

Experimental Design based on Compound
Optimization Criteria: Essays on Design Construction,
Analysis, and Applications

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Anna Errore

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Christopher J. Nachtsheim, Adviser

November 2021

ACKNOWLEDGEMENTS

This thesis represents the accomplishment of my Ph.D. degree at the University of Minnesota, but it is also a symbol of a more comprehensive and longer journey. My life in the past several years has been very challenging, even when rewarding; my personal and professional paths have interacted heavily and significantly; serious problems have slowed down and hurt my professional progress, but I worked hard to move forward despite the hardships, and I have several people to thank for this accomplishment. These people have been, and will continue to be, critically important for my results; most importantly they gave me the strength to always keep fighting all the challenges that life has put on my path.

First, I want to specifically thank my advisor, Chris Nachtsheim, who has supported and encouraged me throughout this journey, has taught me precious lessons, and has been on my corner and helped me to never give up.

I thank my thesis committee members, Dennis Cook, Karen Donohue, and William Li, who are wonderful people, educators, and scholars. I had the invaluable opportunity to learn from them, and the honor to work with each of them on different projects discussed in this dissertation.

In the past years, I had the opportunity to work with many other people. I thank Bradley Jones, who I profoundly esteem as a scholar and a person; Mark Albrecht, involved in one

of the projects of this thesis; Stefano Barone, the advisor of my first Ph.D. degree in Industrial Engineering, who introduced me to the academic world and allowed me to first come to the Carlson School as a visiting student. I have worked with many distinguished scholars and amazing people, I am still collaborating with many of them on ongoing research projects, and I thank them all; I hope to have the privilege to keep working with them in the future as well.

I thank all the professors I had at the University of Minnesota, and particularly KK Sinha, Rachna Shah, and Kevin Linderman; Steve Huchendorf for his mentorship on teaching; my fellow Ph.D. students, the staff members from the Department of Supply Chain and Operations, and the Ph.D. office of the Carlson School of Management. I am grateful for every interaction I had, every opportunity to share and grow.

Many other people in my life have had a more or less direct impact on my academic journey as well. My family and friends have been essential to my accomplishments, especially in the most difficult times. My parents, my sister, and my fiancé have endured a very difficult situation, and I could not have done anything at all without them.

I find it hard to properly show all my gratitude, in this short section and a few words, to all these people for their essential role in life, especially because I usually dislike public statements of private feelings. I hope to acknowledge them every day with my actions and words, and with my humble ways to attempt to make them proud.

DEDICATION

This dissertation is dedicated to my family

ABSTRACT

This thesis focuses on the construction of experimental designs having multiple objectives. It explores theoretical and practical applications of compound design construction and analysis in the context of problems involving statistical and practical optimization criteria. Different essays of this dissertation focus on linear or non-linear models characterizing different experimental settings. The overarching theme of the separate essays is the idea of constructing experimental designs for initial exploration of a given phenomenon - such as in the case of screening experiments - which are created with goals that are not only short-sightedly related to the initial exploratory phase of the experimentation, but that account for the potential subsequent goals of further phases. In this fashion, screening experiments are created with the double goal of initial factor screenings, but also protection from bias induced by higher-order terms not included in the initial model. This idea is applied to both linear and non-linear experimental designs construction and involved in the analysis of such types of designs when doing variables selection. Similarly, in the applied context of two-round procurement auctions, the experimental design created for bidding strategies on the first round of the auction accounts for the impact on profit maximization related to the bidding strategies involved in the second round.

Contents

List of Tables	ix
List of Figures	xi
1 Introduction	1
2 Benefits and Fast Construction of Efficient Two-Level Foldover Designs	11
2.1 Introduction	11
2.2 Motivation for our approach	15
2.3 Constructing EFDs	21
2.4 Results	25
2.4.1 Orthogonal Two-Level EFDs	26
2.4.2 EFDs Versus Margolin’s Designs and the LMS Designs	29
2.4.3 New EFDs	31
2.5 Eliminating Fully Aliased Two-factor Interactions	31
2.6 Simulations	37

2.7	Discussion	42
3	Using Definitive Screening Designs to Identify Active First- and Second-Order Factor Effects	45
3.1	Introduction	45
3.2	Previous Work	49
3.2.1	Designs for Screening Experiments	49
3.2.2	Model Selection	52
3.3	Simulation Study Setup	55
3.4	Simulation Study Results	60
3.4.1	Results for Main-Effects-Only Models	60
3.4.2	Results for Unrestricted Models with Active Second-Order Effects	61
3.4.3	Results for Models Exhibiting Strong Heredity	69
3.4.4	Summary: Performance of DSDs Versus Sparsity Level	75
3.5	Discussion and Conclusions	79
4	Main Effects Designs for Logistic Regression that are Robust to the Presence of Two-factor Interactions	81
4.1	Introduction	81
4.2	Background and Notation	83
4.3	Robust Design via Parameter-Orthogonality	89

4.4	Constructing Robust Designs Numerically	95
4.4.1	Locally Optimal Designs	96
4.4.2	Bayesian Designs	104
4.5	Simulation Study	109
4.6	Conclusions	114
5	Bidding Strategies for Exploration and Exploitation in Two-stage Procurement Auctions with Feedback	117
5.1	Introduction	117
5.2	Bidding Problem Definition	118
5.2.1	Large-Customer Two-stage Auction	118
5.2.2	Placing a bid: the problem of cost consensus	122
5.3	Related literature	127
5.3.1	The Role of Feedback between Auction Stages	127
5.3.2	Regret Theory	129
5.4	Constructing Bidding Designs	131
5.4.1	Design notation	132
5.4.2	Design construction	135
5.4.3	Numerical example	139
5.5	Implementation strategies	142

CONTENTS

viii

References

145

List of Tables

2.1	Efficient two-level foldover designs	27
2.2	Two-level EFD for $m = 7$ and $n = 14$	28
2.3	Compound EFDs from selected cases in Table 2.1	34
2.4	Compromise design for five factors and 14 runs	38
2.5	Simulation study power and Type I error rates	40
2.6	Upper bound on fractional increase in the maximum width of the confidence intervals	43
3.1	Minimum-Run-Size DSD for m Factors	46
3.2	Comparison of Minimum Run Sizes for Central Composite Designs and Orthogonal Definitive Screening Designs	48
3.3	ANOVA Simulation Results for DSD with 10 Factors and $SN = 2$	60
4.1	D-optimal design for six factors and 12 runs	101
4.2	NLEFD for six factors and 12 runs	102

4.3	Comparison of squared bias in main effects parameters estimates resulting from the use of the D-optimal design and the NLEFD.	111
5.1	D-optimal design	140
5.2	Compound design with $\alpha = .7$	142

List of Figures

2.1	Form of correlation cell plot	18
2.2	Correlation cell plots for four alternative designs	19
2.3	Color map of correlations for a two-level EFD	30
2.4	Pareto frontier of criterion values for non-dominated designs	36
2.5	Correlation cell plots for D-optimal EFD and compromise EFD	38
3.1	Interaction and Main Effects Plots for Sensitivity Response for Unrestricted Models	65
3.2	JMP Profile Plot for 10-Factor DSD and Unrestricted True Model	66
3.3	Interaction and Main Effects Plots for Sensitivity Response for Models Following Strong Heredity	71
3.4	Interaction and Main Effects Plots for Specificity Response for Models Following Strong Heredity	72
3.5	JMP Profile Plot for 10-Factor DSD and True Model Following Strong Heredity	73

3.6	Sensitivity vs. Sparsity Ratio for Unrestricted Models	77
3.7	Sensitivity vs. Sparsity Ratio for Models Following Strong Heredity	78
4.1	Locally optimal designs	100
4.2	Bayesian optimal designs	107
4.3	Simulated mean squared error, variance and squared bias - one non-zero two-factor interaction	112
4.4	Simulated mean squared error, variance and squared bias - two non-zero two-factor interaction	113
5.1	Efficiency trace: D-efficiency and Profit-efficiency vs α	141

Chapter 1

Introduction

Defining, analyzing, and implementing solutions to real-world problems requires theory and practice in the field of experimental design to adapt to ever more complex and highly customized circumstances. This demand raises relevant and critical questions that need to be addressed by researchers and practitioners alike. To meet such demands, it is often necessary to go beyond a simple selection of existing experimental designs. The unique characteristics of every specific problem often involve considerations on several different, and frequently contrasting, criteria.

Most circumstances, needs and requirements of any experimental endeavor imply trade-offs between cost and knowledge acquisition. For example, cost considerations translate into the need to keep experimental budget to the minimum, to run and complete the experiments in the shortest time frame, and to require the minimal amount of human and material resources. However, to have effective and valuable results from the experiments, it often necessary to invest more time and resources to acquire initial understanding of the phenomenon that is being investigated, and then refine models, identify optimal conditions

and implement effective solutions. One way to satisfy different, complementing or contrasting needs, is to approach design choices with multi-objective lenses, and consequently, designing the experiments with a multi-criteria method.

Optimal designs created with compound optimization criteria are the overarching theme of the four essays that comprise this dissertation. Experimental Design is an area of research that has experienced tremendous innovation in recent years, and it has significantly broadened its scope and directions for future research. This dissertation, in its entirety, is to be viewed as a work that aims at positioning original research at the intersection of business and statistics. In particular, we aim to advance the theory of experimental design by constructing new designs and developing new construction methods, and providing practitioners and applied researchers with methods of approaching multiple objectives in real-life, complex settings.

When an investigator is interested in studying the relationship between variables, in any given real-world context, designed experiments offer invaluable tools that, if correctly selected, planned, executed, and analyzed, can significantly contribute to product/process improvements, and to tangible benefits in the business context by increasing profits and customer satisfaction, by reducing costs and achieving higher efficiency or effectiveness, depending on the context at hand.

Experimental design is a branch of statistics that has drastically evolved since its introduction, and it has seen significant paradigm shifts that have impacted researchers, practi-

tioners, and educators. In its introduction and for many decades, the established approach to designing an experiment was normative. Designs were created and cataloged, then practitioner selected the existing design that best suited their needs. A critical paradigm shift was the introduction of optimal designs (Fedorov, 1972). In recent years customization has become the dominant approach. Every problem is inherently characterized by a unique set of features and constraints, and the choice of an experimental design must necessarily reflect such unique set of characteristics. Advances in computer processing capabilities, together with analytical and numerical algorithm development, have enabled and driven this paradigm shift.

Optimal designs are usually based on an optimization criterion that best represents the needs of a specific context (Fedorov, 1972). For instance, optimality criteria are used to construct optimal designs to achieve the best performance on a given statistical property of interest. Popular criteria in current use include D-optimality, I-optimality, and A-optimality.

The idea of combining more than one optimization criterion can be accomplished in several ways (Holland-Letz, 2017). One possibility is to use a compound optimal design. Here we employ a linear combination (Jones and Nachtsheim, 2011b), or a weighted geometric average of two criteria. Alternatively a design can be constructed to optimize one criterion subject to a constraint on the efficiency on another criterion. Cook and Wong (1994) show the equivalence of the compound and constrained approaches.

Theoretical and practical considerations usually drive the selection of two optimality criteria subject to economical or technical constraints. Depending on the specific criteria that the experimenter wishes to satisfy when choosing a design, the individual optimal design in one criterion can result quite different than the optimal design in the other criterion. In such cases trade-offs between criteria must be taken into account, and relative importance of the criteria needs to be evaluated. When improvements on one criterion come at the expense of the other criterion, relative efficiencies can be used to evaluate the goodness of any given design in respect to the individual criteria. For instance, D-efficiency is typically used to compute the goodness of any given design in respect to a design that is D-optimal. Similar efficiency measures can be created to represent relative goodness of any design in respect to the design created maximizing any individual criterion. When using relative efficiency measures in a compound optimization criterion, efficiency traces (Jones and Nachtsheim, 2011b) can be valuable visual tools to explore the resulting trade-offs between criteria, and can be used to select a design that best suits the experimenter's needs. For example, selecting the compound design that minimizes the maximum efficiency loss is often a useful approach.

Accounting for more than one objective in the design construction phase can be particularly useful in contexts in which there is high degree of uncertainty about the form of the model. One of such context is that of screening experiments. Screening is frequently the focus in the initial stage of an investigation, when there are a large number of factors of

interest, and it is expected that only a few factors have a substantive impact on the response. Here we not only hope to identify these “active” factors, but we hope to characterize the nature of the relationship between these factors and the response. Traditional screening experiments are small main effects designs in which only the linear main effects of factors are considered. Such designs cannot identify the presence of higher-order terms. Standard choices of screening designs include resolution III fractional factorial design and Plackett-Burman designs. Often, followup designs involving the active factors are employed to refine model understanding. Resolution V fractional factorial designs or response surface designs are popular choices for this followup phase.

Moving beyond the traditional approaches, Jones and Nachtsheim (2011b) introduced the idea of constructing designs with the compound objective of correctly identifying the active main effects, while also minimizing the potential bias resulting from omission of higher-order terms in the model. This approach led to the discovery of the so-called Definitive Screening Designs (DSDs) (Jones and Nachtsheim, 2011a). The introduction of these designs, and this approach in general, are the building block upon which the papers/essays of this dissertation are based.

We use the approach to construct experimental designs based on compound optimization criteria for the construction and analysis of designs for linear models (Chapter 2 and 3) and non-linear models (Chapter 4). We use this approach to further develop theoretical work (Chapter 2, 3, and 4) and an applied real-world case (Chapter 5).

In our more general theoretical and methodological work (Chapter 2, 3, and 4) of this thesis, we have the objective to construct and apply designs that have the two concurrent goals of identifying initial simpler models in the first round of experimentation, while protecting from the potential bias induced by higher order models lurking in the true and more complex relationship between the variables characterizing the phenomena under study. The goal of correctly identifying linear main effects while protecting against bias of interaction terms is for example achieved by constructing designs for linear (Chapter 2 and 3) or non-linear (Chapter 4) models while keeping the size of the experiment sufficiently small.

Finally, we apply a similar multi-objective approach to the real-world business situation of two-round procurement auctions in which the primary interest of the experimentation is to identify competitors bidding behavior while maximizing overall profit and utility of a bidding strategy. In terms of design construction methods, these objectives translate into the compound objective of maximizing a statistical property of the bidding design (such as D-optimality) for the first round of bids submission, while minimizing the losses associated with second-round bids adjustments.

In summary, in this thesis we advance the state of the art in the field of experimental design theory and applications in the following ways:

- Defining and exploring approaches to construct optimal designs based on different, often contrasting, design criteria.
- Applying the above-mentioned approaches to construction of designs based on both

linear and non-linear model assumptions.

- Studying the efficacy of using designs specifically constructed to accomplish different objectives when applying analysis methods for variable selection.
- Investigating the application of a compound optimization design construction to the business context of bidding strategies in two-round auctions.

The field of experimental design is been moving and continues to move in the direction of high customization and adaptation to ever more complex business contexts. Our work of this dissertation, and the resulting future directions of our research, have the goal to embrace this trend and leverage the real-world, ever-changing needs of efficient and effective design choices.

The following four chapters of this manuscript are summarized as follows:

- Chapter 2 is the article published in *Technometrics* on 01/31/2017, co-authored with Bradley Jones, William Li, and Christopher J. Nachtsheim; it is available online at: www.tandfonline.com, doi 10.1080/00401706.2015.1124052.

Jones and Nachtsheim (2011a) introduced a new class of three-level designs called *Definitive Screening Designs* (DSDs). These designs have a number of appealing statistical properties. For example, they provide estimates of main effects that are unbiased by any second-order effects; they require only one more than twice as many runs as there are factors; and they avoid confounding of any pair of second order

effects. These authors later showed how two-level categorical factors could be added to existing DSDs (Jones and Nachtsheim, 2013). Interestingly, the authors never considered the case where all factors are categorical and have only two-levels. This paper is directed toward creating and exploring a class of two-level designs that retain many of the advantages of DSDs. We employ an algorithm for the construction of designs for varying run sizes and numbers of factors, and we characterize the statistical properties of these designs.

- Chapter 3 is the article published in the *Journal of Quality Technology* on 11/21/2017, co-authored with Bradley Jones, William Li, and Christopher J. Nachtsheim; it is available online at: www.tandfonline.com, doi 10.1080/00224065.2017.11917993.

Definitive screening designs (DSDs) were recently introduced by Jones and Nachtsheim (2011a). The use of three-level factors and the desirable aliasing structure of the DSDs make them potentially suitable for identifying main effects and second-order terms in one stage of experimentation. However, as the number of active effects approaches the number of runs, the performance of standard model-selection routines will inevitably degrade. In this paper, we characterize the ability of DSDs to correctly identify first and second-order model terms as a function of the level of sparsity, the number of factors in the design, the signal-to-noise ratio, the model type (unrestricted or following strong heredity), the model-selection technique, and the number of augmented runs. We find that minimum-run-size DSDs can be used to

identify active terms with high probability as long as the number of effects is less than or equal to about half the number of runs and the signal-to-noise ratios for the active effects are above about 2.0. We also find that if minimum-run-size designs are augmented with four or more runs, the number of model terms that can be identified with high probability increases substantially. Among the model-selection methods investigated, we found that both Lasso and the Gauss-Dantzig selector (both based on AICc) can be used to effectively identify active model terms in the presence of unrestricted models. For models following strong heredity, the SHIM method developed by Choi et al. (2010) was the best among methods tested that were designed for the strong-heredity case.

- In Chapter 4 we investigate the problem of constructing experimental designs for non-linear models in which the main effects are robust to the presence of two-factor interactions. In optimal design problems for linear models, foldover designs induce statistical independence between first-order linear effects and two-factor interactions, such that linear main effects are not biased by the presence of the latter. This happens in definitive screening designs (DSDs) (Jones and Nachtsheim, 2011a), and efficient foldover designs (EFDs) (Errore et al., 2017a). In this paper, we aim to achieve the same property in non-linear design settings. We use a compound optimization criterion to explore the trade-offs between design efficiency and parameter-orthogonality in both locally optimal and Bayesian optimal designs for logistic regression models;

and we construct two-level non-linear efficient foldover designs (NLEFDs).

- In Chapter 5, we apply and expand our theoretical and numerical methods in the context of bidding strategies in procurement auctions. Specifically, we employ experimental designs for logistic regression models for bidding strategies in two-round auction. The objectives are: estimating the relationship between pricing - in terms of margins applied - and auction outcomes; maximizing the resulting profit of the strategy applied. In two-rounds auctions, feedback given by the customer on the ranking of the first round rate submissions can be used to explore the competitors positions in any given lane, and consequently exploit the in the second round. The balance between exploration and exploitation results in the tension between decreasing prices to increase probability of winning, while not sacrificing profits unnecessarily.

Chapter 2

Benefits and Fast Construction of Efficient Two-Level Foldover Designs¹

2.1 Introduction

Screening experiments are often useful tools during the early phases of an empirical investigation, when an experimenter has little prior knowledge about the product or process under study. At this stage, subject matter experts may have identified a relatively large set of potential factors that may have causal relationships with the responses of interest. Screening experiments are traditionally small orthogonal, two-level, designs that aim to identify the much-smaller set of active factors. Two factors are orthogonal to each other if the corresponding columns in the design matrix are uncorrelated. We say a design is orthogonal if all ME column pairs are orthogonal. Follow-up experiments employing only the active factors are then used to determine the nature of interactions or other nonlinearities in

¹This work has been published in *Technometrics* (Errore et al., 2017a). To cite this paper: Anna Errore, Bradley Jones, William Li & Christopher J. Nachtsheim (2017) Benefits and Fast Construction of Efficient Two-Level Foldover Designs, *Technometrics*, 59:1, 48-57, DOI: 10.1080/00401706.2015.1124052.

a sequential process.

The primary goal of a screening experiment is to identify active MEs and, secondarily, to identify a few active two-factor interactions (2FIs) if possible. Due to their small cost, many practitioners employ orthogonal MEs plans of two types: (1) resolution III fractional factorial designs, and (2) Plackett-Burman designs. Letting n denote the number of runs in the experiment, up to $n - 1$ factors, can be employed in a design of size n .

Alternatively, various authors have advocated the use of foldover designs for screening applications based on $2m$ runs, where m is the number of factors. These designs may or may not be orthogonal for MEs, but the foldover structure guarantees that estimates of MEs are unbiased by, and statistically independent of, any active 2FIs. Webb (1968) provided more general definitions of even and odd resolution whereby any two-level foldover design with $2m$ runs is resolution IV since MEs are estimable and are unbiased by 2FIs. He also showed that the smallest resolution IV designs must contain at least $2m$ runs, and that foldover designs are always available having $2m$ runs. Margolin (1969) suggested the use of the so-called “ $2^m//2m$ ” designs, which are $2m$ -run designs obtained from fold-overs of m -run weighing designs (more details on weighing designs can be found in Hotelling (1944)). Diamond (1991) showed that Margolin designs are strongly resolvable search designs when the maximum possible number of 2FIs is one, and weakly resolvable when this maximum is two. He also explores (Diamond, 1995) the projection properties of a foldover of a 12-run Plackett-Burman design and shows that the design allows for estimation of up to

two non-zero interactions. Lin et al. (2008), hereafter LMS, discussed the usefulness of non-orthogonal foldover designs, and developed the idea of minimal dependent sets (MDS) and the related MDS-resolution and MDS-aberration criteria for comparing non-orthogonal foldover designs. LMS used these criteria to identify useful non-isomorphic designs for numbers of factors ranging from 4 to 12 and numbers of runs from 8 to 24.

Miller and Sitter (2005) explored the use of Margolin's designs in practice and argued that these designs are superior to standard orthogonal designs for screening applications. In their concluding remarks, they state that non-orthogonal foldover designs "separate the considerations of MEs and 2FIs in such a way that, with an appropriate analysis strategy one is much more likely to correctly identify MEs while losing little in terms of the ability to consider models with a few 2FIs. This suggests that disentangling MEs from 2FIs is more important than strict adherence to orthogonality of MEs in screening situations even when detecting active MEs is a primary goal."

Relatedly, Jones and Nachtsheim (2011a, 2013, 2016) have advocated the use of definitive screening designs (DSDs), which are foldover designs where the factors are either quantitative three-level factors, or they are a mix of two-level categorical and three-level quantitative factors. DSDs are an excellent choice because they provide estimates of curvature for each factor. However, there are times when all factors either must be run at two levels, or when the experimenter may prefer two levels. One example is a marketing choice study where all the factors are the presence or absence of a feature or functionality

in a product or service. Another example arises when there are many categorical factors and some of these have many levels. Here, a common procedure is to choose two levels for each factor that represent the opposite extremes of the expected effect on the response. Alternatively, an experimenter may simply prefer the use of two levels if curvature effects are expected to be small or nonexistent. This is because each factor in a DSD is set to its center value twice (not including overall center-value runs). This leads to a reduction in power for the estimation of linear MEs, and, as Jones and Nachtsheim (2011a) noted, the power for detection of curvature is small when the absolute value of the true curvature coefficient is less than about 2σ . Here, the three runs at the middle level of each factor might be viewed as wasteful. Finally, DSDs are constructed using conference matrices. These only exist for even numbers of rows and columns. Thus, increasing the number of runs in a DSD by using a larger conference matrix requires adding four runs for each incremental increase. As we will illustrate in the sequel, adding runs to efficient two-level foldover designs (EFDs) for extra power can be done by increments of two runs.

The purpose of this article is four-fold.

1. We provide additional support to the argument that non-orthogonal foldover designs can be superior to standard orthogonal alternatives in screening applications, and we advocate the use of such designs for run sizes ranging from $2m$ to $m(m + 1)/2$.
2. We give a fast algorithm for constructing EFDs for $n \geq 2m$ runs.
3. We expand on the class of small, two-level foldover designs using our algorithm.

4. We develop a compromise algorithm that allows the practitioner to choose among many designs making a tradeoff between efficiency of the main effect estimates and correlation of the 2FIs. The most efficient design for estimating the MEs can also have the undesirable property of confounded 2FIs. Using our compromise approach, practitioners can decide just how much efficiency they can sacrifice to avoid confounded 2FIs as well as lowering an omnibus measure of correlation among the 2FIs.

