get much worse if we have a region with large variance, but few observations allocated there. Table IV.7 shows the values of the average average, average minimum and average maximum SPV/PPV values for each of the three cases.

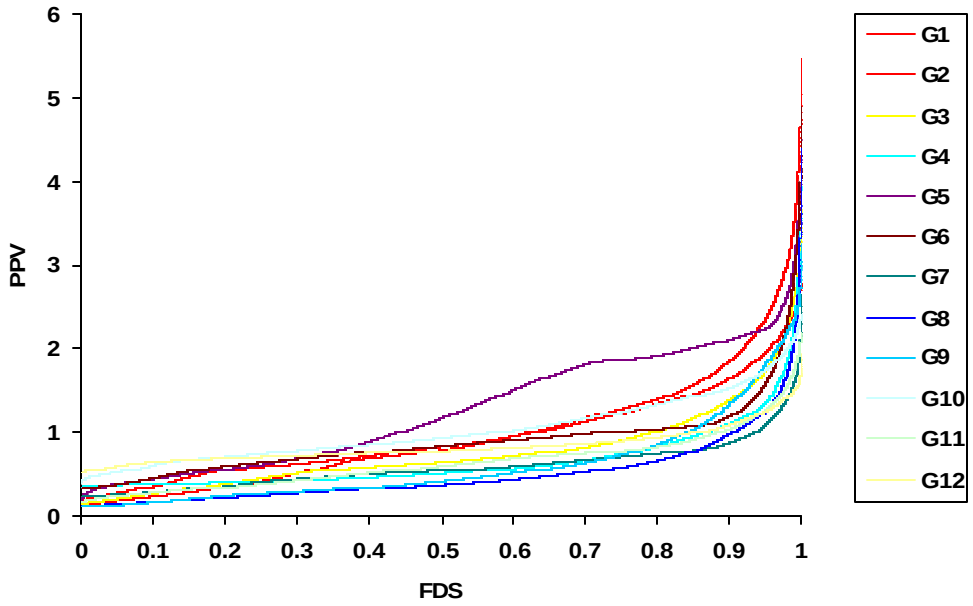

Figure IV.27: FDS Plot (Using PPV) for Cases: G1-G12 with Misjudged Allocation

Table IV.7: Average Minimum, Maximum and Average PPV/SPV Values for Each Allocation

Allocation	Min PPV	Ave PPV	Max PPV	Min SPV	Ave SPV	Max SPV
Equal	0.32	0.77	1.79	3.17	4.43	8.49
Improved	0.43	0.78	1.67	2.80	4.73	11.00
Misjudged	0.29	0.92	3.39	2.88	5.11	14.90

Figures IV.28 and IV.29 show the SPV distributions of the design with the different predicted probability cases assuming an improved allocation and a misjudged allocation, respectively. The performance of the unequal allocation is worse for the average and maximum SPV for both the improved allocation and the misjudged allocation cases. Compared to the equal allocation case in Figure IV.25 there is a 7% increase in the average SPV of the improved allocation case, a 30% increase in the maximum SPV with only a 12% decrease in the minimum SPV. There is a 15% increase in the average SPV of the misjudged allocation case, a 76% increase in the maximum SPV and a 9% decrease in the minimum SPV. We expect the unequal allocations to perform worse than the equal allocations for the SPV values, because the SPV formula is tied to the theoretical aspect

of prediction and the unequal allocation was designed to improve the PPV distribution, not the SPV.

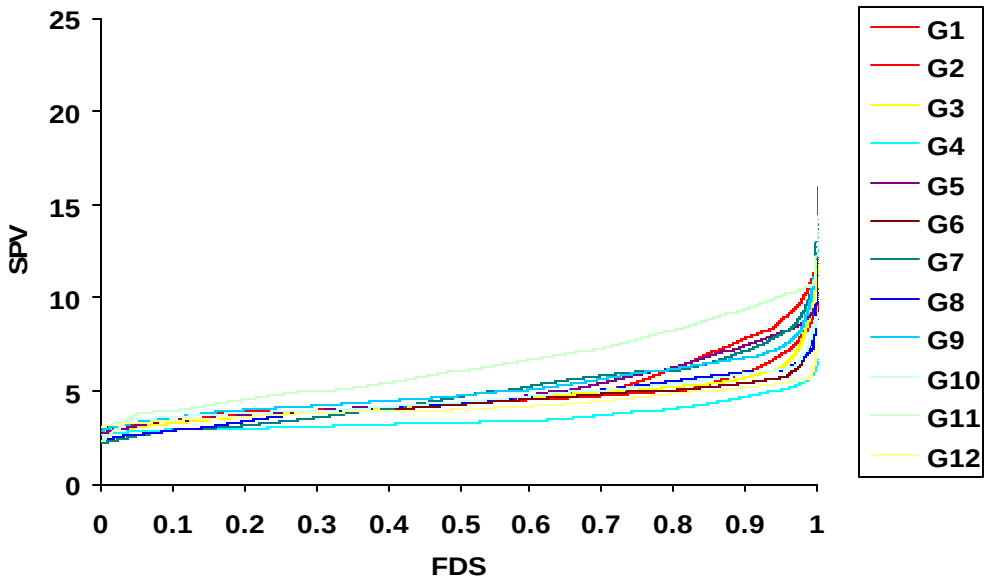

Figure IV.28: FDS Plot (Using SPV) for Cases: G1-G12 with Improved Allocation

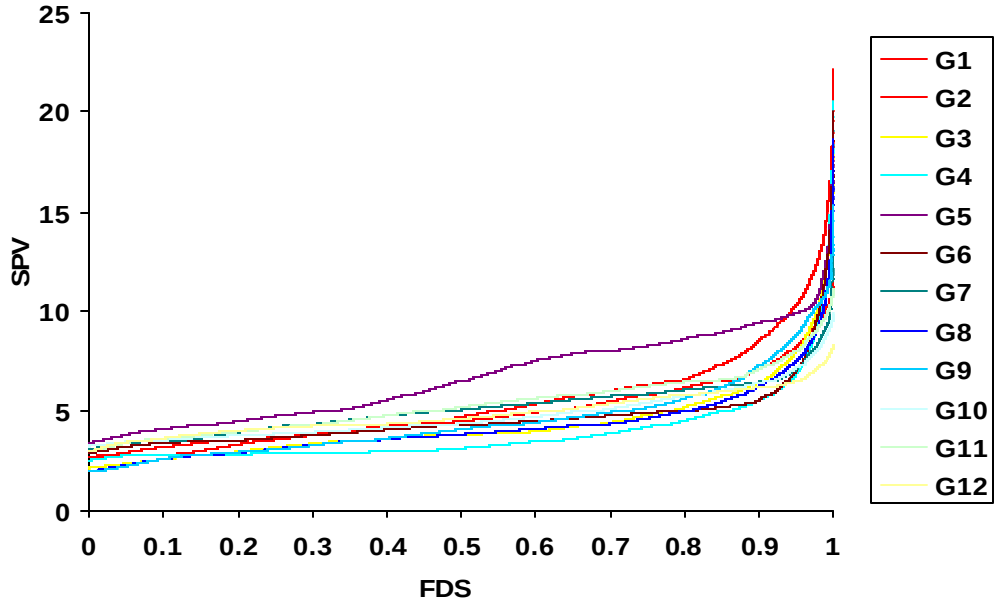

Figure IV.29: FDS Plot (Using SPV) for Cases: G1-G12 with Misjudged Allocation

When the researcher is interested in studying a design's prediction properties for the GLM with a second order linear predictor, it is harder to specify the initial parameter estimates because the surface of the predicted means is more complex. The researcher not

only needs to know the range of values but also the locations of the predicted means. When considering the PPVs, an unequal allocation of observations can improve the FDS curves if it is the correct allocation. However, if it is not the right allocation of observations, then the distribution of the PPVs will be worse than the equal allocation case. Typically this outcome is quite likely given the difficulty of getting all the estimates correct. For the SPV distribution, the unequal allocation whether it is improved or misjudged does not help improve the SPV distribution. Therefore we recommend that as long as the practitioner is not sure of the values and the locations of the predicted means, it is safer to use a design with equal number of observations at all of the design points.

IV.9 Studying Design Robustness for Poisson Data

In this section, we use FDS plots to study design performance for GLM models with Poisson data. After presenting the basic model, we examine different types of parameter misspecifications associated with having a Poisson response.

Consider an example involving an automotive manufacturing company interested in estimating the number of paint defects per area painted. They consider two important factors of the paint, x_1 (viscosity), and x_2 (mixing time), and their interaction on the response. A first order plus interaction linear predictor for the Poisson regression is applied with the canonical log link.

$$\ln(m_i) = b_o + b_1 x_{1i} + b_2 x_{2i} + b_{12} x_{1i} x_{2i} \qquad \text{(IV.12)}$$

The above equation can be rewritten to define the predicted mean count in terms of the parameters, i.e., $m_i = e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i}}$. In this model the variance of a given location of the design is assumed to be proportional to the mean. Hence the SPV is given by $n(\boldsymbol{x}_o) = N\boldsymbol{\mu}_o \boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o$ and the PPV is equal to $r(\boldsymbol{x}_o) = N\boldsymbol{\mu}_o^2 \boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o$ where V=diag(m_i). The chosen design is a 2^2 factorial design, in a square region coded so that each variable is in the range [-1, 1].

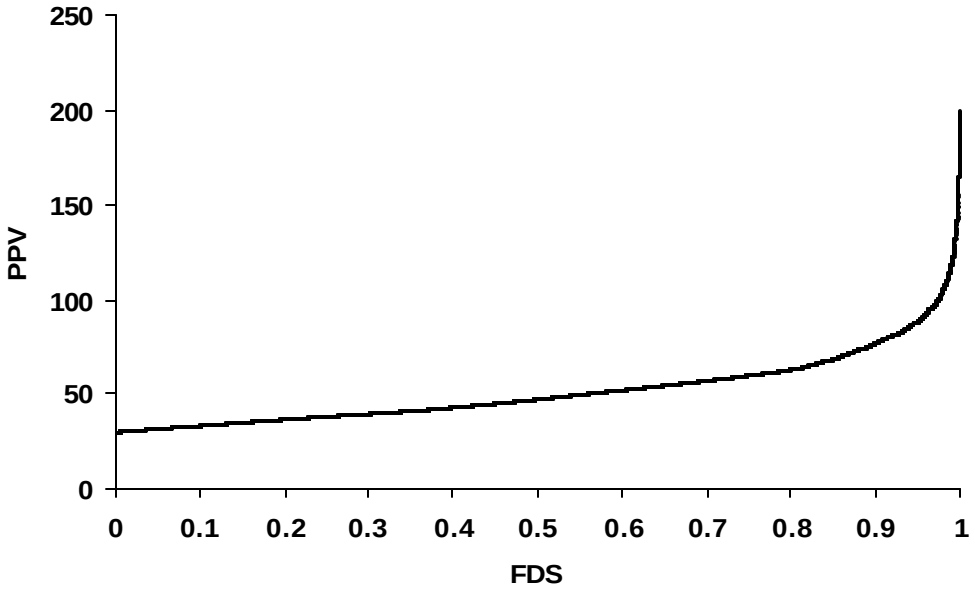

Figure IV.30: FDS Plot (Using PPV) for 2^2 Factorial Design

The practitioner estimates the predicted mean counts at each design location to be (25, 30, 15, 50) for {(-1, -1), (-1, 1), (1, -1), (1, 1)}, respectively. Figures IV.30 and IV.31 show the corresponding FDS plots using the PPV values and the SPV values, respectively. Table IV.8 summarizes the minimum and the maximum SPV and PPV values, the approximate G-efficiency, and the V-average for the factorial design. From Figure IV.30 we see that about 85% of the design space has relatively consistent and small PPV values, but then the slope increases steeply for large PPV values. In Figure IV.31, the design has a maximum SPV value of 4. Therefore, this design is G-optimal for this model.

Table IV.8 Summary of SPV and PPV values, Approximate G-efficiency, and V-average for 2^2 Factorial Design

Min SPV	Max SPV	V-average	G-efficiency
1.06	4.0	1.86	100%

Min PPV	Max PPV
29.76	200

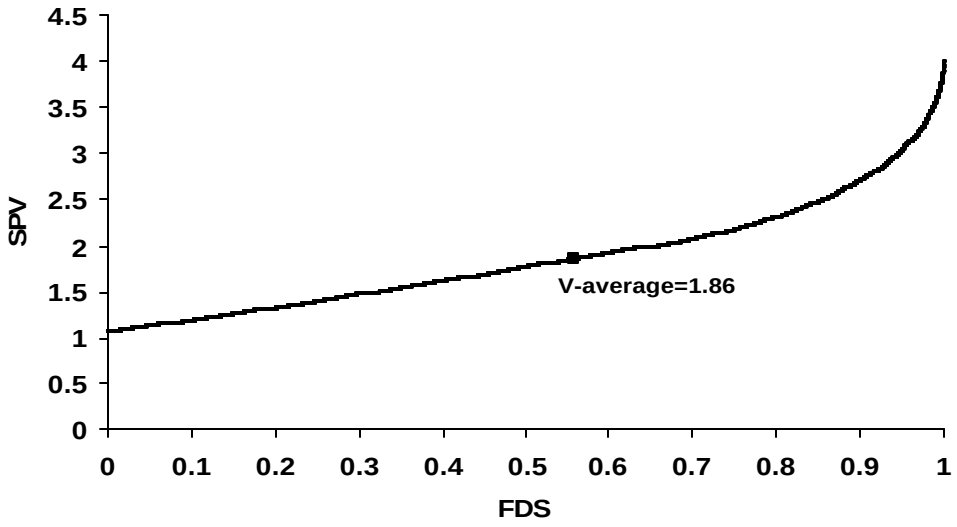

Figure IV.31: FDS Plot (Using SPV) for 2^2 Factorial Design

Similar to binomial regression, for Poisson regression design robustness is examined using the predicted mean counts, $\hat{m}$'s, directly. We do not consider robustness in terms of the $\hat{\beta}$ misspecifications because of the unlikelihood of practitioners thinking of the problem in this form and the issues of bounding the robustness shown in the binomial case. The researcher may consider two types of misspecifications in terms of the predicted mean counts: a shift in the range or a change in the spread of the range by a fixed percentage.

Even though the practitioner gave his/her best guess of the predicted mean counts, it is likely that these guesses are incorrect. Therefore he/she is also interested in quantifying the effect on prediction if the predicted mean counts are changed. Considering the design where the predicted counts, $\hat{m}$'s, are estimated to be $\hat{m}' = (\hat{m}_{-1,-1}, \hat{m}_{-1,1}, \hat{m}_{1,-1}, \hat{m}_{1,1})' = (25,30,15,50)'$, the range of the predicted counts can be shifted either by increasing or decreasing it. When the minimum and the maximum values are decreased by a certain amount, say 10 counts given by the practitioner, the range becomes $(\min \hat{m}_i, \max \hat{m}_i) = (5, 40)$. When the range is increased by 10, it becomes $(\min \hat{m}_i, \max \hat{m}_i) = (25,60)$. The predicted mean counts for the other two locations are

assumed to be the average of the minimum and the maximum predicted counts. Figure IV.32 compares the three cases in terms of SPV values.

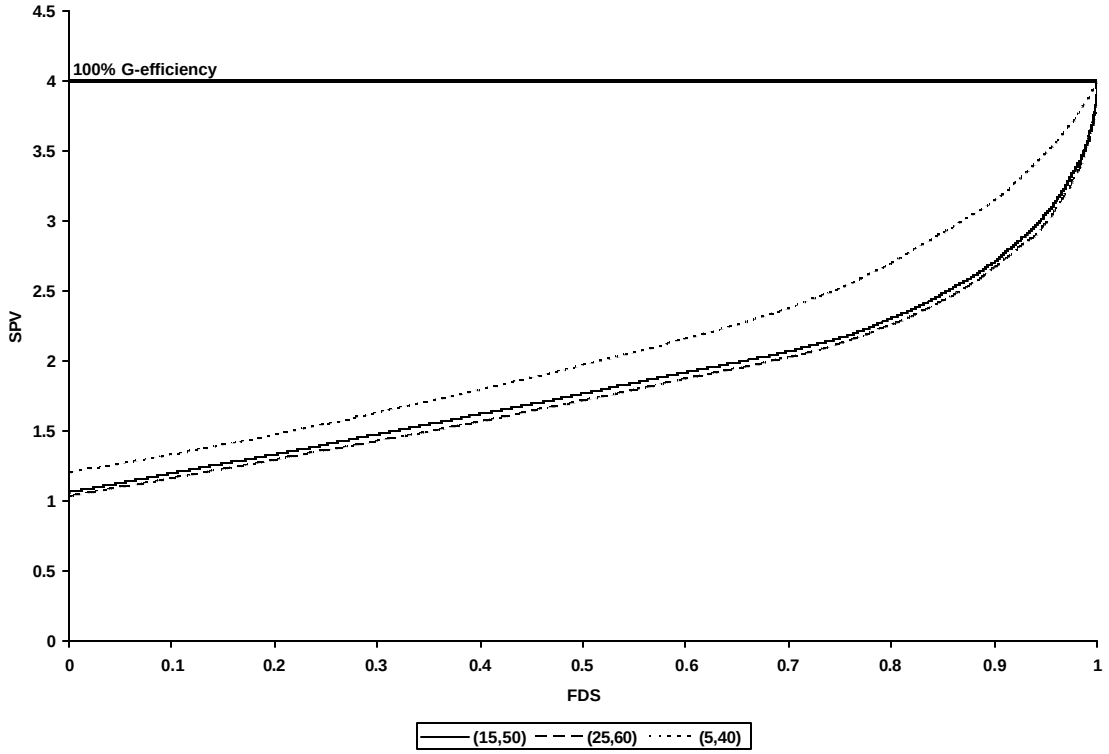

Figure IV.32: FDS Plot (Using SPV) for the Initial Predicted Mean Value, $\hat{m}' = (25,30,15,50)'$, and the Shifted Predicted Mean Count Range

Zahran and Myers (2003) discuss how the distribution of the variances at the design points influence the scaled prediction variance under the GLM framework. For a Poisson random variable the distribution variance at the design points is equal to the distribution mean at the design points since the mean and the variance of a Poisson random variable are the same. If the distribution of the variance at the design points has a small dispersion, then the SPV values are lower compared to the SPV values of the design with a more dispersed distribution variance since we are further away from the assumption of correct variance, for which the design is optimal. The dispersion of the distribution of the SPV is a direct function of the ratio of the maximum variance to the minimum variance at the design points. From Figure IV.32 we see that as the relative difference between the predicted mean counts increases, the SPV values increase. For the initially chosen

predicted mean range (15, 50) the ratio of the maximum variance to the minimum variance is equal to 50/15=3.33. When this range is decreased to (5, 40), the ratio of the maximum variance to the minimum variance increases to 40/5=8. Since the ratio increases, the SPV values increase resulting in a higher FDS curve. Similarly when the range is increased to (25, 60), the relative ratio of the maximum variance to the minimum variance becomes 60/25=2.4 resulting in smaller SPV values compared to the (15, 50) range. This confirms the results of Zahran and Myers (2003). In general, if the researcher is interested in the SPV distribution of a design, the robustness of the design depends on the relative ratio of the minimum and the maximum variances at the design points.

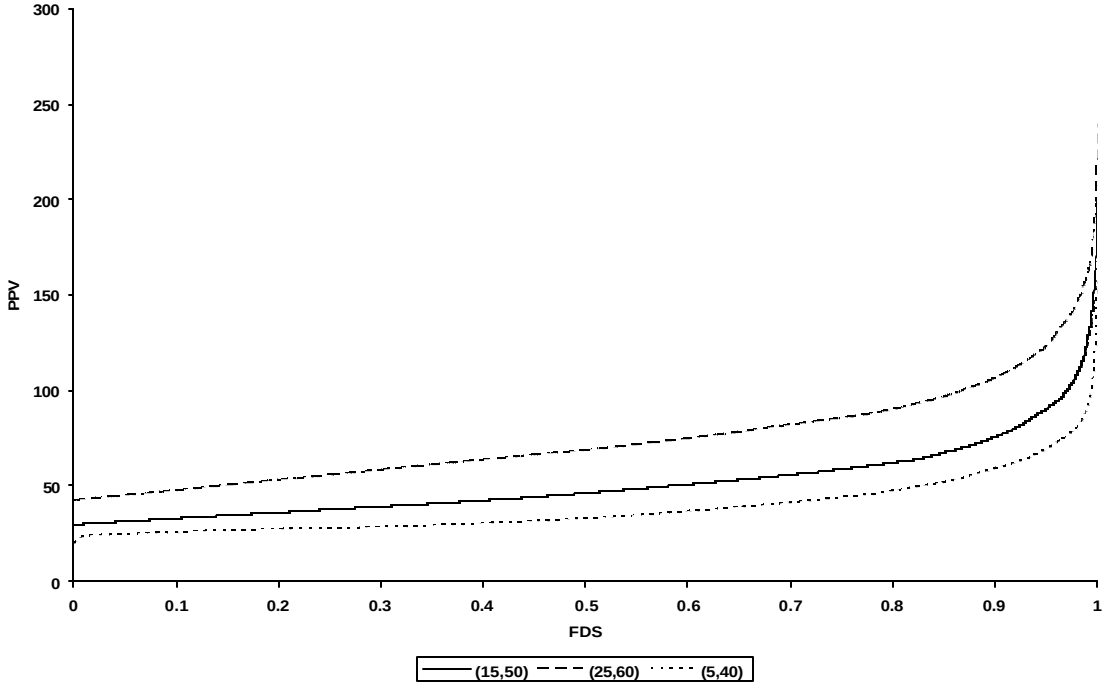

Figure IV.33: FDS Plot (Using PPV) for the Initial Predicted Mean Value, $\hat{m}' = (25, 30, 15, 50)'$, and the Shifted Predicted Mean Count Range

Figure IV.33 shows the FDS plots for the PPV values. The interpretation of the performance of the PPV values is different than the SPV values. The performance of the design's actual prediction is directly tied to the values of the variances (predicted mean counts for the Poisson case) at each design point. Since the form of the PPV is on the same scale as the variance of the estimated mean response, the design with higher variances at the design points will have higher PPV values. As the variance at any design

point increases, the PPV increases. As seen from Figure IV.33 since the (25, 60) range has the highest variances, it has the highest FDS curve. Similarly, the (5, 40) range has the lowest curve since it has the lowest variances at each design point. The higher the predicted mean counts at each design point, the higher the PPV values.

Another type of misspecification that can be considered is to examine the effect on the SPV and PPV values when the spread of the range of the predicted mean counts is increased or decreased by a fixed percentage. Considering the example where the predicted mean counts, were $\hat{m}' = (\hat{m}_{-1,-1}, \hat{m}_{-1,1}, \hat{m}_{1,-1}, \hat{m}_{1,1})' = (25,30,15,50)'$, the practitioner believes that his initial guesses may be off at most by a fixed percentage of 40%. 40% corresponds to $(50\text{-}15) \times 0.4 = 14$ counts. Keeping the average predicted count the same, we increase or decrease the range by 7 counts from each end producing ranges of (15+7, 50-7) = (22, 43) and (15-7, 50+7) = (8, 57), respectively. The predicted mean counts for the other two locations are assumed to take the average of the minimum and the maximum predicted mean counts.

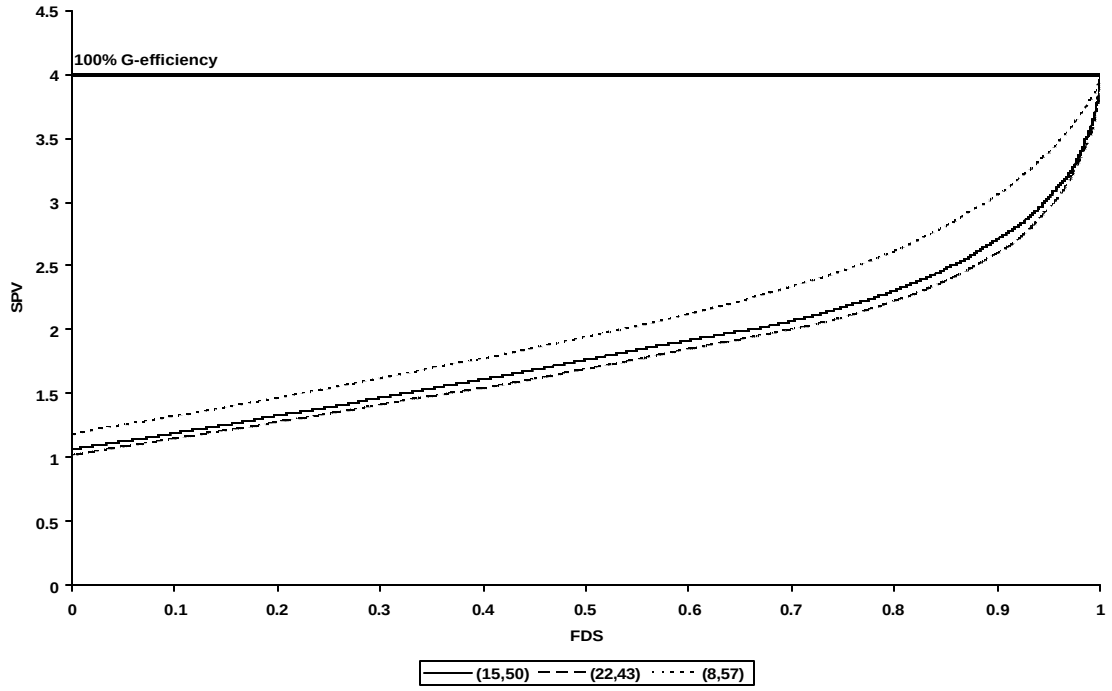

Figure IV.34: FDS Plot (Using SPV) for the Initial Predicted Mean Value, $\hat{m}' = (25,30,15,50)'$, and the Increased/Decreased Predicted Mean Count Range

Figure IV.34 shows the FDS plots of the design with this misspecification type using the SPV values. Similar to the previous misspecification type, the robustness of the design depends on the ratio of the maximum and the minimum variances at the design points for the SPV values. As the ratio decreases to (43/22)=1.95 from (50/15)=3.33, the SPV values decrease, and as the ratio increases to (57/8)=7.13, the SPV values increase. As seen from Figure IV.34 the smaller the relative difference between means the smaller the SPV values.

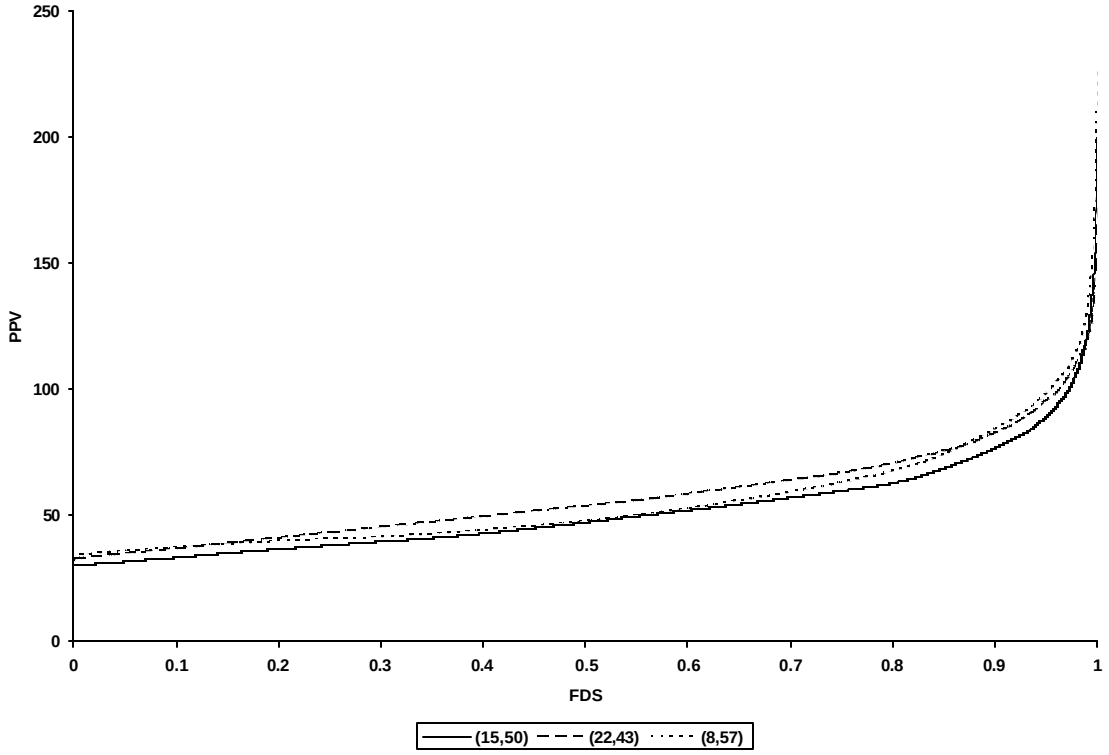

Figure IV.35: FDS Plot (Using PPV) for the Initial Predicted Mean Value, $\hat{m}' = (25,30,15,50)'$, and the Increased/Decreased Predicted Mean Count Range

Figure IV.35 shows the FDS plots comparing the three cases using the PPV values. Once again for the PPV values, we need to consider the values of the variances at each design point. Compared to the misspecification where the range of the predicted means was either decreased/increased by a certain amount, this misspecification is more complicated. When the three ranges are considered, we see that there are intersections

between the three FDS curves. This is due to the overlapping of the variances for the three ranges. We see that while a range may have the smallest minimum variance compared to the other two ranges, it may not have the smallest maximum variance. Since some variances of one range fall within the variances of another range, there are intersections within the FDS curves.

The method of changing the range by a fixed amount could also be applied to the first type of misspecification where a 40% shift in range would result in an upward shift of range of (22, 57) and a downward shift of range of (8, 43). The third type of misspecification, the misspecification of the location of the predicted mean values mentioned in Section IV.7, can also be adapted to the Poisson case.

Overall, changing the ranges of the predicted means has a different effect on the robustness of the design for SPV and PPV measures. The effect on the SPV is more closely related to the ratio of the variances, while for the PPV, the absolute magnitude of the variance plays a larger role.

IV.10 GLM vs. Data Transformation to Stabilize Variance

In this section, we consider the Poisson example to compare the prediction properties of a design using a GLM to the same design where we analyze the data using a transformation of the response. When data that do not follow a normal distribution are observed, a common procedure is to transform the response to try to achieve a stable variance. This way the researcher is still able to use ordinary least squares to fit a linear model to the transformed data. Even though, a transformation does not always guarantee stable variance and normally distributed random errors, this method is widely used. The researcher needs to be very careful while using this approach, as it does not use the natural variance of the data's distribution like a GLM does and one needs to be careful about the assumptions he/she makes regarding the error structure while using a transformation. The main question in choosing between the two approaches is not about which produces better design characteristics but rather about the advantages/ disadvantages that come with the type of analysis being used. The transformation can

never perform better than the GLM on the analysis side. While doing the analysis, the researcher needs to take into account the natural variance of the data's distribution. If the range of variances is not very large, then the transformation can perform quite adequately. For more details, see Myers and Montgomery (2003, page 281). Moreover, Myers and Montgomery give examples of some cases where the GLM performs better in terms of precision of estimation and prediction.

To be able to compare a design with transformed data to the same design with a GLM on the same scale in terms of prediction, we need to obtain the SPV/PPV values of the design with the transformed data on the same scale as the design with the GLM. The transformed SPV of $\hat{y}(\boldsymbol{x}_o)$ at $\boldsymbol{x}_o$ turns out to be identical to the untransformed case or $N\boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o$. The transformed PPV of $\hat{y}(\boldsymbol{x}_o)$ at a location in the design space looks like $N m_o \boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o$. See Appendix-C for more details.

Consider that the practitioners in the Poisson example are interested in how the design would perform if the Poisson response was transformed, instead of using the recommended GLM approach. The variance stabilizing transformation corresponding to a Poisson response is the square root transformation. When the transformation is applied to the data, the homogeneous variance model looks like $\mathrm{E}(\sqrt{y_i}) \cong \boldsymbol{x}_i' b$. The square root transformation on the response y_i, will result in a constant variance of y_i, i.e., $\mathrm{var}(\sqrt{y_i}) \cong s^2 = 1/4$ (Myers and Montgomery, page 243).

Figure IV.36 shows the FDS plot of 2^2 factorial using SPV values where the data are both to be analyzed assuming a GLM and the variance stabilizing transformation. As seen from Figure IV.36 the transformation performs better throughout the whole design space (except at the four design corners) but the difference is very small. The design with the transformation will always be G- and V-optimal for a first order model and the first order model with interaction. Therefore, the design with a GLM will always have higher SPV values compared to the design with the transformation. The only exception is when the

design with the GLM has the same value of variances (predicted mean counts for Poisson case) at each design point. Then the design with the GLM will have the same SPV values as the design with the transformation. However, recall that the main issue in choosing between these two approaches should be the analysis not which technique offers better design properties.

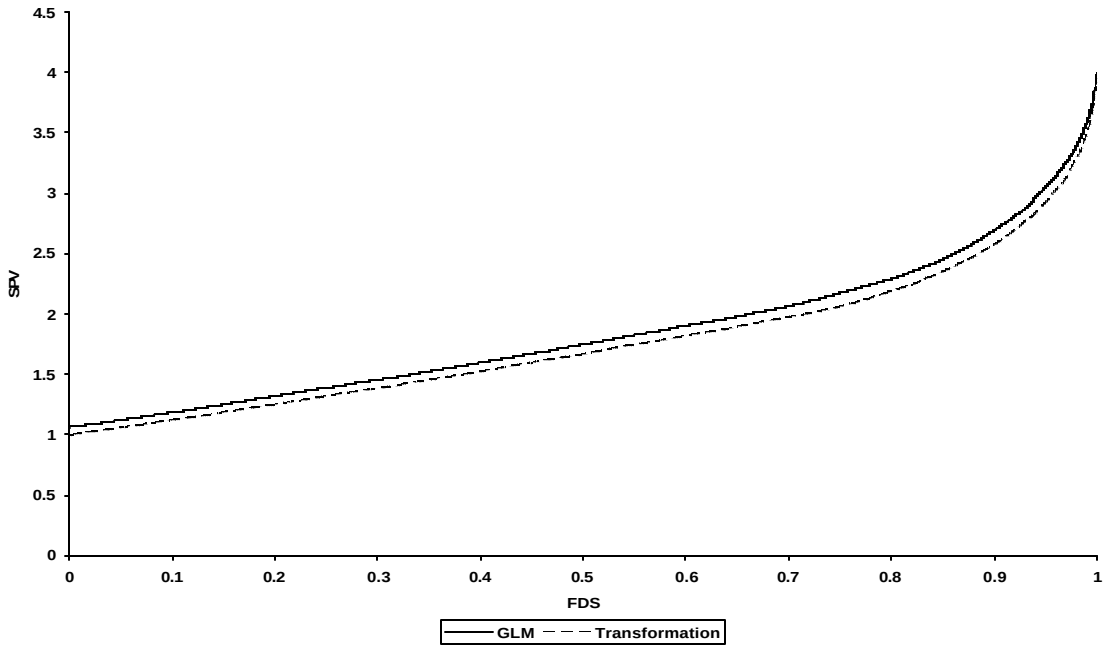

Figure IV.36: FDS Plot (Using SPV) of the 2^2 Factorial Design for the GLM and the Transformation

For the 2^2 factorial design, the two curves perform similarly when the researcher is interested in actual prediction. Figure IV.37 shows the FDS plot of the two cases using the PPV values for the 2^2 factorial design with a Poisson response. As seen from Figure IV.37, the approach using the variance stabilizing transformation performs better for the initial 78% of the total design space and at FDS=1 but its PPV values are higher for the rest of the fraction of design space.

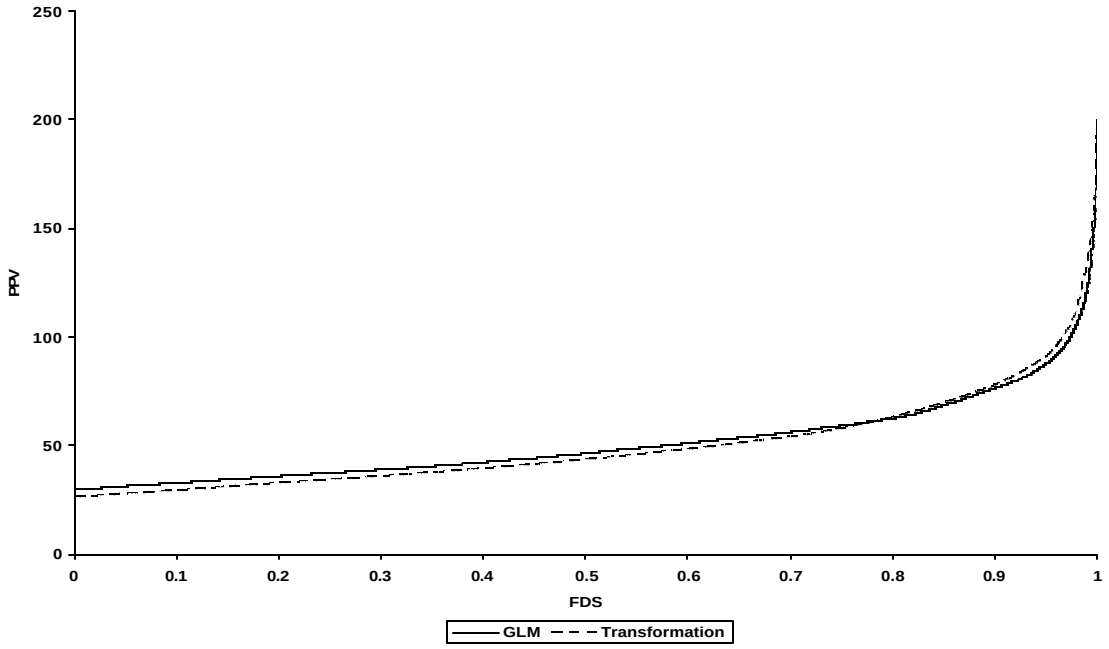

Figure IV.37: FDS Plot (Using PPV) of the 2^2 Factorial Design for the GLM and the Transformation

As the number of terms in the model increases, it is harder to obtain G- or V-optimal designs. Therefore, the design with the transformation may not necessarily perform as well as the design under the GLM framework. To study this, we used the 3^2 factorial design with a second order linear predictor used in Section IV.7. Again the Poisson distribution was assumed. The GLM with the second order linear predictor looks like $m_i = e^{(b_o + b_1 x_{1i} + b_2 x_{2i} + b_{12} x_{1i} x_{2i} + b_{11} x_{1i}^2 + b_{22} x_{2i}^2)}$. The practitioner was asked to give an estimate of the predicted mean counts at each design location, and he gave (134, 12, 33, 7, 4, 11, 10, 74, 8) for {(1, 1), (1, -1), (-1, 1), (-1, -1), (-1, 0), (1, 0), (0, -1), (0, 1), (0, 0)}, respectively. Figure IV.38 shows the FDS plot of the 3^2 factorial design using SPV values where both a GLM and a transformation are assumed. As seen from Figure IV.38, the SPV surface of the design is now more complicated. The design under the GLM framework performs better 67% of the total design space. However towards the end of the fraction of design space, the design with the transformation performs better.

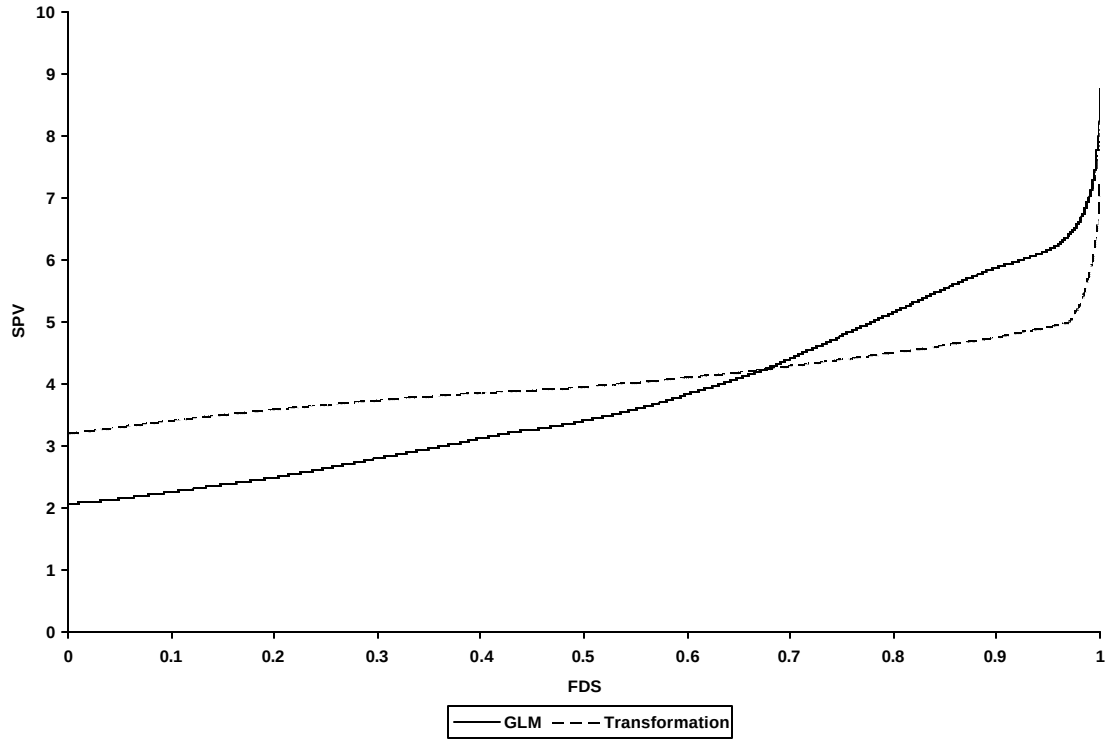

Figure IV.38: FDS Plot (Using SPV) of the 3^2 Factorial Design for the GLM and the Transformation

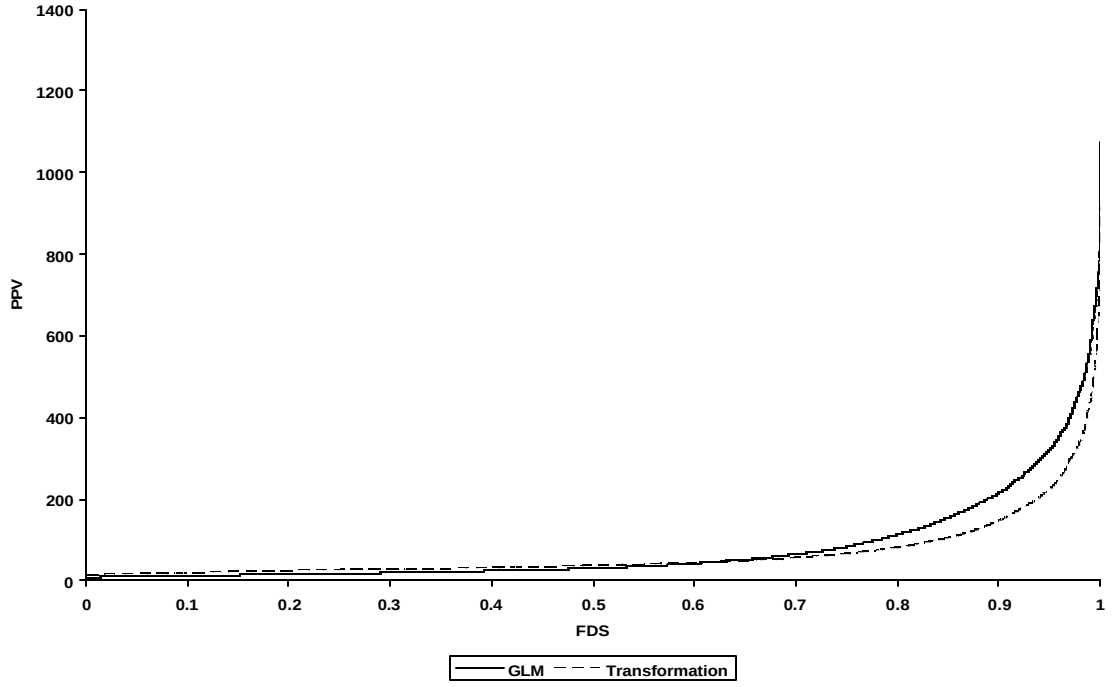

Figure IV.39: FDS Plot (Using PPV) of the 3^2 Factorial Design for the GLM and the Transformation

Figure IV.39 shows the performance of the 3^2 factorial design under the GLM framework and the transformation using PPV values. As seen from Figure IV.39 the design under the GLM framework performs better for 60% of the total design space.

As the number of terms in the model increases, the surface of the SPV and the PPV distributions may become even more complex. If the researcher is interested in comparing prediction for both types of analyses before the data are collected, FDS plots are very useful tools in terms of studying the performance of a design. However, we again emphasize that analysis, not design considerations, should be more influential on making this choice.

IV.11 Conclusion

Initially, FDS plots were developed to compare designs with standard linear models. However, recently an interest developed for examining response surface designs for the GLM case. In this dissertation, we adapted FDS plots to compare designs for the GLM case under a variety of misspecification scenarios. FDS plots are easy to implement and are practical to study designs for the GLM case.

The FDS plot plots the SPV values across the fraction of the design space. For the GLM case, the formula for the SPV needs to be adapted since it depends on the initial parameter estimates when non-normal data are used. In addition to the SPV, a new term called the penalized prediction variance (PPV) is introduced. The difference between the SPV and the PPV is that the SPV has a variance term divided out of the estimated mean response of its formula. The PPV becomes important when the researcher is interested in actual prediction. We plot the PPV values across the fraction of design space to examine the behavior of the prediction variance of a design.

We used the 2^2 factorial design and some space-filling designs with a first order plus interaction model for the linear predictor in the case of the logistic regression to illustrate how the adapted FDS plot can compare designs given a set of initial parameter estimates. One concern of the experimenter is that the initial parameter estimates are rarely correct.

Even though it is an educated guess, the experimenter should also be interested in how robust SPV/PPV values are to parameter misspecifications. Two general types of misspecifications are analyzed: in terms of the initial parameter estimate, $\hat{\beta}$, and the predicted mean value, $\hat{\mu}$. The misspecification in terms of $\hat{\beta}$'s is not suitable for every situation since this may not be an intuitive metric for the practitioner. Even though they are useful considering the PPV distribution of a design, they may be misleading when a design's SPV distribution is of interest.

Several types of misspecifications are available when the predicted mean values are considered. These cases are examined in detail using the 2^2 factorial design with binomial and Poisson data. When the SPV distribution of a design is considered, the robustness to changes in the predicted means is linked to the ratio of maximum to minimum variance of the SPV distribution at the design points. As the ratio increases, the SPV values increase and similarly as the ratio decreases, the SPV values decrease. For the binomial case, assuming every predicted mean value at the design points is greater than 0.5, when the predicted mean range is decreased, the ratio of the maximum to the minimum variance becomes smaller, resulting in a more stable SPV distribution with lower SPV values. However, as the range increases the ratio of the maximum to the minimum variance increases and since the predicted means are closer to 1, the SPV values are larger with a steeper FDS curve. When the PPV distribution for the binomial case is of interest, the actual size of the variances at each design point is important. Assuming all the predicted mean values at the design points are greater than 0.5, as the predicted mean range increases, the highest predicted mean increases decreasing the variance at that design point and the lowest predicted mean decreases increasing the variance at that design point. Therefore, at the initial FDS values, the PPV values decrease and towards the end of the fraction of design space, the PPV values increase. For the mid-FDS values, it is hard to conclude results because it depends on the distribution of the predicted means at those design locations. Shifting the range up and down results in similar conclusions to increasing and decreasing the range.

The third type of misspecification for the binomial case was considered when the predicted locations of the predicted means are changed. If the design is symmetric, changing the location of the predicted means will not change the SPV/PPV performance of the design. This type of misspecification is important for non-symmetric designs. Putting more observations at the locations with larger variances can stabilize the PPV distribution. However, if the locations of the predicted means are guessed incorrectly, then putting more observations at the locations with smaller variances may result in a steeper PPV distribution. Therefore, the practitioner needs to be cautious while making decisions about the allocations of observations. Since the SPV at location x_o is divided by the variance at that location, the results related to SPV for different misspecifications are hard to interpret. However, since the 2^2 factorial design for the GLM with a first order linear predictor is G- and V-optimal, any unequal allocation of observations will result in a worse SPV distribution.

For the Poisson case, two types of misspecifications were studied based on the predicted mean values. In contrast to the binomial case, when the predicted mean count range is kept fixed and shifted up, the ratio of the maximum to the minimum variance at the design points decreases and the SPV values are lower. As the range is shifted down, the ratio is larger and the SPV values increase. For the PPV values, when the range is shifted down, the predicted mean counts are smaller and hence the PPV values become smaller as well. As the range is shifted up, the predicted mean counts increase and the PPV values become larger. When the width of the range of predicted means is changed, the SPV values increase or decrease depending on the change of the ratio of the maximum to the minimum variance at the design points.

The PPV and the SPV values for GLMs are examined simultaneously for the two types of misspecifications. Since the two measures of prediction variance are not fixed multiples of each other as they were in the standard linear models case, their curves frequently look different from each other. The FDS plots obtained using SPV values have the same scaling as the G-efficiency while the FDS plots obtained using PPV values do not. Examining the two measures together gives the researcher a better and more detailed tool

to choose the best design possible. The FDS plots with SPV values can be used to study a design's theoretical properties such as the scaled prediction variance, G-efficiency and V-averages, while the FDS plots with PPV values helps the practitioner study the actual prediction variance properties of a design. Considering the SPV values for the 2^2 factorial design with a first order linear predictor, as the $\hat{p}_i$ values approach 0 or 1, the G-efficiencies decrease. Considering the PPV distribution, the location of the minimum and the maximum PPVs are directly tied to the locations of the actual values of the predicted means within the design space. Moreover, as long as an appropriate predicted mean range is specified, the design with equal allocation of observations performs better than other types of allocations in terms of G- and V- efficiency.

In addition to a first order model linear predictor, we also considered a second order linear predictor under the logistic regression. We used FDS plots with the SPV/PPV values as well as contour plots to analyze the standard 3^2 factorial design in terms of its prediction properties. Simulations considered changing the allocation of the observations at each design point and the effects on the SPV/PPV distributions are compared. In conclusion, equal allocation of the number of observations at each design point gives the most consistent and robust performance.

In the last section of Chapter IV we studied a design's prediction performance for both an analysis with GLM and with a variance stabilizing data transformation. As the number of terms in the model increases, it becomes harder to obtain G- or V-optimal designs. Therefore, the design with the transformation may not necessarily perform as well as the design under the GLM framework. It is important to remember that using the GLM approach always gives superior analysis results than the design with the data transformation, and this should likely be the determining factor in how to proceed with the analysis.

Chapter V

FRACTION OF DESIGN SPACE PLOTS FOR EXAMINING MIXTURE DESIGN ROBUSTNESS TO MEASUREMENT ERRORS

V.1 Introduction

In this chapter, we study another proposed adaptation of the FDS plots, examining mixture design robustness to measurement errors in production. These errors are also called mixing measurement errors because they appear during the production of mixture experiments. Mixing measurement errors can appear due to inaccurate measurements of the individual component amounts and are not an intentional part of the original designed experiment. Steiner and Hamada (1997) and Hamada, Martz and Steiner (2002) have examined a mixture design's robustness to mixing measurement errors for specifying the correct optimum. In this dissertation, we use FDS plots to examine design robustness to mixing measurement errors for scaled prediction variance. By taking the possibility of mixing measurement errors into consideration before running the actual experiment, the practitioner may be better able to select a design with good prediction properties.