The remainder of the article is organized as follows. In Section 2.2, we motivate the use of two-level foldover designs with run sizes ranging from $2m$ to $m(m + 1)/2$. In Section 2.3, we give a simple construction method that leads to two-level designs whose estimated MEs are completely independent of second-order effects. In Section 2.4, we use the algorithm of Section 2.3 to characterize a class of two-level foldover designs, and in Section 2.5 we introduce a compound optimization procedure that leads to refinements for some of our designs. In Section 2.6, the results of a small simulation study are described that demonstrate the efficacy of the proposed designs for identifying active MEs and a few 2FIs. Conclusions and discussion are contained in Section 2.7.

2.2 Motivation for our approach

As noted, the primary goal of a screening experiment is the identification of active effects. A secondary goal is often the identification of a few FIs if possible. If cost is no object, we would use an orthogonal resolution V (or larger) experiment to provide independent

estimates of all MEs and 2FIs. However, run size restrictions lead to the use of resolution III, resolution IV, or other designs. The particular choice of design may depend on the degree of correlation that the experimenter will tolerate: (1) among MEs; (2) between MEs and 2FIs; and (3) among 2FIs. The trade-offs that the experimenter must contemplate can be effectively represented in a correlation cell plot.

Figure 2.1 gives the form of a correlation cell plot, emphasizing the three important areas of interest. Let $|r_{ij}|$ denote the magnitude of the correlation between the two columns in the design matrix corresponding to the i th and j th treatment effects. These magnitudes are represented by the degree of shading in the cell, with $|r| = 0$ corresponding to no shading, and $|r| = 1$ corresponding to black. The square area of cells at the top left shows the absolute correlation for a main effect with another main effect. Orthogonal two-level screening designs have $|r| = 0$ off the diagonal in this area designated the “ME/ME area.” Cells in the rectangular top right area show $|r|$ between a main effect column and a 2FI column. We call this the “ME/2FI area.” Cells in the square area at the bottom right show $|r|$ for pairs of 2FIs. We call this the “2FI/2FI area.” For a number of runs on the order of twice the number of factors, screening design trade-offs involve placing varying importance on having zero entries in these three areas.

For example, requiring that a screening design be orthogonal for MEs and 2FIs means that all the off-diagonal elements in the ME/ME, ME/2FI, and 2FI/2FI areas are zero. Such two-level screening designs have the minimum possible variance for each ME and 2FI - a

desirable characteristic. To make the discussion specific, we choose the design setting for six factors. A design having a resolution of at least V is necessary to achieve our goal that all MEs and 2FIs are orthogonal. The smallest such design for six factors has resolution VI and requires $n = 32$ runs. The correlation cell plot for this design is shown in Figure 2.2(a). As indicated by the plot, this is an excellent design, but it may be larger than necessary, especially if the goal is screening. Let SN ratio = $|\beta|/\sigma$ denote the signal-to-noise ratio, where β is a parameter in the ME + I model. For two-sided t -tests of the significance of any main effect given that all other MEs are in the model, the power to detect any effect having an SN ratio greater than or equal to 1.0 is approximately 0.9997.

Suppose now that the experimenter is willing to relax on the requirement that all 2FI columns be uncorrelated to obtain a smaller design. We note that if the number of runs is less than the number of MEs and interactions plus one (here $1 + m(m+1)/2 = 22$), the off-diagonal elements in the 2FI/2FI area cannot all be zero regardless of how desirable that would be. Regular fractional factorial designs with fewer than $m(m+1)/2$ runs have the property that $|r|$ is either 0 or 1 for every element in the 2FI/2FI area. If $|r| = 1$ in the 2FI/2FI area, then a pair of 2FIs are confounded. The correlation cell plot for the 2^{6-2} resolution IV design is shown in Figure 2.2(b). In this case, there are 18 off-diagonal cells that are black, indicating the nine pairs of 2FIs are confounded. If two effects are confounded, then there is no data-driven way to determine which effect is causing any observed change in the response of interest. Even though the sample size has been reduced

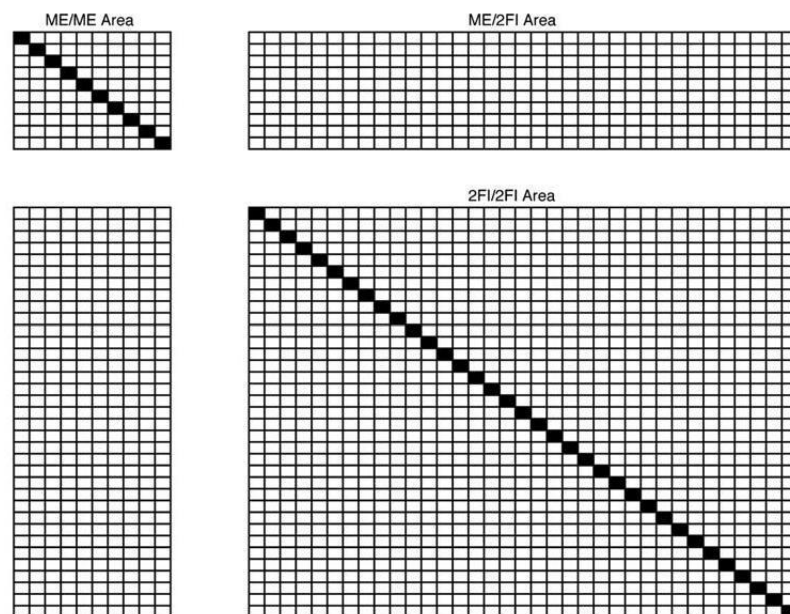


Figure 2.1: Form of correlation cell plot with areas of interest.

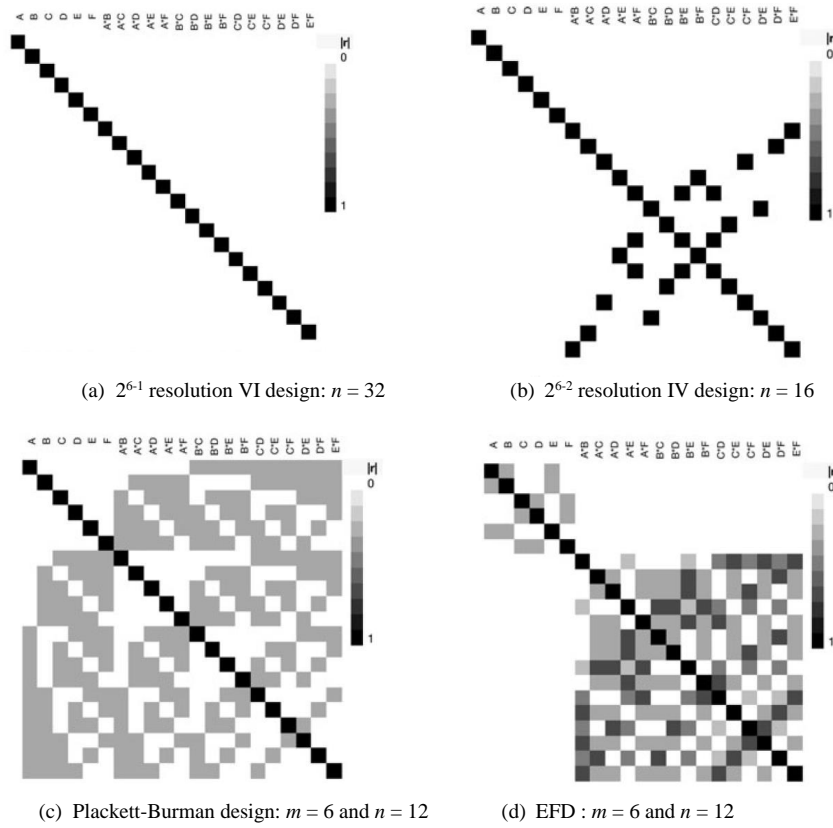


Figure 2.2: Correlation cell plots for four alternative designs for six factors.

by 50% from the resolution VI design, the power for identifying active MEs of the same magnitude as previously reported is still in excess of 0.94.

Further reduction of the sample size is still possible, while maintaining orthogonality of the MEs, by choosing a 12-run Plackett-Burman design. The correlation cell plot for this design is shown in Figure 2.2(c). All correlations in the ME/ME area are still zero, but now many of the correlations in the ME/2FI area (in addition to the 2FI/2FI area) are nonzero. The reduction in sample size has come at the expense of virtually no reduction in power; however, a new risk has arisen that the estimates of the MEs may be biased if 2FIs exist. In addition to potentially degrading the accuracy of the ME estimates, the presence of active 2FIs can increase the probabilities of Type I or Type II errors. For this design, the power for testing MEs with $SN \geq 2$ is still larger than 0.999 under the assumption that no 2FI is active.

It is possible to construct designs having all zero entries in the ME/2FI area by assuring that each row in the design has a matching row where changing the sign of every element in the first row yields the second row. Two such rows are called a foldover pair and the designs are called foldover designs. A desirable property of foldover designs is that an active 2FI will not bias the estimate of any main effect. We discuss the construction of EFDs in the next section, but for now we simply illustrate the characteristics of one six-factor EFD for $n = 12$. The correlation cell plot is shown in Figure 2.2(d). As expected, the ME/2FI area consists of zeros but we have introduced some nonzero correlations in the ME/ME area.

With $SN = 2$, the power is 0.998. Thus, we are still virtually assured of finding important MEs, and we have guaranteed that our estimates will not be contaminated by the presence of any active 2FIs.

As noted previously, Miller and Sitter (2005) advocated the use of non-orthogonal foldover designs for screening, and they developed a model selection procedure that directly exploits the orthogonality of MEs with 2FIs. In a first stage, they use an all-possible regressions procedure to identify the active set of MEs. Due to the low levels of correlation among MEs and the high power provided by these designs, this phase is highly likely to be successful in identifying important MEs. Then, in a second stage, they again use all possible regressions to identify the best model that contains the set of active MEs and any 2FIs that involve at least one factor determined to be active. Their procedure is based on the assumption of weak heredity. We think that this can be an effective procedure, with or without the assumption of weak heredity.

2.3 Constructing EFDs

In this section, we give a simple methodology for constructing EFDs. Consider the linear main effect model (ME)

$$y_i = \beta_0 + \sum_{j=1}^m \beta_j x_{ij} + \varepsilon_i \quad i = 1, \dots, n \quad (2.1)$$

where m is the number of factors, the parameters β_0, \dots, β_m are unknown constants (of which many are zero by the sparsity of effects assumption), and the $\{\varepsilon_i\}$ are iid $N(0, \sigma^2)$.

Similarly, the MEs plus bi-linear interactions model (ME + I) is

$$y_i = \beta_0 + \sum_{j=1}^m \beta_j x_{ij} + \sum_{j=1}^{m-1} \sum_{k=j+1}^m \beta_{jk} x_{ij} x_{ik} + \varepsilon_i \quad i = 1, \dots, n. \quad (2.2)$$

In matrix form, the ME model is $\mathbf{Y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$, where \mathbf{X}_1 is the $n \times (m + 1)$ model matrix for the intercept term and the linear MEs in $\boldsymbol{\beta}_1$. Similarly, the ME + I model in matrix form is $\mathbf{Y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, where \mathbf{X}_2 is the $n \times t$ model matrix for the interactions terms, $t = m(m - 1)/2$, and $\boldsymbol{\beta}_2$ is the vector of interaction effects. If we assume that the reduced model $\mathbf{Y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$ is used for ordinary least squares estimation, it is well known that $E\{\hat{\boldsymbol{\beta}}_1\} = \boldsymbol{\beta}_1 + \mathbf{A} \boldsymbol{\beta}_2$, where $\mathbf{A} = (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{X}_2$ is the alias matrix. If either $\mathbf{A} = \mathbf{0}$ or $\boldsymbol{\beta}_2 = \mathbf{0}$, then the estimate of $\boldsymbol{\beta}_1$ is unbiased.

To construct the EFDs proposed in this article, we start with a two-level design \mathbf{X} , and then impose a foldover structure such that

$$\mathbf{X}_1 = \begin{bmatrix} \mathbf{1} & \mathbf{X} \\ \mathbf{1} & -\mathbf{X} \end{bmatrix} \quad (2.3)$$

With this structure, it is straightforward to show that

$$\mathbf{A} = (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{X}_2 = \begin{bmatrix} \frac{2}{n} \mathbf{1}' \mathbf{X}_1 \\ \mathbf{0} \end{bmatrix}$$

In summary, the construction method of (2.3) guarantees that no confounding will exist between MEs and 2FIs. Some confounding between the intercept term and 2FIs will at times be present. However, when n is a multiple of 4 (i.e., m is even), designs may exist such that $\mathbf{A} = \mathbf{0}$.

In general, a design is said to be an m -factor two-level EFD if the design is chosen to maximize either $|\mathbf{X}_1' \mathbf{X}_1|$ (D criterion) or to minimize $\text{Trace} \left[(\mathbf{X}_1' \mathbf{X}_1)^{-1} \right]$ (A criterion), where \mathbf{X}_1 is constrained to follow the structure implied by (2.3). Following Jones and Nachtsheim (2011a), we use the coordinate exchange algorithm (Meyer and Nachtsheim, 1995) to search for optimal designs. The resulting designs may or may not be globally optimal for the specified criterion, but we use multiple random starts in an effort to identify a best design. One might question the inclusion of the intercept term in the chosen criterion, since interest here should focus on the estimates of the treatment effects. We note that if we write $\mathbf{X}_1 = [\mathbf{1}, \mathbf{S}]$, then it follows easily that maximization of $|\mathbf{X}_1' \mathbf{X}_1|$ is equivalent to maximization of $|\mathbf{S}' \mathbf{S}|$, because mean of each column of \mathbf{S} is zero. Similarly, minimization of $\text{Trace} \left[(\mathbf{X}_1' \mathbf{X}_1)^{-1} \right]$ is equivalent to minimization of the $\text{Trace} \left[(\mathbf{S}' \mathbf{S})^{-1} \right]$. Thus, it does not matter whether or not the intercept term is included in the D or A criteria.

In the following section, we compare our designs, where applicable, to the MDS designs constructed by LMS, who reported on the A-efficiency of their designs, and to Margolin's $2^m/2m$ designs. A lower bound on the D-efficiency of a design having design matrix \mathbf{X} , denoted $D_{\text{eff}}(\mathbf{X})$, is given by

$$D_{\text{eff}}(\mathbf{X}) = \frac{1}{n} |\mathbf{X}'\mathbf{X}|^{\frac{1}{m+1}}. \quad (2.4)$$

Similarly, the A-efficiency of a design having design matrix \mathbf{X} is given by:

$$A_{\text{eff}}(\mathbf{X}) = \frac{m}{n} \left\{ \text{Trace} \left[(\mathbf{X}'\mathbf{X})^{-1} \right] - 1/n \right\}^{-1}. \quad (2.5)$$

The above definition of A-efficiency is the same as that used by LMS. The D and A efficiencies of a design having design matrix $\mathbf{X}_{(1)}$ relative to another design with design matrix $\mathbf{X}_{(2)}$ are given, respectively, by the ratios:

$$D_R(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}) = D_{\text{eff}}(\mathbf{X}_{(1)}) / D_{\text{eff}}(\mathbf{X}_{(2)})$$

and

$$A_R(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}) = A_{\text{eff}}(\mathbf{X}_{(1)}) / A_{\text{eff}}(\mathbf{X}_{(2)}).$$

We explore the characteristics of the class of two-level EFDs in the next section. In a number of cases, the designs produced by our algorithm coincide with those constructed

previously by LMS. We note two differences with LMS. First, they employed a different criterion for evaluating designs. Using an exhaustive search, LMS evaluated each design on the basis of the size of the minimal dependent set. Roughly speaking, a minimal dependent set is the minimal number of columns in a design, such that if one column is removed from that set, the remaining set of columns has full rank. We employ D- or A-optimality of the MEs design. Second, because they used an exhaustive search, their method cannot be used for finding designs having more than 12 factors or more than 24 runs. Our algorithm scales easily. For example, our 13-factor, 30-run design was constructed in just a few seconds on a standard PC.

2.4 Results

We construct two-level EFDs for $3 \leq m \leq 13$ and for selected even values of $n \geq 2m$. The results are summarized in Table 2.1. We compare our designs with alternative designs, if existing, in the table. We also show the D-efficiency of the obtained design, the average absolute correlation between MEs, and the correlation distributions in the last three columns of the table, respectively. In the following sections, we identify some notable features of the proposed designs.

2.4.1 Orthogonal Two-Level EFDs

One attractive feature of the proposed two-level EFDs is that they can be orthogonal when n is a multiple of four. There are several such designs shown in Table 2.1, where $n = 8, 16, 20, 24$, and 32. For $n = 8$, our algorithm found the full factorial 2^3 design for $m = 3$ and the maximum resolution fractional factorial 2_{IV}^{4-1} for $m = 4$.

For $n = 16$, Sun et al. (2008) obtained and cataloged all non-isomorphic orthogonal designs for all values of m . A careful check of the catalog of Sun et al. (2008) reveals that most of the 16-run designs are of resolution III, which do not satisfy the requirement that two-level EFDs have the alias structure (2.3). For the cases of $m = 6, 7$, and 8 considered in Table 2.1, there is only one resolution IV design that is EFD-eligible. Our algorithm found the corresponding designs in all cases. It is interesting to consider the case of $m = 5$ factors, even though for this case the run size (16) exceeds our recommended upper limit of $m(m + 1)/2 = 15$. There are two regular fractional factorial designs. One has the foldover structure and is resolution IV; the other is resolution V but it does not have the foldover structure. Thus, it is not surprising that our algorithm did not produce the resolution V design. Of course, in practice, we would always prefer a resolution V design if it exists.

For $(n, m) = (32, 13)$, the EFD is the same as the minimum aberration 32-run design given in Wu and Hamada (2008).

The cases for $n = 20$ and 24 are also interesting. For $m = 9$ and $n = 20$, we did not find an orthogonal design. This should come as no surprise, as Sun, Li, and Ye (2008)

Table 2.1: Efficient two-level foldover designs

m	n	EFD found	D-efficiency	Ave abs corr	Corr distribution
3	6	Equivalent to Margolin	0.88	0.33	# of [.33] = [3]
	8	Full factorial	1	0	
4	8	2_{IV}^{4-1}	1	0	
5	10	Equivalent to Margolin	0.95	0.2	# of [.20] = [10]
	12	New	0.93	0.13	# of [0, .33] = [6, 4]
	14	New	0.95	0.14	# of [.14] = [10]
	16	2_{IV}^{5-1}	1	0	
6	12	Equivalent to Margolin	0.92	0.13	# of [0, .33] = [9, 6]
	14	New	0.92	0.14	# of [.14] = [15]
	16	2_{IV}^{6-2}	1	0	
7	14	Better than Margolin	0.89	0.18	# of [.14, .43] = [18, 3]
	14	(Margolin)	(0.77)	(0.43)	(# of [.43] = [21])
	16	2_{IV}^{7-3}	1	0	
8	16	2_{IV}^{8-4}	1	0	
9	18	New	0.94	0.12	# of [.11, .55] = [35, 1]
	20	New	0.95	0.09	# of [0, .20] = [20, 16]
	22	New	0.95	0.11	# of [.09, .27] = [33, 3]
	24	Nonregular OA	1	0	
10	20	Equivalent to Margolin	0.95	0.09	# of [0, .20] = [25, 20]
	22	New	0.94	0.11	# of [.09, .27] = [41, 4]
	24	Nonregular OA	1	0	
11	22	New	0.92	0.13	# of [.09, .27] = [47, 8]
	24	Nonregular OA	1	0	
12	24	Nonregular OA	1	0	
	24	Nonregular OA	1	0	
13	26	Equivalent to Margolin	0.98	0.08	# of [.08] = [78]
	28	New	0.96	0.07	# of [0, .14] = [42, 36]
	30	New	0.95	0.09	# of [.07, .20] = [66, 12]
	32	2_{IV}^{13-8}	1	0	

Table 2.2: Two-level EFD for $m = 7$ and $n = 14$

Run	A	B	C	D	E	F	G
1	-1	1	-1	1	-1	-1	-1
2	1	1	1	-1	1	-1	-1
3	-1	1	-1	-1	-1	-1	1
4	-1	1	1	1	-1	1	1
5	1	1	-1	-1	-1	1	-1
6	-1	-1	1	-1	-1	1	-1
7	-1	1	-1	-1	1	1	-1
8	1	-1	1	-1	1	1	1
9	-1	-1	-1	1	-1	1	1
10	1	-1	1	1	1	1	-1
11	1	-1	-1	-1	1	-1	-1
12	-1	-1	1	1	1	-1	1
13	1	1	-1	1	1	-1	1
14	1	-1	1	1	-1	-1	1

showed that all 20-run orthogonal designs have resolution less than IV. However, the 20×9 EFD obtained does not sacrifice much in the way of orthogonality. It has a D-efficiency of 0.95 and the average absolute correlation between columns of 0.09. The last column of Table 2.1 shows that, among the 36 pairs of columns, 20 have a correlation of 0, and 16 have a correlation of 0.20. We prefer the EFD in this case.

For $n = 24$, and $m = 9, 10, 11$, and 12 , the EFDs found are all fold-overs of the corresponding 12-run Plackett-Burman designs with m factors, respectively. These results are consistent with those in Miller and Sitter (2001), who advocated the use of foldovers of 12-run Plackett-Burman designs.

2.4.2 EFDs Versus Margolin's Designs and the LMS Designs

Margolin's approach requires that $n = 2m$. In all such cases considered in Table 2.1, our approach found the designs that are equivalent to or better than Margolin's designs. We call two designs equivalent if they have the same D-efficiency and correlation distributions. For $m = 3, 5, 6, 10,$ and 13 , the EFDs constructed are equivalent to those of Margolin's. For instance, when $m = 10$ and $n = 20$, the EFD has $D_{\text{eff}} = 0.95$, and among the 45 pairs of columns, 20 have zero correlation while 25 have a correlation of 0.2. Overall, the lack of orthogonality appears to be moderate.

There is one case in which we obtained better results than Margolin's approach. For $m = 7$ and $n = 14$, our design has a D-efficiency of 0.89 and an average absolute correlation of 0.18. This represents substantial improvement over Margolin's $2^7//14$ design, which has $D_{\text{eff}} = 0.77$ and an average absolute correlation of 0.43. This design is shown in Table 2.2.

Figure 2.3 provides a correlation cell plot (for correlations among MEs and 2FI columns) for the 14×7 EFD. We can compare this design with some alternative designs for studying seven factors, in which both MEs and 2FIs are of interest. A resolution III fractional factorial design requires only eight runs, but each main effect is confounded with three 2FIs. Similarly, a non-regular Plackett-Burman design requires 12 runs, but each main effect exhibits complex aliasing with 15 2FIs. The 14×7 EFD of Table 2.2, like other EFDs proposed here, performs well in terms of run size, orthogonality, and confounding structure.