V.2 Definition of Absolute and Relative Error

As with Steiner and Hamada (1997) paper, two types of mixing measurements errors in the component amounts are discussed in this dissertation: relative (multiplicative) error and absolute (additive) error. The error is defined to be the difference between the actual amount and the desired component amount. The magnitude of the absolute error remains the same regardless of the amount of ingredient used. An example of an absolute error might be if the measurement equipment is systematically off by a certain amount causing the actual amount to be higher than the desired amount. The magnitude of the difference between the actual and the desired amount is not dependent on the magnitude of the desired amount. The magnitude of the relative error is proportional to the size of the intended component amount. See Sirohi and Radha-Krishna (1991, page 30) for more information about measurement errors.

Let A_i represent the desired mixture amount and e_i represent the error in the mixture amount for component *i*. Then the actual component proportions are given by

$$x_i(rel) = \frac{A_i(1+e_i)}{\sum_{j=1}^{q} A_j(1+e_j)} \quad \text{and} \quad x_i(abs) = \frac{(A_i+e_i)}{\sum_{j=1}^{q}(A_j+e_j)}, \tag{V.1}$$

for $i=1,\ldots,q$. See Steiner and Hamada (1997) for more details.

In this dissertation, we assume that we begin with $\sum_{i=1}^{q} A_i = 1$ for the intended design point, which means that we assume that the total amount of the mixture intended is equal to 1. We also assume that the same error structure (absolute or relative, and the same range of error amount) holds for all mixture component proportions. Moreover, we consider the errors to be either uniformly distributed from $U(-a_i, a_i)$ or normally distributed from $N(0, s_{E_i}^2)$, although other distributions could be used. The choice between using absolute error or relative error is a decision made by the practitioner based on understanding of the process. It can incorporate knowledge from historical data or expert judgment to choose the error distribution and the size of the error.

To illustrate, for a three component mixture experiment assume a design point of A=(0.5, 0.3, 0.2) is part of a larger design. For example, if we have assumed absolute errors from a uniform distribution with bounds [-0.1, 0.1], then a particular instance of the design point, A, might be obtained for the absolute error in measurement of E=(+0.1, -0.05, +0.07). Then the actual design point, A^*, becomes A^*=(0.5+0.1, 0.3-0.05, 0.2+0.07)=(0.6, 0.25, 0.27). Due to the absolute error affecting each component proportion, the sum of the proportions has been changed from 1 to 1.12 (=1+0.1-0.05+0.07). Therefore A^* needs to be rescaled. The standardized A^*, becomes $A^{**} = \frac{A^*}{1.12} = (\frac{0.6}{1.12}, \frac{0.25}{1.12}, \frac{0.27}{1.12})$ = (0.536, 0.223, 0.241). The change in each design point will affect the actual design's prediction performance relative to the intended design.

Unlike absolute error, relative error is proportional to the size of the intended component amount. To illustrate, we begin with the same design point, A = (0.5, 0.3, 0.2). The actual component proportions for A with relative errors from the uniform [-0.1, 0.1] becomes, $R^* = \left(\frac{0.5*(1+0.1)}{1.049}, \frac{0.3*(1-0.05)}{1.049}, \frac{0.2*(1+0.07)}{1.049}\right) = (0.524, 0.272, 0.204)$

where $1.049 = 0.5*(1+0.1)+0.3*(1-0.05)+0.2*(1+0.07)$ by equation V.1.

Steiner and Hamada (1997) studied the effect of mixing measurement errors on the mixture proportions. The authors provided techniques for determining mixture proportions that are robust to measurement errors. They used a loss function to evaluate the robustness of a mixture design to mixing measurement errors. Hamada, Martz and Steiner (2002) showed a Bayesian approach to account for absolute mixing measurements errors. Using a simulation study and a glass mixture experiment, they showed that the Bayesian approach gave better model variance estimates leading to better prediction. In this chapter, we adapt FDS plots to account for measurement errors in mixture experiments. Using this approach, the researcher is able to obtain information about a mixture design's robustness to measurement error before the data are collected.

V.3 FDS Plots Adapted for Mixing Measurement Errors

Given a design and an expected amount of error, either quantified in absolute or relative terms, the researcher would like to be able to see the effect of measurement error on SPV under actual experimentation. The single curve of FDS plots, which shows the SPV if the design were run exactly as intended, is supplemented with two additional curves. These show the ranges of values for the SPV distributions in the design space that might be obtained if the experiment were run with some errors. These curves give us a lower and an upper bound on what may be expected for prediction precision.

Given the intended design space and the type of misspecification, new designs are generated through simulation by sampling the amount of error for each component for each design point which mimics the assumed misspecification. These new designs are used to obtain a (1-α) bound for the intended design's SPV values. For each generated

design, the corresponding FDS curve is obtained. At each fraction of the design space, the $\frac{a}{2}$ and $(1-\frac{a}{2})$ quantiles are obtained from all the simulated FDS curves. The (1-a) upper and lower bounds for the FDS plot are obtained by connecting these quantiles. This way, the researcher is able to see how robust a design's prediction properties are to misspecifications in its design points.

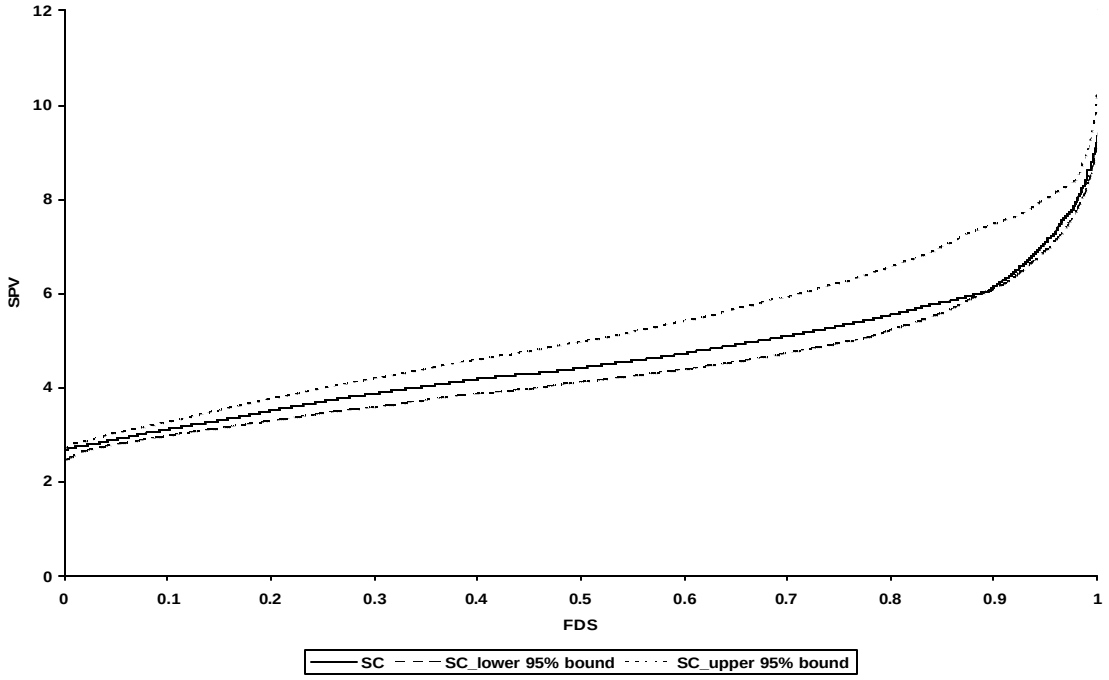

Figure V.1: FDS Plot of the Simplex-centroid Adapted for Absolute Measurement Error from U(–0.1, 0.1)

Table V.1: Summary of SPV values, Approximate G-efficiencies and V-averages for the Adapted Simplex-centroid with Absolute Error from U(–0.1, 0.1)

	Min SPV	Max SPV	V-average	G-efficiency
95% Lower Bound	2.43	9.41	4.36	74%
SC	2.68	9.34	4.56	75%
95% Upper Bound	2.70	10.24	5.2	68%

Figure V.1 shows an example of an adapted FDS plot of the augmented simplex-centroid design (Figure III.10) assuming the special cubic model. The absolute errors are chosen

from a uniform distribution, $U(-0.1, 0.1)$. Table V.1 gives a summary of the SPV values, the approximate G-efficiencies, and the V-averages.

From Figure V.1 we see that the 95% empirical interval curves around the original SPV values do not resemble a standard confidence interval. The original design is closer to the lower bound curve except at the initial and final fractions of the design space. This is to be expected because arbitrary errors tend to worsen prediction precision rather than improve it. The original design still has the lowest maximum SPV value because any change in the location of the design points will result in a less desirable design. The intended points in the design space were chosen to obtain good prediction properties, and errors from those locations are likely to reduce prediction.

Figure V.2 shows the same design in Figure V.1 with additional intervals for 80% and 90% levels, since this allows us to gain an understanding of the tails of the SPV distribution under the assumed error structure. As seen from Figure V.2, there is relatively little difference between the three sets of bounds, in other words there is a small spread of the SPV values at the tails of the distribution.

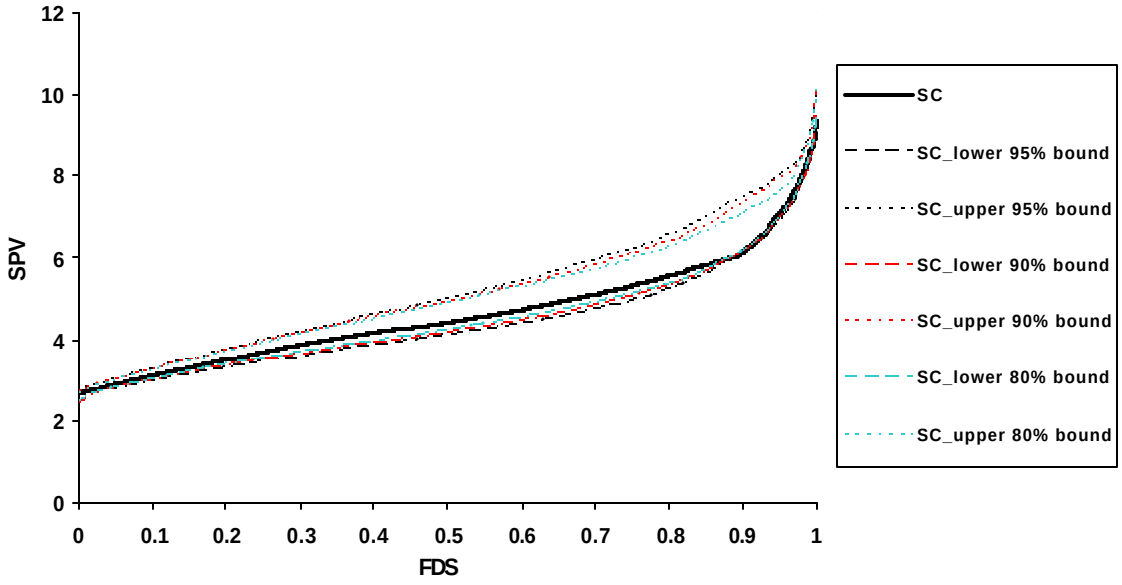

Figure V.2: FDS Plot of the Simplex-centroid with 80%, 90% and 95% Bounds with Absolute Error from $U(-0.1, 0.1)$

FDS plots are also useful for comparing mixture designs for these type of measurement errors. Figure V.3 shows the comparison of two mixture designs: the augmented simplex-centroid and the simplex lattice (Figure III.10) assuming the special cubic model. The absolute error sampled from a uniform distribution, U(–0.1, 0.1), and a 95% bound are considered. As seen from Figure V.3 and from Section III.3.3, the simplex-centroid performs better than the simplex lattice for the entire design space. Under uniform error misspecification, the intersection of the lower bound of the simplex lattice and the upper bound of the simplex-centroid indicates that some of the most advantageously misspecified simplex lattice designs perform better than some of the worst misspecified simplex-centroid designs. For most of the fraction of design space, the bands along the intended simplex-centroid FDS curve are slightly wider than the bands for the intended simplex lattice. However, towards the end of the fraction of design space the upper 95% curve for the simplex lattice becomes further apart from its initial curve compared to the difference between the simplex-centroid curve and its 95% band. This indicates that the simplex lattice is less robust to mixing measurement error than the simplex-centroid for larger SPV values.

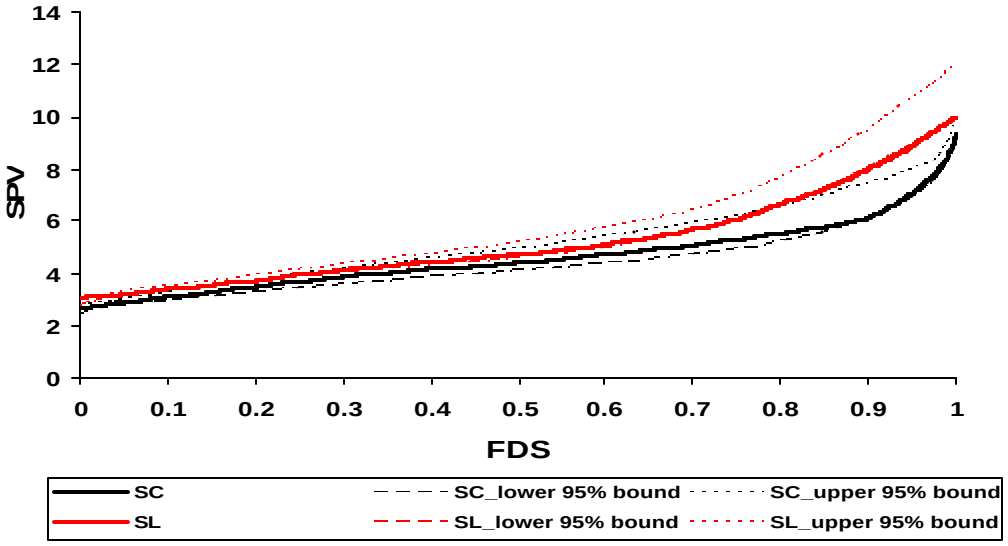

Figure V.3: Comparison of the Augmented Simplex-centroid and the Simplex Lattice for Robustness to Absolute Error from U(–0.1, 0.1) Using FDS Plots

As the absolute error range for the uniform distribution increases, the difference between the upper 95% bound and the initial simplex lattice curve becomes even larger. This can be observed from Figure V.4 where we assume the absolute error rate can be as large as 0.2.

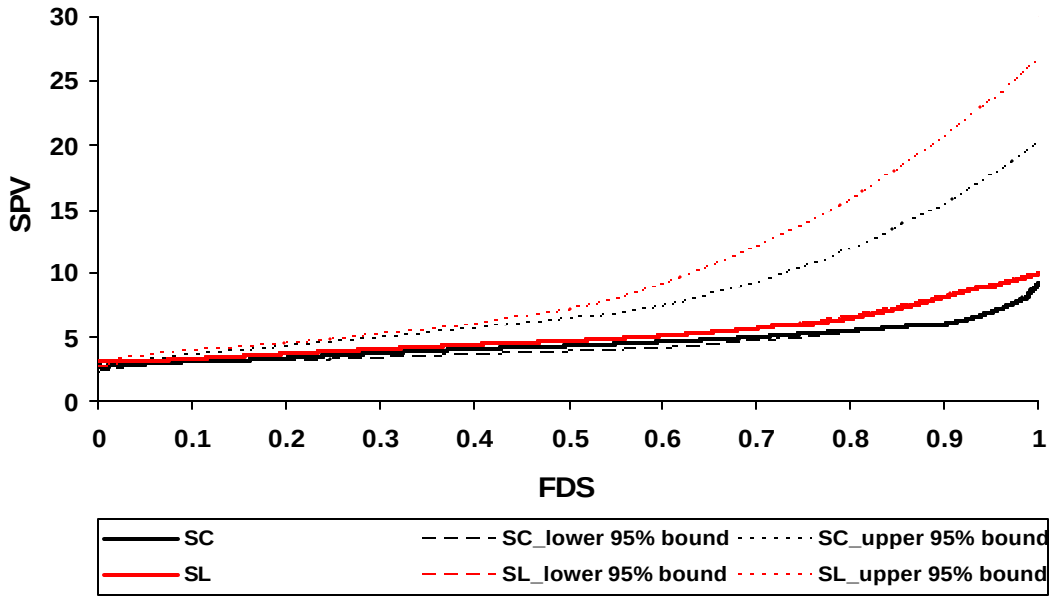

Figure V.4: Comparison of the Augmented Simplex-centroid and the Simplex Lattice for Robustness to Absolute Error from $U(-0.2, 0.2)$ Using FDS Plots

The second error type considered is relative error. For some mixture experiments, relative error becomes of interest if the size of the error is expected to be proportional to the size of the intended amount of each ingredient. Hence, for relative error, each ingredient has a distinctive range of misspecification for each design point. Figure V.5 shows the adapted FDS plot for the augmented simplex-centroid design where the relative errors are chosen from a uniform distribution, $U(-0.2, 0.2)$. Table V.2 gives a summary of the SPV values, the approximate G-efficiencies, and the V-averages. As seen from Figure V.5, the upper 95% bound has a higher maximum SPV compared to the initial SPV curve. Both the upper and the lower SPV bands have values close to the intended SPV curve indicating a robust design given the specific misspecification.

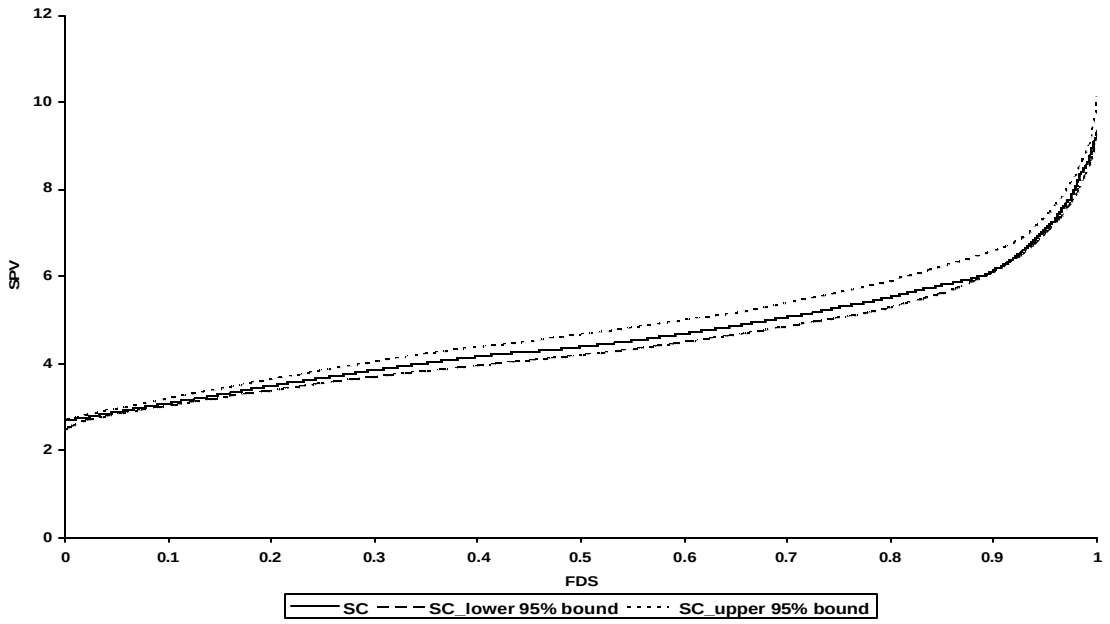

Figure V.5: FDS Plot of the Simplex-centroid Adapted for Relative Measurement Error from U(–0.2, 0.2)

Table V.2: Summary of SPV values, Approximate G-efficiencies and V-averages for the Adapted Simplex-centroid with Relative Error from U(–0.2, 0.2)

	Min SPV	Max SPV	V-average	G-efficiency
95% Lower Bound	2.49	9.33	4.44	75%
SC	2.68	9.34	4.56	75%
95% Upper Bound	2.69	10.16	4.85	69%

Figure V.6 shows the augmented simplex-centroid with additional intervals of the initial SPV values. From Figure V.6 we see there is little difference between the three interval bands, i.e., the distribution of the SPV for a wide variety of error choices is stable. In terms of the spread of the bounds, there is a slight difference between the relative and the absolute error cases.

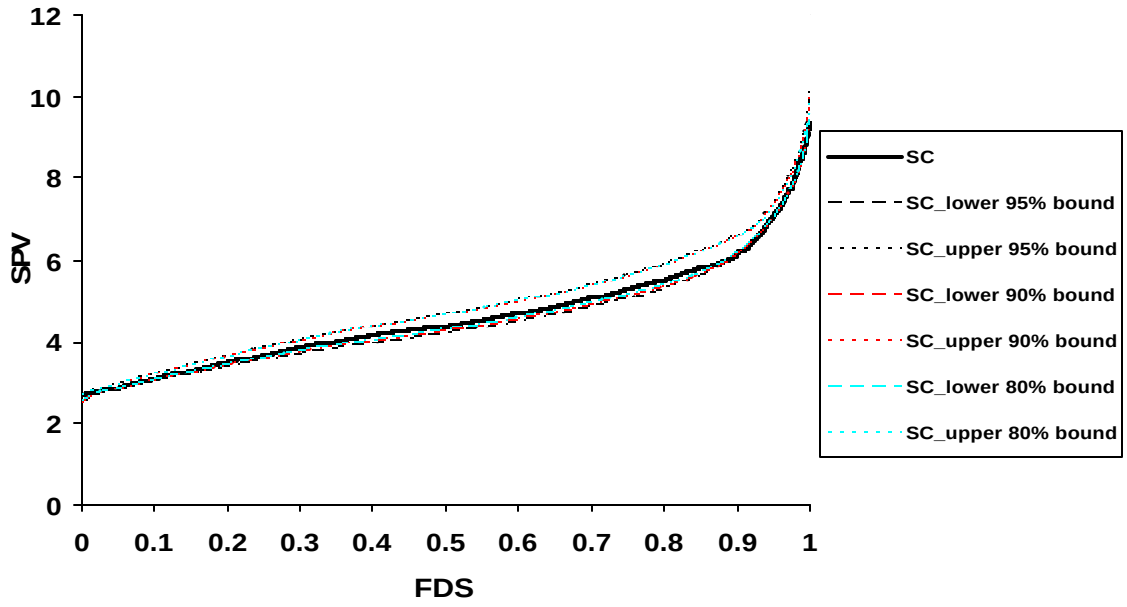

Figure V.6: FDS Plot of the Simplex-centroid with 80%, 90% and 95% Bounds and Relative Error from $U(-0.2, 0.2)$

Similar to the absolute error case, FDS plots can be used for comparing mixture designs for robustness to relative measurement error. Figure V.7 shows the comparison of two mixture designs: the augmented simplex-centroid and the simplex lattice assuming the special cubic model. The relative error is sampled from a uniform distribution, $U(-0.2, 0.2)$. As seen from Figure V.7 the simplex centroid performs better than the simplex lattice for the entire design space. Similar to the absolute error case, the difference between the upper 95% interval curve and the initial design curve at the high FDS values for the simplex-centroid is less than it is for the simplex lattice indicating a more robust design to relative measurement error at the end of the fraction of design space.

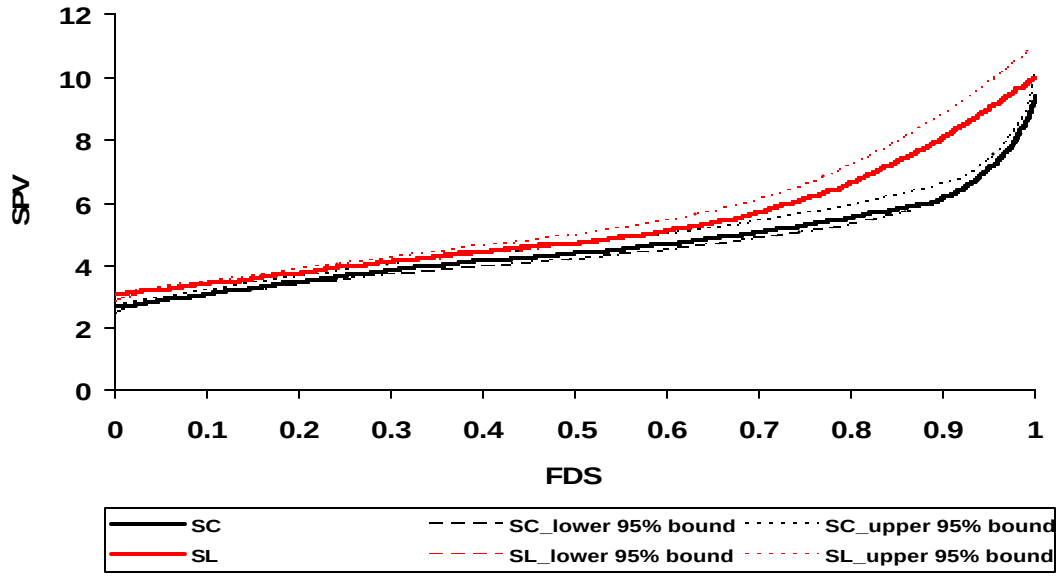

Figure V.7: Comparison of the Augmented Simplex-centroid and the Simplex Lattice for Robustness to Relative Error from $U(-0.2, 0.2)$ Using FDS Plots

Figure V.8 shows the comparison of the two designs for a relative error sampled from $U(-0.5,0.5)$. As seen from Figure V.8 as the relative error range increases, the gap between the upper 95% bound curve and the initial design curve at the end of the fraction of design space increases more, especially for the simplex lattice design.