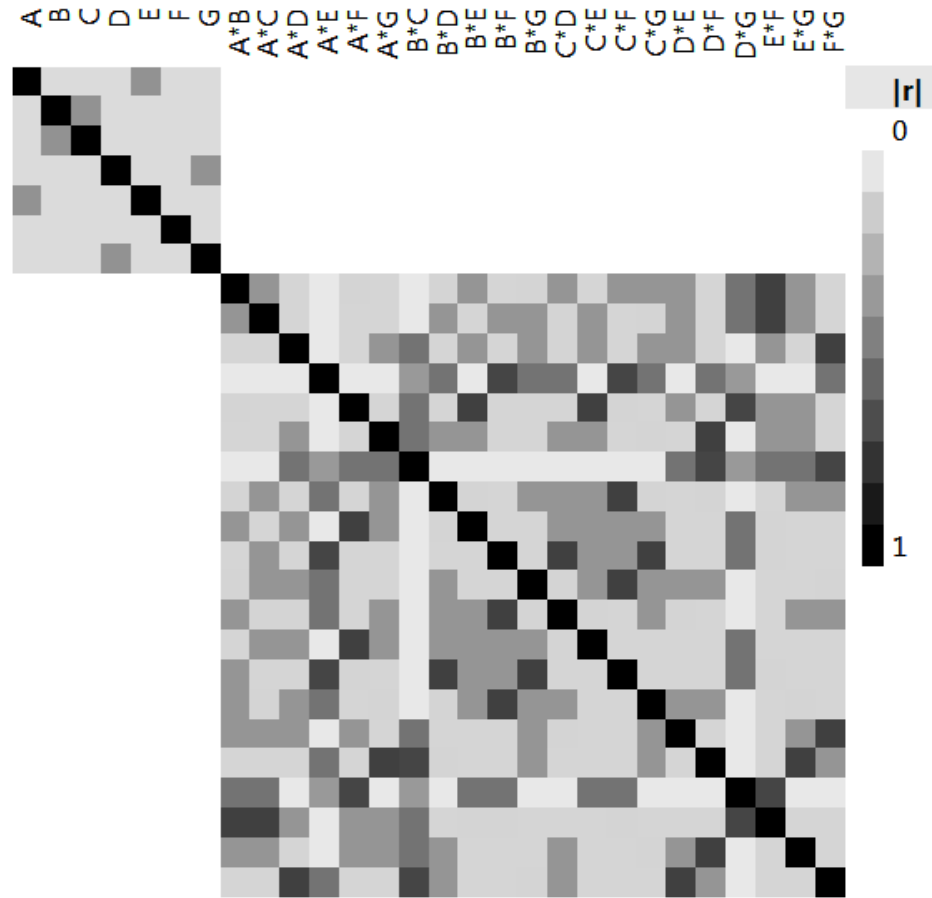


Figure 2.3: Color map of correlations for a two-level EFD with $m = 7$ and $n = 14$

2.4.3 New EFDs

Nine of the 27 EFDs summarized in Table 2.1 are identified as “New,” in the sense that there are no published, maximally D-efficient designs that satisfy the requirement of (2.3). Consider, for instance, the case for $m = 9$ and $n = 18$, for which no existing design is available. (Note that Margolin’s approach is not applicable to all cases for $n = 2m$.) The EFD obtained has $D_{\text{eff}} = 0.94$. The average absolute correlation is 0.12. Again, the loss of orthogonality is moderate. In exchange, it has completely eliminated any aliasing between MEs and 2FIs.

2.5 Eliminating Fully Aliased Two-factor Interactions

In the previous sections, we considered the optimization of $|\mathbf{X}'_1\mathbf{X}_1|$ for the MEs model subject to the constraint that the design is an EFD. As illustrated in Table 2.1, this approach results in a number of designs that perform quite well with respect to the D-efficiency of the MEs model. In addition to estimating the MEs, an aim of investigators using these designs might be to estimate a small number of non-negligible 2FIs without ambiguity. To accomplish this goal, it is necessary that no two columns of \mathbf{X}_2 be identical. Such pairs of interactions are confounded, which means that their effects cannot be separated using any data-driven methodology. One potential limitation of the class of EFDs summarized in Table 2.1 is that they include some designs in which some 2FIs are confounded with

other 2FIs. In this section, we develop a methodology for constructing EFDs that avoids confounding of pairs of second-order interactions.

To break the confounding between 2FI pairs, when it exists, we use a multiple objective (or compound) function optimization as advocated by Jones and Nachtsheim (2011b) and as implemented in the procedure described by Jones (2013). The procedure involves creating designs that maximize a weighted average of two criteria. The first criterion is approximate D-efficiency or approximate A-efficiency:

$$C_1 = \frac{|\mathbf{X}'_1 \mathbf{X}_1|^{\frac{1}{m+1}}}{n} \quad \text{or} \quad C_1 = \frac{m}{n} \left\{ \text{Trace} \left[(\mathbf{X}' \mathbf{X})^{-1} \right] - 1/n \right\}^{-1} \quad (2.6)$$

The second criterion, C_2 , seeks to find small values of the off-diagonal elements of the $t \times t$ information matrix $\mathbf{M} = \mathbf{X}'_2 \mathbf{X}_2$, where \mathbf{X}_2 is comprised of the $m(m-1)/2$ interaction columns. Let $\{c_k\}$, for $k = 1, \dots, g$ denote the set of $g = t(t-1)/2$ elements of \mathbf{M} that lie above the diagonal. The second (maximization) criterion is given by the inverse of the L_r norm for $r \geq 1$:

$$C_2 = \left(\sum_{k=1}^g |c_k|^r \right)^{\frac{-1}{r}} \quad (2.7)$$

We discuss the choice of r below. For $r \geq 1$, this second criterion penalizes designs having pairs of columns in \mathbf{X}_2 with large covariances. The overall objective function is to maximize, for a specified weight w , ($0 \leq w \leq 1$), a weighted average of the above two

criteria:

$$C_w = wC_1^s + (1 - w)C_2^s$$

where C_1^s and C_2^s are scaled values of C_1 and C_2 . To find the scaling, following Jones (2013), we first maximize C_1 . Let C_1^{\max} be the resulting D-efficiency. We then optimize C_2 yielding C_2^{\max} . Subsequently, for any design, the scaled criteria are given by

$$C_i^s = \frac{C_i}{C_i^{\max}}$$

for $i = 1, 2$. To produce our compromise designs, we repeatedly choose w at random from $(0, 1)$ and find the design that optimizes C_w . We sort the designs by efficiency and then choose the first design in the resulting list that has no confounded 2FIs.

For criterion (2.7), we initially chose $r = 1$. However, with this choice we did not consistently eliminate confounding among all 2FI pairs. We observed that minimizing the average absolute covariance was not sufficient, and that minimization of the maximum absolute covariance was of greater relevance to the goal of eliminating confounded pairs. Since the minimax criterion results for $r \rightarrow \infty$, we simply experimented with larger values of r , finding that the choice of $r = 4$ was sufficient for the designs explored in this article.

Of the EFDs produced in Table 2.1, 11 had 2FI pairs fully aliased. We applied the compound optimization algorithm in each of these cases. We used the A criterion in (2.6) rather than the D criterion to facilitate comparisons with the MDS designs of LMS. LMS

Table 2.3: Compound EFDs from selected cases in Table 2.1

m	n	EFD		LMS		Relative efficiency of EFD
		Average $ r $	Max $ r $	Average $ r $	Max $ r $	
5	14	0.25	0.75	0.25	0.75	100.00
6	14	0.28	0.75	0.28	0.75	100.00
6	16	0.19	0.58	0.24	0.77	100.00
7	16	0.22	0.58	0.26	0.77	100.00
8	16	0.24	0.58	0.24	0.58	100.00
9	18	0.25	0.80	0.25	0.80	100.00
9	20	0.21	0.82	0.23	0.82	120.74
9	22	0.21	0.83	0.23	0.69	101.42
10	20	0.22	0.82	0.24	0.82	129.62
10	22	0.22	0.83	0.23	0.69	104.25
11	22	0.23	0.78	0.24	0.83	117.35
12	24	0.23	0.33	0.23	0.51	107.33

produced 10 alternative designs for each combination of sample size and number of factors and gave the A-efficiency for each design. Thus, for comparing our designs, we use the relative A-efficiency of our design with the LMS design having the largest A-efficiency. The results are summarized in Table 2.3. The EFD designs in Table 2.3 were obtained by sorting the designs from our compromise algorithm in descending order of A-efficiency. Then we chose the first design in the list having no pairs of confounded 2FIs. This is a simple approach and removes the potential confusion of the novice when confronted with possibly hundreds of alternatives. Note that in every case, the average value of $|r|$ for the EFD design is less than or equal to that for the LMS design. Also, the values reported for the maximum $|r|$ for the EFD designs often compare favorably with those of the LMS

designs.

We illustrate the use of our compound optimization procedure by applying to the case with $m = 9$ factors and $n = 22$ runs. To do so, we provide a plot of the two criteria for designs on the Pareto frontier of non-dominated designs in Figure 2.4. In the context of compound optimization of designs, one design dominates another if it is better for one criterion and at least as good for another. When constructing many designs, it is unnecessary to store a dominated design. At the end of the optimization procedure, the set of designs that remain form what is called a Pareto front. The point plotted at the upper left corresponds to the D-optimal EFD while the point in the lower right corresponds to the design having the maximum value of the secondary criterion C_2 . Note that of 10,000 designs generated with the previously described random weights, w , we found only four designs on the Pareto front.

The design we prefer corresponds to the point at the top right. Its D-efficiency is 0.995 while its efficiency with respect to the secondary criterion is 0.99. For this design, all the factor columns have a pairwise correlation of $1/11$ th with other columns. The maximum correlation among pairs of 2FI columns is $7/15(0.47)$. There are 126 such pairs. There are 252 pairs of 2FI columns having correlations of $4/15(0.27)$ and a final 252 pairs of 2FI columns having correlations of $1/10$. By contrast, the D-optimal EFD confounds three pairs of 2FIs. It also has 18 pairs of interactions with correlations greater than one-half. On the other hand, its MEs have standard errors that are about 5% smaller. In addition, the average

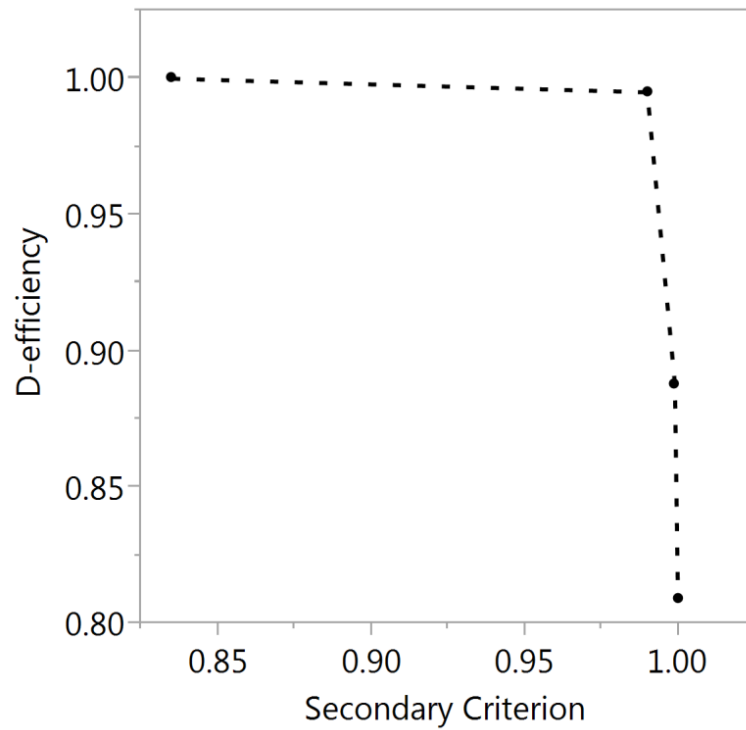


Figure 2.4: Pareto frontier of criterion values for non-dominated designs for $m = 9$ and $n = 22$

correlation of 2FI pairs is also about 5% smaller. Correlation cell plots for the D-optimal EFD and the compromise EFD are shown in Figure 2.5(a) and Figure 2.5(b), respectively. The dark cells (i.e., high absolute correlations) in Figure 2.5(a) have been eliminated in Figure 2.5(b) with very little change to the magnitudes of the absolute correlations for the main effect columns.

2.6 Simulations

In this section we apply a slight modification of the analytical approach of Miller and Sitter (2005) in two cases to show the capability of EFDs to identify all the important MEs in addition to several 2FIs.

The first simulation study makes use of the compromise design with five factors and 14 runs shown in Table 2.4. Nine of the 10 pairwise absolute correlations among the MEs are $1/7$. The last is $3/7$. For the most D-efficient EFD, all 10 pairwise absolute correlations are $1/7$. However, this design has three pairs of 2FIs that are fully confounded. By contrast, the design of Table 2.4 has no pairwise absolute correlation among the 2FI that is larger than $3/4$. The fact that this design does not confound 2FIs means that it has the potential to allow practitioners to reliably identify a few 2FIs. Our simulation study provides the circumstances under which this potential becomes reality.

We fielded our simulation study as a full factorial design to investigate the ability of this design to identify active MEs and active 2FIs. The factors employed in the study were:

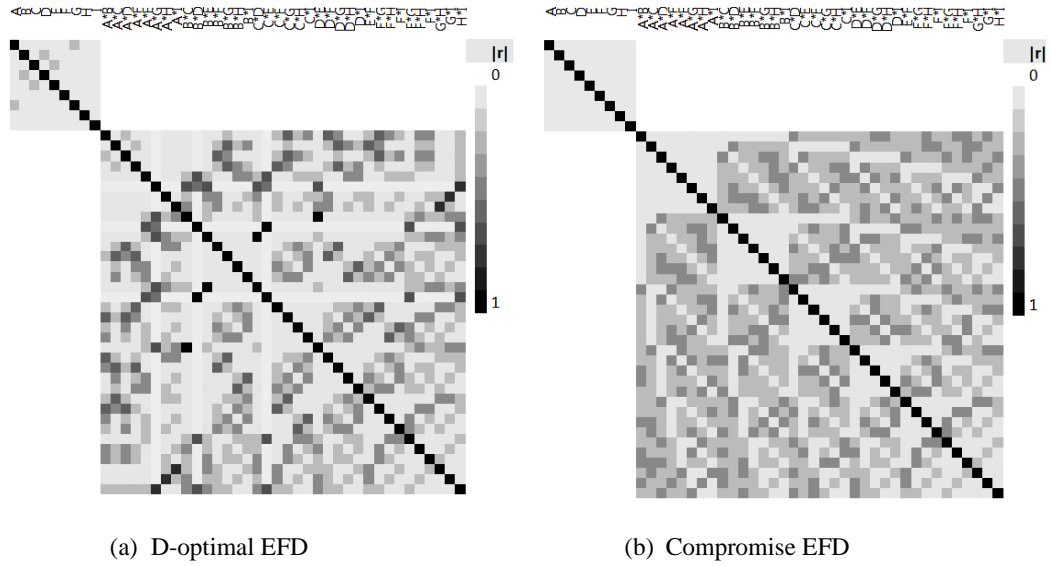


Figure 2.5: Correlation cell plots for D-optimal EFD and compromise EFD for $m = 9$ and $n = 22$

Table 2.4: Compromise design for five factors and 14 runs

1	1	1	-1	-1
1	1	1	-1	1
1	-1	-1	-1	1
1	-1	1	1	1
-1	1	1	-1	1
-1	1	-1	-1	1
-1	-1	1	-1	-1
-1	-1	-1	1	1
-1	-1	-1	1	-1
-1	1	1	1	-1
-1	1	-1	-1	-1
1	-1	-1	1	-1
1	-1	1	1	-1
1	1	-1	1	1

1. The number of active MEs, having levels 0, 1, 2, 3, 4, and 5.
2. The number of active 2FIs, having levels 0, 1, 2, 3, and 4.
3. The SN ratio of active effect, having levels 2.0 and 3.0.

We simulated 1000 response vectors for each of the $6 \times 5 \times 2 = 60$ design cases. To generate a response vector, we chose the required number of active factors randomly from the five factors. We chose the required number of active 2FIs randomly from the 10 possible 2FIs. For an SN ratio of 2.0, we generated the coefficient for each active effect by adding 2.0 to an exponentially distributed random number. We chose the sign of each coefficient randomly. We added a standard normal random number to the expected value of the response for each factor setting.

We analyzed each simulated experiment using a modification of the method suggested by Miller and Sitter (2005). Since their design employed $2m$ runs where the number of factors was m , the number of degrees of freedom available for fitting MEs was m . As a result, their first stage-strategy used all possible subsets to identify the active MEs. Of course, with such a strategy, they could not identify all the MEs if all the MEs were active because in that case there would be no degrees of freedom for error. For our design with five factors and 14 runs, there are seven degrees of freedom in the subspace of the MEs model and seven degrees of freedom in the subspace of the intercept and 2FIs ignoring third- and higher-order effects. That means that there are two degrees of freedom in the MEs subspace available for estimating the error variance. We then use the error variance

Table 2.5: Simulation study power and Type I error rates for five active main effects and one to four active two-factor interactions

SNR	Active ME	Active 2FIs	Power ME	Power 2FIs	Type I 2FIs
2	5	1	0.99	0.97	0.049
2	5	2	0.98	0.90	0.073
2	5	3	0.99	0.87	0.096
2	5	4	0.98	0.81	0.102

estimate to perform t -tests on each of the five main effect estimates. For the 2FIs, we do all subsets up to models having four 2FIs but we stop early if the residuals for a model with fewer terms do not have a significantly larger variance than the variance estimate obtained from the analysis of the MEs. While Miller and Sitter (2005) limited their consideration to the set of 2FIs involving MEs found in the first stage, we did not restrict the choice of 2FIs in our subsets regression.

We calculate the power for a group of terms (e.g., MEs or 2FIs) as the number of correctly identified active effects in that group of terms divided by the total number of active effects. Since the MEs analysis is independent of the analysis of 2FIs, the power for identifying MEs is relatively constant and greater than 0.98 on average. As expected, the power for identifying active 2FIs drops as the number of active 2FIs increases. However, even for four active 2FIs, the power exceeded 0.8. Type I error rates for the null case MEs were at the nominal rate of $\alpha = 0.05$. For 2FIs, the Type I error rates increased as the number of active 2FIs increased to a maximum of about 0.1.

The worst case for identification of MEs was for simulations having SN = 2 and all five active MEs. Table 2.5 shows the results of this worst case for active 2FIs ranging from one

to four. We see that by adding four extra runs beyond the minimum $2m = 10$ advocated by Miller and Sitter (2005), the compromise design has considerable power for reliable identification of any main effect and up to four 2FIs.

We also performed simulations using a compromise design having 13 factors and 30 runs. Note that this design also has an extra four runs above the minimum $2m = 26$. For this study, we considered SN ratios of 1.0 and 2.0. We allowed the number of active MEs to range from 10 to 13 and the number of active 2FIs to range from zero to three. Thus, the simulation design was based on $2 \times 4 \times 4 = 32$ treatment combinations, each replicated 1000 times. Since there are 91 possible 2FIs and the amount of computation can become prohibitive, we limited the subsets regression to stop with three-term models.

The power for detecting MEs averaged 0.94 for SN ratios of 1.0 and 0.9996 for SN ratios of 2.0. The power for identifying 2FIs decreased with the number of active MEs. When the SN ratio was 1.0, the minimum power for 2FIs was about 0.7. For the case having all MEs active, it also identified the correct 3 of 91 2FIs with power equal to 0.96. By contrast, the 13-factor regular fractional factorial design requires two more runs (32) and is resolution IV. Each 2FI is confounded with at least four other 2FIs, so unique identification of any active 2FI is impossible.

2.7 Discussion

For screening designs, there are four particularly desirable features: orthogonality of the MEs, orthogonality of MEs and 2FIs, orthogonality of 2FIs with each other, and a small run size. It is often the case that these characteristics cannot be satisfied simultaneously, and trade-offs are necessary. For example, Plackett-Burman designs have small run sizes and are orthogonal, but they exhibit complex aliasing of MEs by 2FIs. Resolution III fractional factorial designs are also orthogonal for MEs, but some MEs and 2FIs are confounded. Resolution IV fractional factorial designs are another orthogonal alternative; however, 2FIs are confounded with other 2FIs.

Two-level EFDs represent a different kind of tradeoff. These are small designs that sacrifice orthogonality of MEs for complete independence between MEs and 2FIs. This approach was introduced by Webb (1968), Margolin (1969), and others, and has been strongly advocated by Miller and Sitter (2005), who studied the robustness properties of Margolin's designs. Two-level EFDs improve upon and extend the class of Margolin-type designs. In addition, using the methodology of Section 2.5, they can be constructed in such a way that 2FIs are not completely confounded with other 2FIs. This provides the experimenter with the opportunity to identify not only key MEs, but also active 2FIs in the presence of sparsity.

Non-orthogonal MEs suffer two related consequences: the parameter estimates have

Table 2.6: Upper bound on fractional increase in the maximum width of the confidence intervals for main effects compared to the width of a confidence interval for an ideal orthogonal design

m	n	Upper Bound On Fractional Increase in Maximum Standard Error for Main Effects
5	10	0.05
	12	0.09
	14	0.08
6	12	0.10
	14	0.15
7	14	0.14
9	18	0.21
	20	0.05
	22	0.07
10	20	0.05
	22	0.07
11	22	0.10
13	26	0.02
	28	0.04
	30	0.06

wider confidence intervals than an orthogonal design based on the same number of runs, and the power to detect an active main effect is reduced, again in comparison to an orthogonal design of the same size. Consider the cases of non-orthogonal EFDs in Table 2.6. This table shows the fractional increase in the maximum width of the confidence intervals for the MEs compared to the those for an orthogonal design-assuming one exists. In most of these cases, no such design exists.

We recommend EFDs to practitioners who are limited to two-level factors, where a few 2FIs are likely to be active, and who have the resources to field a design having $2m$ to $m(m + 1)/2$ runs. These designs are virtually guaranteed to identify the important MEs in an unbiased fashion, and the separation of MEs and 2FIs into separate, orthogonal spaces improves the likelihood of identifying the active 2FIs when an analysis procedure such as that recommended by Miller and Sitter (2005), or the modified version advocated here in Section 2.6, is employed. Use of designs based on the compromise procedure can further enhance the prospects for correctly identifying a few active interactions. We also prefer $2(m + 2)$ runs to the minimal run size if possible. The extra runs tend to lessen the degree of correlation between pairs of 2FIs. Further research into analysis and model selection for these designs taking account of their structure is ongoing.

Chapter 3

Using Definitive Screening Designs to Identify Active First- and Second-Order Factor Effects¹

3.1 Introduction

Response Surface Designs (RSDs) are used for optimizing or fine tuning an experimental system when a few key, active factors (generally between two and six) have been identified. Traditionally, practitioners employ central composite designs (CCDs) and Box-Behnken designs (BBDs) in response surface experiments because they allow the experimenter to estimate all quadratic effects and two-factor interactions efficiently when it is assumed that third-order effects are small or negligible. For m factors, there are $(m + 1)(m + 2)/2$ model effects that need to be estimated, namely the intercept term, the m linear main effects, the m quadratic main effects, and the $m(m - 1)/2$ two-factor interactions. As a result, the required

¹This work has been published in Journal of Quality Technology (Errore et al., 2017b). To cite this paper: Anna Errore, Bradley Jones, William Li & Christopher J. Nachtshiem (2017) Using Definitive Screening Designs to Identify Active First- and Second-Order Factor Effects, Journal of Quality Technology, 49:3, 244-264. DOI: 10.1080/00224065.2017.11917993

Table 3.1: Minimum-Run-Size DSD for m Factors

Foldover Pair	Run (i)	Factor Levels				
		$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	\cdots	$x_{i,m}$
1	1	0	± 1	± 1	\cdots	± 1
	2	0	∓ 1	∓ 1	\cdots	∓ 1
2	3	± 1	0	± 1	\cdots	± 1
	4	∓ 1	0	∓ 1	\cdots	∓ 1
3	5	± 1	± 1	0	\cdots	± 1
	6	∓ 1	∓ 1	0	\cdots	∓ 1
\vdots	\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
m	$2m - 1$	± 1	± 1	± 1	\cdots	0
	$2m$	∓ 1	∓ 1	∓ 1	\cdots	0
Centerpoint	$2m + 1$	0	0	0	\cdots	0

run sizes for CCDs and BBDs grow quickly with the number of factors and, for this reason, most practitioners do not attempt to implement RSDs for more than five or six factors. Our experience with model selection for RSDs indicates that it is rarely necessary to use more than a small fraction of the terms in the full model to adequately approximate the true surface. If the investigator suspects that the number of active second-order effects is small, then using a design with a smaller number of runs would be attractively economical. The class of definitive screening designs (DSDs), introduced by Jones and Nachtsheim (2011a), represents one such option. The purpose of this research is to evaluate the effectiveness of DSDs in identifying the active and inactive first- and second-order effects, when used as alternatives to standard response surface designs.

DSDs provide a new approach for screening experiments that allows the experimenter potentially to identify all the active first- and second-order effects in one stage of experi-

mentation. A typical DSD is comprised of m pairs of foldover rows in which one factor is kept at its center value and the others are at the extremes. A final row has all factors set to their center values. For m factors, these designs require (a minimum of) $n = 2m + 1$ runs and have the structure described in Table 3.1, where $x_{i,j}$ denotes the setting of the j th factor for the i th run. Properties of DSDs include the following:

1. The set of all linear and quadratic main effects is estimable and uncorrelated.
2. Main effects are statistically independent of two-factor interactions.
3. Two-factor interactions are not confounded with any other two-factor interactions, although they are correlated.
4. DSDs having six or more factors project to efficient response surface designs in any three factors. DSDs having 18 or more factors similarly project to efficient response surface designs in any four factors.

Jones and Nachtsheim (2011a) used numerical methods to construct DSDs and found that these designs were orthogonal for 4, 6, 8, and 10 factors. Jones and Nachtsheim (2011a) noted that if m is odd and an orthogonal DSD exists for $m + 1$ factors, an orthogonal design for m factors can be obtain by dropping any column from the orthogonal design for $m + 1$ factors at a cost of two extra runs. They also noted that multiple “fake” factors could be used to increase the sample size and the resultant power for tests of factor effects. Xiao et al. (2012) showed how to construct DSDs by vertical concatenation of an $m \times m$

Table 3.2: Comparison of Minimum Run Sizes for Central Composite Designs and Orthogonal Definitive Screening Designs for Four Through Eight Factors

Design Type	Number of factors				
	4	5	6	7	8
CCD	25	27	45	79	81
DSD	9	13	13	17	17

conference matrix \mathbf{C} , $-\mathbf{C}$, and a center-value run. Jones and Nachtsheim (2013) showed how to augment a DSD with two-level categorical factors. In this study, we will assume that all factors are quantitative, and so can be run at three levels.

DSDs can be thought of as supersaturated designs for second-order models. In the usual definition, a supersaturated design is one in which the number of factors exceeds the number of rows. This concept usually applies to two-level designs in which the number of factor effects, m , is greater than n . The same idea applies when the parameters of interest are the $p = (m + 1)(m + 2)/2$ terms from a full second-order response surface model. It is known that two-level supersaturated designs can be used to identify active factors if the number of such factors is sufficiently small. Rules of thumb generally require that the number of active factors be no greater than roughly $n/4$ to $n/2$ (cf., Draguljić et al. (2014)). The savings that can result from the use of a DSD in place of a standard response surface design can be immense. Table 3.2 provides a comparison of run sizes for minimal-run CCDs and DSDs for four through eight factors.

Our purpose here is to characterize the ability of DSDs to correctly identify the active

(and inactive) response surface model effects as a function of number of factors, the level of sparsity, the model selection method, the types of active second-order effects, the type of model (unrestricted or following strong heredity), the number of augmented runs, and the number of factors in the design. We do so using the simulation study described in Section 3.3.

The remainder of the paper is organized as follows. In Section 3.2, we provide an overview of the related literature on the construction and analysis of saturated and super-saturated designs for screening experiments. In Section 3.3, we describe our study design. In Section 3.4, we summarize the results of our simulation study. In Section 3.5, we conclude with a discussion and suggestions for practitioners.

3.2 Previous Work

3.2.1 Designs for Screening Experiments

Screening is often the first step in an investigation when there is a large number of potential causal factors in a system but only a few are expected to influence the response of interest. Traditionally, screening experiments have been based on small main-effects orthogonal designs, such as resolution III regular fractional factorial designs or nonregular designs such as Plackett-Burman designs. These designs are effective choices when all second-order effects are small in relation to main effects. However, if substantive two-factor interactions

do exist, the ability of the analyst to identify them will be limited due to the aliasing structure of the designs. With resolution III fractional factorial designs, some main effects are directly confounded with two-factor interactions. With Plackett-Burman designs, the complex aliasing between main effects and two-factor interactions leads to biased estimates of main effects in the presence of active two-factor interactions and limits the ability of the analyst to identify such effects (Miller and Sitter, 2005; Jones and Nachtsheim, 2011b; Errore et al., 2017a). Moreover, these traditional designs employ only two levels for each factor and therefore do not permit the estimation of quadratic effects. For these reasons, follow-up experiments are frequently required to obtain a complete analysis.

The use of supersaturated designs (SSDs) in screening applications has been increasing in recent years. An SSD can be characterized as any design having fewer runs than model terms of interest. Thus, for instance, a design can be unsaturated when only main effects are of interest but supersaturated when interactions are also to be investigated (Draguljić et al., 2014). In the case of DSDs, the designs are saturated for the main linear and quadratic effects but supersaturated for the estimation of two-factor interactions. A useful property of any DSD is that linear main effects are orthogonal to two-factor interactions and quadratic main effects. This property enhances our ability to estimate main effects unbiasedly and to identify active second-order effects in the presence of effects sparsity. Moreover, if three or fewer main effects are active, DSDs with 13 runs or more project to highly efficient RSDs, and model selection becomes much easier because the full quadratic model is estimable.

Of course, successful use of SSDs depends critically on the level of sparsity. Sparsity, also called the Pareto principle in experimental design, refers to the observation that the number of relatively important effects in a factorial experiment is generally small (Box and Meyer, 1986). In addition to its level, sparsity can assume different qualitative forms. The principle of hierarchy states that the lower-order effects are more likely to be active than the higher-order effects. Moreover, a model possesses the strong heredity property if a two-factor interaction is present in the model only if both main effects of the factors involved in the interaction are also present. A model follows the weak heredity property if a two-factor interaction appears only if at least one of the main effects is present (Wu and Hamada, 2008). The assumption of strong heredity can significantly impact sequential approaches to experimentation. For example, if in a screening experiment involving factors A through F, factors A, B, and C are found to have statistically significant main effects, while the main effects of factors D, E, and F are not statistically significant, subsequent investigations can focus on factors A, B, and C without any concerns about missing interactions such as AD or EF. Heredity can also provide advantages in analyzing data from experiments with complex aliasing patterns, enabling experimenters to identify likely interactions without resorting to high-resolution designs (Chipman et al., 1997).

Li et al. (2006) analyzed a very large collection of full factorial designs and found that the general principles of sparsity, heredity, and hierarchy tend to hold in practice. With respect to heredity, if both main effects involved in an interaction effect are active, Li

et al. (2006) found that the probability that the interaction is active is 0.33; if one main effect is active, the probability that the factor is involved in a two-factor interaction is 0.045; and finally, if neither of two main effects is present, the probability that the two factors are involved in an active two-factor interaction is 0.0048. Moreover, they found that the probability of a main effect to be active is 0.40. Chipman et al. (1997) considered probabilities of 0.25 for strong heredity effects, 0.10 for weak heredity effects, 0.01 for no heredity effects. In the simulation study to follow, we consider models that possess the strong heredity property and models that possess neither weak nor strong heredity. We refer to the latter class of models as “unrestricted” models. We now turn to the analysis of DSDs and related designs.

3.2.2 Model Selection

Jones and Nachtsheim (2011a, 2016) suggest the use of stepwise procedures for model selection when using DSDs. However, a comprehensive investigation of the best methods for analysis of these designs has not yet been conducted. The literature does, however, address the analysis of saturated and supersaturated designs. Most previous studies concern two-level designs, but they explore variable selection methods that are easily adapted to the analysis of DSDs.

The literature on variable selection methods proposed for the analysis of supersaturated designs is diverse. Wu (1993) and Lin (1993) suggested the use of forward selection

methods to identify active main effects. Westfall et al. (1998) utilized a modified forward model selection. Lu and Wu (2004) suggested a stepwise selection based on staged dimensionality reduction. Chipman et al. (1997) proposed a Bayesian variable selection method, while Beattie et al. (2002) used a two-stage Bayesian approach for model selection. Li and Lin (2002) introduced a variable selection method via nonconvex penalized least squares that employed an iterative ridge regression. Holcomb et al. (2003) proposed contrast-based methods. Zhang et al. (2007) proposed a method based on partial least squares. Phoa et al. (2009) used a simulation study to examine the application of the Gauss-Dantzig selector described by Candes and Tao (2007). Marley and Woods (2010) performed a simulation study that compared two classes of supersaturated designs and three methods for variable selection: forward stepwise regression (Miller, 2002), Gauss-Dantzig selector (Candes and Tao, 2007; Phoa et al., 2009), and their proposed model averaging procedure.

For screening designs, Draguljić et al. (2014) presented a comprehensive comparison of screening strategies. They conducted a simulation study that compared the use of supersaturated designs and group screening, in conjunction with several variable selection methods: Lasso, smoothly clipped absolute deviation (SCAD), Gauss-Dantzig selector, simulated annealing model search (SAMS), Bayesian model selection, and Maximum A Posteriori (MAP) estimation.

Variable selection methods that directly address principles of sparsity, heredity, and hierarchy were also considered in the literature. Yuan et al. (2007) proposed a modification

of least-angle regression (LARS) (Efron et al., 2004) to account for the empirical principles. Choi et al. (2010) proposed a method called strong heredity interaction model (SHIM) that automatically enforces the strong heredity constraints in the penalty function. Later, Li and Zhu (2014) proposed a similar method, called the weak heredity interaction model (WHIM), to restrict the search for models that follow weak heredity.

Our purpose here is not to identify a “best” method for the analysis of DSDs, but rather to evaluate the ability of DSDs to identify the active model terms as a function of sparsity when using proven methods for analysis. For this reason, we initially considered six methods for model selection with unrestricted models. Four of these methods employed forward stepwise with the following stopping criteria: Bayesian information criterion (BIC), the Akaike information criterion (Akaike, 1992) corrected for small samples (AICc), and the p -value threshold for $p = 0.05$ and $p = 0.20$. Also included were the Lasso method and the Gauss-Dantzig selector, with best subset determined by AICc. Early experience with the stepwise methods led to the elimination of the BIC criterion and p -value threshold criterion for $p = 0.2$ because both methods led to significant overfitting. This left four methods for evaluation in the simulation study for unrestricted models. For the strong heredity case, we considered SHIM and two modified versions of forward stepwise.

3.3 Simulation Study Setup

We state once more that the main purpose of the simulation study is to determine the number of active main and second-order effects that the analysis of data from a DSD of a given size can reliably identify, as a function of various characteristics of the design and the level and type of sparsity.

We constructed the simulation study as a designed experiment with all possible combinations of the factors below:

1. Number of factors, m , in the DSD. This design factor assumed three levels: 8, 10 and 12.
2. Number of fake factors employed by the DSD. The numbers of fake factors used was zero, two, and four. This resulted in three different run sizes for each number of factors, namely, the minimum-run-size DSD, the minimum-run-size DSD plus four runs, and the minimum-run-size DSD plus eight runs, respectively.
3. Number of active main effects, m_a . The number of active main effects varied in steps of size two: (i) from 2 to m when only main effects are active, (ii) from 2 to $m - 2$ for unrestricted models having second-order effects, and (iii) from 4 to $m - 2$ for models following strong heredity having second-order effects.
4. Proportion of active second-order effects, denoted p_{so} . This design factor assumed

three levels: 0 (first-order models), 0.5, and 1.0. So, for example, if the number of active effects is $m_a = 4$, and $p_{so} = 0.5$, the number of active second-order effects is given by the product, $m_a \times p_{so} = 2$. These numbers span a range of cases we might see in practice. Generally, the number of main effects is greater than the number of second-order effects, so that $p_{so} \leq 1$.

5. Signal to noise ratio (SN) of the active effects. We considered three levels of signal-to-noise ratio for active effects, where SN is defined as $|\beta|/\sigma$: $SN \geq 1$, $SN \geq 2$, and $SN \geq 3$.
6. Mix of second-order terms. The mix of second-order terms assumes three qualitative levels: (1) two-factor interactions only, where the $p_{so} \times m_a$ second-order terms are drawn at random from the population of two-factor interactions, (2) a random mix, where the $p_{so} \times m_a$ second-order terms are chosen at random from the population of all two-factor interaction terms and all pure quadratic terms, and (3) pure quadratics only, where the $p_{so} \times m_a$ second-order terms are chosen at random from the population of all quadratic terms.
7. Model type. The model type is either unrestricted or follows strong heredity.
8. Model selection method. There are either three or four variable selection methods, depending on model type. For models employing unrestricted choice of second-order effects, we used the following four model-selection methods:

- (a) Forward stepwise with a p -value threshold of 0.05, denoted SW/ p -value
- (b) Forward stepwise based on minimum AICc, denoted SW/AICc
- (c) Lasso based on minimum AICc, denoted Lasso/AICc
- (d) Gauss-Dantzig selector based on minimum AICc, denoted Dantzig/AICc

For models following strong heredity, we used the following three model-selection methods:

- (a) Forward stepwise with a p -value threshold of 0.05, denoted SW/ p -value
- (b) Forward stepwise based on minimum AICc, denoted SW/AICc
- (c) SHIM, also denoted SHIM

For models following strong heredity, stepwise forward selection is conducted in two stages. In the first stage, stepwise selection of main effects is performed. The subgroup of second-order effects that follow strong heredity is considered in a second stage. Similarly, SHIM is a generalization of Lasso that applies a penalized least-squares criterion to the selection of main effects and two-factor interactions, enforcing strong heredity. We modified the algorithm to extend the same mechanism to the heredity between main effects and quadratic terms. The modification was accomplished in a simple fashion by treating any quadratic term as an interaction between a factor and itself.

Like Lasso, SHIM is a variable selection method based on parameter shrinkage, and it requires that two tuning parameters, denoted λ_1 and λ_2 , be chosen. These two tuning

parameters are used, respectively, in the selection of active main effects and second-order terms. We tune the performance of the method by executing it over a grid (λ_1, λ_2) pairs. Specifically we use a 10-by-10 grid where each λ can take value 10^p , where p is evenly spaced from 1 to 4. These 100 grid points translate into 100 applications of SHIM on every iteration of our simulation loop. For every (λ_1, λ_2) pair, we save the model terms selected and resulting AICc value. The final model chosen is the one exhibiting minimum AICc among the 100 fits.

For clarity, we describe the setup for one run of the study. The model is selected as follows. Suppose the number of factors in the DSD is eight and there are four active main effects. First, the four active main effects are selected at random from the eight factors. Let the proportion of second-order effects be $p_{so} = 0.5$, so that the number of second-order effects is two. Let the type of second-order effects be a mix of interaction and quadratic terms, so that, if the model type is unrestricted, the two active second-order effects are chosen at random from among the 28 possible interactions and eight possible quadratic terms. If the model type for the run calls for a model exhibiting strong heredity, then the two active effects are chosen at random from among the six allowable two-factor interactions and four allowable quadratic terms.

To generate a response vector, given the form of the model as just described, we proceed as follows. If the SN of active effects is two, we generated the coefficient for each active effect by adding 2.0 to an exponentially distributed random number. We also randomly

generated the sign of the coefficient so that, on average, half the coefficients are positive and half are negative. The expected value of the response vector was computed next, and an n -vector of $N(0, 1)$ random variates was then added to the expected response vector to obtain the observed response vector. For each variable selection method, we counted the number of active terms selected and the number of inactive terms not selected and computed three performance measures:

1. Sensitivity: The ratio of the number of selected active effects over the total number of active effects
2. Specificity: The ratio of the number of unselected inactive effects over the total number of inactive effects
3. Proportion correct: The ratio of the sum of the number of selected active effects and the number of unselected inactive effects over the total number of parameters

The definitions of sensitivity and specificity match those of Choi et al. (2010). For every analysis method, we performed variable selection for 500 randomly generated model/response-vector pairs.

We also recorded the sparsity ratio, defined as the ratio of the total number of active effects to the number of runs in the design. This measure is used to summarize our results in the next section.

Table 3.3: ANOVA Simulation Results for DSD with 10 Factors and SN = 2

Order of model	Sum of squares	R^2	ANOVA Model df
Main effects only	4849.26	0.514	11
Up to 2-factor interactions	6418.25	0.680	58
Up to 3-factor interactions	6576.90	0.697	155
Up to 4-factor interactions	6661.37	0.706	251
Up to 5-factor interactions	6672.40	0.707	287

3.4 Simulation Study Results

In this section, we summarize our results by first briefly looking at the identifiability of main-effects-only models, then turning to separate discussions of second-order models by model type.

3.4.1 Results for Main-Effects-Only Models

When second-order terms are not present, the true model is identified with probability near one for any of the methods except SW/ p -value, which breaks down as m_a approaches m . The results for methods other than SW/ p -value should not be surprising given the orthogonality of main effects and the power analyses reported by Jones and Nachtsheim (2011a) for identifying active main effects. Most of the following remarks pertain to identifying active effects when second-order terms are present in the true model.

3.4.2 Results for Unrestricted Models with Active Second-Order Effects

For each fixed number of factors and model type, the study design is a full factorial design for the number of active factors, the proportion of active second-order effects ($p_{so} = 0.5$ or $p_{so} = 1.0$), the analysis method, the mixture of second-order terms, and the number of augmented runs. (We note that the “zero” level for the proportion of active second-order effects was covered separately in the previous subsection.) For example, for $m = 10$, the study design is a $4 \times 2 \times 4 \times 3 \times 3$ full factorial design with 500 replicates. Thus, analysis of variance models were fit to the sensitivity and specificity responses for each number of factors. We found for both responses and for each number of factors, many high-order terms, including five-factor interaction terms, were statistically significant. For this reason, a simple and precise explanation of the results in terms of main effects and low-order interactions is not possible. However, for the $m = 10$ factor case, Table 3.3 shows the proportion of variation in the sensitivity response (R^2) as the order of the model increases. From Table 3.3, we observe that the proportion of variation in the sensitivity response explained by the main effects model is 0.514. Adding all two-factor interaction terms increased R^2 by 0.166 to 0.680. After this, the addition of higher order interaction terms had little effect on R^2 . Going to the full factorial model, which requires an additional 229 degrees of freedom, increased R^2 by only 0.027. Thus the second-order model gave a good

approximation to the actual model, and an examination of the two-factor interaction plots can provide useful (though approximate) insight. Similar conclusions were also reached in the case for 8- and 12-factor DSDs and for both the specificity and sensitivity responses.

Sensitivity Response

Figure 3.1 contains plots of all two-factor interactions, with main effects plots provided on the diagonal, for the sensitivity response in the 10-factor case. We note that, even though nearly all two-factor interactions are statistically significant, many are not of practical significance; for most of the plots, the lines are approximately parallel. We enumerate below key findings for main effects and interactions.

- Examining the main effect plots along the diagonal, it is clear that the two factors that had the greatest impact are the number of active main effects and the proportion of active second-order effects - in other words, the degree of sparsity of the true model. As the number of active main effects increases and/or the number of active second-order effects increases, the sensitivity drops. The worst sensitivity for DSDs with any number of factors occurred, as expected, when the number of active main effects was largest and the number of active second-order effects was equal to the number of active main effects.
- The next two most important factors are the model-selection method and the number of augmented runs. Overall, Dantzig/AICc and Lasso/AICc led to the highest

sensitivities with Dantzig/AICc having a slight edge. The two stepwise approaches exhibited considerably less sensitivity. The ability of the model-selection routines to identify active effects clearly increases with the number of augmented runs. Overall, sensitivities increase from 0.68 to 0.78 when four runs are added and from 0.68 to 0.83 when eight runs are added. This result suggests that experimenters should routinely consider the addition of at least four runs if their objectives include identifying second-order effects.

- Overall, the factor having the least effect is the mix of second-order effects. The plot shows that quadratic effects are slightly more difficult to detect than interaction effects, although there is very little difference between the sensitivities for all-interaction models and models comprised of both interactions and quadratic terms.

In terms of interactions, we call attention to four cases in Figure 3.1.

- As shown in the third panel of the first row and the third panel of the first column, SW/ p -value and Stepwise/AICc, whose sensitivities overall are about the same, are affected differently by the number of active factors. The performance of SW/ p -value is as good as any of the methods for two active factors, but its performance degrades rapidly as the number of active factors increases. Conversely, SW/AICc performs relatively badly for two active factors, but it is best (by a very slight margin) for eight factors.

- As shown in the fourth panel of the third row and/or the third panel of the fourth row, SW/AICc seemed to fail, relative to the other methods, to identify quadratic effects in pure quadratic models.
- As shown in the fifth panel of the third row and/or the third panel of the fifth row, SW/AICc is relatively unaffected by augmenting additional runs, a very troubling result for that method.
- Finally, as shown in the fifth panel of the first row and/or the first panel of the fifth row, the decline in sensitivities as the number of active main effects increases (i.e., as sparsity decreases) is substantially less when the design is augmented by four or eight runs.

The discussion above does not mention the effect of signal-to-noise ratio, the number of factors in the design, or the effect of assuming strong heredity on observed sensitivities.

In general we note the following:

- Conditional on having the same number of active main effects, DSDs with more factors (and therefore more runs) had higher sensitivity, as expected.
- Increasing the signal-to-noise ratio from 1.0 to 2.0 had a very strong and positive impact on sensitivity as expected. However, increasing the signal-to-noise ratio beyond 2.0 had a very small effect.

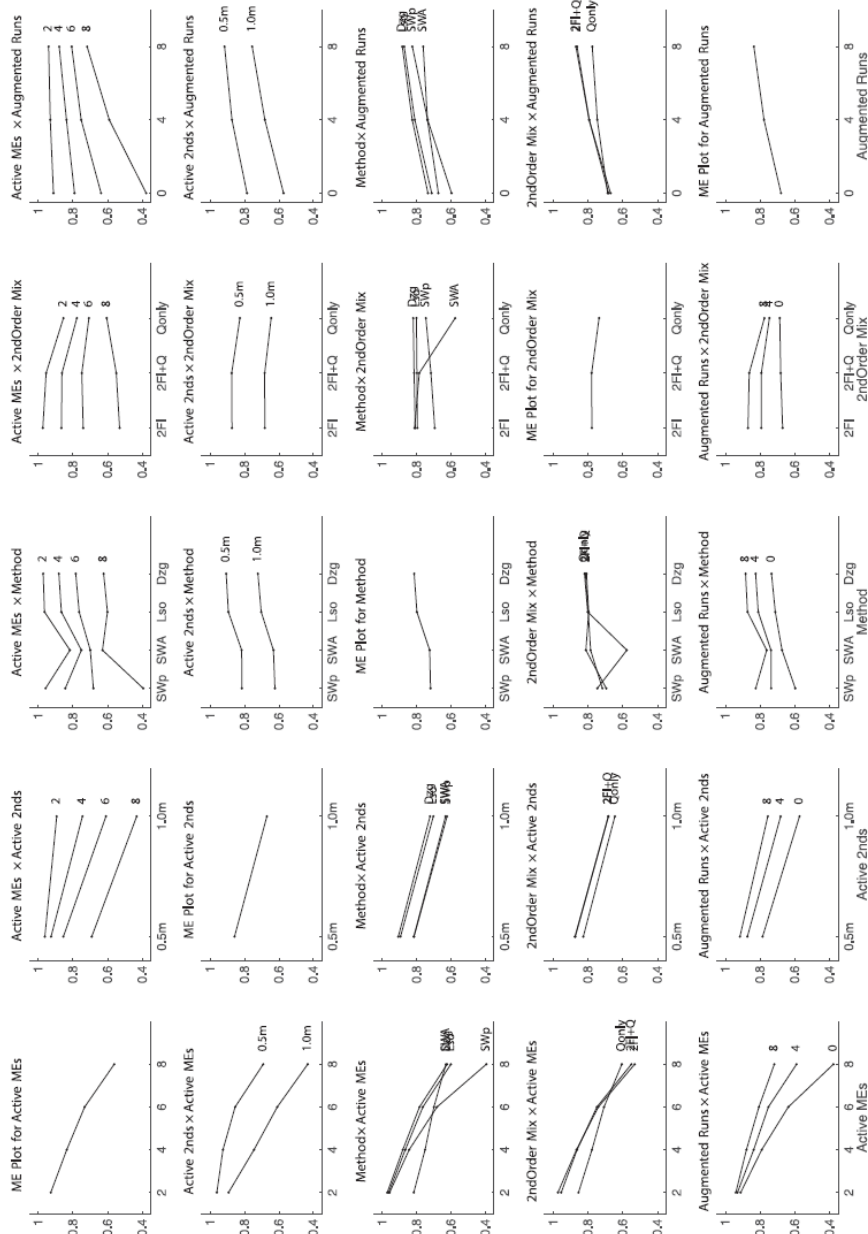


Figure 3.1: Interaction and Main Effects Plots for Sensitivity Response for Unrestricted Models with $m = 10$ Factors.