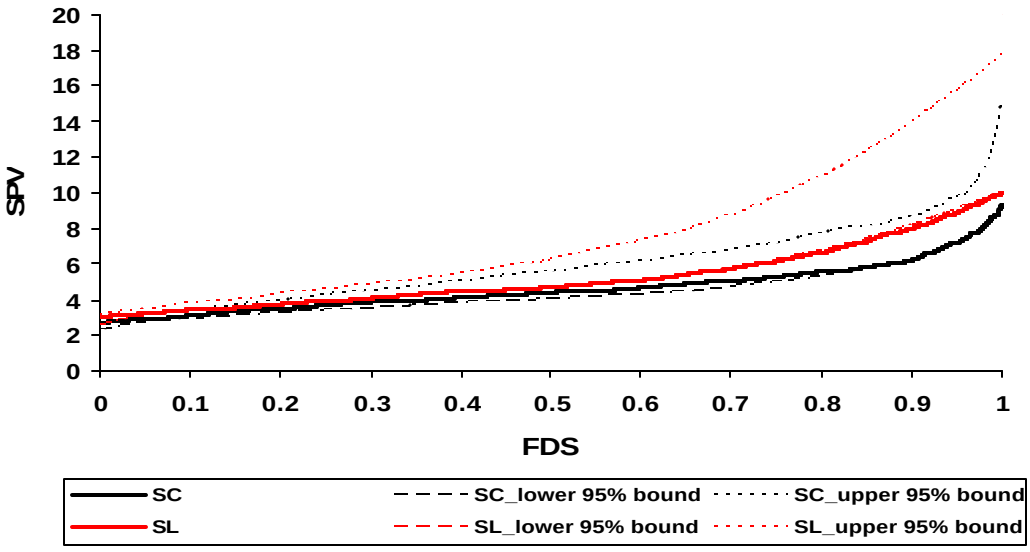

Figure V.8: Comparison of the Augmented Simplex-centroid and the Simplex Lattice for Robustness to Relative Error from $U(-0.5,0.5)$ Using FDS Plots

As we have indicated at the beginning of this section, the mixing measurement errors can be sampled from both the uniform distribution and the normal distribution. To be able to study the differences/similarities between the two error distributions on the same scale, the variance of the normally distributed errors were chosen to be equal to the variance of the uniformly distributed errors, i.e., if the errors are generated from a $U(-0.1, 0.1)$, then we can equate the two distributions roughly by setting the variance of the normally distributed errors equal to $\frac{(0.1+0.1)^2}{12} = 0.0\overline{3} \cong (0.058)^2$, the variance of a uniformly distributed error. This scaling provides the two error distributions to have approximately similar spreads.

Figure V.9 shows the FDS plot for the simplex-centroid design where the absolute errors are sampled from $N(0,(0.058)^2)$. As the bounds become wider, the band becomes wider for most of the fraction of design space which is expected. Compared to the absolute errors sampled from a uniform distribution, there is a slightly wider range of different SPV values across the sampled error choices for the normal case.

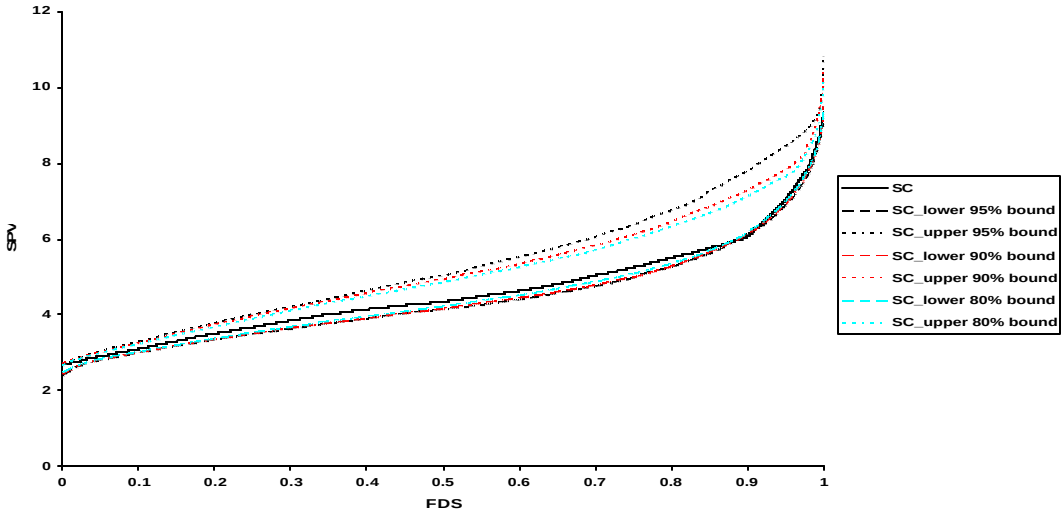

Figure V.9: FDS Plot of the Simplex-centroid with 80%, 90% and 95% Bounds and Absolute Error from $N(0,(0.058)^2)$

Figure V.10 shows the comparison of the simplex lattice and the simplex-centroid for an absolute error structure sampled from $N(0,(0.058)^2)$. From Figure V.10 we see that the difference between the upper 95% bound and the initial design curve using the normally distributed absolute errors is larger than the case using the uniformly distributed absolute errors for the simplex lattice design. Similar to Figures V.3, V.4, V.7 and V.8 the difference between the upper 95% bound and the initial design curve for the simplex lattice is larger compared to the simplex-centroid at the end of the fraction of design space.

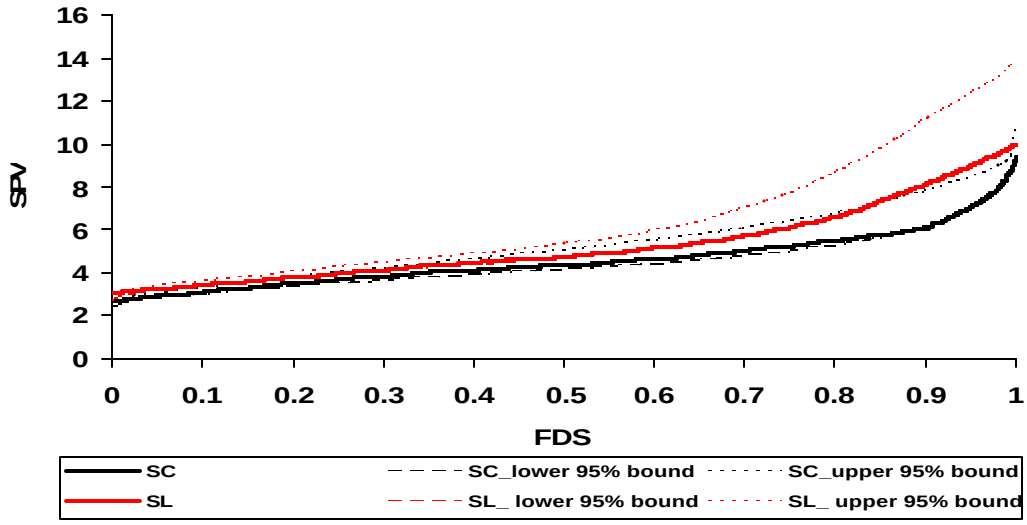

Figure V.10: Comparison of the Augmented Simplex-centroid and the Simplex Lattice for Robustness to Absolute Error from $N(0,(0.058)^2)$ Using FDS Plots

Finally, we compare the two designs using relative errors sampled from a normal distribution, $N(0,(0.115)^2)$ (having the same variance as a $U(-0.2, 0.2)$) in Figure V.11. Relative errors sampled from $N(0,(0.115)^2)$ cause less variability to the component proportions compared to absolute error sampled from $N(0,(0.058)^2)$ because of the difference in the definition of the two error terms. This can be seen from Figures V.10 and V.11 by looking at the range of SPV values for both designs. From Figure V.11 we see that the simplex-centroid is more robust than the simplex lattice in terms of relative

error sampled from a normal distribution, which is the case for all combinations of error types and error distributions.

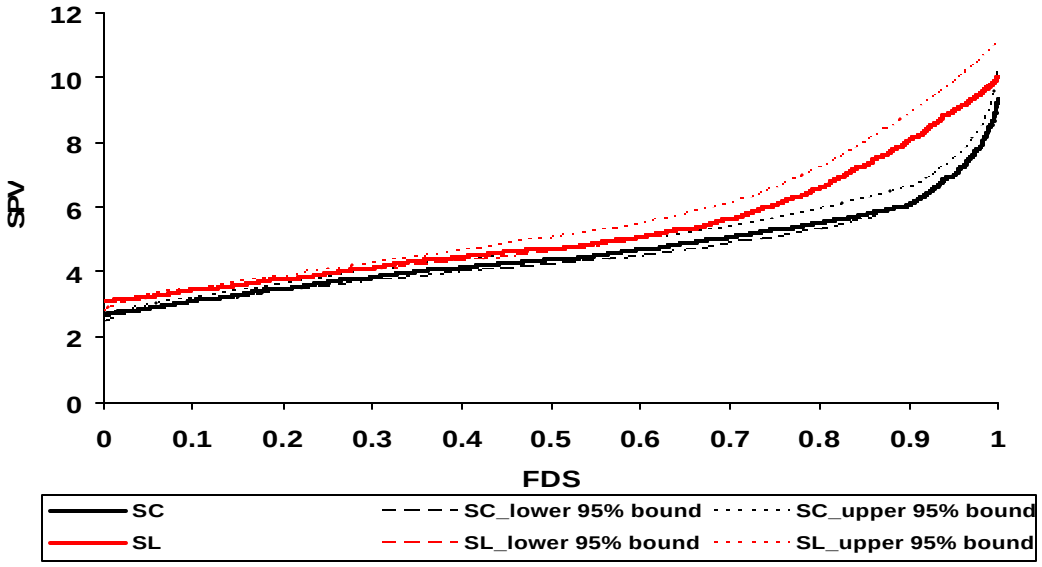

Figure V.11: Comparison of the Augmented Simplex-centroid and the Simplex Lattice for Robustness to Relative Error from $N(0,(0.115)^2)$ Using FDS Plots

V.4 Conclusion

In this chapter, we examined the adaptation of the FDS plots for mixing measurement errors. We considered two types of measurement errors: absolute error and relative error. Both error types were generated either from a uniform or a normal distribution. We used FDS plots to examine a mixture design's robustness to measurement error in terms of its effect on the design's prediction variance by adding measurement error to component proportions. Upper and lower intervals based on simulated designs from the assumed error distribution were obtained around the initial design's SPV values.

Based on our study, it appears that for a mixture design, if only absolute error is considered and the effect of uniformly and normally distributed errors are studied, then SPV bounds corresponding to normally distributed errors are slightly wider around the initial design's FDS curve than the SPV curves corresponding to uniformly distributed

errors. When relative error is considered, the results are similar to the results obtained from the absolute errors for both error distributions.

FDS plots were also used to compare mixture designs for robustness to measurement error. Specifically, the augmented simplex-centroid and the simplex lattice designs for 3 factors assuming the special cubic model were compared. For both simplex-centroid and simplex lattice designs, the upper and the lower SPV bounds corresponding to absolute error are wider around the initial design's FDS curve compared to SPV bounds corresponding to relative error. As the SPV values become large, this difference becomes more apparent. The augmented simplex-centroid performs better and is more robust to the two types of measurement errors sampled from either a uniform distribution or a normal distribution than the simplex lattice.

Chapter VI

MIXTURE EXPERIMENTS FOR GENERALIZED LINEAR MODELS

VI.1 Introduction

In this chapter of the dissertation, we study mixture experiments under the GLM framework. This situation might occur if the response from a mixture experiment involves studying proportions of defects produced or number of defects per unit area. Other examples of a mixture experiment for a GLM might arise when the researcher is interested in the effects of a toxic gas mixture on respiratory disease or the effect of a chemical mixture on failure time. When a GLM is considered, it is important to remember that certain changes that will occur related to the non-constant variance, the shape of the response surface, and the interpretation of the coefficients. Little research has been done in the literature concerning the combination of mixture experiments and non-normal data. Chen, Li and Jackson (1996) analyzed quantal response data from a mixture experiment using logistic regression. In this dissertation, using contour and FDS plots, we will study the prediction variance properties of the standard mixture designs under the GLM structure.

We begin by considering a mixture experiment with non-normal data. In this dissertation, we assume the linear predictor of the GLM will be in the form of a Scheffe′ model. For example, for a three component mixture design with a response in the form of binomial data, an appropriate link function between the mean and the linear predictor is the canonical logistic link. This link function is a transformation on the population mean relating it to the linear predictor. Therefore the first order linear predictor for the logit looks like:

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i}\,. \qquad \text{(VI.1)}$$

Similar to standard response surface designs, when a GLM is used with a mixture experiment, the constant variance assumption is no longer valid. The variance becomes a

function of the mean, and hence the prediction variance is a function of the number of observations, the design location, and the model parameters.

In addition, the shape of the surface for the response and hence the variance form will need to be estimated. For the binomial case, the surface of the estimated logits is linear, because the linear predictor is a first order Scheffe′ model. However, frequently the practitioner is interested in studying the surface of the predicted probabilities directly. The equation for obtaining the predicted probability values from the model parameters is as follows:

$$p_i = \frac{\exp(b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i})}{1 + \exp(b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i})} \qquad \text{(VI.2)}$$

When the predicted probabilities are plotted against the mixture components, the surface is no longer linear, because of the nonlinear form of the model. This change in the surface shape becomes even further accentuated for the second order Scheffe′ model as the linear predictor. For a 3-component mixture design with binomial data and a logistic link, the GLM using a second order linear predictor is as follows:

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \sum_{j=1}^{3} b_j x_{ji} + \sum\sum_{j<k}^{3} b_{jk} x_{ji} x_{ki} \; . \qquad \text{(VI.3)}$$

Recall, there are no pure quadratic terms in this model, because they are absorbed into $b_j x_{ji}$ and $b_{jk} x_{ji} x_{ki}$ through $x_i^2 = x_i * x_i = x_i\left(1 - \sum_{\substack{j=1 \\ j \neq i}}^{q} x_j\right)$.

The interpretation of the coefficients is different from a standard response surface design with non-normal data due to the mixture constraints. Each b_j represents the expected response on the logit scale at component x_i. The $b_j x_{ji}$ terms represent the linear blending on the logit scale and the $b_{jk} x_{ji} x_{ki}$ terms represent the departures from linear blending on the logit scale. The parameter estimates are used in order to define the relationship between the effect of the changes in the proportions of the mixture

components and the expected predicted probabilities. Ultimately, these parameters are used to transform the information into the predicted probabilities, $\hat{p}_i$'s.

To examine the prediction properties of mixture designs, several graphical methods have been developed. Piepel and Anderson (1992), and Piepel, Anderson, and Redgate (1993a, b) developed VDGs to compare mixture designs with constraints. Vining, Cornell, and Myers (1993) used VDGs to plot the prediction variance along Cox rays. Goldfarb, Montgomery, Borror, and Anderson-Cook (2003) developed three-dimensional VDGs to study the prediction variance properties of mixture experiments with process variables. Khuri, Harrison, and Cornell (1999) used quantile plots to compare designs for constrained mixture regions. Goldfarb, Anderson-Cook, Borror and Montgomery (2003) considered FDS plots for non-regular regions for mixture experiments and mixture experiments with process variables. In this dissertation, we adapt FDS plots to examine the prediction variance properties of mixture experiments for the GLM case.

VI.2 Example: Gasoline Mixture with Binomial Data

Consider a gasoline company, which produces gasoline mixtures which are combinations of three types of gasoline: x_1, x_2 and x_3. The researchers are interested in calculating the probability of a non-defect in engine performance for a variety of mixtures.

Figure VI.1 shows the chosen design space for the experiment, and the design, a {3, 2} simplex lattice for *q=3* (Cornell, 2002) with an equal number of observations at each design point. The logistic link is used with a first order Scheffe′ model, since based on scientific knowledge the assumed relationship is not overly complex. The practitioner is asked to estimate the probability of a non-defect at each corner and these are shown in Figure VI.1. Interest is in the chosen design's prediction properties with the given predicted probabilities. From the practitioner's estimates, the parameter estimates are calculated as $\hat{\boldsymbol{\beta}}' = (-2, 1.5, -3.3)'$. Therefore $\text{logit}(\hat{p}_i) = \ln(\frac{\hat{p}_i}{1-\hat{p}_i}) = -2x_1 + 1.5x_2 + -3.3x_3$

in model form. To illustrate the use of the model, for $x_1 = 1,\ x_2 = 0$ and $x_3 = 0$ $\hat{p}_{(1,0,0)} = 0.12$ from $\hat{p}_{(1,0,0)} = \dfrac{\exp(-2)}{1+\exp(-2)}$.

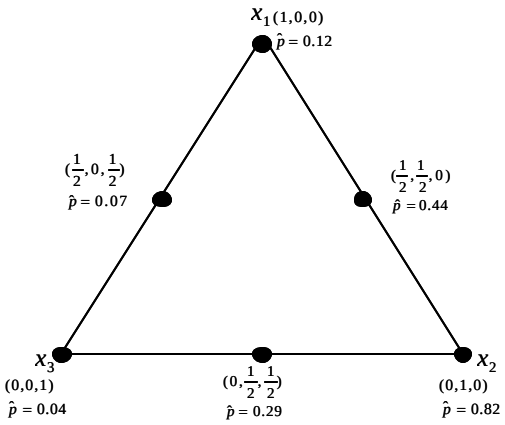

Figure VI.1: {3, 2} Simplex Lattice for the GLM Case with a First Order Linear Predictor

The three mixture components x_1, x_2 and x_3 constrained by $\sum_{i=1}^{3} x_i = 1$ are transformed into 2 independent variables, w_1 and w_2, to obtain the surface plots and the contour plots in two dimensions.

$$(w_1, w_2)' = 2\,(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)\,\frac{1}{\sqrt{6}}\begin{bmatrix} 2 & 0 \\ -1 & \sqrt{3} \\ -1 & -\sqrt{3} \end{bmatrix} \qquad \text{(VI.4)}$$

where $\tilde{x}_i = 3x_i - 1$. See Cornell (2002) for more details. Figure VI.2 shows the location of the design points, expressible in the new system (w_1, w_2) where $(w_1, w_2) = (0,0)$ is the centroid of the (x_1, x_2, x_3) mixture space.

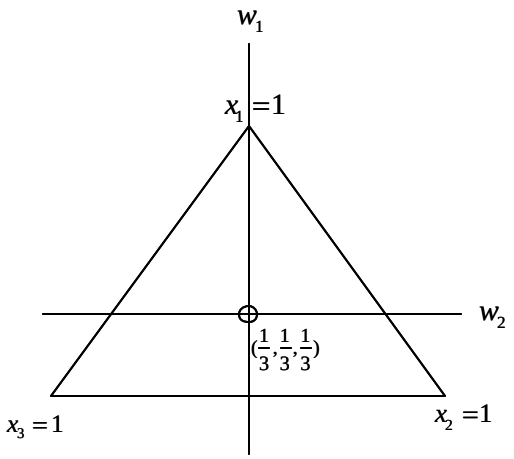

Figure VI.2: The Triangular Region in (w_1, w_2)

The surface plot of the assumed underlying $\text{logit}(\hat{p}_i)$, versus the mixture components using the same parameter estimates is shown in Figure VI.3. The $\text{logit}(\hat{p}_i)$s may be of interest to the experimenter if there is interest in the log odds ratio. As seen from Figure VI.3, there is no curvature in the surface for $\text{logit}(\hat{p}_i)$, because $\text{logit}(\hat{p}_i)$ is modeled by a first order linear model.

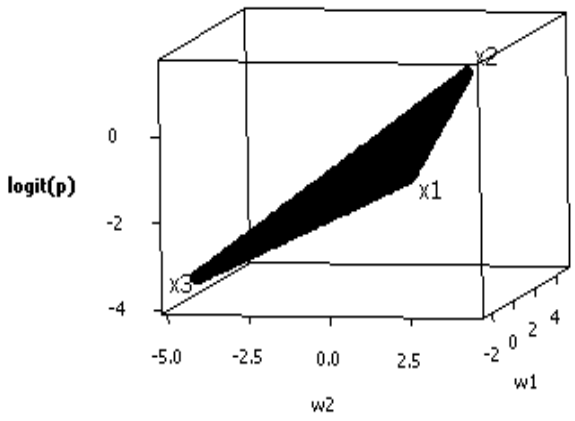

Figure VI.3: The Surface Plot of the Logit$(\hat{p}_i)$s for the First Order Linear Predictor

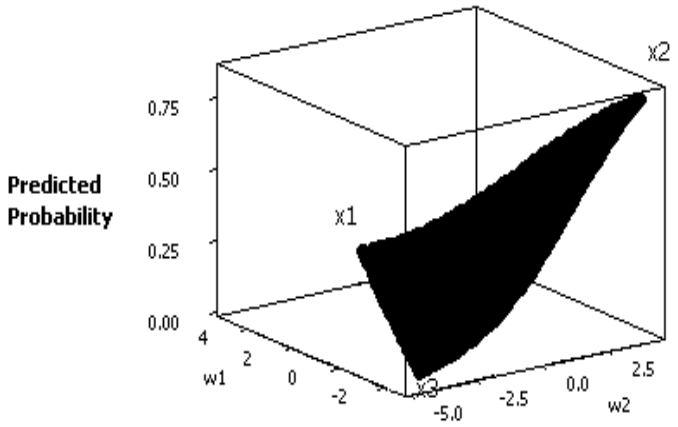

Figure VI.4: The Surface Plot of the Predicted Probabilities for the First Order Linear Predictor

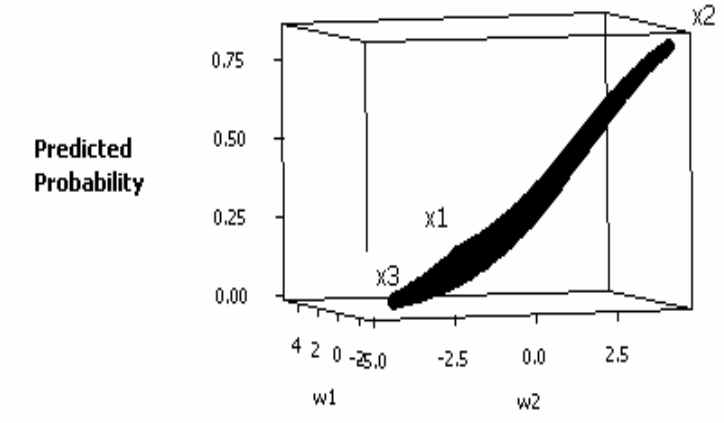

Figure VI.5: Cross Section of the Surface Plot of the Predicted Probabilities for the First Order Linear Predictor

Figure VI.4 shows the surface plot of the estimated predicted probabilities of a non-defect against the mixture components. From Figure VI.4, we see that there is now curvature on the surface due to the nonlinear model, $p(\boldsymbol{x}_i) = \frac{1}{1+e^{-x_i'\beta}}$. As seen from Figure VI.4, the predicted probabilities of non-defects are higher at the x_2 corner. When we rotate the surface of the predicted probabilities, we see that the curve is in the form of a logistic function. By definition, the logistic function is restricted to lie between 0 and 1. The observed range in this example is (0.07, 0.82), and it takes on the familiar s-shaped logistic curve as depicted in Figure VI.5.

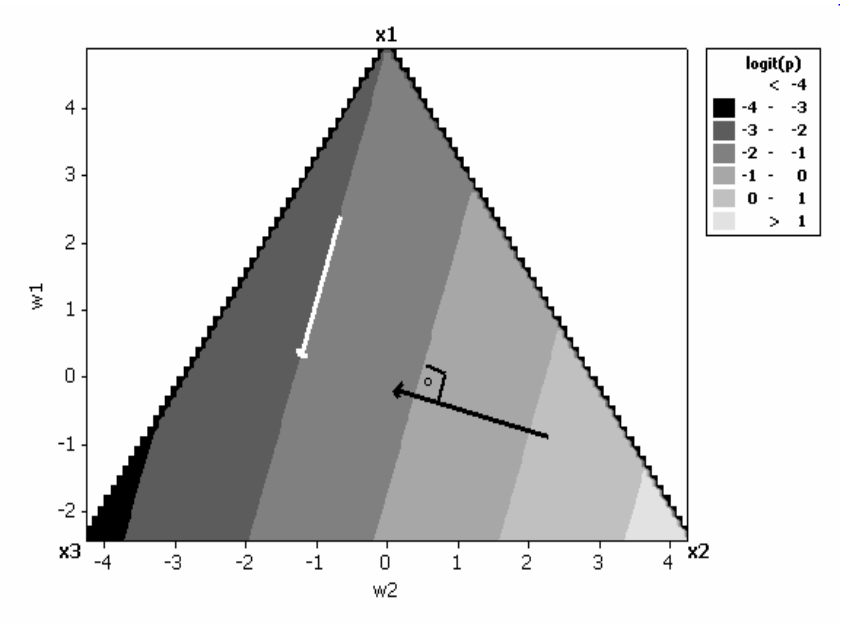

Figure VI.6: Contour Plot of the Logit($\hat{p}_i$)s for the First Order Linear Predictor

Figure VI.6 shows the contour plot for logit($\hat{p}_i$), where contour lines for equally spaced logit($\hat{p}_i$) values are parallel and evenly spaced, since the logit($\hat{p}_i$)s are fit by a first order model, which represents a flat plane. As one moves parallel to the contour lines, the logit($\hat{p}_i$) values remain the same. As one moves perpendicular to the contours, the logit($\hat{p}_i$) values change at a fixed rate, and if plotted would fall on a straight line.