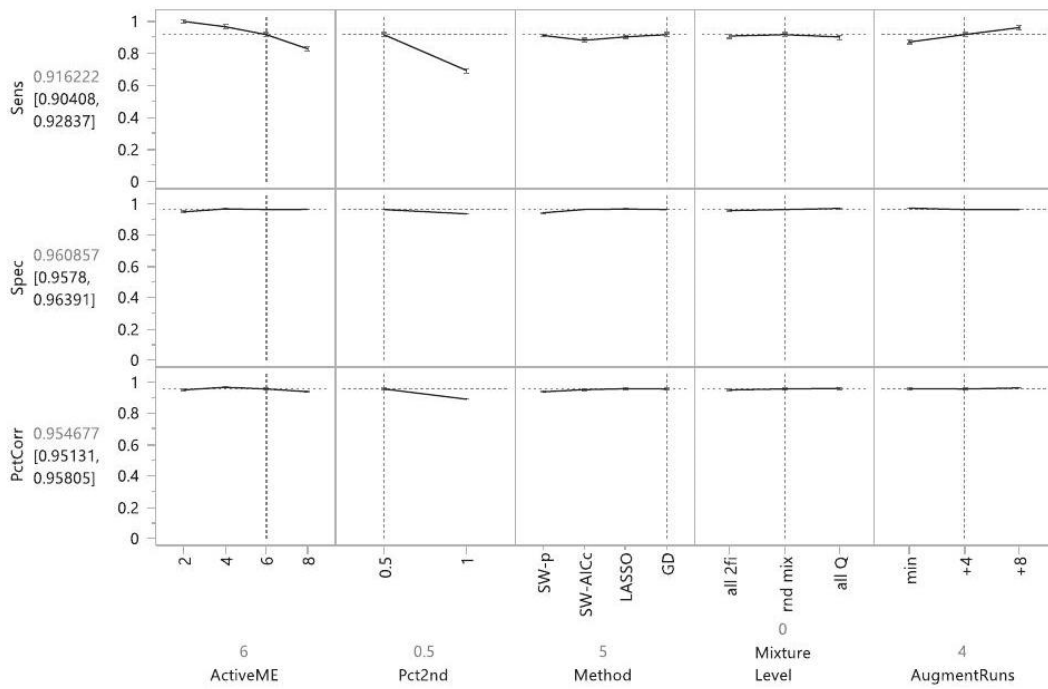


Figure 3.2: JMP Profile Plot for 10-Factor DSD and Unrestricted True Model. The factor settings (vertical dotted lines) correspond to six active main effects, proportion of active second-order effects equal to 0.50, Gauss-Dantzig method, a random mix of second-order-effect types, and four augmented runs.

- For unrestricted models, Lasso/AICc and Gauss-Dantzig/AICc perform the best, both in terms of sensitivity and as we show below, for specificity.

Specificity Response

Specificities for all four model selection methods under all of the various design conditions were uniformly high. Mean specificities for SW/ p -value, SW/AICc, Lasso/AICc, and Dantzig/AICc were 0.937, 0.942, 0.958, and 0.953, respectively. Also, the specificity response was relatively unaffected by any of the design factors from a practical point of view and, for this reason, we have omitted an interaction/main effects plot in the form of Figure 3.1. As was the case with the sensitivity response, Lasso/AICc and Dantzig/AICc are slightly better than the two stepwise procedures in terms of specificity. Also, overall, there is a slight decline (about 0.01) in specificity if the design is augmented by four runs and about a 0.02 decline if the design is augmented by eight runs. Because the factor effects on specificity are so slight relative to their effects on sensitivity, the factor effects on the proportion correct response are largely determined by the sensitivity response.

Performance of DSDs When Analyzed Using Dantzig/AICc

The interaction plots in Figure 3.1 show averages for combinations of two factors over all other experimental factors including model selection methods. To give a sense of how well the designs perform when Dantzig/AICc is used, Figure 3.2 provides JMP prediction profiler plots of the study factors for cases corresponding to the 10-factor DSD with unre-

stricted models for the three responses of interest. In a prediction profile plot, a conditional effects plot is displayed for each of the experimental factors. For a given factor, the predicted response is plotted as a function of the value of the factor, conditional on all of the levels chosen (as indicated by the vertical dashed lines and printed below the horizontal axis) for each of the other factors. The predicted value for each response is shown in the panel to the left of the plots. For example, we observe from Figure 3.2 that, when (i) the number of active main effects is six, (ii) the proportion of active second-order effects is 0.50 (so that the total number of active second-order effects is three), (iii) the variable selection method is the Gauss-Dantzig selector, (iv) the second-order effects are a mix of interactions and square terms, and (v) there are four augmented runs, the mean sensitivity, specificity, and proportion correct are 0.916, 0.961, and 0.955, respectively. Here the number of active terms in the true model is 9, so that the sparsity ratio is $9/25 = 0.36$. Sensitivity falls below 0.80 when the proportion of second-order terms is moved to $p_{so} = 1.0$ so that the number of second-order terms is six. We note that the sparsity ratio in this case is $12/25 < 0.5$. Notice that neither the model-selection method nor the mixture of second-order term types has much effect on specificity or the proportion correct, although the marginal superiority of the Gauss-Dantzig selector is apparent.

3.4.3 Results for Models Exhibiting Strong Heredity

The path we took for the analysis of the simulation results for models following strong heredity mirrors that taken for unrestricted models. We consider the sensitivity and specificity responses in turn. We note that results presented are likely to be a bit better than one would observe in practice. This is because we only consider as possibilities response models that follow strong heredity. In practice this is not likely to hold all of the time (Li et al., 2006).

Sensitivity response

As was the case for the unrestricted models, the ANOVA for the sensitivity response showed that five-factor (and all lower level) interactions were statistically significant ($p < 0.0001$). But as was previously the case, the model comprised of main effects and all two-factor interactions accounted for most of the variation explained by the full model. In this case, the two-factor interaction model accounted for 0.67 of the variation in specificity. The five-factor interaction model accounted for an additional 0.06. For this reason, we again use the main-effects and two-factor interaction plots to facilitate interpretation. Figure 3.3 provides these plots for the sensitivity response. Key findings follow.

- Examining the main effect plots along the diagonal, we see that the results for number of main effects, the proportion of active second-order effects, the second-order mix, and the number of augmented runs are qualitatively the same as we found for the

sensitivity response for unrestricted models.

- The main effects plot for the methods reveals that SHIM is significantly better than the two stepwise procedures. Overall, the sensitivities of SHIM, SW/ p -value, and SW/AICc are 0.829, 0.736, and 0.753, respectively.

In terms of interactions, we call attention to two cases.

- As shown in the third panel of the first row and the third panel of the first column, SHIM is essentially unaffected by the number of active main effects, whereas SW/AICc and SW/ p -value are negatively impacted by increases to the number of active effects, especially the latter.
- As shown in the fourth panel of the third row and/or the third panel of the fourth row, SW/ p -value improves with the level of pure quadratic effects in the model, whereas the opposite occurs for SHIM and SW/AICc. SHIM's performance in particular is hurt by the presence of pure quadratics. SHIM sensitivity drops from 0.96, for all-interaction models, to 0.66 for pure quadratic models.

Specificity Response

Unlike for the case of unrestricted models, the model-selection routines used for the case of strong heredity do not exhibit similar specificities. The main-effects/interactions plot for the specificity response is shown in Figure 3.4. Whereas SHIM is the clear winner in

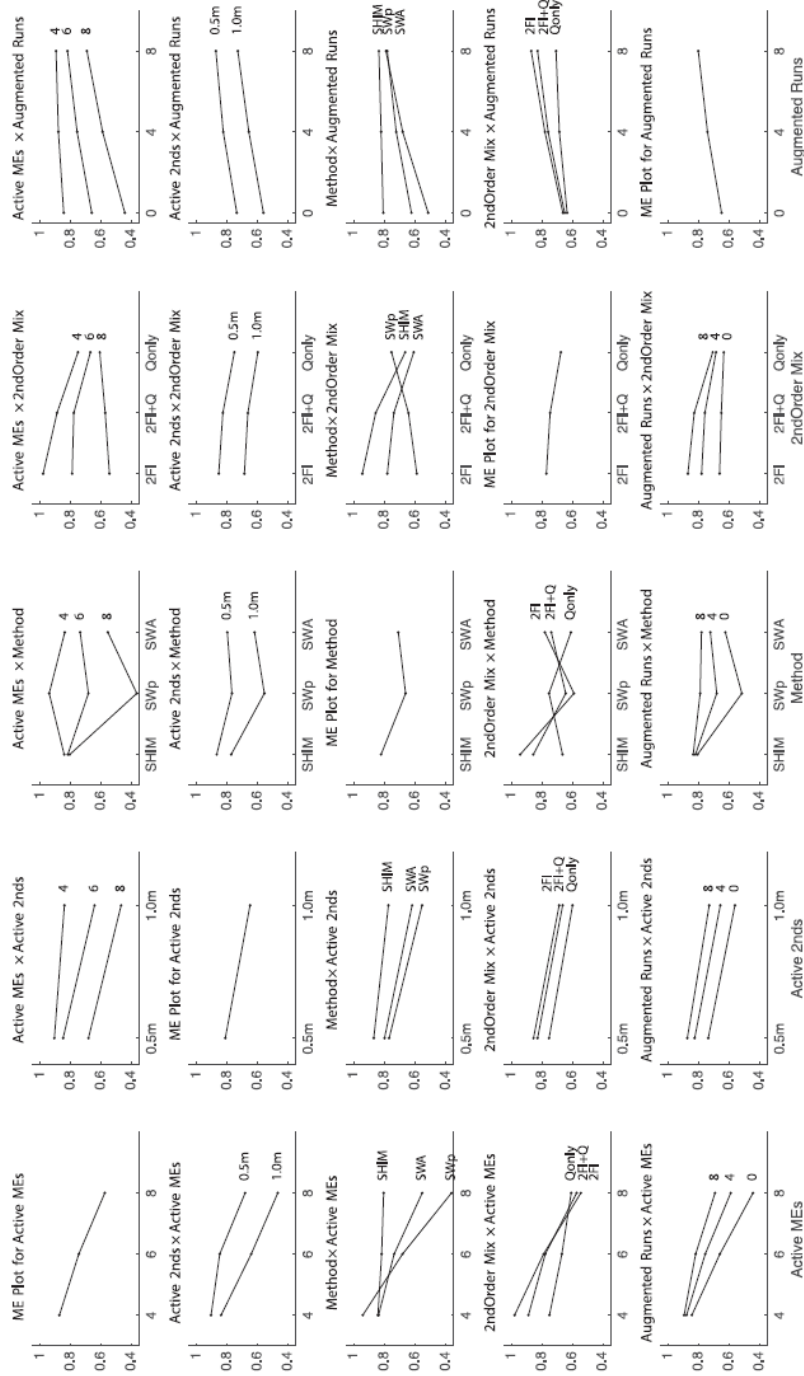


Figure 3.3: Interaction and Main Effects Plots for Sensitivity Response for Models Following Strong Heredity with $m = 10$ Factors.

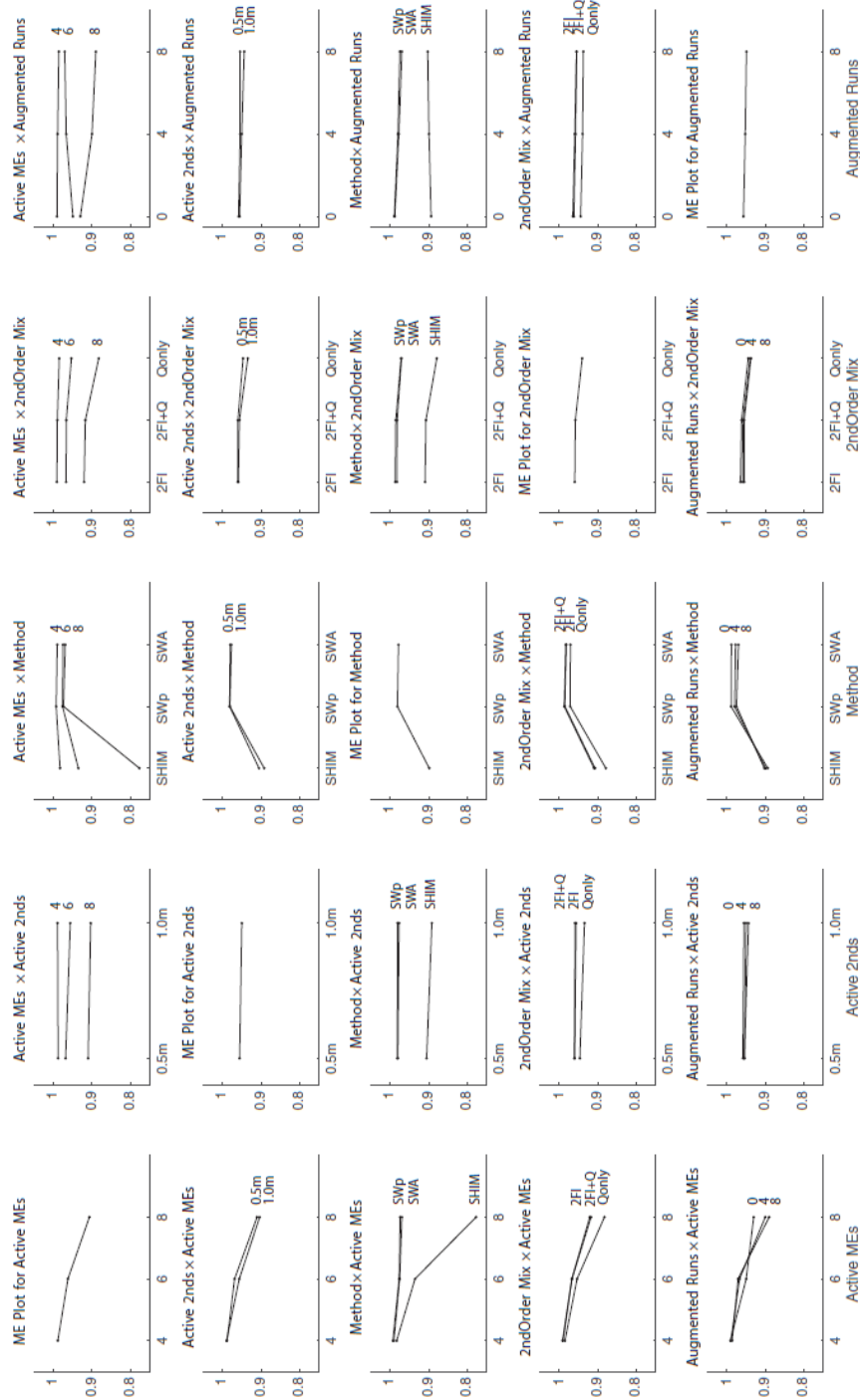


Figure 3.4: Interaction and Main Effects Plots for Specificity Response for Models Following Strong Heredity with $m = 10$ Factors.

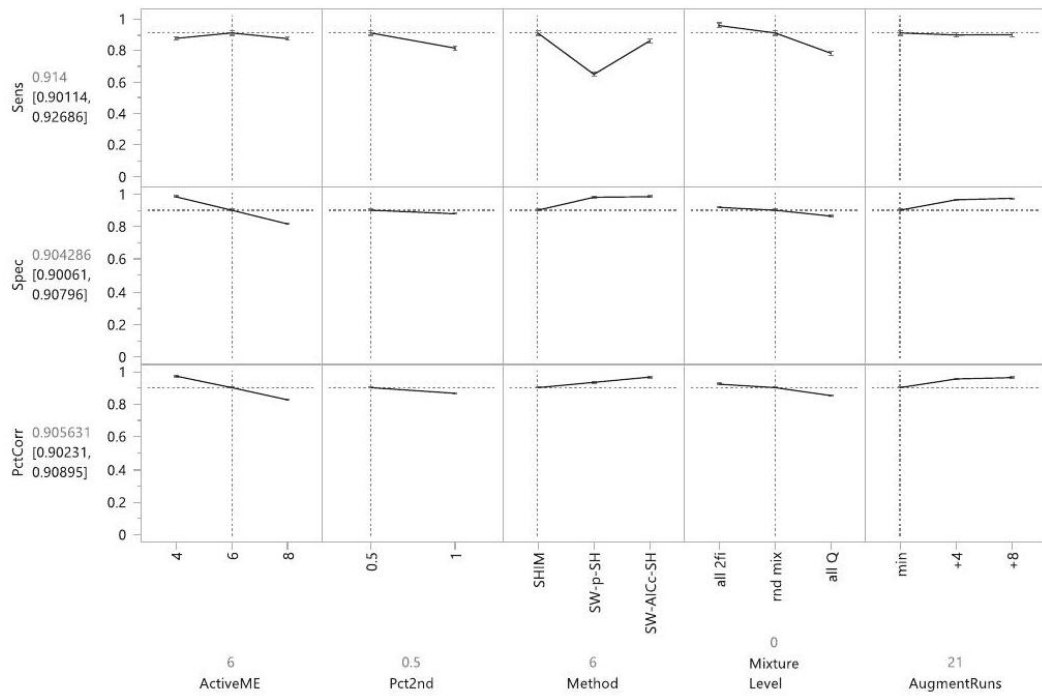


Figure 3.5: JMP Profile Plot for 10-Factor DSD and True Model Following Strong Heredity. The factor settings (vertical dotted lines) correspond to six active main effects, proportion of second-order effects equal to 0.5, SHIM method, a random mix of second-order-effect types, and no augmented runs.

terms of sensitivity, the main-effects plot for model-selection methods in Figure 3.4 (panel in row 3, column 3) reveals that the opposite is true for specificity. As shown there, the average specificities for SHIM, SW/ p -value, and SW/AICc are 0.894, 0.989, and 0.984, respectively. SHIM's specificity problem gets worse as the number of active main effects increases. As shown in the panel in row 3 and column 1 of Figure 3.4, SHIM's specificities for $m_a = 4$ or $m_a = 6$ are competitive with those for SW/ p -value and SW/AICc. A steep drop in specificity occurs for $m_a = 8$. These results show that there is a trade-off among the model-selection strategies designed for models following strong heredity. SHIM is clearly the best in terms of sensitivity, whereas SW/ p -value, and SW/AICc are better for specificity.

Performance of DSDs When Using SHIM for Model Selection

As noted above, we believe that, in practice, good performance on sensitivity is more important than good performance on specificity. For this reason, we further consider the use of SHIM if a practitioner wishes to consider only models following strong heredity. JMP profile plots for models exhibiting strong-heredity are shown in Figure 3.5. There, the settings correspond to six active main effects, the proportion of second-order effects is 0.50, the method is SHIM, there is a random mix of second-order-effect types, and there are no augmented runs. For this case, the mean sensitivity is 0.914. As shown previously in Figure 3.2, for a similar case with unrestricted models (and the Gauss-Dantzig selector)

the sensitivity was 0.916. This demonstrates that, for models exhibiting strong heredity, sensitivity remains comparable. However, the profile plot for mixture level reveals the drop in sensitivity for pure quadratic models.

The results described in the preceding paragraphs for 10 factors are representative of what we observed for 8 and 12 factors.

3.4.4 Summary: Performance of DSDs Versus Sparsity Level

Although we recorded sensitivity, specificity, and proportion correct as response variables, our major focus in this simulation study was on sensitivity, as it corresponds to the power of the designs for identifying active effects. We believe that missing active effects in a screening experiment is generally more undesirable than including a few extra, inactive effects. More comprehensive summaries of our results are given by the various plots in Figure 3.6 and Figure 3.7, which pertain respectively to unrestricted model results when employing the Gauss-Dantzig selector and strong heredity model results when employing SHIM. In each figure, panel rows one to three pertain to no augmented runs, four augmented runs, and eight augmented runs, respectively. Panel columns one to three pertain to models comprised only of main effects, models comprised of main effects and second-order terms with $p_{so} = 0.5$, and models comprised of main effects and second-order terms with $p_{so} = 1.0$. A logistic regression smooth is superimposed in each plot. The plot symbols give the number of active main effects with one exception: “0” is the symbol used for $m_a = 10$. For a visual

reference, we have added a horizontal sensitivity threshold of 0.70, following Choi et al. (2010), and a vertical sparsity threshold of 0.50.

Consider first the case of unrestricted models in Figure 3.6. Clearly, for models comprised only of main effects, the three panels in the first column demonstrate that the likelihood of identifying main effects is always near 1.0, irrespective of the sparsity ratio and the level of augmentation. The center column of panels shows that, when the number of second-order effects is half the number of main effects and the sparsity ratio is less than 0.50, sensitivity is consistently above about 0.80. Panel (b) shows the results for the minimum run size DSD. For this case, it is critical that the sparsity ratio be less than 0.50. In contrast, when the DSD is augmented by four runs or eight runs (see panels (e) and (h)), sensitivities never fall below about 0.75 and 0.80, irrespective of sparsity level. It is essentially for this reason that we recommend using at least four augmented runs when the objective is to identify second-order effects. Finally, the panels in column three summarize results for the most difficult case, when the number of second-order effects is equal to the number of active main effects. Here sensitivities stay above the 0.70 threshold as long as sparsity remains above about 0.45. The plot symbols reveal that the problems arise when the number of active main effects are 6, 8, or 10.

Consider now the case where strong heredity is assumed and the selection method is SHIM, summarized in Figure 3.7. With the exception of the all-main-effects case in panel column 1, results reveal a decline in sensitivities relative to the unrestricted case in Figure 6

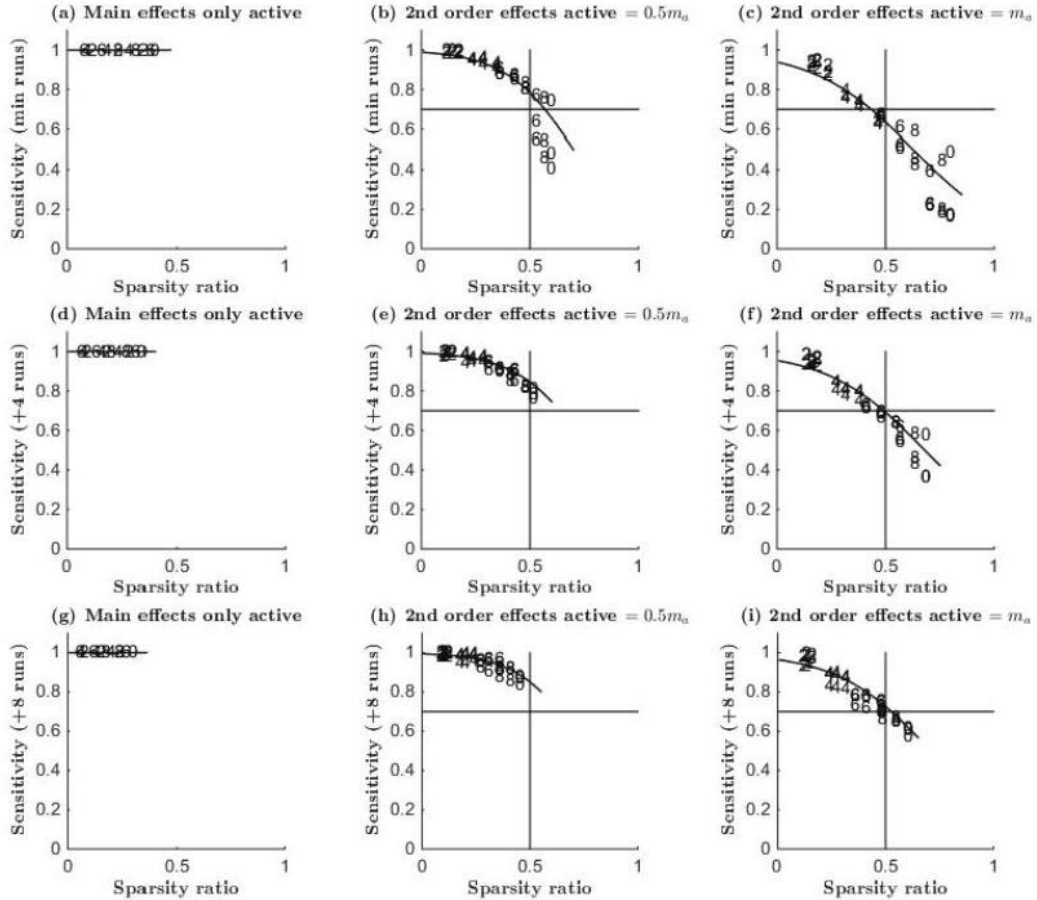


Figure 3.6: Sensitivity vs. Sparsity Ratio for Unrestricted Models. Rows one to three pertain to zero augmented runs, 4 augmented runs, and 8 augmented runs, respectively. Columns one to three pertain to models comprised only of main effects; models comprised of main effects and second-order terms, where the number of second-order terms is half the number of main effects; and models comprised of main effects and second-order terms, where the number of second-order terms is equal to the number of main effects.

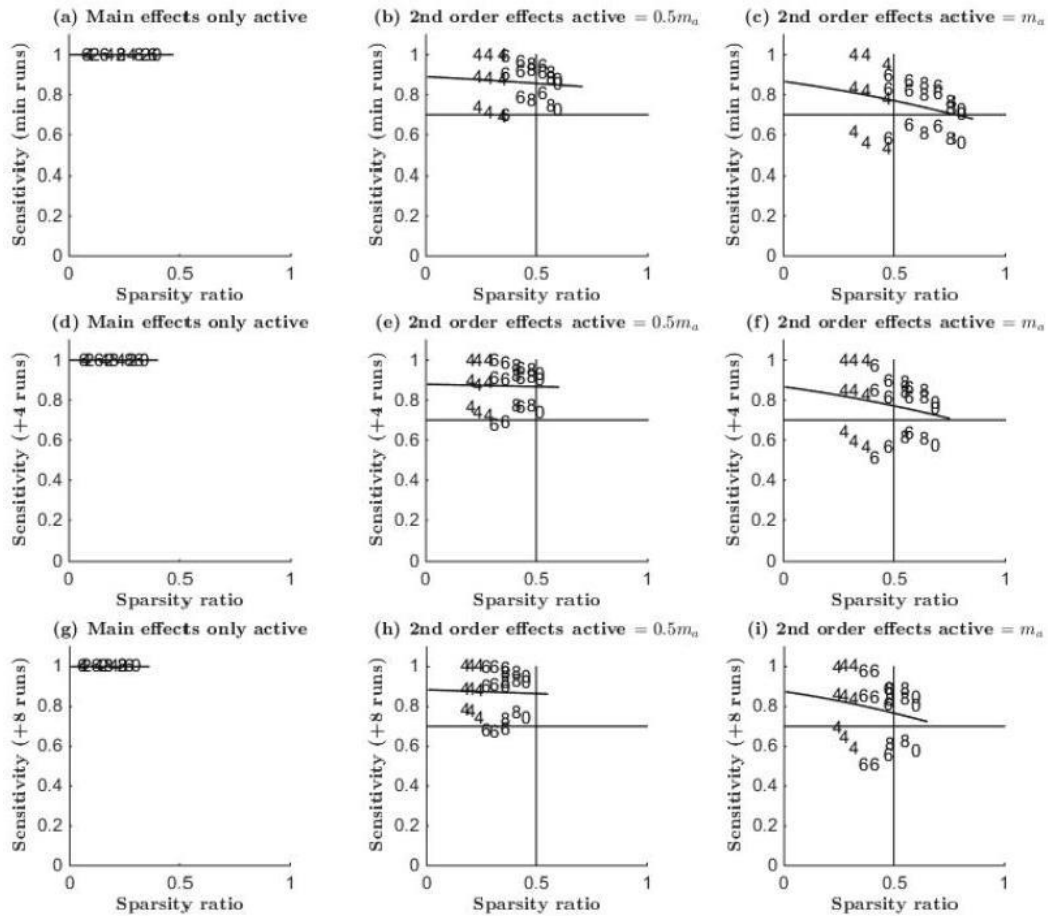


Figure 3.7: Sensitivity vs. Sparsity Ratio for Models Following Strong Heredity. Rows one to three pertain to zero augmented runs, 4 augmented runs, and 8 augmented runs, respectively. Columns one to three pertain to models comprised only of main effects; models comprised of main effects and second-order terms, where the number of second-order terms is half the number of main effects; and models comprised of main effects and second-order terms, where the number of second-order terms is equal to the number of main effects.

when the sparsity ratio is less than 0.5. This is counter to expectations: because the model space is so much smaller than for the unrestricted case, we expected that the correct model would be easier to identify. Notice that, when the number of active second-order effects is half the number of active main effects, obtaining sensitivities in excess of about 0.70 require that the sparsity ratio be less than about 0.30. When the number of second-order effects is equal to the number of active main effects (panel column 3), we are never able to meet the 0.70 threshold and, surprisingly, the performance of SHIM does not improve with the level of augmentation. Although this is not a positive result, we note that these particular models are not likely to arise frequently in practice. For example, the empirical study by Li et al. (2006) suggested that only about 11% of possible interactions are likely to be active with seven or fewer factors. For $m = 6$ factors, there are 15 potential two-factor interactions, so the expectation would be that on the order of two interactions might be present.

3.5 Discussion and Conclusions

In this paper, we examined the efficacy of DSDs when used to identify active first- and second-order effects in screening experiments. We conclude that, for the levels of sparsity frequently encountered in practice, the designs can be used effectively for this purpose. For minimum-run-size DSDs, when the sparsity ratio (number of active effects divided by run size) is less than about 0.50, model selection via either the Lasso/AICc or Gauss-

Dantzig/AICc is an effective strategy. If the analyst finds that the number of active terms is such that the sparsity ratio exceeds 0.50, results must be viewed with caution. In such a situation, we recommend augmenting the design so that all second-order effects involving factors found to be active can be simultaneously estimated. This “0.50 sparsity ratio” conclusion may not be entirely surprising; it is consistent with results reported by Marley and Woods (2010) for the analysis of (first-order) supersaturated designs. If a minimum-run-size DSD is augmented by four runs, results improve dramatically. For the models considered in our simulation with $p_{so} = 0.5$, the levels of sparsity never reached a point where model selection, with either Dantzig/AICc or Lasso/AICc broke down. For this reason, we recommend the use of two fake factors to augment minimum-run-size DSDs by four runs as a general practice.

The results just summarized pertain to unrestricted models. For models following strong heredity, SHIM can be used to identify active effects as long as the sparsity ratio is less than about 0.30.

This study did not consider DSDs with two-level categorical factors (Jones and Nachtsheim, 2013). It is known that, for these designs, correlations among two-factor interactions and between quadratic effects and two-factor interactions can be higher than in the all-continuous-factors case. For this reason, we would expect a general decline in resulting sensitivities relative to results reported here. The extent of the expected decline is an open question.

Chapter 4

Main Effects Designs for Logistic Regression that are Robust to the Presence of Two-factor Interactions

4.1 Introduction

The importance of the foldover technique in design of experiments (DOE) is firmly established. For example, most textbooks in DOE highlight the fact that folding over a resolution III screening design leads to a resolution IV design in which main effects estimators are statistically independent of the estimators of two-factor interaction effects. One major advantage of foldover designs is that when a first-order model is used for estimation, main effects estimators will not be biased by the presence of any active second-order effects that have been omitted from the model. Definitive Screening Designs (Jones and Nachtsheim, 2011a) are foldover designs in which main effects are orthogonal to each other, orthogonal to all second-order effects, and in which interactions are not confounded with each other. Other recent work in foldover designs includes McLeod and Brewster (2008), Ai

et al. (2010), Elsworth and Qin (2015), Errore et al. (2017a) Elsworth and Fang (2019), and Yang and Li (2019). Our objective here is to investigate the utility of two-level foldover and related designs when the response is dichotomous and logistic regression is used for analysis.

Experimental designs for logistic regression are, of course, used in a wide variety of applications. Examples include dose-response experiments and discrete choice experiments. Our work here has been motivated, in particular, by applications of discrete choice experiments. These designs are widely used in marketing research to investigate which product features drive consumer choices. When choice sets in discrete-choice experiments are dichotomous, the responses are typically analyzed using logistic regression. Frequently, standard main-effects designs are employed in discrete choice studies in an effort to keep the surveys short and avoid subject fatigue. Very little work has focused on designs for estimating two-factor interactions and/or model robustness to the presence of active two-factor interactions. Exceptions include Street and Burgess (2004), Yu et al. (2008), and Li et al. (2013). In this paper, we consider the question: Under what circumstances do the advantages associated with foldover designs for linear models carry over to the non-linear design case? For example, does the use of a folded over main-effects design for logistic regression eliminate potential biasing of main effects estimators from active second-order effects? And if such properties do not carry over, can we construct designs that capture all or most of the advantages associated with foldover designs in the linear models case? In

what follows, we make use of the property of parameter-orthogonality - in the sense of Cox and Reid (1987) - to construct designs that retain most or all of these advantages.

The remainder of the paper is structured as follows. In Section 4.2 we provide background information regarding current approaches to non-linear design and we introduce notation. In Section 4.3 we review the property of parameter-orthogonality (Cox and Reid, 1987) and indicate how it can be used to construct main-effects designs that are robust to the potential presence of active second-order effects. We also give conditions under which foldover designs lead to parameter orthogonality. In Section 4.4, we describe a numerical method for constructing locally optimal and Bayesian D-optimal designs having high degrees of parameter-orthogonality. In Section 4.5, we describe the results of a simulation study that evaluates the performance of our designs; and, finally, in Section 4.6 we summarize our findings and conclusions.

4.2 Background and Notation

Design criteria for non-linear regression models usually depend on the model's unknown coefficients, and when this is true it is necessary to make assumptions about the values of such coefficients. Several approaches have been proposed in the literature to face this conundrum—and each approach has its advantages and disadvantages. A *locally optimal design* can be constructed employing a prior guess of the values of the parameters. This approach is simple, but it can result in inefficient designs if the true parameters differ appre-

ciably from the prior values. Another approach is provided by the *Bayesian optimal design* approach, which accounts for the uncertainty of the parameters by integrating the design criterion over a prior distribution of such parameters. This integration can be computationally cumbersome, but it allows for the construction of designs that may be more efficient for a larger space of potential model parameter values than locally optimal designs.

Consider the classical settings of a logistic regression model in which the mean binary response Y is distributed according to a logistic model with logit link, for which the probability mass function can be written as follows:

$$p_i(\mathbf{x}_i, \boldsymbol{\beta}) = \frac{e^{\mathbf{f}^T(\mathbf{x}_i)\boldsymbol{\beta}}}{1 + e^{\mathbf{f}^T(\mathbf{x}_i)\boldsymbol{\beta}}} = \frac{1}{1 + e^{-\mathbf{f}^T(\mathbf{x}_i)\boldsymbol{\beta}}}, \quad (4.1)$$

where \mathbf{x}_i is the $m \times 1$ vector of factor settings for the i -th run, $\mathbf{f}^T(\mathbf{x}_i)$ is the $1 \times p$ row vector of model terms for the i -th run of the experiment, and $\boldsymbol{\beta}$ is the $p \times 1$ vector of unknown parameters. Let x_{ij} denote the j -th element of \mathbf{x}_i . The linear main effects (*ME*) model can be written as:

$$\mathbf{f}_{ME}^T(\mathbf{x}_i)\boldsymbol{\beta} = \beta_0 + \sum_{j=1}^m \beta_j x_{ij} \quad i = 1, \dots, n. \quad (4.2)$$

Similarly, the main effects plus interactions model (*ME+I*) is as follows:

$$\mathbf{f}_{ME+I}^T(\mathbf{x}_i)\boldsymbol{\beta} = \beta_0 + \sum_{j=1}^m \beta_j x_{ij} + \sum_{j=1}^{m-1} \sum_{k=j+1}^m \beta_{jk} x_{ij} x_{ik} \quad i = 1, \dots, n. \quad (4.3)$$

Let \mathbf{d} denote the $n \times m$ design matrix

$$\mathbf{d} = \begin{bmatrix} x_{11} & \dots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{nm} \end{bmatrix}. \quad (4.4)$$

We partition the $ME+I$ model as follows:

$$\mathbf{f}_{ME+I}^T(\mathbf{x}_i)\boldsymbol{\beta} = \mathbf{f}_1^T(\mathbf{x}_i)\boldsymbol{\beta}_1 + \mathbf{f}_2^T(\mathbf{x}_i)\boldsymbol{\beta}_2, \quad i = 1, \dots, n, \quad (4.5)$$

such that $\mathbf{f}_1^T(\mathbf{x}_i) = (x_{i1}, \dots, x_{im})$, $\boldsymbol{\beta}_1$ is the vector of main effects (intercept omitted), $\mathbf{f}_2^T(\mathbf{x}_i) = (1, x_{i1}x_{i2}, x_{i1}x_{i3}, \dots, x_{i(m-1)}x_{im})$ is the row vector consisting of a one followed by the $t = m(m-1)/2$ interaction terms, and $\boldsymbol{\beta}_2^T = (\beta_0, \beta_{1,2}, \beta_{1,3}, \dots, \beta_{m-1,m})$ is the row vector consisting of the intercept and the t interaction effects. Notice that by this partition of the coefficients we are separating the terms of major interest from those of lesser importance to the experimenter. This is not to say they are assumed to be negligible but that the experimenter is primarily focused on the estimation of the linear main effects. In what follows, the coefficients in $\boldsymbol{\beta}_1$ are of primary interest, whereas the coefficients in $\boldsymbol{\beta}_2$ are of secondary

interest.

Let $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$, where

$$\mathbf{X}_1 = \begin{bmatrix} \mathbf{f}_1^T(\mathbf{x}_1) \\ \vdots \\ \mathbf{f}_1^T(\mathbf{x}_n) \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2 = \begin{bmatrix} \mathbf{f}_2^T(\mathbf{x}_1) \\ \vdots \\ \mathbf{f}_2^T(\mathbf{x}_n) \end{bmatrix}. \quad (4.6)$$

For a design \mathbf{d} , the Fisher information matrix corresponding to a model \mathbf{f} is:

$$\mathbf{I}(\mathbf{d}, \boldsymbol{\beta}) = \mathbf{X}^T \mathbf{W}(\mathbf{d}, \boldsymbol{\beta}) \mathbf{X}, \quad (4.7)$$

where $\mathbf{W}(\mathbf{d}, \boldsymbol{\beta}) = \text{diag}[\mathbf{P}(\mathbf{d}, \boldsymbol{\beta}) - \mathbf{p}(\mathbf{d}, \boldsymbol{\beta})\mathbf{p}^T(\mathbf{d}, \boldsymbol{\beta})]$, with $\mathbf{P}(\mathbf{d}, \boldsymbol{\beta}) = \text{diag}(p_1, \dots, p_n)$, $\mathbf{p}(\mathbf{d}, \boldsymbol{\beta}) = (p_1, \dots, p_n)^T$, and $p_i = p(\mathbf{x}_i, \boldsymbol{\beta})$ is as given in (4.1). A design \mathbf{d} is D-optimal if it maximizes the logarithm of the determinant of the information matrix in (4.7):

$$\phi_D = \log(|\mathbf{I}(\mathbf{d}, \boldsymbol{\beta})|). \quad (4.8)$$

As noted above, since $\boldsymbol{\beta}$ is unknown, the information matrix in (4.7) cannot be used as a direct measure of information provided by the design, unless assumptions are made regarding the parameters in the model. Let $\boldsymbol{\beta}^0$ denote a prior estimate of the parameter vector that can be used to compute the information matrix in (4.7). Similarly, when $\boldsymbol{\beta}$ is partitioned,

the prior values for β_1 and β_2 , are denoted respectively by β_1^0 and β_2^0 .

There is an extensive body of existing literature pertaining to the construction of experimental designs for non-linear models. Literature reviews include Ford et al. (1989), Chaloner and Verdinelli (1995), and Khuri et al. (2006). A major takeaway from the existing literature is that optimal designs for generalized linear models have often been provided as a solution to specific special cases, but there are no general theoretical results.

A study by Box and Lucas (1959) represented one of the first papers to apply optimal design theory when the response model is presumed to be non-linear. In this paper, the optimality criterion was the maximization of the determinant of the information matrix, and the optimization was carried out for a specific value of the unknown parameters. This approach is labeled *local optimality*, as introduced by Chernoff (1953). Later work explored the construction of locally optimal designs for logistic regression or other non-linear models, using D-optimality and other standard design criteria.

A different approach was proposed by advocates of the Bayesian paradigm, which accounts for uncertainty in the prior parameter estimates. Abdelbasit and Plackett (1983), for instance, discussed several approaches for constructing optimal designs with sequential or Bayesian approaches. They raised concerns about the robustness of the designs in the case of a poor initial estimate on the parameters. Chaloner and Larntz (1989) introduced the Bayesian D-optimality criterion (discussed below) and provided equivalence theorems for Bayesian criteria. They constructed Bayesian D-optimal designs for the logistic regression

case and carried out efficiency comparisons with locally optimal designs. Ponce de Leon and Atkinson (1991) introduced the idea of T-optimal designs, which were related to discrimination between rival models. This was a criterion that utilized a Bayesian approach, but it required the true model to be an unknown element from a set of known models. Other work in this stream of literature includes Chaloner (1993); DuMouchel and Jones (1994); Sebastiani and Settimi (1997, 1998).

Other scholars focused on the choice and comparison between different design construction criteria. This approach includes the work of Sitter and Wu (1993); Sitter and Torsney (1995); Heise and Myers (1996); Dette and Sahn (1998); Mathew and Kumar Sinha (2001); Dette et al. (2004); Li and Majumdar (2008); Yang and Huang (2011); Yang and Mandal (2015); Kabera et al. (2012). Recent work by Haines et al. (2018) discussed interaction effects in a two-predictor logistic regression case for approximate D-optimal designs. Woods et al. (2006) introduced the idea of compromise designs in order to account for several sources of model uncertainty, such as uncertainty about the link function, the linear predictors and the model parameters. Woods and Lewis (2011) introduced a design criterion that accounts for such uncertainty.

Our work joins this latter stream of literature in that it aims to mitigate the effects of model uncertainty on the estimators of the main effects. The next section describes how to achieve model robustness via parameter-orthogonality between effects of primary and secondary interest.

4.3 Robust Design via Parameter-Orthogonality

Consider the full model in (4.5) in a linear regression case. It is well known that if the reduced model $y_i = \mathbf{f}_1^T(\mathbf{x}_i)\boldsymbol{\beta}_1 + \epsilon$ is employed for estimation, then the estimators of the parameters in $\boldsymbol{\beta}_1$ will be biased as follows:

$$E(\hat{\boldsymbol{\beta}}_1) = \boldsymbol{\beta}_1 + \mathbf{A}\boldsymbol{\beta}_2, \quad (4.9)$$

where the so-called alias matrix \mathbf{A} is given by:

$$\mathbf{A} = (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{X}_2. \quad (4.10)$$

That is, estimators of the main effects may be biased by the presence of active two-factor interactions, depending on the design employed. Suppose that \mathbf{d} has a foldover structure, such that \mathbf{d} can be written:

$$\mathbf{d} = \begin{bmatrix} \mathbf{d}_1 \\ -\mathbf{d}_1 \end{bmatrix}, \quad (4.11)$$

where \mathbf{d}_1 is $n/2 \times m$.

Let $\mathbf{X}_1(\mathbf{d}_1)$ denote the $n/2 \times m$ design matrix for \mathbf{d}_1 corresponding to the primary (i.e.

main) effects in \mathbf{f}_1 . Then, from (4.6),

$$\mathbf{X}_1 = \begin{bmatrix} \mathbf{X}_1(\mathbf{d}_1) \\ -\mathbf{X}_1(\mathbf{d}_1) \end{bmatrix}. \quad (4.12)$$

Similarly, let \mathbf{X}_2 denote the design matrix for the secondary effects in \mathbf{f}_2 . It is straightforward to show that

$$\mathbf{X}_2 = \begin{bmatrix} \mathbf{1} & \mathbf{X}_1 \\ \mathbf{1} & \mathbf{X}_1 \end{bmatrix}, \quad (4.13)$$

where \mathbf{X}_1 denotes the columns corresponding to the interaction effects. It follows immediately that $\mathbf{A} = \mathbf{0}$, since $\mathbf{X}_1^T \mathbf{X}_2 = \mathbf{0}$.

Efficient Foldover Designs (EFDs) (Errore et al., 2017a) take advantage of this property. A two-level design is said to be an EFD if the design maximizes the D-optimality criterion, $|\mathbf{X}_1^T \mathbf{X}_1|$, with \mathbf{X}_1 constrained to follow the structure implied by (4.11). Notice that, for estimability, we require the \mathbf{X}_1 have full column rank, which implies that $n = 2n_1 \geq 2m$.

For non-linear models, an analytical expression for the expected bias, as provided in (4.9) for the linear case, is not available and another approach is required. In what follows, we adopt the concept of parameter-orthogonality as developed by Cox and Reid (1987). Let $\boldsymbol{\beta}' = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2)$ denote a partition of the p elements of $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}'_1 = (\beta_{1_1}, \dots, \beta_{1_{p_1}})$, $\boldsymbol{\beta}'_2 = (\beta_{2_1}, \dots, \beta_{2_{p_2}})$, and $p_1 + p_2 = p$. Following Cox and Reid (1987), the two parameter spaces

are orthogonal if the elements of the information matrix satisfy the following equality:

$$I(\beta_{1_i}, \beta_{2_j}) = \frac{1}{n} E \left(\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_{1_i}} \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_{2_j}}; \boldsymbol{\beta} \right) = \frac{1}{n} E \left(- \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_{1_i}} \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_{2_j}}; \boldsymbol{\beta} \right) = 0 \quad (4.14)$$

where $l(\boldsymbol{\beta})$ is the log-likelihood, $i = 1, \dots, p_1$, $j = 1, \dots, p_2$, and $I(\beta_{1_i}, \beta_{2_j})$ refers to the information per observation. When this equality holds for all $\boldsymbol{\beta}$ in the parameter space, the resulting property is called *global orthogonality* of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. If it holds only at one parameter value $\boldsymbol{\beta}^0$, the resulting property is called *local orthogonality* (Cox and Reid, 1987).

In what follows, we explore the application of parameter-orthogonality to the construction of main-effects designs for logistic regression, when $\boldsymbol{\beta}_1$ corresponds to the main effect parameters and $\boldsymbol{\beta}_2$ corresponds to the intercept term and potential two-factor interaction effects. In this situation, global parameter orthogonality between $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ is achieved when $\mathbf{X}_1^T \mathbf{W} \mathbf{X}_2 = \mathbf{0}$ for all values $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. (Henceforth \mathbf{W} is used in lieu of $\mathbf{W}(\mathbf{d}, \boldsymbol{\beta})$ for brevity). In particular we seek to characterize the value of foldover designs, if any, when employed in logistic regression contexts. We note that, in the linear case, foldover designs automatically lead to parameter orthogonality since for such designs $\mathbf{X}_1^T \mathbf{X}_2 = \mathbf{0}$.

Consider the situation in which a main-effects design for logistic regression is to be conducted, but there is concern about potential bias from active two-factor interactions. It is frequently assumed that second-order effects will be negligible, or at least small relative

to main effects. This leads naturally to the assumption that $\beta_2^0 = \mathbf{0}$. This assumption implies that the experimenter expects the secondary terms to be negligible or absent, and the assumption relates to the prior value of the coefficients on which to construct the non-linear design. This does not imply that the true coefficient will be in truth not active. The superscript 0 identifies the prior values and not the true parameters. The theorem that follows addresses the following question: if $\beta_2^0 = \mathbf{0}$, will the use of a foldover design achieve parameter-orthogonality in logistic regression?