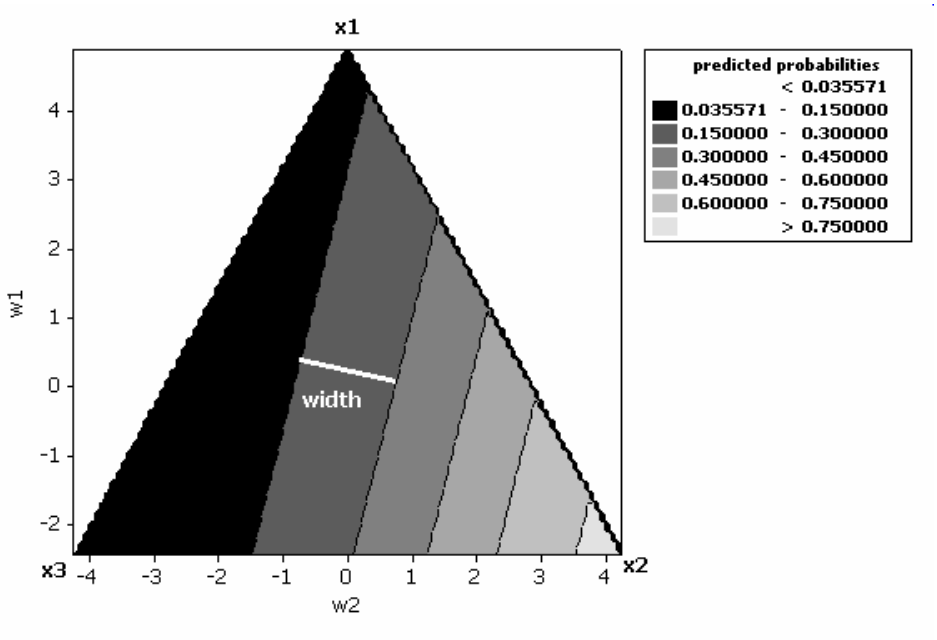

Figure VI.7: Contour Plot of the Predicted Probabilities for the First Order Linear Predictor

From the contour plot of the predicted probabilities for the first order model in Figure VI.7, we see that even though the contour lines are parallel to each other for equal spacing of probability values, the distance between contour lines has now changed. This difference between the widths of the contour areas is a result of the non-linear logistic curve as shown on the surface plot in Figure VI.5. When the change in the distance between the contour lines is large, this corresponds to a smaller slope and a flatter section of the logistic function, whereas when the difference in the widths of the contour areas is small, this corresponds to a region of the logistic function with a steeper slope.

We consider the same example, but now assume that the relationship between the response and the mixture amounts is best summarized with a GLM with a second order Scheffe′ linear predictor and a logistic link. The simplex lattice design is an appropriate choice for the design space and the chosen model, because it supports the fitting of a second order model. Since the simplex lattice is a well-known standard design for the linear models case, we investigate it further under the GLM framework. We begin by considering a simplex lattice (D1) with an equal numbers of observations at each design

point. The design is shown in Figure VI.8 with the predicted probabilities given by the practitioner.

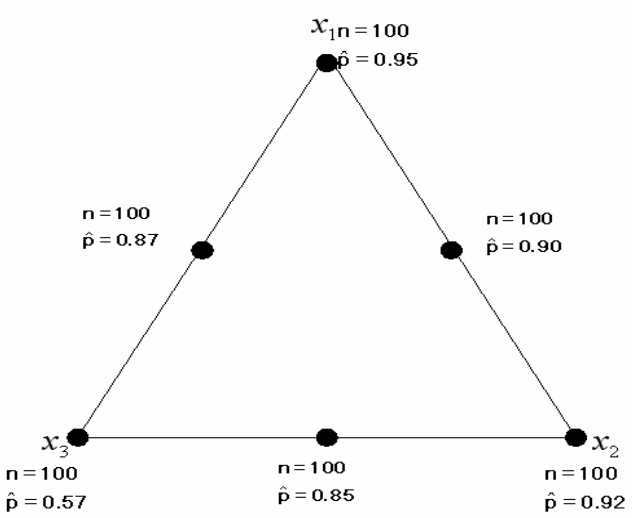

Figure VI.8: Simplex Lattice (D1) for the GLM Case with a Second Order Linear Predictor

It is very unlikely that the practitioner will provide precise guesses of the predicted probabilities because the model is now more complex compared to the previous example. The practitioner not only has to provide predicted probabilities at the pure blends but also at the binary blends. Therefore, it is unlikely that these probability estimates are correct. That is why in the following sections we investigate the robustness of the design to the misspecifications of the predicted probabilities. From the practitioner's estimates of the rates of defects at various mixtures, the parameter estimates are calculated as $\hat{\boldsymbol{\beta}}' = (3,\ 2.5,\ 0.3,\ -2,\ 0.9, 1.4)'$ and the model can be written as $\text{logit}(\hat{p}_i) = \ln(\frac{\hat{p}_i}{1-\hat{p}_i}) = 3x_1 + 2.5x_2 + 0.3x_3 + -2x_1x_2 + 0.9x_1x_3 + 1.4x_2x_3$. These values will be used to obtain information about the distribution of the SPV/PPV values to study the chosen design's prediction properties.

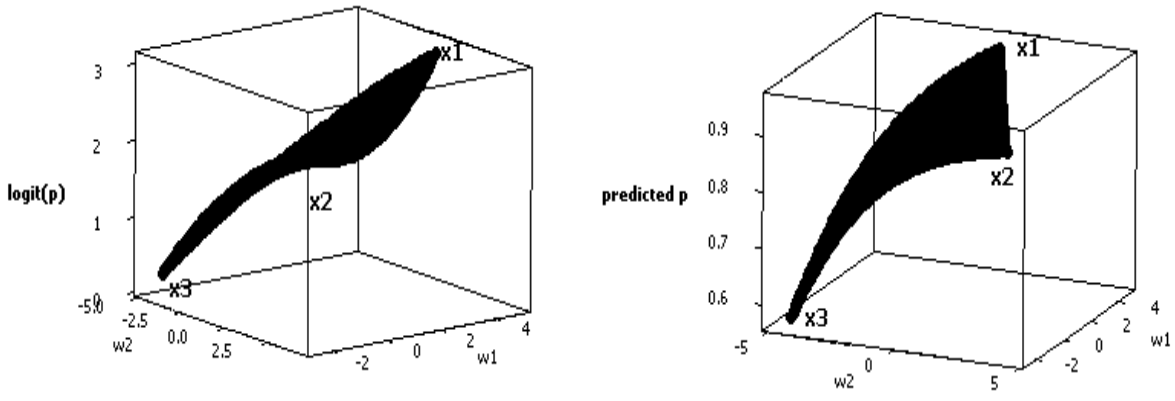

Figure VI.9: The Surface Plot of the Logit($\hat{p}_i$)s and the Predicted Probabilities for the Simplex Lattice Design

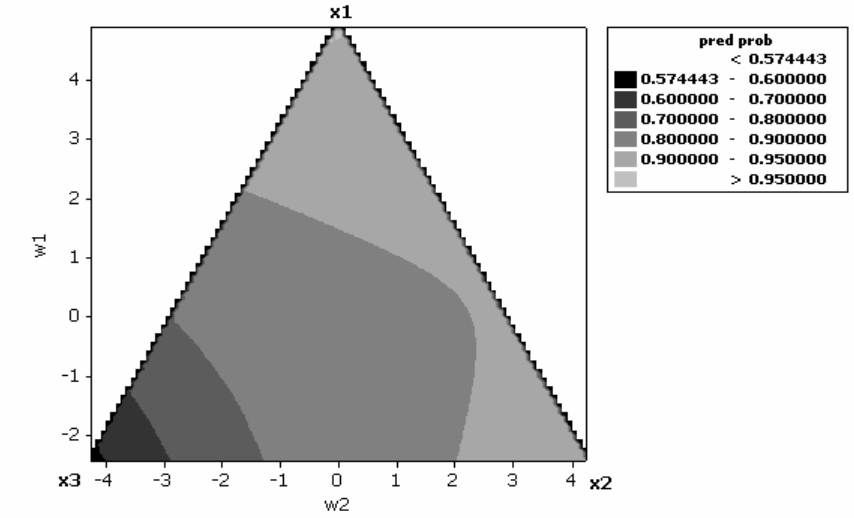

Figure VI.10: Contour Plot of the Predicted Probabilities for the Simplex Lattice Design

Figure VI.9 shows the surface plot for logit($\hat{p}_i$) and the predicted probabilities. Since the second order Scheffe′ model is used, the shape of the surface of the logits is no longer linear. In addition, because the logistic function is a monotonic transformation of all possible numbers to the range [0, 1], the minimum and the maximum values occur at the

same design locations in both plots. Compared to the surface plot of the predicted probabilities for the first order linear predictor, this surface plot has even more curvature. Figure VI.10 shows the contour plot for the predicted probabilities, with the predicted probabilities of non-defects higher when the x_1 and the x_2 types are present in higher proportions compared to x_3. The predicted probability is the highest for the pure blend of x_1.

So far we have been looking at the actual predicted means, but we are also interested in the prediction capabilities of the design. Therefore we will look at the plots of the SPV and PPV distributions throughout the design space. The SPV and PPV values are functions of the predicted means and therefore of the model form. The SPV and the PPV for a mixture design with a generalized linear model are the same as in Chapter IV where $n(\boldsymbol{x}_o) = N \text{ var}(y(\boldsymbol{x}_o))\boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o$ and $r(\boldsymbol{x}_o) = N \text{ var}(y(\boldsymbol{x}_o))^2 \boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o$, respectively. Both terms are dependent on the number of runs, the design, the model and the parameter estimates. The only difference is due to the constraint of a mixture design, the linear predictor now has the Scheffe′ model form.

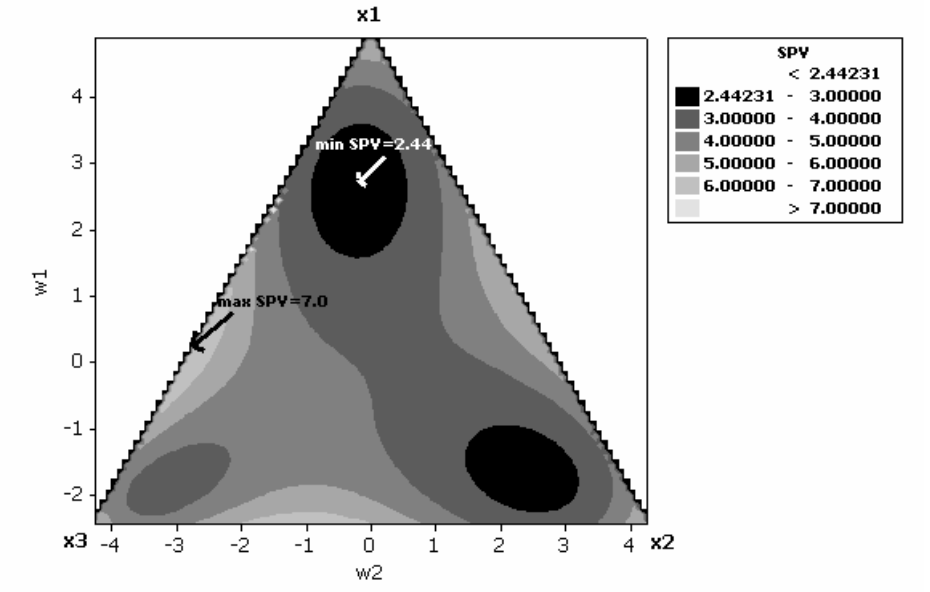

Figure VI.11: Contour Plot of SPVs for the Simplex Lattice Design

In Figure VI.11, the initial parameter estimates, $\hat{\boldsymbol{\beta}}' = (3,\ 2.5,\ 0.3,\ -2,\ 0.9, 1.4)'$, are used to obtain the contours of the SPV values for the simplex lattice design. The SPV values range between 2.44 and 7. Even though for most designs, the maximum SPV values are obtained at the design corners, the maximum SPV occurs along an edge. Based on the predicted probabilities, this would not have been easy to anticipate. As seen from Figure VI.11, most of the design space has SPV between 3 and 5.

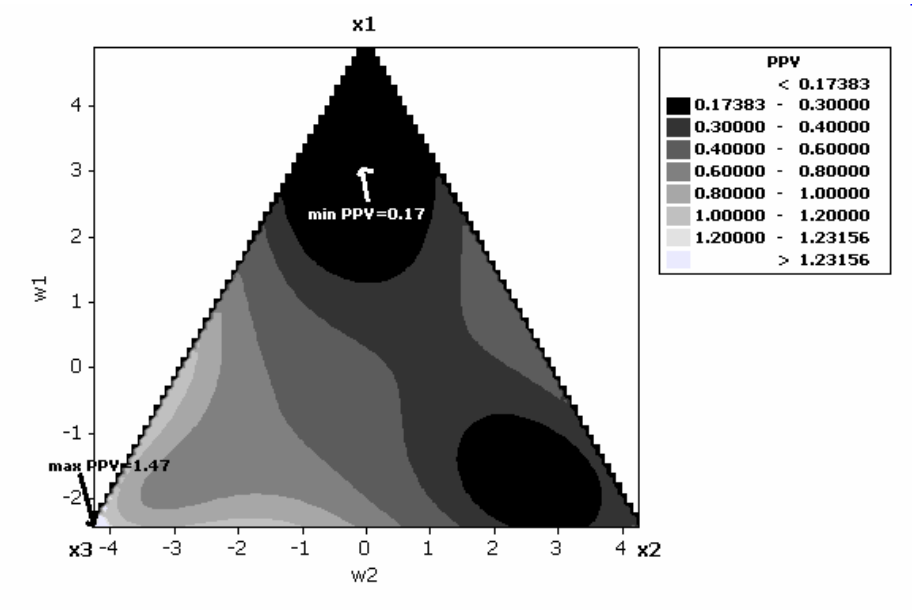

Figure VI.12: Contour Plot of PPVs for the Simplex Lattice Design

Figure VI.12 shows the contour plot of the PPVs. The PPV values are quite different compared to the SPV values due to the extra variance term in the PPV formula. They inform the researcher about a design's actual prediction. The PPV values range between 0.17 and 1.47. The variance of a binomial random variable increases as the probability of a success/failure gets closer to 0.5. Therefore, the maximum PPV is located at the x_3 corner where the predicted probability is the closest to 0.5 (0.57) in the design space. Contour plots are an excellent tool for examining mixture designs for the GLM case, but they are hard to construct and view in higher dimensions. However, FDS plots are straightforward for any design with any number of factors.

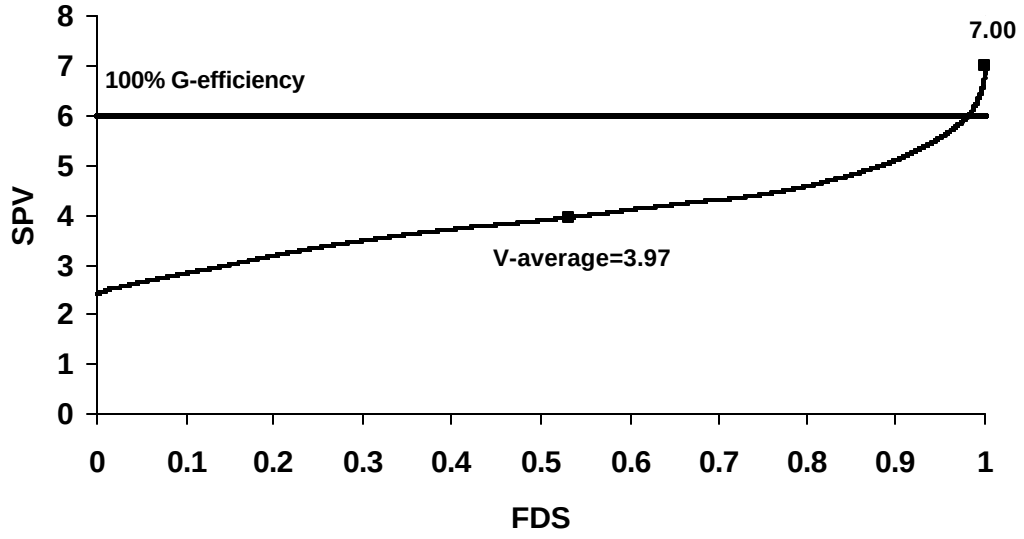

Figure VI.13: FDS Plot (using SPVs) for the Simplex Lattice Design

Figure VI.13 shows the FDS plot using SPVs for D1. For small SPV values, the FDS curve increases relatively slowly indicating many similar smaller SPV values while for larger SPV values, the slope increases at a faster rate showing that there are small regions where the SPV values are considerably larger than the majority of the region. The V-average is 3.97 for this design which corresponds to 53% of the fraction of design space giving a slightly skewed SPV distribution. The FDS plot informs the researcher about what percentage of the design space has an SPV at or below a given contour value. The G-optimal value for this design for the chosen model is 6, the number of parameters in the model. Therefore, the design is 86% G-efficient. Figure VI.14 shows the FDS plot using the PPV values. The PPV curve increases at a faster rate for larger PPV values. When the SPV and the PPV curves are compared, overall the PPV curve has a steeper slope. The FDS plots using the SPVs and the PPVs have different interpretations. The FDS plot with the SPVs is important in terms of comparing the design's prediction properties to the theoretical optimum. The SPV values for the GLM are still on the same scale as the G- and the V-optimal values used for standard linear models. Therefore, the researcher is still able to compare the design to the theoretical optimum. The FDS plot with the PPV is important for the practitioner who is interested in actual prediction. By

examining an FDS plot with the PPV values, the researcher is able to obtain important information about a design's actual prediction properties.

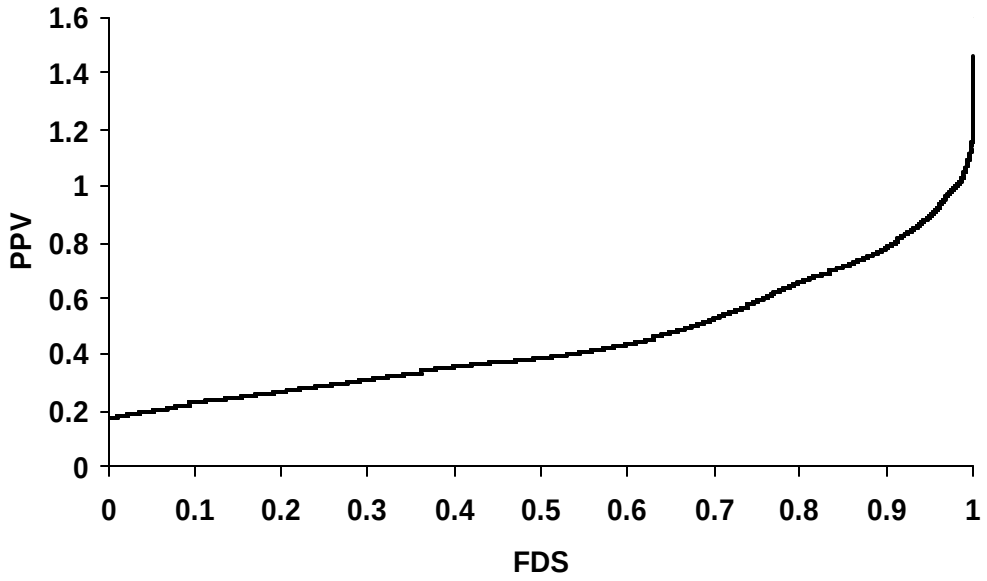

Figure VI.14: FDS Plot (using PPVs) for the Simplex Lattice Design

VI.3 Improving a Chosen Design

FDS plots can also be used to suggest improvements for a chosen design. The experimenter may be interested in allocating the number of observations at each design location to improve the estimation at the corners where it is thought to have larger variances based on the estimated proportion of non-defects. Therefore, he/she might want to compare the simplex lattice design, D1, to an adapted design, D2, where there are more observations at the design locations for the predicted probabilities that are closer to 0.5 (and hence the prediction variance is expected to be larger) and less observations at the locations where the predicted probabilities of non-defects are closer to 0 and 1. D1 is an equal allocation design, whereas this is not true for D2. Since the variance at each design point is $s_i^2 = p_i(1-p_i)/n_i$ where n_i is the number of observations at that point, D2 selects sample sizes at each location to make $\hat{s}_i^2$ as close to constant as possible.

Table VI.1 shows the number of observations and the estimates of the variances of the predicted probabilities at each design location for D1 and D2. As seen from Table VI.1

the estimates become stable once the allocation of the observations is changed. The largest number of observations, 212, in bold corresponds to the predicted probability closest to 0.5 and the smallest allocation of 41 corresponds to the predicted probability (0.95) that is closest to 1.

Table VI.1: Observation Allocations and Estimates of the Variance of the Predicted Probabilities for D1 and D2

Location	(1, 0, 0)	(0, 1, 0)	(0, 0, 1)	(1/2, 1/2, 0)	(1/2, 0, 1/2)	(0, 1/2, 1/2)
$\hat{p}_i$	**0.95**	0.92	**0.57**	0.90	0.87	0.85
N (D1)	100	100	100	100	100	100
$\hat{s}_i^2$ (D1)	0.00048	0.00074	0.00245	0.00090	0.00113	0.00128
N (D2)	**41**	63	**212**	77	97	110
$\hat{s}_i^2$ (D2)	0.00116	0.00117	0.00116	0.00117	0.00117	0.00116

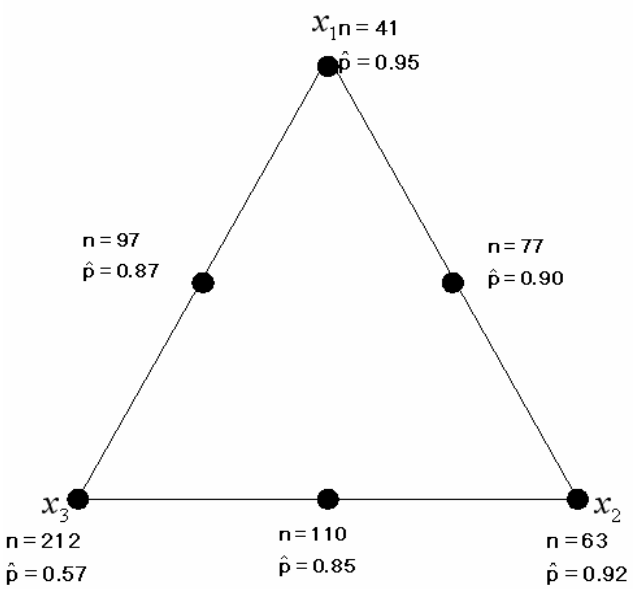

Figure VI.15: Simplex Lattice (D2) for the GLM Case

Figure VI.15 shows the design space for D2. From Figure VI.15, we see that the x_3 corner has the highest number of observations to account for the high variability. The

other design locations are adapted accordingly keeping the total number of observations constant.

Figure VI.16 shows the contour plots for both designs using the PPV values. The locations of the best and worst prediction values have changed. The maximum PPV is no longer at the x_3 corner. The distribution of the PPVs is more stable for D2, since it has more area with the same color indicating a more stable PPV distribution.

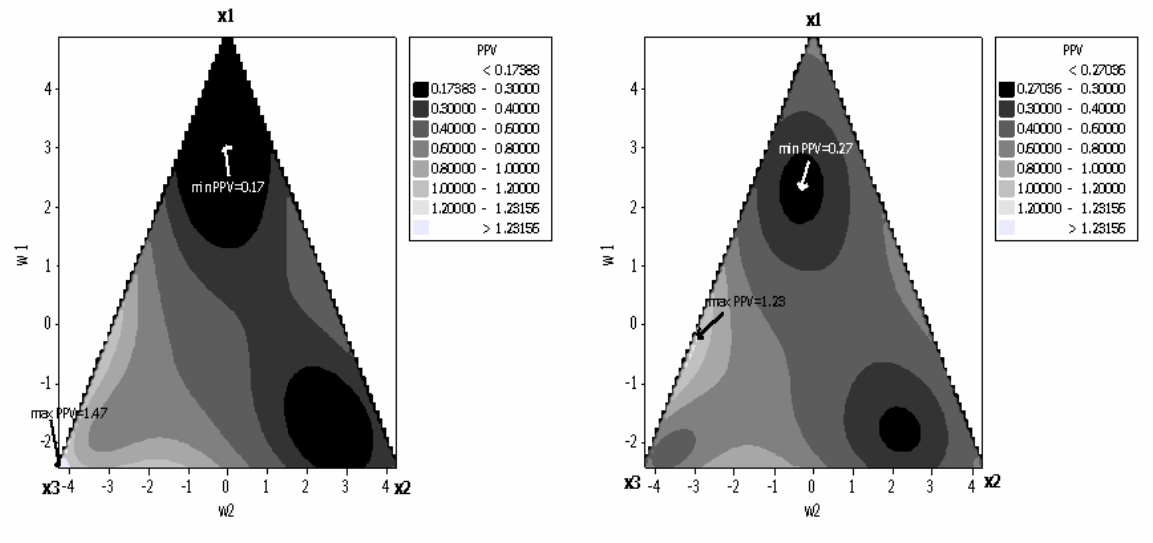

Figure VI.16: Contour Plots (using PPVs) for D1 and D2, respectively

Figure VI.17 shows the FDS plots for D1 and D2 using the PPV values. The FDS curve for D2 is more horizontal than the FDS curve for D1 indicating a more stable PPV distribution. The aim of the practitioner was to obtain a more stable prediction distribution by changing the number of observations at each location, and it has been achieved. Even though D2 has a higher minimum, overall it has a more consistent PPV distribution with a lower maximum. These plots are based on knowing the parameter values, which is typically not the case. Therefore in the following section we discuss the implications if these estimates were not correct.

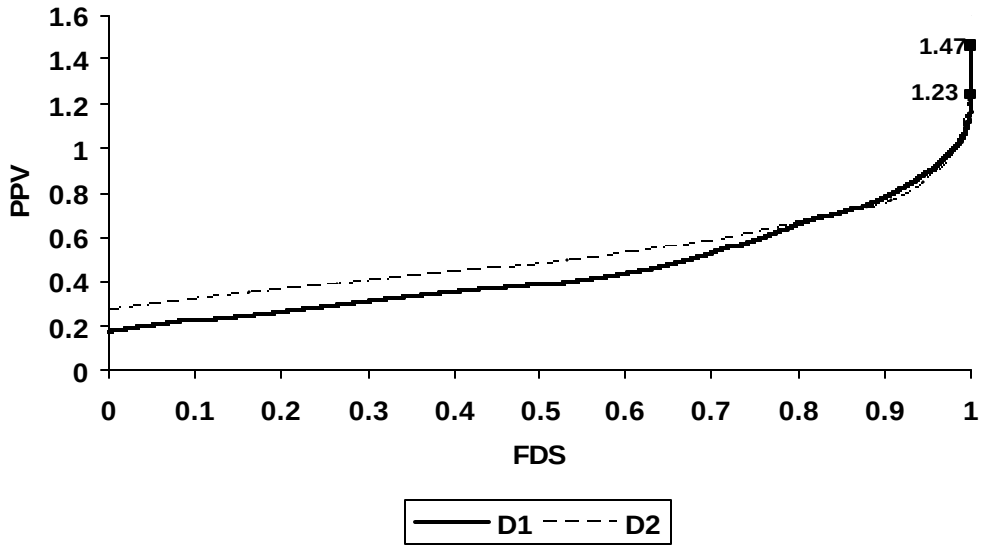

Figure VI.17: FDS Plot (using PPVs) for Comparing D1 and D2

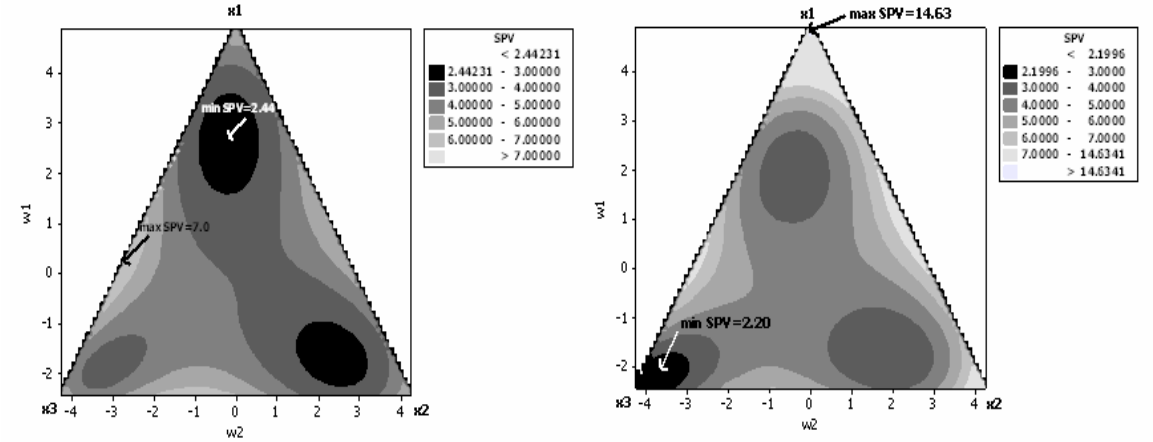

Figure VI.18: Contour Plots (using SPVs) for D1 and D2, respectively

Figure VI.18 shows the contour plots for both cases using the SPV values. From Figure VI.18, we see that the locations of the minimum and the maximum SPVs have been shifted. The maximum SPV is no longer at an edge, instead it is at the x_1 corner and the minimum SPV has been shifted inside the design space. For most of the design space D1 has more low SPV ranges than D2 indicating lower SPV values. However, D2 has lower SPVs at the x_3 corner which corresponds to a very small fraction of the total design space. When the high SPV ranges are considered, D2 has more high SPV values than D1.

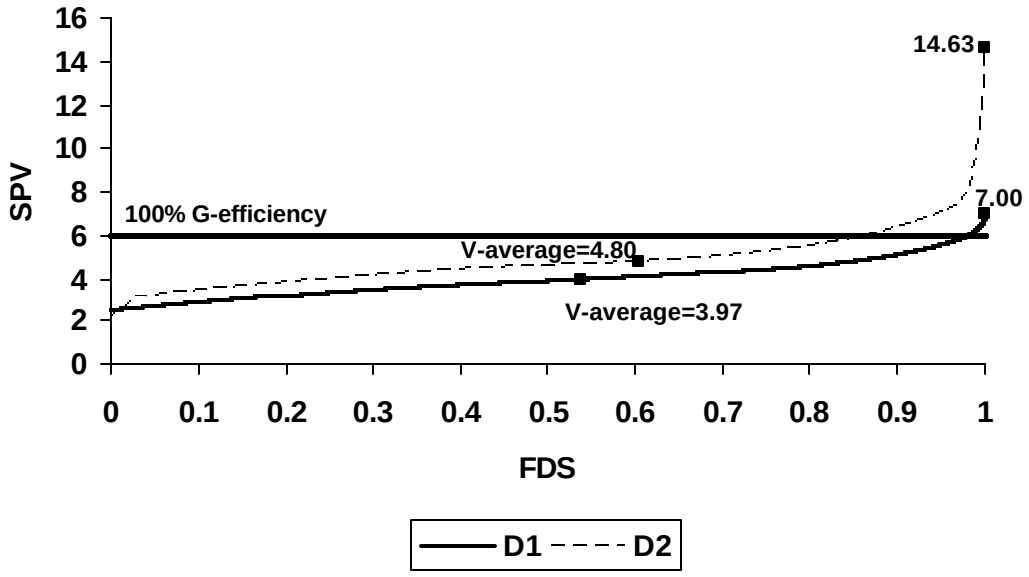

Figure VI.19: FDS Plot (using SPVs) for D1 and D2

Figure VI.19 shows the FDS plots using SPVs for both designs. As we see from Figure VI.19, D1 performs better for most of the fraction of design space except for the smallest 1% of SPV values. Table VI.2 summarizes the minimum and the maximum SPV and PPV values, the approximate G-efficiencies, and the V-averages for D1 and D2. D1 has an overall lower SPV average (3.97) than D2 (4.80). The SPV values for D2 increase very rapidly at the end of the fraction of design space resulting in a 41% G-efficiency.

Table VI.2: Summary of SPV and PPV values, Approximate G-efficiencies, and V-averages for D1 and D2

Design	V-average	G-efficiency	Min SPV	Max SPV	Min PPV	Max PPV
D1	3.97	86%	2.44	7.00	0.17	1.47
D2	4.80	41%	2.20	14.63	0.27	1.23

From the Figures VI.17 and VI.19, we can see that the two designs D1 and D2 perform quite differently depending on which criterion the researcher wishes to focus on. For the practitioner interested in optimizing the prediction of the rate of non-defects throughout the region, D2 is a better choice with a more stable range of PPV values throughout the

design space, and with a lower maximum value. The flatness of the line reflects the stability of the design across the entire space. The equal allocation design, D1, has better prediction in the majority of the design space, but it has a small region of poorer prediction at the corner of the design space with probability of non-defects close to 0.60. However, from a theoretical standpoint, D1 has better G-efficiency with a smaller maximum SPV value. Clearly in the GLM case, the distinction between theoretical and practical assessment of the designs is more complicated than in the standard linear model case for mixture experiments.

VI.4 Robustness to Predicted Means

Similar to Chapter IV, FDS plots can also be used to examine a mixture design's robustness to predicted means for the GLM case. As mentioned in the previous section, the experimenter is unlikely to know the predicted means precisely. Therefore the comparison of competing mixture designs should take into account the robustness to these predicted means. The estimates will never be exact, so a design that is relatively robust to changes in the predicted means would be highly desirable. FDS plots help the researcher study robustness by looking at the distribution of the SPV/PPV values.

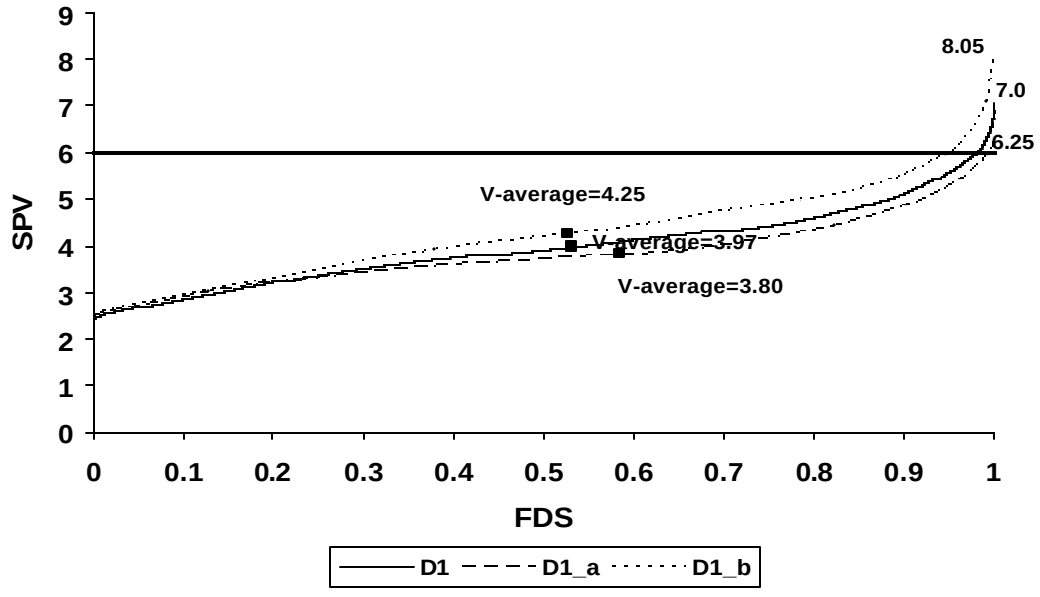

Figure VI.20: FDS Plot (using SPVs) for D1, D1_a and D1_b

The researcher may be interested in how robust the design is to the shifts of the range of the predicted means. Trying to quantify how much the original probability estimates could be misspecified, the practitioner decides to decrease the predicted probabilities by 0.10 where the range becomes $(\min \hat{p}_i, \max \hat{p}_i) = (0.47, 0.85)$ and when they are increased by 0.03, the range becomes $(\min \hat{p}_i, \max \hat{p}_i) = (0.60, 0.98)$. The first scenario is denoted by D1_a, and the second scenario is denoted by D1_b. Figure VI.20 shows the FDS plots of the two cases using SPV values.

From Figure VI.20 we see that compared to D1, D1_a has lower SPV values for most of the fraction of the design space whereas D1_b has higher SPV values for the whole design space. As the SPV values increase towards the end of the fraction of design space, the distance between the three curves at a single SPV value increases. At FDS=1, D1_b has the highest SPV. Decreasing the predicted means at each design location has decreased the SPV values and increasing the predicted means has increased the SPV values. As the predicted probabilities approach 1, the G-efficiency of the design decreases.

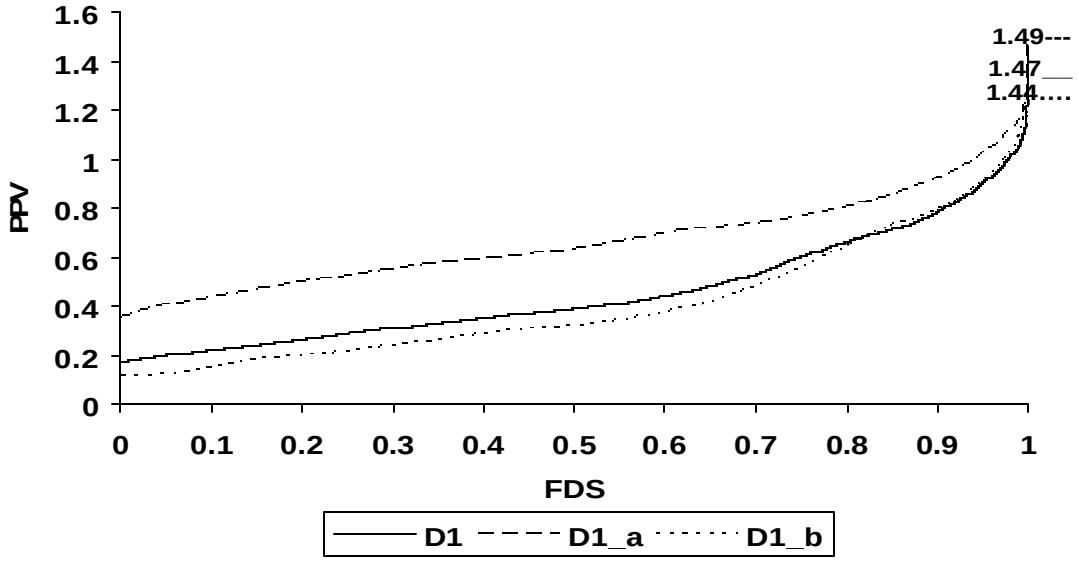

Figure VI.21: FDS Plot (using PPVs) for D1, D1_a and D1_b

Figure VI.21 shows the FDS plots of the three cases using the PPV values. Increasing the predicted means has decreased the PPV values whereas decreasing them has increased the PPVs. Even though the PPV values for D1_a and D1_b are very different than the PPV values for D1, all three cases have similar maximum PPV values. For a practitioner who is interested in the robustness of this design for the worst case prediction variance, this design is quite robust.

The types of misspecifications of the predicted means described in Chapter IV could also be examined for the mixture experiments for the GLM case.

VI.5 Conclusion

In this chapter, we adapted mixture designs to the generalized linear models case. The non-constant variance, the model form and the interpretation of the coefficients were illustrated by an example of a 3-component mixture using a simplex lattice design for logistic regression.

FDS plots combined with contour plots were used to examine mixture designs in terms of their prediction properties. Contour plots are mostly used for designs of up to 3-4 factors. As the number of factors increases, they become less efficient in terms of viewing the changes in the SPV/PPV values. However, FDS plots remain direct, as they can be obtained for any number of factors.

In addition, FDS plots combined with contour plots are helpful to improve a mixture design's prediction properties. We studied an example of a simplex lattice for the binomial case with a second order Scheffe′ linear predictor where the researcher was interested in improving the prediction at the locations where it was thought to have high variability. Improving the allocation of the observations results in a more stable PPV distribution.

Misspecifications of the predicted means were also examined. Other types of misspecifications as defined in Chapter IV could easily be adapted to the mixture

experiments for the GLM case. The PPV and the SPV values for the GLM case were studied simultaneously for the design performance in terms of misspecification. Examining the SPV and PPV values together helps the researcher choose a best design. The differences between the SPV and the PPV values emphasize the importance of choosing the most natural measure of design performance for a given situation.

Chapter VII

SUMMARY & FUTURE WORK

The research consists of four main contributions. First, we adapt FDS plots to study robustness of various designs to the underlying model form. Cases involving three- and four-factor response surface designs having cuboidal regions and three-factor mixture experiments were considered to compare various designs in terms of their robustness to model misspecification. Several types of reduced models for a second order response surface design are considered to show the sensitivity of the prediction variance to the reductions from the second order model. In addition to calculating the G-efficiency of a design from an FDS plot, one can obtain an approximate value of the average SPV. FDS plots are scaled by a design's G-optimal bound for the chosen model. This approach helps the researcher compare possible models for a chosen design on the same scale, therefore obtain accurate results on the robustness of the design. Moreover, the researcher can easily augment the designs with additional points to obtain more robust designs. This approach can also be extended to designs on spherical regions.

The second part of the dissertation adapts FDS plots to the generalized linear models case assuming the canonical link. In addition to considering the form the SPV for the GLM case, penalized prediction variance (PPV) is proposed. These two measures are important for different purposes. SPV values can be used to study a design's theoretical properties whereas PPV values become important when the researcher is interested in actual prediction. Since both SPV and PPV are dependent on the estimated predicted means which are unlikely to be estimated correctly, several misspecification types are considered. Effects of the misspecifications on the design's SPV/PPV distributions are displayed using FDS plots. These cases are examined in detail using the 2^2 factorial design with binomial and Poisson data. In addition to designs with GLMs having first order model plus interaction linear predictors, designs having GLMs with second order linear predictors are studied and the extra complexity of prediction variance is encountered using FDS plots. In addition to assuming the canonical link, other link

functions could be studied. Designs with other types of response distributions, such as gamma and exponential distributions could also be examined using the adapted FDS plots. The robustness to the choice of the link function could also be studied using FDS plots as well.

Robustness of mixture experiments to measurement error is examined by adapting the FDS plots. Relative and absolute errors are considered as the measurement errors and these are generated from either a normal distribution or a uniform distribution. Two mixture design types: the simplex lattice and the simplex-centroid for three components are compared. For three factors the simplex-centroid performs better and is more robust to measurement error than the simplex lattice assuming a special cubic model.

Finally, mixture experiments with a non-normal response modeled with a GLM are studied using FDS plots and contour plots. The model form for a mixture experiment with GLM is considered. The prediction variance properties of a mixture design with GLM are studied using FDS plots and contour plots. The robustness of mixture experiments to misspecifications of the predicted means is studied using a binomial example. The process of improving a mixture experiment's prediction properties is shown using FDS plots. In this research, we assume the only restrictions on the component proportions of the mixture experiments are $0 \le x_i \le 1 \ \ \forall i = 1,2,...,q$ and $\sum_{i=1}^{q} x_i = 1$. The methodology could be extended to mixture experiments with additional constraints. It could also be extended to mixture experiments with process variables. Mixture experiments with process variables for the GLM case are a common situation encountered in the industry. The prediction properties of such designs with such models could be studied using FDS plots.

Chapter VIII

REFERENCES

Abdelbasit, K. M. and Plackett R. L. (1983). "Experimental Design for Binary Data". *Journal of the American Statistical Association* 78, pp. 90-98.

Andrea-Rendon, J., Montgomery, D. C., and Rollier, D. W. (1997). "Design of Mixture Experiments Using Bayesian D-Optimality". *Journal of Quality Technology* 29, pp. 451-463.

Andrews, D. F. (1971). "Sequentially Designed Experiments for Screening out Bad Models with F Tests". *Biometrika* 58, pp. 427-432.

Atkinson, A. C. and Federov, V. V. (1975). "The Design of Experiments for Discriminating Between Two Rival Models." *Biometrika* 62, pp. 57-70.

Bingham, D. R. and Li, W. (2002). "A Class of Optimal Robust Parameter Designs". *Journal of Quality Technology* 34, pp. 244-259.

Block, R. M. and Mee, R.W. (2001). "Some New Second-Order Designs". Technical Report 01-04, Department of Statistics, University of Tennessee.

Borkowski, J. J. (1995). "Spherical Prediction Variance Properties of Central Composite and Box-Behnken Designs". *Technometrics* 37, pp. 399-410.

Borkowski, J. J. (2003). "A Comparison of Prediction Variance Criteria for Response Surface Designs". *Journal of Quality Technology* 35, pp. 70-77.

Borkowski, J. J. and Valeroso, E. S. (2001). "Comparison of Design Optimality Criteria of Reduced Models for Response Surface Designs in the Hypercube." *Technometrics* 43, pp. 468-477.

Borror, C. M., Montgomery, D. C., and Myers R. H. (2002). "Evaluation of Statistical Designs for Experiments Involving Noise Variables". *Journal of Quality Technology* 34, pp. 54-70.

Box, G. E. P. and Behnken, D. W. (1960). "Some New Three-Level Designs for the Study of Quantitative Variables". *Technometrics* 2, pp. 455-475.

Box, G. E. P. and Draper, N. R. (1959). "A Basis for the Selection of a Response Surface Design". *Journal of the American Statistical Association* 54, pp.622-654.

Box, G. E. P. and Draper, N. R. (1963). "The Choice of a Second Order Rotatable Design". *Biometrika* 50, pp. 335-352.

Box, G. E. P. and Draper, N. R. (1974). "On Minimum-Point Second-Order Designs". *Technometrics* 16, pp. 613-616.

Box, G. E. P.;and Draper, N. R. (1975). "Robust Designs". *Biometrika* 62, pp. 347-352.

Box, G. E. P. and Draper, N. R. (1987). *Empirical Model Building and Response Surfaces*. John Wiley & Sons, New York, NY.

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

Box, G. E. P. and Hunter, J. S. (1957). "Multi-factor Experimental Designs for Exploring Response Surfaces". *Annals of Mathematical Statistics* 28, pp. 195-241.

Box, G. E. P. and Wilson, K. B. (1951). "On the Experimental Attainment of Optimum Conditions". *Journal of the Royal Statistical Society*, Series B, 13, pp. 1-45.

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

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

Chaudhuri, P. and Mykland, P. A. (1993). "Nonlinear Experiments: Optimal Design and Inference Based on Likelihood". *Journal of the American Statistical Association* 88, pp. 538-546.

Chen, J. J., Li, L., and Jackson, C. D. (1996). "Analysis of Quantal Response Data from Mixture Experiments". *Environmetrics* 7, pp. 503-512.

Cheng, S.-W. and Wu, C. F. J. (2001). "Factor Screening and Response Surface Exploration". *Statistica Sinica* 11, pp. 553-604.

Chipman, H. (1996). "Bayesian Variable Selection with Related Predictors". *The Canadian Journal of Statistics* 24, pp. 17-36.

Cook, R. D. and Nachtsheim, C. J. (1982). "Model-Robust, Linear Optimal Designs". *Technometrics* 24, pp. 49-54.

Cornell, J. A. (1986). "A Comparison Between Two Ten-point Designs for Studying Three-component Mixture Systems". *Journal of Quality Technology* 18, pp.1-15.

Cornell, J. A. (1988). "Analyzing Data from Mixture Experiments Containing Process Variables: A Split-Plot Approach". *Journal of Quality Technology* 20, 2-23.

Cornell, J. A. (1995). "Fitting Models to Data from Mixture Experiments Containing Other Factors". *Journal of Quality Technology* 27, pp. 13-33.

Cornell, J. A. (2002). *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data.* Third Edition. John Wiley & Sons, New York, NY.

Cornell, J. A. and Gorman, J. W. (1984). "Fractional Design Plans for Process Variables in Mixture Experiments". *Journal of Quality Technology* 16, pp. 20-38.

Cornell, J. A. and Gorman, J. W. (2003). "Two New Mixture Models: Living with Collinearity but Removing its Influence". *Journal of Quality Technology* 35, pp. 78-88.

Cox, D. R. (1961). "Tests of Separate Families of Hypotheses". *Proceedings of the Fourth Berkeley Symposium* 1, pp. 105-123.

Cox, D. R. (1971). "A Note on Polynomial Response Functions for Mixtures". *Biometrika* 58, pp.155-159.

Draper, N. R. (1985). "Small Composite Designs". *Technometrics* 27, pp. 173-180.

DuMouchel, W. and Jones, B. (1994). "A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model." *Technometrics* 36, pp. 37-47.

Fang, K. T. (1980). "Experimental Design by Uniform Distribution." *Acta Mathematica Applicatae Sinica* 3, pp. 363-372.

Giovannitti-Jensen, A. and Myers, R. H. (1989). "Graphical Assessment of the Prediction Capability of Response Surface Designs". *Technometrics* 31, pp.159-171.

Goldfarb, H. B., Anderson-Cook C. M., Borror, C. M., and Montgomery, D. C. (2004). "Fraction of Design Space Plots to Assess the Prediction Capability of Mixture and Mixture-Process Designs". *Journal of Quality Technology* 36, pp. 169-179.

Goldfarb, H. B., Borror, C. M., Montgomery, D. C., and Anderson-Cook C. M. (2004). "Three-Dimensional Variance Dispersion Graphs for Mixture-Process Experiments". *Journal of Quality Technology* 36, pp. 109-124.

Gorman, J. W. and Cornell, J. A. (1982). "A Note on Model Reduction for Experiments with Both Mixture Components and Process Variables'". *Technometrics* 24, pp. 243-247.

Hamada, M., Martz, H. F., and Steiner, S. H. (2002). "Accounting for Mixture Errors in Analyzing (Analytical) Mixture Experiments". *Los Alamos National Laboratory Report LA-UR-02-5484.*

Hartley, H. O. (1959). "Smallest Composite Designs for Quadratic Response Surfaces". *Biometrics* 15, pp. 611-624.

Heredia-Langner, A., Montgomery, D. C., Carlyle, W. M., and Borror, C. M. (2004). "Model-Robust Optimal Designs: A Genetic Algorithm Approach". *Journal of Quality Technology* 36, pp. 263-279.

Hill, P. D. H. (1978). "A Review of Experimental Design Procedures for Regression Model Discrimination". *Technometrics* 20, pp. 15-21.

Hoke, A. T. (1974). "Economical Second-Order Designs Based on Irregular Fractions of the 3^n Factorial". *Technometrics* 16, pp. 375-384.

Hunter, W. G. and Reiner, A. M. (1965). "Designs for Discriminating between Two Rival Models". *Technometrics* 7, pp. 307-323.

Khuri, A. I. (2001). “An Overview of the Use of Generalized Linear Models in Response Surface Methodology”. *Nonlinear Analysis* 47, pp. 2023-2034.

Khuri, A. I. (2003). “Current Modeling and Design Issues in Response Surface Methodology: GLMs and Models with Block Effects”. *Statistics in Industry*, pp.209-229. Handbook of Statistics 22, North-Holland, Amsterdam.

Khuri, A. I., Harrison, J. M., and Cornell, J. A. (1999). “Using Quantile Plots of the Prediction Variance for Comparing Designs for a Constrained Mixture Region: An Application Involving a Fertilizer Experiment”. *Applied Statistics* 48, pp. 521-532.

Khuri A. I., Kim H. J., and Um Y. (1996). “Quantile Plots of the Prediction Variance for Response Surface Designs”. *Computational Statistics & Data Analysis* 22, pp. 395-407.

Khuri A. I. and Lee J. (1998). “A Graphical Approach for Evaluating and Comparing Designs for Nonlinear Models”. *Computational Statistics and Data Analysis* 27, pp. 433-443.

Kiefer, J. (1958). “On the Nonrandomized Optimality and Randomized Nonoptimality of Symmetrical Designs”. *Annals of Mathematical Statistics* 29, pp. 675-699.