Theorem 4.1

If β_1^0 is arbitrary and $\beta_2^0 = \mathbf{0}$, then any design with a foldover structure achieves parameter-orthogonality. That is, $\mathbf{X}_1^T \mathbf{W} \mathbf{X}_2 = \mathbf{0}$, for all β_1^0 when $\beta_2^0 = \mathbf{0}$. □

Proof 4.1

In model (4.5), the matrix of weights, \mathbf{W} , is the matrix of which the diagonal elements are as follows:

$$w_{ii} = p_i(1 - p_i) \quad i = 1, \dots, n; \quad (4.15)$$

and

$$p_i = \frac{e^{\mathbf{f}^T(\mathbf{x}_i)\beta}}{1 + e^{\mathbf{f}^T(\mathbf{x}_i)\beta}} = \frac{1}{1 + e^{-\mathbf{f}^T(\mathbf{x}_i)\beta}} = \frac{1}{1 + e^{-\mathbf{f}_1^T(\mathbf{x}_i)\beta_1 - \mathbf{f}_2^T(\mathbf{x}_i)\beta_2}}, \quad (4.16)$$

Thus we can rewrite the i -th diagonal element of \mathbf{W} as:

$$\begin{aligned}
w_{ii} &= p_i(1 - p_i) \\
&= \frac{1}{1 + e^{-\mathbf{f}_1^T(\mathbf{x}_i)\beta_1 - \mathbf{f}_2^T(\mathbf{x}_i)\beta_2}} \left(1 - \frac{1}{1 + e^{-\mathbf{f}_1^T(\mathbf{x}_i)\beta_1 - \mathbf{f}_2^T(\mathbf{x}_i)\beta_2}} \right) \\
&= \frac{1}{1 + e^{-\mathbf{f}_1^T(\mathbf{x}_i)\beta_1 - \mathbf{f}_2^T(\mathbf{x}_i)\beta_2}} \left(\frac{1 + e^{-\mathbf{f}_1^T(\mathbf{x}_i)\beta_1 - \mathbf{f}_2^T(\mathbf{x}_i)\beta_2} - 1}{1 + e^{-\mathbf{f}_1^T(\mathbf{x}_i)\beta_1 - \mathbf{f}_2^T(\mathbf{x}_i)\beta_2}} \right) \\
&= \frac{e^{\mathbf{f}_1^T(\mathbf{x}_i)\beta_1 + \mathbf{f}_2^T(\mathbf{x}_i)\beta_2}}{(1 + e^{\mathbf{f}_1^T(\mathbf{x}_i)\beta_1 + \mathbf{f}_2^T(\mathbf{x}_i)\beta_2})^2},
\end{aligned} \tag{4.17}$$

Suppose now that a design \mathbf{d} has a foldover structure as in (4.11), and assume the rows are arranged in foldover-pair order. Let \mathbf{x}_{jj} and $-\mathbf{x}_{jj}$ denote the two runs of the j th mirror image pair, for $j = 1, \dots, n/2$. Also, let w_{jj} and w_{-jj} denote the corresponding elements of the diagonal of \mathbf{W} . Then: $\mathbf{f}_1^T(\mathbf{x}_{jj}) = -\mathbf{f}_1^T(-\mathbf{x}_{jj})$, and $\mathbf{f}_2^T(\mathbf{x}_{jj}) = \mathbf{f}_2^T(-\mathbf{x}_{jj})$. It follows that:

$$w_{jj} = \frac{e^{\mathbf{f}_1^T(\mathbf{x}_{jj})\beta_1 + \mathbf{f}_2^T(\mathbf{x}_{jj})\beta_2}}{(1 + e^{\mathbf{f}_1^T(\mathbf{x}_{jj})\beta_1 + \mathbf{f}_2^T(\mathbf{x}_{jj})\beta_2})^2}, \tag{4.18}$$

and

$$w_{-jj} = \frac{e^{-\mathbf{f}_1^T(\mathbf{x}_{jj})\beta_1 + \mathbf{f}_2^T(\mathbf{x}_{jj})\beta_2}}{(1 + e^{-\mathbf{f}_1^T(\mathbf{x}_{jj})\beta_1 + \mathbf{f}_2^T(\mathbf{x}_{jj})\beta_2})^2}, \tag{4.19}$$

and, when β_2 is zero:

$$\begin{aligned}
 w_{-jj} &= \frac{e^{-\mathbf{f}_1^T(\mathbf{x}_{jj})\beta_1}}{(1 + e^{-\mathbf{f}_1^T(\mathbf{x}_{jj})\beta_1})^2} = \frac{e^{-\mathbf{f}_1^T(\mathbf{x}_{jj})\beta_1}}{\left(1 + \frac{1}{e^{\mathbf{f}_1^T(\mathbf{x}_{jj})\beta_1}}\right)^2} = \\
 &= \frac{e^{-\mathbf{f}_1^T(\mathbf{x}_{jj})\beta_1}}{\left(\frac{e^{\mathbf{f}_1^T(\mathbf{x}_{jj})\beta_1} + 1}{e^{\mathbf{f}_1^T(\mathbf{x}_{jj})\beta_1}}\right)^2} = \frac{e^{\mathbf{f}_1^T(\mathbf{x}_{jj})\beta_1}}{(1 + e^{\mathbf{f}_1^T(\mathbf{x}_{jj})\beta_1})^2} = w_{jj},
 \end{aligned} \tag{4.20}$$

This implies that pairs of diagonal elements of \mathbf{W} are identical. Thus,

$$\begin{aligned}
 \mathbf{X}_1^T \mathbf{W} \mathbf{X}_2 &= \\
 &= \sum_{j=1}^{n/2} w_{jj} (\mathbf{f}_1(\mathbf{x}_{jj}) \mathbf{f}_2^T(\mathbf{x}_{jj}) + \mathbf{f}_1(-\mathbf{x}_{jj}) \mathbf{f}_2^T(-\mathbf{x}_{jj})) \\
 &= \sum_{j=1}^{n/2} w_{jj} ((\mathbf{f}_1(\mathbf{x}_{jj}) - \mathbf{f}_1(\mathbf{x}_{jj})) \mathbf{f}_2^T(\mathbf{x}_{jj})) = \mathbf{0}_{m \times m_2},
 \end{aligned} \tag{4.21}$$

□

as required.

Since this applies for all β_1 , we view the resulting designs as *partially local optimal designs* to distinguish them from fully local non-linear designs in which all of β is specified.

One implication of Theorem 4.1 is that one can easily search for partially local optimal designs within the space of foldover designs and compare the efficiency of the resulting design to a structure-free D-optimal design. Then the tradeoffs, if any, between D-efficiency and the potential bias can be explored. In the following section we will construct both efficient foldover designs and structure-free compound designs and compare their perfor-

mance.

Theorem 4.1 establishes $\beta_2^0 = \mathbf{0}$ as a sufficient condition for parameter-orthogonality when a foldover design is used. It is not, however, a necessary condition. Example 4 of Section 4.1.4 provides a counterexample. Interestingly, our empirical results suggest that, for any $\beta_2^0 \neq \mathbf{0}$, if a parameter-orthogonal design exists, such design is always a foldover. Examples of this are provided in Sections 4.1.2 and 4.1.4. Finally, for some small design sizes, in some cases of $\beta_2^0 \neq \mathbf{0}$, a parameter-orthogonal design cannot be found.

In the next section we summarize two construction algorithms that we employ to compute robust designs for logistic regression, and we discuss several examples, with accompanying numerical results.

4.4 Constructing Robust Designs Numerically

In this section, we investigate the construction of efficient main effects designs for logistic regression that are either parameter-orthogonal or nearly so. We explore the tradeoff between design efficiency and parameter-orthogonality for both locally optimal and Bayesian designs. In both cases, the construction algorithms are described, and several examples are provided.

4.4.1 Locally Optimal Designs

We implement a compound optimization approach (Cook and Wong, 1994) that considers both the efficiency of the first-order model and the degree of parameter-orthogonality between first- and second-order effects achieved. We also construct D-optimal designs with an induced foldover structure. Notice that henceforth we refer to the D-optimal design as the design that is constructed by the sole criterion of D-efficiency. Such design may not be unique, and designs with the same D-efficiency may have different parameter-orthogonality, nonetheless, a construction algorithm that only employs the D-efficiency criterion will not discriminate between such designs.

Our compound optimization approach is similar to that given by Jones and Nachtsheim (2011b) for the construction of minimal alias designs. For $0 \leq \alpha \leq 1$, let:

$$\phi_\alpha(\mathbf{d}, \boldsymbol{\beta}) = (1 - \alpha)\phi_D^s(\mathbf{d}, \boldsymbol{\beta}) + \alpha\phi_{PO}^s(\mathbf{d}, \boldsymbol{\beta}), \quad (4.22)$$

where $\phi_D^s(\mathbf{d}, \boldsymbol{\beta})$ is the D-efficiency of the design \mathbf{d} (Jones, 2013), and where $\phi_{PO}^s(\mathbf{d}, \boldsymbol{\beta})$ is a scaled measure of the degree of parameter-orthogonality, which we define as follows. Let \mathbf{D} denote a D-optimal design. Then, the D-efficiency of \mathbf{d} is given by

$$\phi_D^s(\mathbf{d}, \boldsymbol{\beta}) = \left[\frac{|\mathbf{X}'_1(\mathbf{d}, \boldsymbol{\beta})\mathbf{W}\mathbf{X}_1(\mathbf{d}, \boldsymbol{\beta})|}{|\mathbf{X}'_1(\mathbf{D}, \boldsymbol{\beta})\mathbf{W}\mathbf{X}_1(\mathbf{D}, \boldsymbol{\beta})|} \right]^{1/p}. \quad (4.23)$$

Let $\mathbf{C}(\mathbf{d}, \boldsymbol{\beta}) = \mathbf{X}_1' \mathbf{W} \mathbf{X}_2 = [c_{ij}]$ and let $\phi_{PO}(\mathbf{d}, \boldsymbol{\beta})$ denote the average of the absolute values of the elements in $\mathbf{C}(\mathbf{d}, \boldsymbol{\beta})$:

$$\phi_{PO}(\mathbf{d}, \boldsymbol{\beta}) = \left(\sum_i \sum_j |c_{ij}| \right) / (m \times t). \quad (4.24)$$

The smaller this value, the more robust is the design to the presence of two-factor interactions. Notice that this criterion is scale dependent unless standardization is performed. Criterion (4.24) is proportional to the L_2 norm of $\mathbf{C}(\mathbf{d}, \boldsymbol{\beta})$. Other measures, such as the L_2 or L_∞ norms could also be employed, depending on the degree to which the experimenter wishes to penalize larger absolute values of the $\{|c_{ij}|\}$. In what follows, we restrict attention to (4.24).

Clearly, a design with $\phi_{PO}(\mathbf{d}, \boldsymbol{\beta}) = 0$ achieves local parameter-orthogonality. We employ a scaled version of $\phi_{PO}(\mathbf{d}, \boldsymbol{\beta})$ denoted $\phi_{PO}^s(\mathbf{d}, \boldsymbol{\beta})$ as follows:

$$\phi_{PO}^s(\mathbf{d}, \boldsymbol{\beta}) = 1 - \frac{\phi_{PO}(\mathbf{d}, \boldsymbol{\beta})}{\phi_{PO}^{max}}, \quad (4.25)$$

where $\phi_{PO}^{max} = \max_{\mathbf{d}, \boldsymbol{\beta}} \phi_{PO}(\mathbf{d}, \boldsymbol{\beta})$ and we refer to $\phi_{PO}^s(\mathbf{d}, \boldsymbol{\beta})$ as the level of *parameter-orthogonality efficiency*. Clearly, $0 \leq \phi_{PO}^s(\mathbf{d}, \boldsymbol{\beta}) \leq 1$, and $\phi_{PO}^s(\mathbf{d}, \boldsymbol{\beta}) = 1$ when \mathbf{d} is parameter-orthogonal. It can be shown that the maximum of $\phi_{PO}^s(\mathbf{d}, \boldsymbol{\beta})$ over \mathbf{d} and $\boldsymbol{\beta}$ is $\phi_{PO}^{max} = n/4$. Thus from (4.25) we have $\phi_{PO}^s(\mathbf{d}, \boldsymbol{\beta}) = 1 - 4\phi_{PO}(\mathbf{d}, \boldsymbol{\beta})/n$.

Our compound construction algorithm performs the following steps:

1. Construct the D-optimal design. This is also the design that maximizes (4.22) for $\alpha = 0$.
2. For evenly spaced values of $\alpha \in (0, 1)$, construct a design that maximizes (4.22). We refer to these designs as compound-optimal designs.
3. Construct the design that maximizes the parameter-orthogonality. This design is the numerical solution to (4.22) when $\alpha = 1$.

We also adopt a second approach to find the D-optimal design, subject to the constraint that the design is a foldover design. We refer to optimal designs found in this fashion as non-linear efficient foldover designs (NLEFDs), building on the EFDs of Erre et al. (2017a). Our NLEFD algorithm basically performs design construction using a simple modification of the coordinate exchange algorithm (Meyer and Nachtsheim, 1995). This algorithm constructs D-optimal designs, subject to the constraint that the design is a foldover design.

In all of the following examples, we show our results from the construction of designs with six factors and 12 runs. All the examples are exact designs on a design region of all two level factors at possible values -1 and 1 . We report plots of the D-efficiencies and parameter orthogonality efficiencies for the compound designs for $0 \leq \alpha \leq 1$ and we refer to these plots as *efficiency trace plots*, in the sense of Jones and Nachtsheim (2011b).

Example 1: $\beta_1^0 = \beta_2^0 = \mathbf{0}$

The first example that we investigated is a locally optimal design constructed with prior estimates of all parameters all equal to zero; that is, $\beta_1^0 = \beta_2^0 = \mathbf{0}$. In Figure 4.1(a), the efficiency trace plot of the compound designs constructed with our first algorithm and the D-optimal robust foldover design constructed with our second algorithm are shown. Weights α are represented on the x-axis and efficiencies on the y-axis. The asterisk and the circle on the right-hand side of the graphs represent D-efficiency and $\phi_{PO}(d)$ (degree of parameter-orthogonality), respectively, of the NLEFD. This latter design is also marked as “Fold” in the plots. A D-optimal design and an NLEFD for this example are shown in Table 4.1 and Table 4.2, respectively.

A D-optimal design does not achieve parameter-orthogonality. However, as shown in Figure 4.1(a), as α increases from 0 to 0.3, the designs constructed maintain 100% D-efficiency while achieving a slightly higher degree of parameter-orthogonality. As α increases from 0.3 to 0.9, the compound designs achieve higher degrees of parameter-orthogonality with small losses in D-efficiency. As indicated by the asterisk and the circle, the NLEFD achieves parameter-orthogonality with about 90% D-efficiency. Notice that, in this example, by Theorem 4.1, any foldover design will lead to parameter-orthogonality. Moreover, having both sets of parameters priors assumed equal to zero, the non-linear model design construction reduces to the linear model design case, as highlighted by Cox (1988).

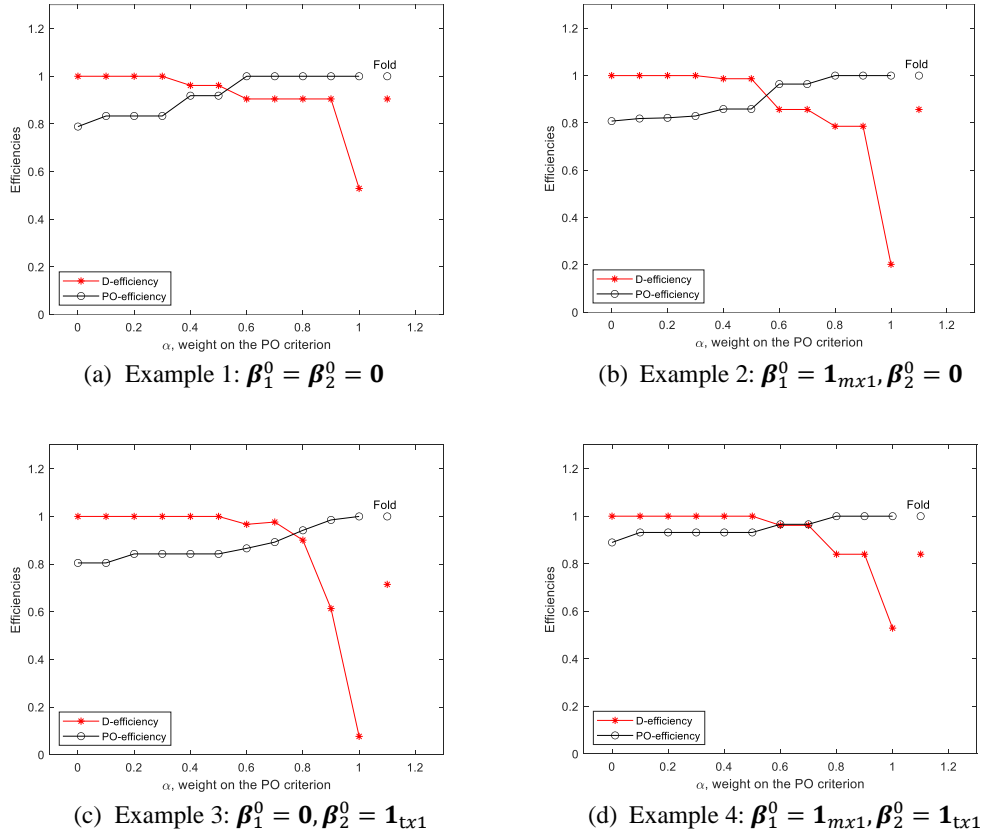


Figure 4.1: Locally optimal designs with $m = 6, n = 12$: (a) Example 1: $\beta_1^0 = \beta_2^0 = \mathbf{0}$; (b) Example 2: $\beta_1^0 = \mathbf{1}_{m \times 1}$ and $\beta_2^0 = \mathbf{0}$; (c) Example 3: $\beta_1^0 = \mathbf{0}$ and $\beta_2^0 = \mathbf{1}_{t \times 1}$; (d) Example 4: $\beta_1^0 = \mathbf{1}_{m \times 1}$ and $\beta_2^0 = \mathbf{1}_{t \times 1}$. Weights α are represented on the x-axis and efficiencies on the y-axis. The asterisk and the circle on the right-hand side of the graphs represent D-efficiency and parameter-orthogonality efficiency, respectively, of the NLEFD. This latter design is also marked as “Fold” in the plots.

Table 4.1: D-optimal design for six factors and 12 runs

X1	X2	X3	X4	X5	X6
1	1	1	1	-1	-1
1	1	1	-1	-1	1
1	1	-1	1	1	-1
1	1	-1	-1	1	1
1	1	-1	-1	-1	-1
1	-1	-1	1	1	1
1	-1	-1	1	-1	-1
1	-1	-1	-1	-1	1
-1	1	-1	1	-1	-1
-1	1	-1	-1	1	-1
-1	1	-1	-1	-1	1
-1	-1	1	-1	1	-1

Example 2: $\beta_1^0 \neq \mathbf{0}$ and $\beta_2^0 = \mathbf{0}$

Next, we investigated a locally optimal design constructed on non-zero prior estimates of the parameters related to the main effects, so $\beta_1^0 \neq \mathbf{0}$, but $\beta_2^0 = \mathbf{0}$. In particular, we assume $\beta_1^0 = \mathbf{1}_{m \times 1}$. Figure 4.1(b) shows the efficiency trace plot of this example. The general pattern is similar to Example 1. The D-optimal design does not achieve parameter-orthogonality. The efficiency of the compound designs for the parameter-orthogonality criterion increases with α , with little loss of D-efficiency. The design constructed with weight $\alpha = 1$ does not account for D-efficiency, by construction, so the resulting design has low D-efficiency. The NLEFD is parameter-orthogonal and 85% D-efficient. As was the case for Example 1, $\beta_2^0 = \mathbf{0}$, so that Theorem 4.1 applies.

Table 4.2: NLEFD for six factors and 12 runs

X1	X2	X3	X4	X5	X6
1	1	1	1	-1	-1
1	1	1	-1	1	-1
1	1	-1	-1	-1	-1
1	-1	1	1	1	-1
1	-1	-1	1	1	1
1	-1	-1	-1	-1	1
-1	1	1	1	1	-1
-1	1	1	-1	-1	-1
-1	1	-1	-1	-1	1
-1	-1	1	1	1	1
-1	-1	-1	1	-1	1
-1	-1	-1	-1	1	1

Example 3: $\beta_1^0 = \mathbf{0}$ and $\beta_2^0 \neq \mathbf{0}$

In this example, we take $\beta_1^0 = \mathbf{0}$ and $\beta_2^0 \neq \mathbf{0}$. In particular, we take $\beta_2^0 = \mathbf{1}_{t \times 1}$. Figure 4.1(c) shows the efficiency trace plot of this example. The general pattern is again similar to the previous examples. The D-optimal design does not achieve parameter-orthogonality; the NLEFD achieves parameter-orthogonality with D-efficiency equal to 71.5%.

Example 4: $\beta_1^0 \neq \mathbf{0}$ and $\beta_2^0 \neq \mathbf{0}$

Finally, we consider an example with non-zero prior estimates for all parameters, main effects, two-factor interactions and intercept. Here we take $\beta_1^0 = \mathbf{1}_{m \times 1}$ and $\beta_2^0 = \mathbf{1}_{t \times 1}$. Figure 4.1(d) shows the efficiency trace plot of this example. Again, the D-optimal design does not achieve parameter-orthogonality. The efficiency of the compound designs for the

parameter-orthogonality criterion increases as expected as α increases, initially with no loss of D-efficiency, for $\alpha < 0.6$, then with an increased loss. The compound designs achieve parameter-orthogonality when $\alpha = 0.8$, with about 84% D-efficiency. This design and the following are all foldover designs, though the foldover structure was not enforced. The NLEFD achieves parameter-orthogonality with 84% D-efficiency.

Findings on Locally Optimal Designs

Based upon the above results and more not reported here, we make the following observations:

1. As expected the D-efficiencies of the compound designs decrease as α increases, whereas the degree of parameter-orthogonality increases. The trace plots depict the tradeoffs and can be used by the experimenter to choose a suitable compromise between D-efficiency and potential bias.
2. Generally, the D-optimal design does not achieve parameter-orthogonality. This is to be expected because the D-optimality criterion does not account for the secondary terms in the model.
3. Frequently, a sequence comprised of designs that are equivalent in terms of D-efficiency, but have different degrees of parameter-orthogonality, is produced. In such case, the efficiency trace plots can be used to discriminate between designs having the same D-efficiency, and select the design with the highest degree of parameter-orthogonality.

Such design is the one constructed with the minimum weight α that provides a design with the highest degree of parameter-orthogonality with the same value of D-efficiency.

4. The designs constructed with $\alpha = 1$, optimizing only on the parameter-orthogonality criterion, do not achieve high values of D-efficiencies in all cases, as expected.
5. The NLEFD frequently achieves parameter-orthogonality with high levels of D-efficiency.

These results pertain to locally-optimal designs. We now consider methods for constructing robust Bayesian D-optimal main effects designs.

4.4.2 Bayesian Designs

In order to incorporate the uncertainty about the assumed values of the parameters, we also employ our construction methods with a Bayesian approach. We employ Bayesian D-optimality criterion for non-linear design Chaloner and Verdinelli (1995), that can be expressed as follows:

$$\phi_D(\mathbf{d}, \pi) = \int \log |\mathbf{I}(\mathbf{d}, \boldsymbol{\beta})| \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}, \quad (4.26)$$

where $\pi(\boldsymbol{\beta})$ is the prior distribution on the full parameter vector $\boldsymbol{\beta}$. Chaloner and Verdinelli (1995) refer to designs that maximize (4.26) as Bayesian D-optimal designs.

In order to approximate the calculation of the integral in (4.26), we employ a quadrature approach as advocated by Gotwalt et al. (2009). These authors modify spherical-radial integration rules proposed in Monahan and Genz (1997) in order to provide a fast numerical approximation to (4.26). The objective is achieved by reducing the integral to a summation of evaluations of the log-determinant of the information matrix at specific spherical points with two radii and a center point (Gotwalt, 2010).

Assume that the prior density $\pi(\boldsymbol{\beta})$ takes the form of multivariate normal density with mean $\boldsymbol{\beta}^0$ and prior variance-covariance matrix given by

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{bmatrix}, \quad (4.27)$$

where $\boldsymbol{\Sigma}_1 = \text{Var}(\boldsymbol{\beta}_1)$, $\boldsymbol{\Sigma}_2 = \text{Var}(\boldsymbol{\beta}_2)$ and $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21} = \mathbf{0}$. The approximate value of $\phi_D(\mathbf{d}, \pi(\boldsymbol{\beta}))$ is as follows:

$$\phi_D(\mathbf{d}, \pi(\boldsymbol{\beta})) \approx \omega_0 \log|\mathbf{I}(\mathbf{d}, \boldsymbol{\beta}^0)| + \sum_i \sum_j \omega_{ij} \log|\mathbf{I}(\mathbf{d}, \boldsymbol{\beta}^0) + \mathbf{L}\mathbf{V}_{ij}|, \quad (4.28)$$

where \mathbf{L} is the lower triangular Cholesky root of $\boldsymbol{\Sigma}$, and ω_0 and ω_{ij} are the weights assigned to the mean and the spherical points \mathbf{V}_{ij} . The i index refers to the number of spheres and the j index refers to the sphere's points used in the quadrature approximation. For further

details on this procedure, refer to Gotwalt et al. (2009).