Kiefer, J. and Wolfowitz, J. (1959). “Optimum Designs in Regression Problems”. *Annals of Mathematical Statistics* 30, pp. 271-294.

Kowalski, S. M., Cornell, J. A., and Vining, G. G. (2002). “Split-Plot Designs and Estimation Methods for Mixture Experiments with Process Variables”. *Technometrics* 44, pp. 72-79.

Kowalski, S. M., Cornell, J. A., and Vining, G. G. (2000). “A New Model and Class of Designs for Mixture Experiments with Process Variables”. *Communications in Statistics, A-Theory and Methods* 29, pp.2255-2280.

Li, C. and Nachtsheim, C.J. (2000). "Model-Robust Factorial Designs". *Technometrics* 42, pp. 345-352.

Lin, H., Myers, R. H., and Ye, K. (2000) "Bayesian Two-stage Optimal Design for Mixture Models". *Journal of Statistical Computation and Simulation* 66, pp. 209-231.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* Second Edition. Chapman and Hall, New York, NY.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code". *Technometrics* 21, pp. 239-245.

Minkin S. (1987). "Optimal Designs for Binary Data". *Journal of the American Statistical Association* 82, pp.1098-1103.

Montgomery, D.C. (2001). *Design and Analysis of Experiments.* Fifth Edition. John Wiley & Sons, New York, NY.

Montgomery, D.C. and Voth, S. R. (1994). "Multicollinearity and Leverage in Mixture Experiments". *Journal of Quality Technology* 26, pp. 96-108.

Myers, R. H. (1999). "Response Surface Methodology - Current Status and Future Directions". *Journal of Quality Technology* 31, pp. 30-44.

Myers, R. H., Vining, G. G, Giovannitti-Jensen, A., and Myers, S. L. (1992). "Dispersion Properties of Second-Order Response Surface Designs". *Journal of Quality Technology* 24, pp. 1-11.

Myers, R. H. and Montgomery, D. C. (2002). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments.* Second Edition. John Wiley & Sons, New York, NY.

Myers, R. H., Montgomery, D. C., and Vining G. G. (2002). *Generalized Linear Models: With Applications in Engineering and the Sciences.* First Edition. John Wiley & Sons, New York, NY.

Myers, R. H., Montgomery, D. C., Vining, G. G., Borror, C. M., and Kowalski S. M. (2004). "Response Surface Methodology: A Retrospective and Literature Study". *Journal of Quality Technology* 36, pp. 53-77.

Myers, W. R., Myers, R. H., Carter, W. H. Jr., and White K. L. Jr. (1996). "Two-stage Designs for the Logistic Regression Model in Single-agent Bioassays". *Journal of Biopharmaceutical Statistics* 6, pp. 283-301.

Nelder, J. A. and Wedderburn, R. W. M. (1972). "Generalized Linear Models". *Journal of the Royal Statistical Society, Series A* 135, pp. 370-384.

Notz, W. (1982). "Minimal Second-Order Designs". *Journal of Statistical Planning and Inference* 6, pp. 47-58.

Ozol-Godfrey A., Anderson-Cook, C. M., and Montgomery D. C. (to appear, 2005). "Fraction of Design Space Plots for Examining Model Robustness". *Journal of Quality Technology*.

Park Y., Richardson D. E., Borror C. B., Ozol-Godfrey A., Anderson-Cook C. M., and Montgomery D. C., "Prediction Variance Properties of Second-Order Response Surface Designs for Cuboidal Regions", submitted to *Journal of Quality Technology*.

Piepel, G. F. and Anderson, C. M. (1992). "Variance Dispersion Graphs for Designs on Polyhedral Regions". *Proceedings of the Section on Physical and Engineering Sciences*, American Statistical Association, Alexandria, VA, pp.111-117.

Piepel, G. F. Anderson, C. M; and Redgate, P. E. (1993a). "Variance Dispersion Graphs for Designs on Polyhedral Regions - Revisited". *Proceedings of the Section on Physical and Engineering Sciences,* American Statistical Association, Alexandria, VA, 102-107.

Piepel, G. F., Anderson, C. M, and Redgate, P. E. (1993b). "Response Surface Designs for Irregularly-Shaped Regions" (Parts 1, 2, and 3). *Proceedings of the Section on Physical and Engineering Sciences,* American Statistical Association, Alexandria, VA, pp. 205-227.

Piepel, G. F. and Cornell, J.A. (1994). "Mixture Experiment Approaches: Examples, Discussion, and Recommendations". *Journal of Quality Technology* 26, pp. 177-195.

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

Rendon-Andere, J., Montgomery, C. M., and Rollier, D. A. (1997). "Design of Mixture Experiments Using Bayesian D-Optimality". *Journal of Quality Technology* 29, pp. 451-463.

Robinson, K. S. and Khuri, A. I. (2003). "Quantile Dispersion Graphs for Evaluating and Comparing Designs for Logistic Regression Models". *Computational Statistics & Data Analysis* 43, pp. 46-62.

Roth, P. M. (1965). *Design of Experiments for Discriminating Among Rival Models.* Ph. D. Thesis, Princeton University.

Rozum, M. A. and Myers, R. H. (1991). "Variance Dispersion Graphs for Cuboidal Regions". Paper Presented at Joint Statistical Meetings, American Statistical Association, Atlanta, GA.

Rubenstein, R. Y. (1982). "Generating Random Vectors Uniformly Distributed Inside and on the Surface of Different Regions". *European Journal of Operational Research* 10, pp. 205-209.

Scheffe′, H. (1958). "Experiments with Mixtures". *Journal of the Royal Statistical Society,* Series B, 20, pp. 344-360.

Scheffe′, H. (1963). "Simplex-Centroid Designs for Experiments with Mixtures". *Journal of the Royal Statistical Society,* Series B, 25, pp. 235-263.

Sirohi, R. S. and Radha-Krishna, H. C. (1991). *Mechanical Measurements.* Third Edition. Halstead Press/John Wiley & Sons, New York, NY.

Sitter, R. R. (1992). "Robust Designs for Binary Data". *Biometrics* 48, pp. 1145-1155.

Smith, W. F. and Beverly, T. A. (1997). "Generating Linear and Quadratic Cox Mixture Models". *Journal of Quality Technology* 29, pp. 211-224.

Snee, R. D. and Rayner, A. A. (1982). "Assessing the Accuracy of Mixture Model Regression Calculations". *Journal of Quality Technology* 14, pp. 647-79.

Soares, J. F. and Wu, C. F. J. (1985). "Optimality of Random Allocation Design for the Control of Accidental Bias in Sequential Experiments." *Journal of Statistical Planning and Inference* 11, pp. 81-87.

Steiner, S. H. and Hamada, M. (1997). "Making Mixtures Robust to Noise and Mixing Measurement Errors." *Journal of Quality Technology 29*, pp. 441-450.

Trinca, L. A. and Gilmour, S. G. (1998). "Variance Dispersion Graphs for Comparing Blocked Response Surface Designs". *Journal of Quality Technology* 30, pp. 349-364.

Tsutakawa, R. K. (1980). "Selection of Dose Levels for Estimating a Percentage Point of a Logistic Quantal Response Curve". *Applied Statistics* 29, pp. 25-33.

Vining, G. G. (1993). "A Computer Program for Generating Variance Dispersion Graphs". *Journal of Quality Technology* 25, pp. 45-58.

Vining, G. G., Cornell, J. A., and Myers, R. H. (1993). "A Graphical Approach for Evaluating Mixture Designs". *Journal of the Royal Statistical Association, Series C* 42, pp.127-138.

Vining, G. G. and Myers, R. H. (1991). "A Graphical Approach for Evaluating Response Surface Designs in Terms of the Mean Squared Error of Prediction". *Technometrics* 33, pp. 315-326.

Wedderburn, R. W. M. (1974). "Generalized Linear Models Specified in Terms of Constraints". *Journal of the Royal Statistical Society, Series B* 36, pp. 449-454.

Wu, C. F. J. (1981). "On the Robustness and Efficiency of Some Randomized Designs". *The Annals of Statistics* 9, pp. 1168-1177.

Wu, C. F. J. and Hamada M. (2000). *Experiments: Planning, Analysis, and Parameter Design Optimization*. First Edition. John Wiley & Sons, New York, NY.

Wynn, H. P. (1972). "Results in the Theory and Construction of D-optimum Experimental Designs". *Journal of the Royal Statistical Society, Series B* 34, pp. 133-147.

Zacks, S. (1977). "Problems and Approaches in Design of Experiments for Estimation and Testing in Nonlinear Models". *Preceedings of the 4th International Symposium on Multivariate Analysis*, ed. P. R. Krishnaiah, Amsterdam: North-Holland, pp. 209-223.

Zahran, A. R (2002). “On the Efficiency of Designs for Linear Models in Non-regular Regions and the Use of Standard Designs for Generalized Linear Models”. Dissertation, Department of Statistics, Virginia Polytechnic Institute and State University.

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

Zahran, A. R. and Myers, R. H. (2003). “Use of Standard Factorial Designs with Generalized Linear Models”. Technical Report 03-04, Department of Statistics, Virginia Polytechnic Institute and State University.

Appendix-A Visual Basic for Applications Code

Visual Basic for Applications in Excel was used for programming. Macros were created using functions, sub procedures, class modules and forms. Examples of macros are given including some screen shots and coding.

A.1 Example of a Macro Named SPVMacro

This macro allows the viewer to input a design matrix, corner points of the design space, the number of factors and the number of points to be sampled (n) into a dialog box, and the macro outputs two Excel objects. The macro randomly chooses the n sample points by sampling uniformly from the design space, and it calculates the SPV for each point for three types of models: the second order model, the first order model with interaction and the first order model. Since the probability of sampling the extreme of all factors simultaneously is very low, the macro also calculates the SPVs for the provided corner points. It then sorts all the SPVs from smallest to highest and obtains the FDS plot by plotting these sorted values versus the values $\frac{1}{n},\frac{2}{n},.....\frac{n}{n}$. The outputs are a graph of an FDS plot for the three types of models for the chosen design and an Excel sheet of the SPV values calculated at the sample points.

A.1.i Screen Shots from the SPVMacro

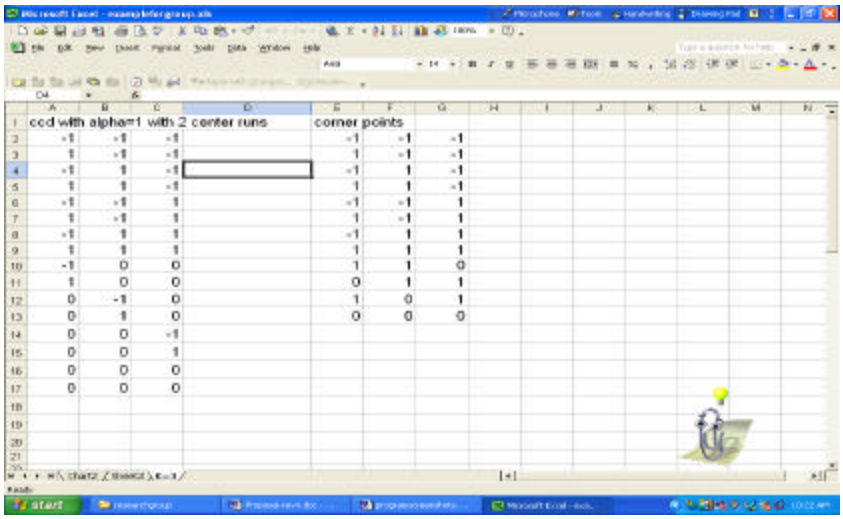

Figure A.1.1: Screen Shot of Excel Sheet Where Data are Entered

Figure A.1.1 shows the form of the data to be used by the macro, while Figure A.1.2 shows the dialog box to be filled in by the user.

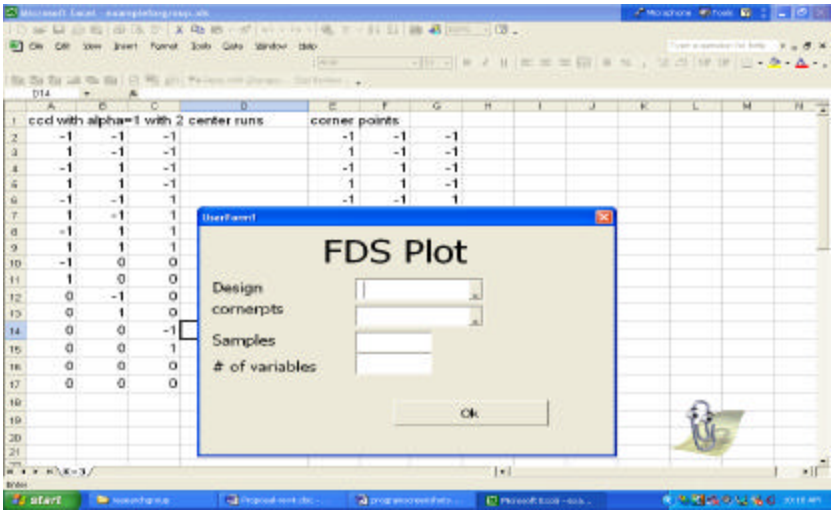

Figure A.1.2: Screen Shot of Excel Sheet with the Dialog Box

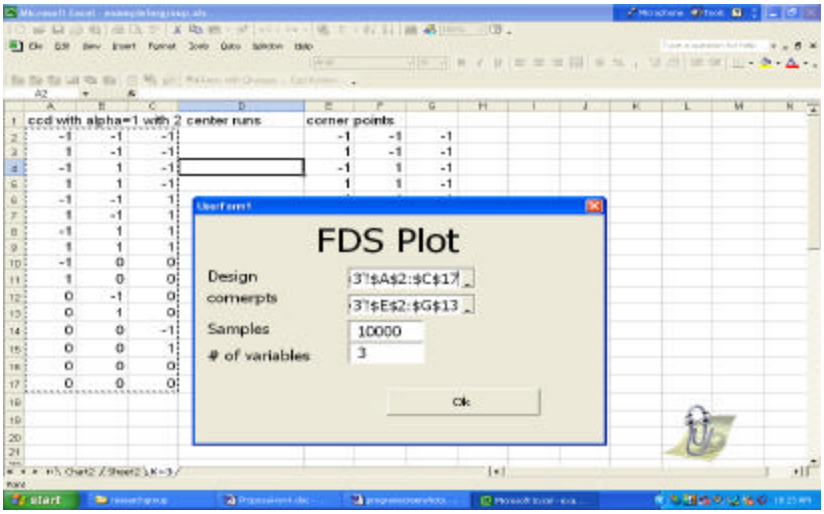

Figure A.1.3: Screen Shot of Excel Sheet where Data are Being Entered into the Dialog Box

Figure A.1.3 shows how data can be entered by highlighting cells in the spreadsheet. A common value for the number of samples is 10000 with the corresponding macro running time of 15 seconds. Figure A.1.4 is one of the two outputs of the macro.

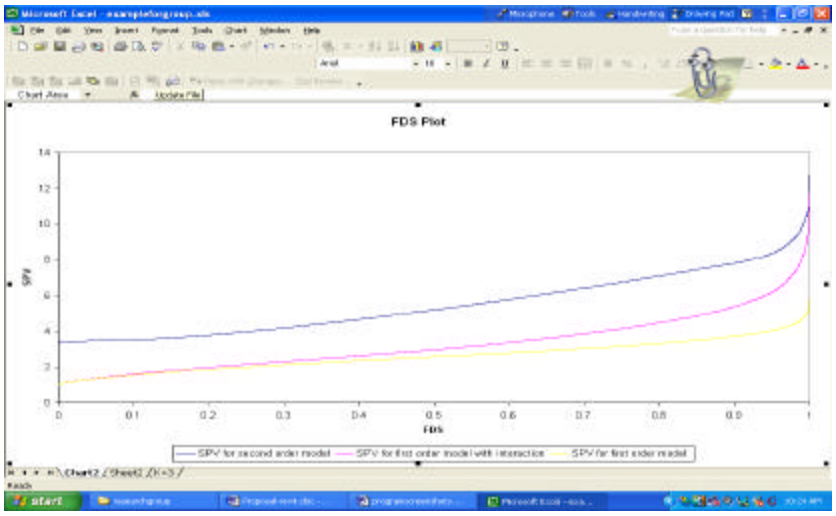

Figure A.1.4: Screen Shot of an Excel Plot Displaying the FDS Plot

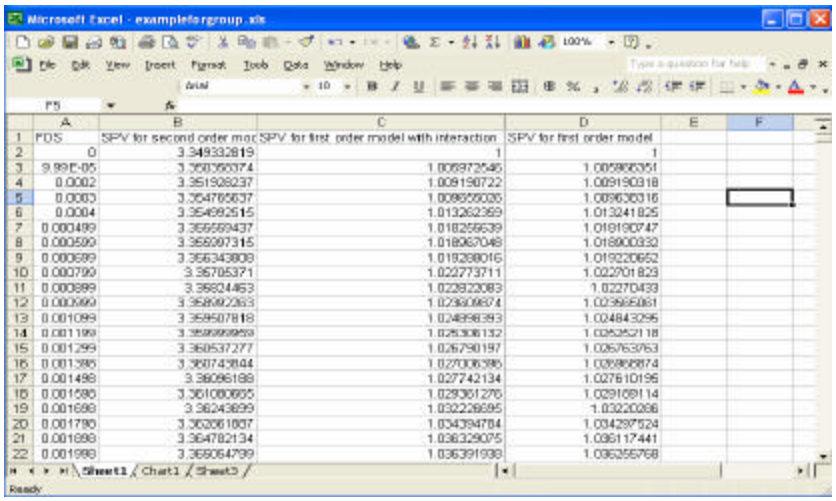

Figure A.1.5: Screen Shot of an Excel Sheet Displaying the FDS values and SPVs for Each Model

Figure A.1.5 shows the numerical output of the macro which would allow flexible graphing in any other graphics package or for new specialized graphs to be created.

A.2 Example of a Macro Named FDS_GLM Macro

This macro is used to obtain an FDS plot (using PPV values) of a design adapted for the GLM case (assuming a binomial response). It allows the viewer to input an X matrix, the number of observations at each design location, three predicted mean ranges, the number

of points to be sampled and additional locations in the design space into a dialog box, and the macro outputs two Excel objects. The outputs are a graph of three FDS curves for the design corresponding to three different predicted mean values and an Excel sheet of the SPV values for each case calculated at the sampled points.

A.2.i Screen Shots from the FDS_GLM Macro

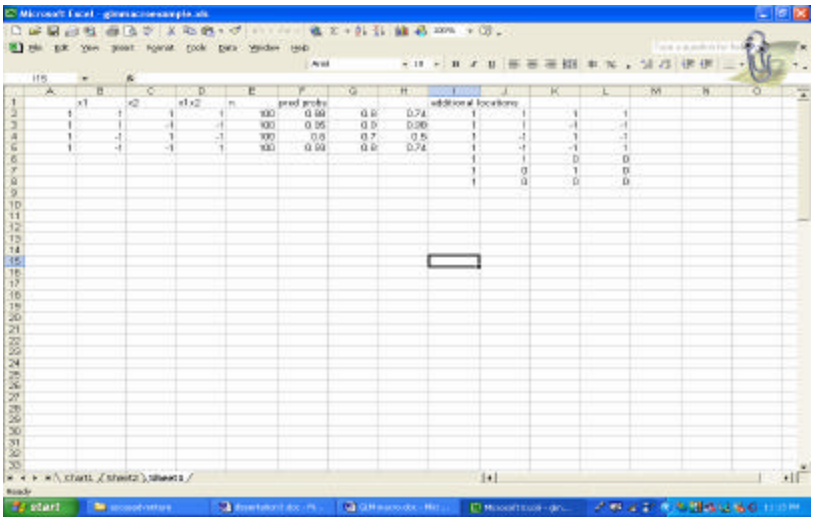

Figure A.2.1: Screen Shot of Excel Sheet Where Design Information is Entered (GLM)

Figure A.2.1 shows the form of the data to be used by the macro, while Figure A.2.2 shows the dialog box to be filled in by the user.

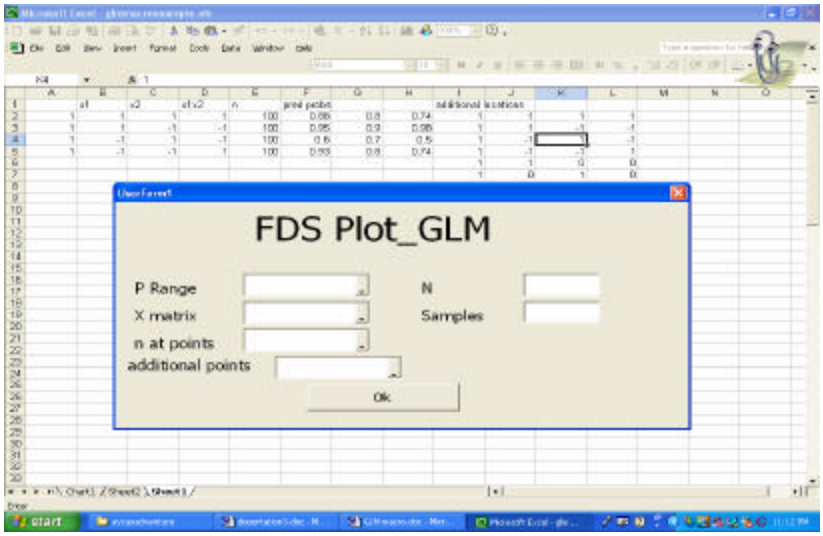

Figure A.2.2: Screen Shot of Excel Sheet with the Dialog Box (GLM)

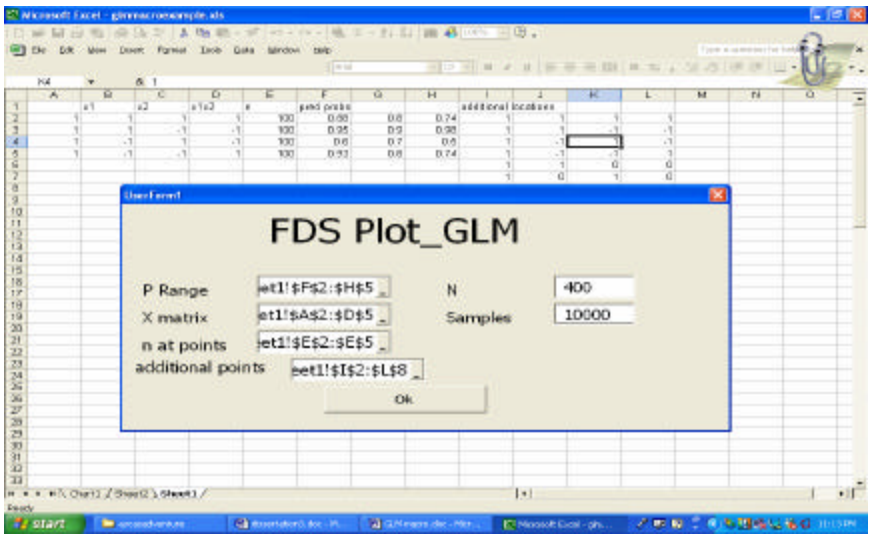

Figure A.2.3: Screen Shot of Excel Sheet where Data are Being Entered into the Dialog Box (GLM)

Figure A.2.3 shows how data can be entered by highlighting cells in the spreadsheet. A common value for the number of samples is 10000 with the corresponding macro running time of 7 seconds. Figure A.2.4 is one of the two outputs of the macro.

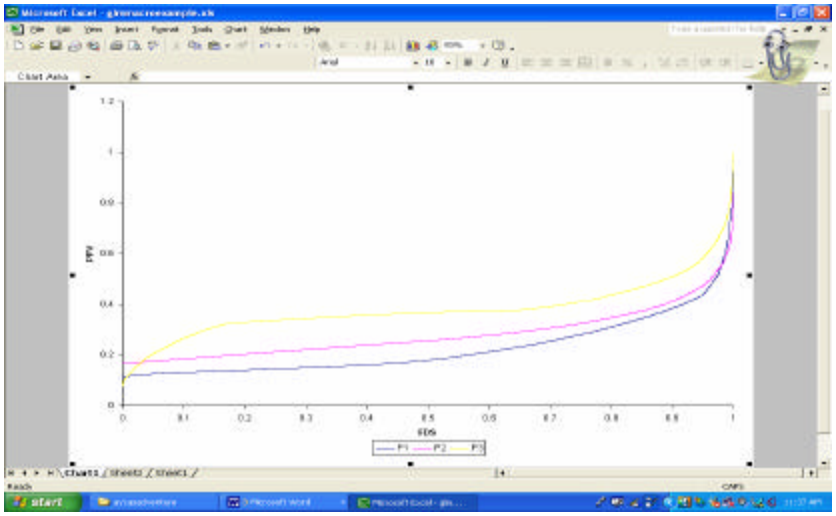

Figure A.2.4: Screen Shot of an Excel Plot Displaying the FDS Plot (GLM)

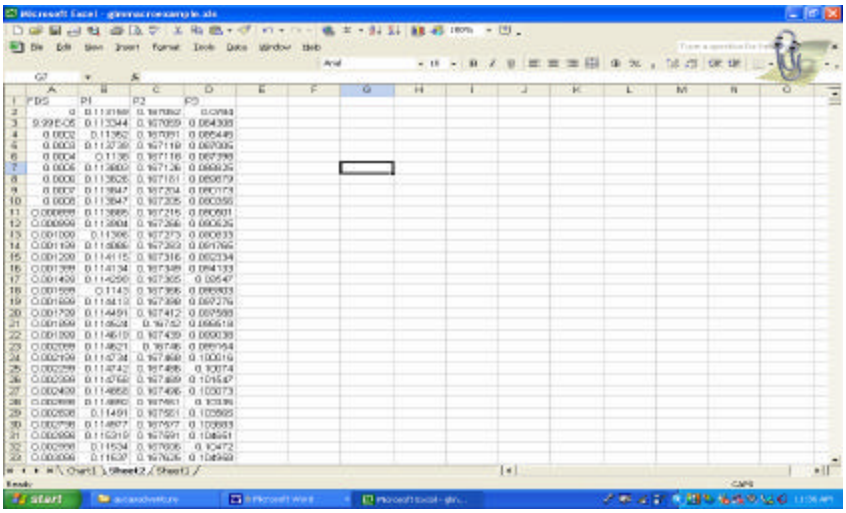

Figure A.2.5: Screen Shot of an Excel Sheet Displaying the FDS values and SPVs of a Design with for Different Predicted Mean Ranges (GLM)

Figure A.2.5 shows the numerical output of the macro which would allow flexible graphing in any other graphics package or for new specialized graphs to be created.

A.3 Example of a Macro Named FDS_Measurement Error Macro

This macro is used to obtain an FDS plot of a design adapted for mixing measurement error (assuming absolute errors from a normal distribution). It allows the viewer to input a design matrix, the number of mixture components, the standard deviation of the error distribution, the number of generated designs, the $(1-a)\%$ quantiles, the number of points to be sampled, and additional locations in the design space into a dialog box, and the macro outputs two Excel objects. The outputs are a graph of three FDS curves for a design with three different predicted mean values and an Excel sheet of the SPV values for each case calculated at the sample points.

A.3.i Screen Shots from the FDS_Measurement Error Macro

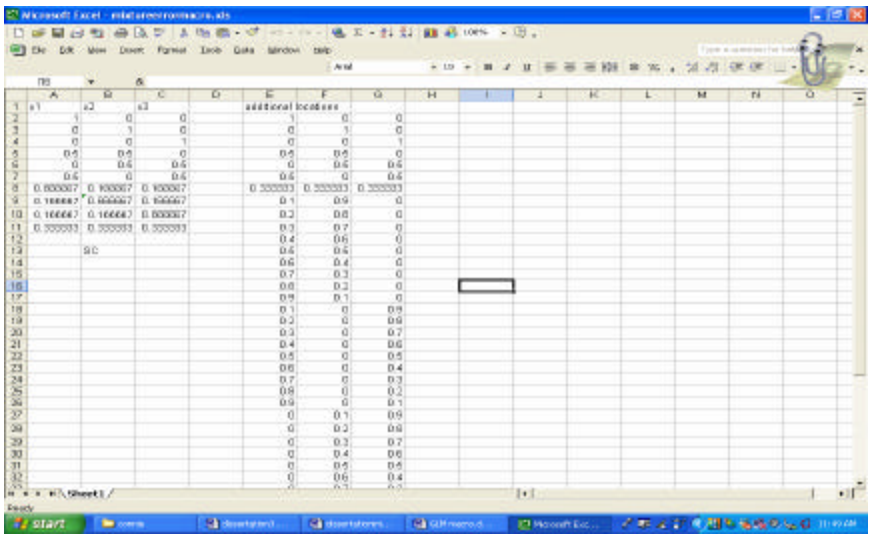

Figure A.3.1: Screen Shot of Excel Sheet Where Design Information is Entered (Measurement Error)

Figure A.3.1 shows the form of the data to be used by the macro, while Figure A.3.2 shows the dialog box to be filled in by the user.

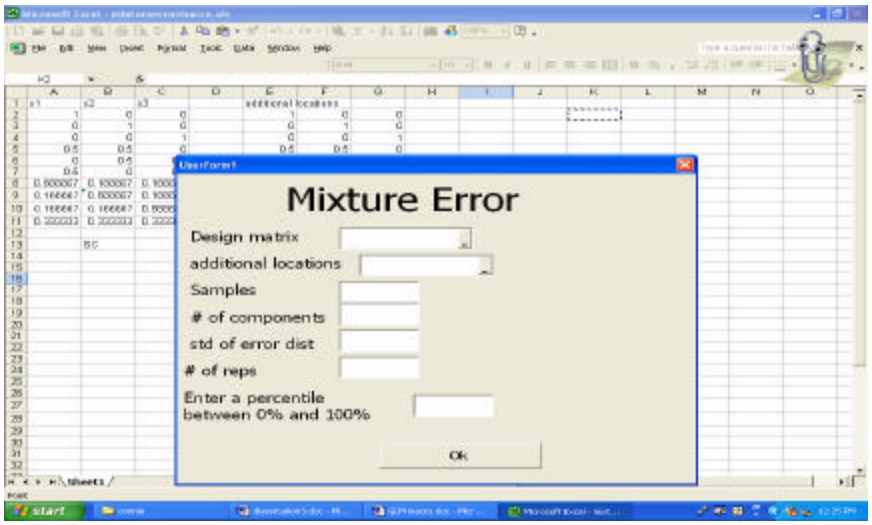

Figure A.3.2: Screen Shot of Excel Sheet with the Dialog Box (Measurement Error)

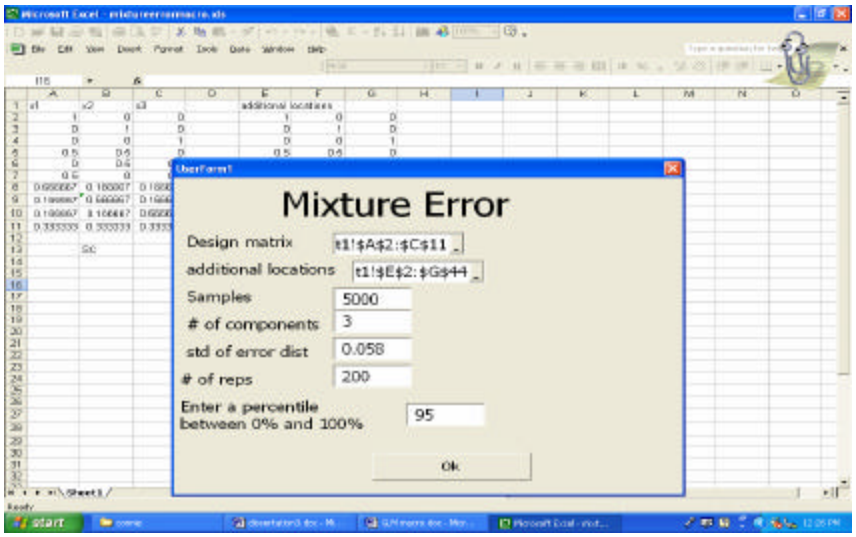

Figure A.3.3: Screen Shot of Excel Sheet where Data are Being Entered into the Dialog Box (Measurement Error)

Figure A.3.3 shows how data can be entered by highlighting cells in the spreadsheet. A common value for the number of samples is 5000 with the corresponding macro running time of 7 seconds. Figure A.3.4 is one of the two outputs of the macro.

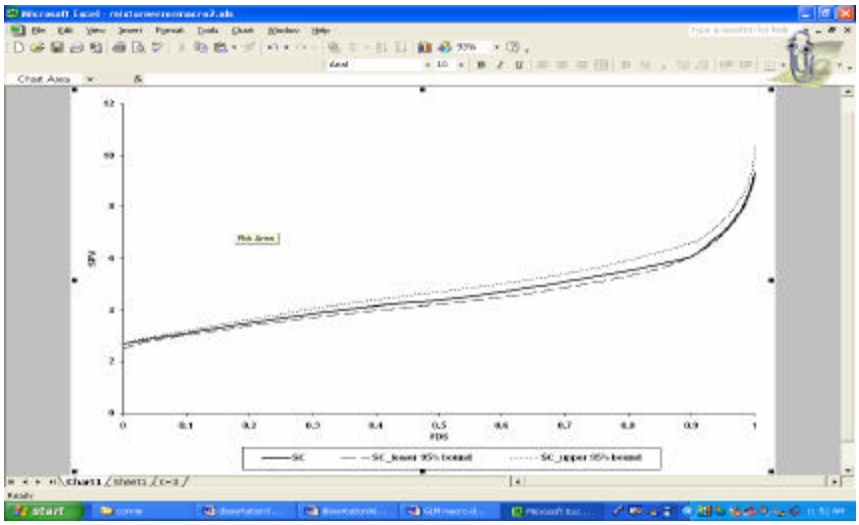

Figure A.3.4: Screen Shot of an Excel Plot Displaying the FDS Plot (Measurement Error)

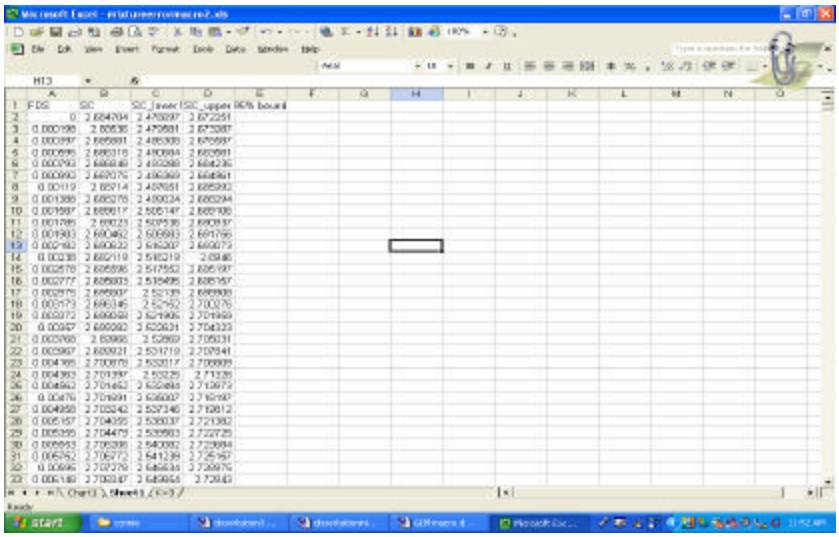

Figure A.3.5: Screen Shot of an Excel Sheet Displaying the FDS values and SPVs of a Design with Measurement Error

Figure A.3.5 shows the numerical output of the macro which would allow flexible graphing in any other graphics package or for new specialized graphs to be created.

A.4 Example of a Sub Procedure and Some Functions

Below is an example of a sub procedure and some functions.

A.4.i Sub Procedure Linear1

```
Sub linear1()
'This module outputs the FDS Plot for a design for the three standard types of models
Dim n_i As Integer
Dim X1 As New Matrix
Dim corner As New Matrix
Dim I As Integer
Dim hold As Double
Dim SPV As New Matrix
Dim SPVx As New Matrix
Dim Label As Variant
Dim Nsamps As Integer
```

```
Dim nsamps2 As Integer
Dim k As Integer
'Dimension Ranges for data
Dim Xrange As Range
Dim corrange As Range
Dim m As Integer
'Read the data into ranges
Set Xrange = Range(SPVForm.Design)
Set corrange = Range(SPVForm.cornerpts)
'Read the data into matrices
Call X1.Read(Xrange)
Call corner.Read(corrange)
'Read constants into variables and round
Nsamps = Round(SPVForm.Samples, 0)
k = Round(SPVForm.VAR, 0)
m = corner.rows
n_i = X1.rows
'Add a new Sheet
Sheets.Add
ActiveSheet.Cells(1, 1) = "FDS"
ActiveSheet.Cells(1, 2) = "SPV for second order model"
ActiveSheet.Cells(1, 3) = "SPV for first order model with interaction"
ActiveSheet.Cells(1, 4) = "SPV for first order model "
nsamps2 = Nsamps + m
' Set the index for the elements/plot
Call SPVx.Init(nsamps2)
For I = 1 To nsamps2
Call SPVx.setElem((I - 1) / (nsamps2 - 1), I, 1)
Next I
dRandSeed (0)
'Generate initial value
```

```
Set SPVx = robustSPV(SPVx, X1, n_i, k, corner)
'Write the results to sheet
Call SPVx.Write2(ActiveSheet.Range("A2"))
Dim Loc1 As Range
Set Loc1 = ActiveSheet.Range("A1:D65001")
Application.ScreenUpdating = False
Charts.Add
ActiveChart.ChartType = xlXYScatterLinesNoMarkers
ActiveChart.SetSourceData Source:=Loc1, _
PlotBy:=xlColumns
ActiveChart.Location Where:=xlLocationAsNewSheet
With ActiveChart
.HasTitle = True
.ChartTitle.Characters.Text = "FDS Plot"
.Axes(xlCategory, xlPrimary).HasTitle = True
.Axes(xlCategory, xlPrimary).AxisTitle.Characters.Text = "FDS"
.Axes(xlValue, xlPrimary).HasTitle = True
.Axes(xlValue, xlPrimary).AxisTitle.Characters.Text = "SPV"
End With
With ActiveChart.Axes(xlCategory)
.HasMajorGridlines = False
.HasMinorGridlines = False
End With
With ActiveChart.Axes(xlValue)
.HasMajorGridlines = False
.HasMinorGridlines = False
End With
ActiveChart.HasLegend = True
ActiveChart.Legend.Select
Selection.Position = xlBottom
ActiveChart.PlotArea.Select
```

```
ActiveChart.Axes(xlCategory).Select
With ActiveChart.Axes(xlCategory)
.MinimumScale = 0
.MaximumScale = 1
.MinorUnitIsAuto = True
.MajorUnitIsAuto = True
.Crosses = xlAutomatic
.ReversePlotOrder = False
.ScaleType = xlLinear
.DisplayUnit = xlNone
End With
End Sub
```

A.4.ii Function Xsecond

```
Function xsecond(x As Matrix) As Matrix
'This function outputs the second order model form for any design
Dim I As Integer
Dim j As Integer
Dim k As Integer
Dim l As Integer
Dim hold As Matrix
Dim hold1 As Matrix
Dim hold2 As Matrix
Dim MultC As Matrix
Dim res As Matrix
Dim y As New Matrix
Call y.Init(x.rows)
For I = 1 To x.rows
Call y.setElem(1, I, 1)
Next I
Set res = y.MultConst(1)
```

```
For j = 1 To x.cols
Set hold = x.GetColumn(j)
Set res = res.ConCatCol(hold)
Next j
For k = 1 To x.cols
For l = k To x.cols
Set hold1 = x.GetColumn(k)
Set hold2 = x.GetColumn(l)
Set MultC = hold1.MultCol(hold2)
Set res = res.ConCatCol(MultC)
Next l
Next k
Set xsecond = res
End Function
```

A.4.iii Function Xfirstint

```
Function xfirstint(x As Matrix) As Matrix
'This function outputs the first order model with interactions for any design
Dim I As Integer
Dim j As Integer
Dim k As Integer
Dim l As Integer
Dim hold As Matrix
Dim hold1 As Matrix
Dim hold2 As Matrix
Dim MultC As Matrix
Dim res As Matrix
Dim y As New Matrix
Call y.Init(x.rows)
For I = 1 To x.rows
Call y.setElem(1, I, 1)
```

```
Next I
Set res = y.MultConst(1)
For j = 1 To x.cols
Set hold = x.GetColumn(j)
Set res = res.ConCatCol(hold)
Next j
For k = 1 To x.cols - 1
For l = k + 1 To x.cols
Set hold1 = x.GetColumn(k)
Set hold2 = x.GetColumn(l)
Set MultC = hold1.MultCol(hold2)

Set res = res.ConCatCol(MultC)
Next l
Next k
Set xfirstint = res
End Function
```

A.4.iv Function Xfirst

```
Function xfirst(x As Matrix) As Matrix
Dim I As Integer
Dim j As Integer
Dim hold As Matrix
Dim MultC As Matrix
Dim res As Matrix
Dim y As New Matrix
Call y.Init(x.rows)
For I = 1 To x.rows
Call y.setElem(1, I, 1)
Next I
Set res = y.MultConst(1)
```

```
For j = 1 To x.cols
Set hold = x.GetColumn(j)
Set res = res.ConCatCol(hold)
Next j
Set xfirst = res
End Function
```

A.4.v Function RobustSPV

```
Function robustSPV(SPV As Matrix, X1 As Matrix, N As Integer, k As Integer, y As
Matrix) As Matrix
'This function calculates the SPV values
Dim Xnew As Matrix
Dim xnewfirst As Matrix
Dim xnewint As Matrix
Dim I As Integer
Dim j As Integer
Dim X2 As Matrix
Dim xfirs As Matrix
Dim xfirsint As Matrix
Dim XX2 As Matrix
Dim xx1 As Matrix
Dim xx1int As Matrix
Dim x3 As Matrix
Dim xcorner1 As Matrix
Dim xcorner2 As Matrix
Dim xcorner3 As Matrix
Dim holdi As Double
Dim hold2 As Double
Dim hold1 As Double
Dim holdcor1 As Double
Dim holdcor2 As Double
```

```
Dim holdcor3 As Double
Dim holdint As Double
Dim m  As Integer
Dim SPVnew2 As New Matrix
Dim SPVnewfirst As New Matrix
Dim SPVnewint As New Matrix
Dim cor1 As Matrix
Dim cor2 As Matrix
Dim cor3 As Matrix
m = y.rows
Call SPVnew2.Init(SPV.rows)
Call SPVnewfirst.Init(SPV.rows)
Call SPVnewint.Init(SPV.rows)
Set xcorner1 = xsecond(y)
Set xcorner2 = xfirstint(y)
Set xcorner3 = xfirst(y)
Set X2 = xsecond(X1)
Set xfirs = xfirst(X1)
Set xfirsint = xfirstint(X1)
Set XX2 = X2.Transpose.Mult(X2, True).Invdll
Set xx1 = xfirs.Transpose.Mult(xfirs, True).Invdll
Set xx1int = xfirsint.Transpose.Mult(xfirsint, True).Invdll
For I = 1 To (SPV.rows - m)
Set x3 = NoC(k)
Set Xnew = xsecond(x3)
hold2 = (N * (Xnew.Mult(XX2, True).Mult(Xnew.Transpose, True).Elem(1, 1)))
Call SPVnew2.setElem(hold2, I, 1)
Set xnewint = xfirstint(x3)
holdint = (N * (xnewint.Mult(xx1int, True).Mult(xnewint.Transpose, True).Elem(1, 1)))
Call SPVnewint.setElem(holdint, I, 1)
Set xnewfirst = xfirst(x3)
```

```
hold1 = (N * (xnewfirst.Mult(xx1, True).Mult(xnewfirst.Transpose, True).Elem(1, 1)))
Call SPVnewfirst.setElem(hold1, I, 1)
Next I
For j = 1 To m
Set cor1 = xcorner1.GetRow(j)
holdcor1 = (N * (cor1.Mult(XX2, True).Mult(cor1.Transpose, True).Elem(1, 1)))
Call SPVnew2.setElem(holdcor1, SPV.rows - m + j, 1)
Set cor2 = xcorner2.GetRow(j)
holdcor2 = (N * (cor2.Mult(xx1int, True).Mult(cor2.Transpose, True).Elem(1, 1)))
Call SPVnewint.setElem(holdcor2, SPV.rows - m + j, 1)
Set cor3 = xcorner3.GetRow(j)
holdcor3 = (N * (cor3.Mult(xx1, True).Mult(cor3.Transpose, True).Elem(1, 1)))
Call SPVnewfirst.setElem(holdcor3, SPV.rows - m + j, 1)
Next j
Call Sort2Matrix(SPVnew2)
Set SPV = SPV.ConCatCol(SPVnew2)
Call Sort2Matrix(SPVnewint)
Set SPV = SPV.ConCatCol(SPVnewint)
Call Sort2Matrix(SPVnewfirst)
Set SPV = SPV.ConCatCol(SPVnewfirst)
Set robustSPV = SPV
End Function
```

Appendix-B

For designs under the GLM framework, it is still possible to obtain information about a design's G- and V-efficiencies through a transformation to the linear models case. The SPV for GLM is defined as follows: $N\frac{\text{var}(\hat{y}(\boldsymbol{x}_o))}{s_o^2}$ at a particular location, $\boldsymbol{x}_o$.

$= N\,\text{var}(y(\boldsymbol{x}_o))\boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o = Ns_o^2\boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o$.

By a transformation of the X matrix we can show that the SPV for GLMs is still in the correct scale as the SPV for linear models.

Let $Z = V^{1/2}X$ where $V^{1/2} = diag(s_i)$. Then $Z' = X'V^{1/2}$ which is a rescaling of the X matrix in a new design space. Each location $\boldsymbol{x}_i$ in the design matrix is rescaled proportional to the size of the variance, s_i. In addition, $\boldsymbol{z}_o' = s_o\boldsymbol{x}_o'$ places the location of interest, $\boldsymbol{x}_o$, into the new design space. The transformed $\boldsymbol{z}_o$ of Z is still in the correct position relative to the $\boldsymbol{x}_i$'s in X.

Therefore,

$Ns_o^2\boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o = Ns_o^2\frac{1}{s_o}\boldsymbol{z}_o'(ZZ)^{-1}\boldsymbol{z}_o\frac{1}{s_o} = Ns_o^2\frac{1}{s_o^2}\boldsymbol{z}_o'(Z'Z)^{-1}\boldsymbol{z}_o = N\boldsymbol{z}_o'(Z'Z)^{-1}\boldsymbol{z}_o$ where the Z matrix is dependent on the unknown parameters since s_i is a function of the parameters.

Appendix-C

The variance stabilizing transformation corresponding to a Poisson response is the square root transformation. When the transformation is applied to the data, the homogeneous variance model looks like E(w_i)=E($\sqrt{y_i}$) $\cong \boldsymbol{x}_i'b$ where we define $w_i = \sqrt{y_i}$.

The square root transformation on the response y_i, will result in a constant variance of y_i, i.e., var($\sqrt{y_i}$) $\cong s^2 = 1/4$ (Myers and Montgomery, page 243).

We are interested in comparing the SPV/PPV values of a design for GLM to the design with a variance stabilizing transformation. The SPV/PPV values of the design obtained directly using the transformation correspond to $w_i = \sqrt{y_i}$ not y_i. This is not a fair comparison unless we obtain the SPV/PPV values of the design for the transformation on the same scale as the original y_i. This makes sure the SPV/PPV values of the design with the GLM and the transformation are on the same scale.

The Delta method states that $\text{var}(g(w)) \cong \text{var}(w)[g'(q)]^2$ where q is the expected mean of the transformed response and $g(q) = q^2$ for Poisson data. Therefore $\text{var}(\hat{g}(\boldsymbol{w}_o)) = \text{var}(\hat{\boldsymbol{w}}_o)[g'(q)]^2 = 4\text{var}(\hat{\boldsymbol{w}}_o)q^2$. Since $q \cong X'b \cong \sqrt{m}$,

$$\text{var}(\hat{g}(\boldsymbol{w}_o)) = 4\text{var}(\hat{\boldsymbol{w}}_o)m_o . \qquad (*)$$

The PPV of the transformed case ($PPV_{\hat{g}(\boldsymbol{w}_o)}$) on the same scale as the PPV of the GLM case ($PPV_{\hat{\boldsymbol{w}}_o}$) is obtained as follows:

$$PPV_{\hat{g}(\boldsymbol{w}_o)} = N\,\text{var}(\hat{g}(\boldsymbol{w}_o)) \overset{by(*)}{=} 4N\,\text{var}(\hat{\boldsymbol{w}}_o)m_o =$$

$4Ns^2\boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o m_o$ since $\text{var}(\hat{\boldsymbol{w}}_o) = s^2\boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o$.

Since the variance of each observation, s^2, for the transformed case is approximately equal to ¼, $4Ns^2\boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o m_o \cong 4N(1/4)\boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o m_o$.

Cancelling out ¼ from $\text{var}(\hat{g}(\boldsymbol{w}_o))$ results in $4N(1/4)\boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o m_o$.

$= N\boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o m_o$.

Recall the definition of the PPV for the GLM case is equal to $Nm_o^2\boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o$. As you see the form of the PPV of the GLM is very similar to the form of the PPV of the transformed case, however, the GLM case has an extra m_o factor.

The SPV of the transformed case $(SPV_{\hat{g}(\boldsymbol{w}_o)})$ on the same scale as the SPV of the GLM case $(SPV_{(\hat{\boldsymbol{w}}_o)})$ is obtained as follows:

$SPV_{\hat{g}(\boldsymbol{w}_o)} = N\dfrac{\text{var}(\hat{g}(\boldsymbol{w}_o))}{\sigma_o^2} = N\dfrac{\text{var}(\hat{g}(\boldsymbol{w}_o))}{m_o}$ since the mean and the variance of a Poisson random variable are the same.

$\overset{by(*)}{=} N\dfrac{4\text{var}(\hat{\boldsymbol{w}}_o)m_o}{m_o} = \dfrac{4\sigma^2\boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o m_o}{m_o}$ since $\text{var}(\hat{\boldsymbol{w}}_o) = \sigma^2\boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o$.

$\cong N\dfrac{4(1/4)\boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o m_o}{m_o}$ since the variance of each observation, σ^2, for the transformed case is approximately equal to ¼.

$= N\boldsymbol{x}_o'(X'X)^{-1}\boldsymbol{x}_o$ which is identical to the SPV for the transformed response, *w*.

Recall the definition of the SPV for the GLM case is equal to $Nm_o\boldsymbol{x}_o'(X'VX)^{-1}\boldsymbol{x}_o$. As you see the form of the SPV of the GLM is very similar to the form of the SPV of the transformed case.

VITA

Ayca Ozol-Godfrey

The author was born on February 3, 1977 in Istanbul, Turkey. She graduated from Uskudar American High School, Istanbul, Turkey in June 1995. In 1999, she graduated from Bogazici University, Faculty of Arts and Sciences, with a B.S. in Mathematics. Her graduate studies in the Department of Statistics started in August 1999 at the Virginia Polytechnic Institute and State University. She worked as a Graduate student in the department and a Consultant at the Statistical Consulting Center. She received her M.S. in Statistics at Virginia Polytechnic Institute and State University in December 2000. She received her Ph.D. in Statistics at the Virginia Polytechnic Institute and State University in December, 2004. She will be joining the Biostatistics and Clinical Technology department at Wyeth Pharmaceuticals as a Senior Biostatistician in January 2005.