In the Bayesian case the computation of relative efficiency of a design in respect to a D-optimal design \mathbf{D} is computed as:

$$\phi_D^s(\mathbf{d}, \pi(\boldsymbol{\beta})) = \exp(\phi_D(\mathbf{d}, \pi(\boldsymbol{\beta})) - \phi_D(\mathbf{D}, \pi(\boldsymbol{\beta}))/p). \quad (4.29)$$

We construct robust Bayesian designs in three cases: (1) $\boldsymbol{\Sigma}_1 \neq \mathbf{0}, \boldsymbol{\Sigma}_2 = \mathbf{0}$, (2) $\boldsymbol{\Sigma}_1 = \mathbf{0}, \boldsymbol{\Sigma}_2 \neq \mathbf{0}$ and (3) $\boldsymbol{\Sigma}_1 \neq \mathbf{0}, \boldsymbol{\Sigma}_2 \neq \mathbf{0}$. In each case, we assume that the prior parameter vectors are independent, so that $\boldsymbol{\Sigma}$ is block-diagonal.

Example 5: $\boldsymbol{\Sigma}_1 \neq \mathbf{0}; \boldsymbol{\Sigma}_2 = \mathbf{0}$

In the first Bayesian design example, we assume $\boldsymbol{\beta}_1^0 = \boldsymbol{\beta}_2^0 = \mathbf{0}$, $\boldsymbol{\Sigma}_1 = \mathbf{I}_m$ and $\boldsymbol{\Sigma}_2 = \mathbf{0}$. This example basically assumes that there is uncertainty about the main-effects coefficients and that the second-order effects are negligible, which corresponds to the partially local optimal design conditions of Example 1 and 2. The main effects are assumed to have a independent standard normal distributions. Figure 4.2(a) shows the efficiency trace plot of this example, and it has a pattern that is qualitatively similar to the case in Example 1.

Example 6: $\boldsymbol{\Sigma}_1 = \mathbf{0}; \boldsymbol{\Sigma}_2 \neq \mathbf{0}$

In the next example, we assume $\boldsymbol{\beta}_1^0 = \boldsymbol{\beta}_2^0 = \mathbf{0}$, $\boldsymbol{\Sigma}_1 = \mathbf{0}$ and $\boldsymbol{\Sigma}_2 = \mathbf{I}_t$. Now, we basically investigate a Bayesian version of the Example 3. Again, the pattern of efficiency plots

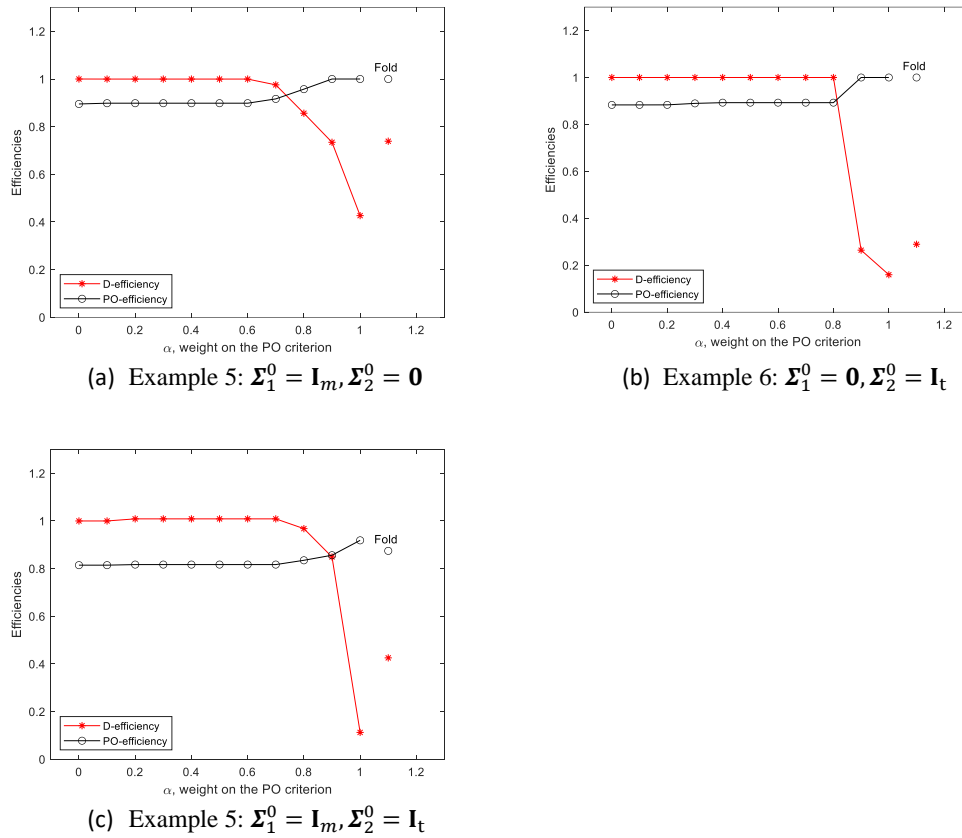


Figure 4.2: Bayesian optimal designs with $m = 6, n = 12$: (a) Example 5: $\Sigma_1 = \mathbf{I}_m; \Sigma_2 = \mathbf{0}$; (b) Example 6: $\Sigma_1 = \mathbf{0}; \Sigma_2 = \mathbf{I}_t$; (c) Example 7: $\Sigma_1 = \mathbf{I}_m; \Sigma_2 = \mathbf{I}_t$. Weights α are represented on the x-axis and efficiencies on the y-axis. The asterisk and the circle on the right-hand side of the graphs represent D-efficiency and parameter-orthogonality efficiency, respectively, of the NLEFD. This latter design is also marked as “Fold” in the plots.

in Figure 4.2(b) is very similar to the locally optimal example. However, in this case the NLEFD does not have high D-efficiency. However, note that this example is somewhat unrealistic in that all main effects are assumed to have zero priors and only two-factor interactions have non-zero prior distributions. This is the reverse of what would typically be assumed in practice.

Example 7: $\Sigma_1 \neq \mathbf{0}; \Sigma_2 \neq \mathbf{0}$

In our final example, we assume $\beta_1^0 = \beta_2^0 = \mathbf{0}; \Sigma_1 = \mathbf{I}_m; \Sigma_2 = \mathbf{I}_t$. The trace plot in Figure 4.2(c) departs from those of the previous examples. None of the designs achieves full parameter-orthogonality. The NLEFD has 42.5% D-efficiency and its degree of parameter-orthogonality is 87.4%. In this example a better choice might be the minimax design found with $\alpha = 0.8$, that has about 85% both D-efficiency and degree of parameter-orthogonality.

Findings on Bayesian Optimal Designs

From the Bayesian approach examples, we observe the following:

1. The general pattern of the efficiency traces holds consistent with the locally optimal case.
2. In Examples 5 and 6, in which only one of the two sets of parameters has non-zero prior variance, the patterns of efficiencies are fairly similar, and the minimax design seems to be a better choice than either one of the two designs constructed on only

one criterion, and the NLEFD.

3. When all parameters' prior distributions employ a non-zero variance, the resulting designs do not achieve parameter-orthogonality.

We note that while the patterns are similar, the efficiencies displayed in Figure 4.1 and Figure 4.2 are not directly comparable. This is a result of the fact that the efficiencies in Figure 4.1 obtained from the frequentist efficiency calculation given in (4.23), whereas those of Figure 4.2 stem from a Bayesian calculation employing (4.26).

4.5 Simulation Study

To investigate the efficacy of our NLEFDs we perform a simulation study in which we compare them to D-optimal designs for main effects. In particular, we consider designs having six factors with 12 runs, as in the cases explored in the previous section.

The study design involves two factors:

1. The design type. We consider two design types:
 - (a) a D-optimal design for the first-order model;
 - (b) an NLEFD.
2. The number and magnitude of non-zero coefficients of the two-factor interactions in the true model. We employ three levels:

- (a) No interaction effects;
- (b) One two-factor interaction effect with a coefficient randomly drawn from a uniform distribution $[-1, 1]$;
- (c) Two two-factor interaction effects with coefficients randomly drawn from a uniform distribution $[-1, 1]$.

In the study, we hold the true value of the main effects coefficients β_1 constant and equal to **0**. We expect that use of the NLEFD will result in significantly less bias in the estimation of the coefficients of the main effects, even in presence of active two-factor interactions.

We simulate responses from the specified logistic regression model for each of the two designs under study. The true model is used to determine $p(\mathbf{x}_i, \boldsymbol{\beta})$, for $i = 1, \dots, 12$, for each design. We then generate Bernoulli random variables for each of the 12 runs in each design, for 100 replicates of each design. Logistic regression is then used to estimate the main effects regression coefficients. This simulation process is repeated $N = 10,000$ times and the average squared bias for each main effects coefficient is determined as follows.

Let $\hat{\beta}_{kj}$ denote the estimated coefficient for the j th main effect obtained in the k -th replicate of the simulation. The mean squared error for the j th main effect estimator is given by:

$$\text{MSE}(\beta_j) = \text{Var}(\beta_j) + \text{Bias}^2(\beta_j), \quad (4.30)$$

Table 4.3: Comparison of squared bias in main effects parameters estimates resulting from the use of the D-optimal design and the NLEFD. Standard errors are given in parentheses.

Main Effect affected	One active interaction		Two active interactions	
	D-opt	NLEFD	D-opt	NLEFD
x_1	0.0721 (0.0335)	2.5×10^{-7} (8.2×10^{-8})	0.0497 (0.0081)	4.4×10^{-7} (6.5×10^{-8})
x_2	0.8108 (0.0381)	6.4×10^{-7} (1.2×10^{-7})	0.0614 (0.0089)	9.95×10^{-7} (1.2×10^{-7})
x_3	0.0565 (0.0291)	7.4×10^{-7} (2.4×10^{-7})	0.0537 (0.0086)	9.7×10^{-7} (1.3×10^{-7})
x_4	0.05076 (0.0126)	2.6×10^{-7} (6.8×10^{-8})	0.06886 (0.0107)	4.3×10^{-7} (5.9×10^{-8})
x_5	0.0708 (0.0377)	6.6×10^{-7} (2.3×10^{-7})	0.0519 (0.0066)	1.6×10^{-6} (1.9×10^{-7})
x_6	0.0508 (0.0138)	3.2×10^{-7} (8.7×10^{-8})	0.0709 (0.0115)	5.2×10^{-7} (8.3×10^{-8})

where the squared bias component is given by:

$$\text{Bias}^2(\beta_j) = \left(\frac{\sum_{k=1}^N \hat{\beta}_{kj}}{N} - \beta_j \right)^2 \quad (4.31)$$

Table 4.3 summarizes the results of this study. We focus here on the amount of bias induced by the hidden presence of two-factor interactions and the role of the design in making the main effects estimation robust to such secondary terms. In the case where the interactions are identically zero, the two sets of designs behave, as expected, in that they both provide unbiased estimates of the main effects (results not reported here). Consider a case in which the true model has one non-zero interaction effect. We should expect here a substantially better performance of the NLEFD relative to the D-optimal design. We can easily see, from the squared bias in Table 4.3, that the NLEFD performs significantly better, in fact providing essentially unbiased estimators of the main effects, while the use

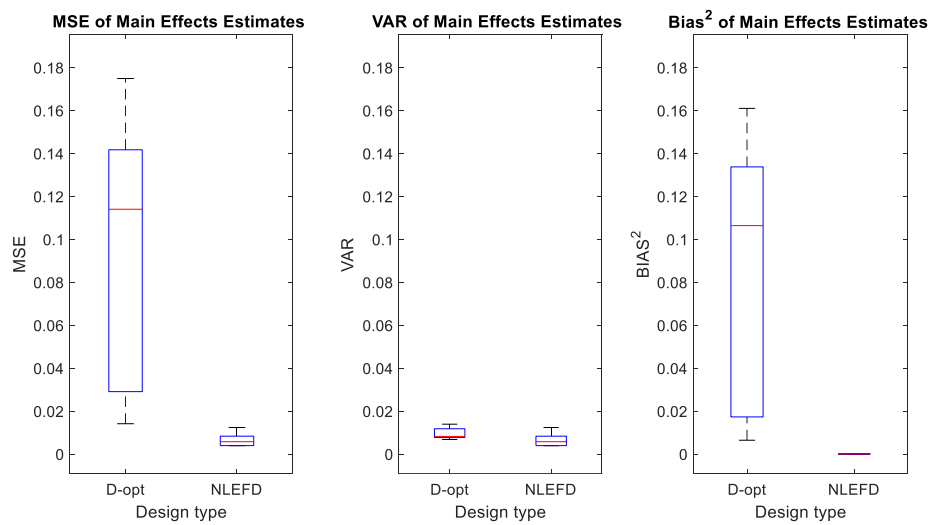


Figure 4.3: Simulated mean squared error, variance and squared bias resulting from the estimation of a main effect for a D-optimal design and an NLEFD when there is one non-zero two-factor interaction effect.

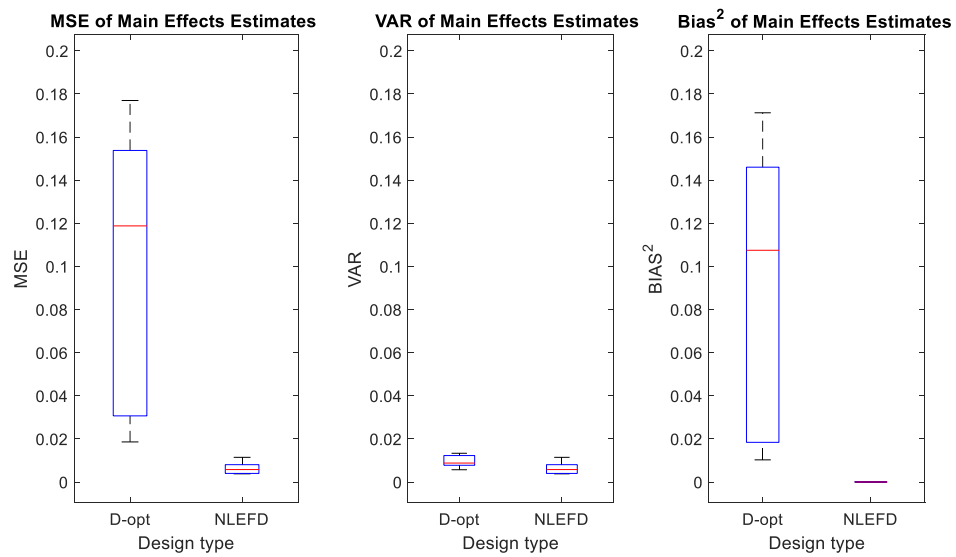


Figure 4.4: Simulated mean squared error, variance and squared bias resulting from the estimation of a main effect for a D-optimal design and an NLEFD when there are two non-zero two-factor interaction effects.

of a D-optimal design results in bias on all of the estimated coefficients. The boxplots in Figure 4.3 depict the mean square error and its decomposition into variance and bias components. The bias resulting when using the D-optimal design is the largest component in the significantly higher MSE of the D-optimal design in respect to NLEFD. Moreover, the simulation results in Figure 4.4 show that the performance improvement of the NLEFD relative to the D-optimal design is consistent as the number of the two-factor interaction effects increases from one to two. A similar effect (not shown here) arises when the sizes of such interaction effects increase.

4.6 Conclusions

In this paper, we have examined ways of constructing main-effects designs for logistic regression that are robust to the presence of one or more active two-factor interactions. We do so by constructing efficient designs that lead to high degrees of parameter orthogonality between main effects and two-factor interactions. In particular, we showed that when the prior estimates of the two-factor interaction coefficients are identically zero, foldover designs are guaranteed to achieve parameter orthogonality. This result motivated our development of non-linear efficient foldover designs (NLEFDs). In addition, we developed numerical methods for constructing efficient designs with high levels of parameter orthogonality for general cases of prior estimates and for Bayesian D-optimality.

In summary, our work has provided the following insights:

1. Under the assumption of $\beta_2^0 = \mathbf{0}$ for the two-factor interactions and intercept, a foldover design is parameter-orthogonal, regardless of the assumptions about the main effects that are of primary interest.
2. Our NLEFDs, which are constructed by forcing the foldover structure while maximizing the D-efficiency, frequently lead to minimal losses of D-efficiency, while achieving substantial gains in parameter-orthogonality.
3. Compound designs constructed by giving sufficient weight to the parameter-orthogonality criterion may achieve higher levels of D-efficiency than the NLEFD, while exhibiting negligible loss in parameter-orthogonality. In such cases the minimax design seems to be a reasonably good choice.
4. Our empirical results suggest that if a parameter-orthogonal design exists (which is true in almost all examined cases), then that design is a foldover design.

This paper is the first of its kind to utilize a measure of parameter-orthogonality for design construction with non-linear models. It is also the first to apply a compound optimization approach to non-linear design construction and to highlight the advantages of a foldover structure for design robustness to model specification in non-linear settings. We note that our results apply only to logistic regression models. Expanding this work to include Poisson, Gamma or other generalized linear models would be of interest. Also, we considered only the main effects model as the model of primary interest, and the interac-

tions terms to be of secondary interest. Examination of other primary/secondary model component pairings would be of further interest.

Chapter 5

Bidding Strategies for Exploration and Exploitation in Two-stage Procurement Auctions with Feedback

5.1 Introduction

In real-life business settings, experimentation is a fundamental tool for a vast variety of applications. From product/process improvement, to marketing strategies, from human resources management to pricing strategies, experimentation allows to investigate, understand and optimize critical relationships between variables of relevant interest. However, there is always a certain cost to experimentation and a tension between how much the acquisition of new or refined knowledge cost and how much such knowledge is desirable and beneficial to the business. For example, experimenting to improve a product performance requires time, resources and potential losses in the short term, but it may significantly improve the product characteristics and profits in the long term. In this work, we investigate the use and potential benefits of experimental design approaches applied to bidding strate-

gies. The context is that of a broker of transportation services who participates in two-round procurement auctions. In such auctions, the uncertainty related to competitors behavior and the operator's behavioral biases lead to sub-optimal bidding strategies, and experimentation can be used to mitigate such effects. In particular, carefully designed experiments can leverage the two-round mechanism to effectively and efficiently explore market prices equilibrium in the first round of the auction (exploration phase), and exploit the acquired or refined knowledge in the final pricing in the second round of the auction (exploitation phase).

The remainder of this chapter is organized as follows: Section 5.2 introduces the context of a large customer bid and the problem of cost consensus; Section 5.3 discusses related literature on the role of feedback and the behavioral implications of anticipated regret in bidding settings; Section 5.4 discusses our proposed design construction approach; and Section 5.5 explores implementation strategies.

5.2 Bidding Problem Definition

5.2.1 Large-Customer Two-stage Auction

A large customer (hereafter referred to as “she”) submits a request for prices (RFP) for a large set of lanes, typically on the order of tens or hundreds of lanes. A lane is a unique origin-destination pair on which a certain type of freight has to be moved over the course

of the year. The customer typically specifies the following information for each lane for which transportation service is requested: origin-destination pair, in terms of zip code or city name; mode of transportation, such as van or refrigerated, all full truckload; miles; expected yearly volume, if available, or an estimate of it. The broker, or focal bidder in our context (hereafter referred to as “he”), must submit a rate per mile (RPM) for each lane in the package or a portion of lanes. The auction is configured as a two-round, sealed-bid auction. After the first round, the customer provides a feedback about the ranking of the individual bids with respect to the other bidders. The broker does not know how many competitors are going to bid on the full package of bids, or perhaps part of it, but he has access to third-party-provided data on average RPM per lane of the past 12 months. In other words, he can predict the expected behavior of the market (his competitors) based on a point estimate of a market average rate and a prior belief on the distribution of the other bidders in the market. Feedback received after the first round is expected to reduce uncertainty about the competitors positions by updating the beliefs on their distribution.

In devising a bidding strategy, the broker takes into account cost uncertainty that relates to the fact that only after acquiring business on an awarded lane can he search for and contract for carrier services to move the awarded freight. For cost estimation the broker can rely on internal, historical cost-distribution data for individual lanes. Similarly, historical values of his own profit for each lane is readily available.

It is well established that the broker is a dominant player in the market and can count

on reputation and scale advantage over his competitors. In other words, reputation, incumbency and location related attributes are considered exogenous. The broker's preference for certain lanes is related to his historical presence on a given lane or route. A complex set of considerations makes a lane more or less desirable from the broker's perspective, and likely they can be simply summarized by a measure of desirability. Such a measure will depend on a variety of factors such as: the density of existing relationships with carriers; total volume already moving between the origin and destination, or cumulative head haul to the same destination and cumulative back haul from the same origin; probability of recovering from a bounce load; and the diversity of available carriers.

For reasons of supplier diversity policy and risk mitigation of the customer, it is also well known to the broker that no more than some percentage, say 20%, of the lanes can be awarded to him, no matter the bidding strategies and prices submitted. So a strategy that aims at winning the most or all lanes is not applicable.

After a first round of the auction, the customer provides feedback to all bidders. Feedback includes a ranking of the bidder's bid, then an opportunity for re-submission of a RPM is given. The bids submitted in the second round can be only less than or equal to those submitted in the first round. After a second submission of the adjusted bids, the auction is terminated by the customer final awards allocation.

The typical feedback is a simple green, yellow, or red light, which represents respectively a competitive, a not-as-competitive, and a non-competitive bid. In other words, the

feedback is not directly actionable, but it gives an information on the need to adjust the bid downward in the second round to increase the probability of winning the lane.

Consider the costs and benefit of employing a certain pricing strategy in this two-round auction scenario. The feedback received after the first round of the auction gives a valuable opportunity to learn about the competitors' rates distribution, however, the cost of learning such information is given by the bound on the bid that is imposed by the first submission (because in the second submission it can be only be decreased or maintained). How much does the bidder value such information? It clearly depends on the contingencies of the specific lanes.

When a lane is very desirable, the information is of more value and the bidder is more willing to 'pay' for it, in the sense of giving up profit in order to win the lane. On the contrary, the less desirable lanes are those for which the value of the information is lower and so the bidder is more concerned with maintaining a high rate and learn less about the competitors. Intuitively then, the higher the desirability of a certain lane, the more the market should be explored in the first round, acquiring relevant information on the competitors via the feedback, and then exploiting it in the second round.

Our question is, how much should the bidder explore in the first round, by distributing margin points to estimate the market equilibrium, versus how much to safeguard for exploitation for the second round of the auction? In the first round of the auction, the main objective of the bidding strategy should be to learn the customer preferences and the com-

petitors strategies, then in the second round the objective becomes to exploit the position acquired in the most desirable lanes and in the overall utility maximization.

5.2.2 Placing a bid: the problem of cost consensus

For a given lane l_i , the operator placing a bid starts from considering two main components that will form his starting rate r_i :

$$r_i = c_i(1 + m_i) = c_i + \pi_i \quad (5.1)$$

where c_i is the cost estimate, m_i is a percent margin applied, π_i would be the profit per load. Before submitting the rate this starting rate will be multiplied by the mileage and expected loads, surcharged for fuel, and finally submitted as an all-in rate.

The cost component is of critical importance. This is a cost estimation based on the historical data and cost forecasting methods unique to the broker in question. This starting cost estimate is a point estimate that is individually calculated for the specific lane i . Existing internal prediction models can be used to identify this cost estimate and a certain level of uncertainty around it. The point estimate c_i can be an average cost, a 50th quantile, or any other desired quantile resulting from the internal predicting model. The lower the cost quantile the more risky it is to assume that value as an estimate of the actual cost that will realize later on the market.

The percent margin component of the starting rate is typical of the lane category, not of the lane itself. In a given lane category, an experienced pricing agent has a good starting guess of what would constitute a fair margin. The idea is that, should the costs be aligned among all the competitors in the auction, such fair margin should be the most likely margin applied by any bidder on lanes of such category. In other words, the rate resulting from applying this margin level to the initial cost estimate should be a competitive rate. Should any bidder apply a much higher margin for his bid, the resulting rate should become not competitive. This general understanding of the problem excludes the possibility of collusion and widespread and unjustified greed across all bidders.

All bidders are assumed to employ a similar cost-plus-margin approach to formulate their rates, and so their bids. If the fair margin approach was the implied strategy of every bidder, then a particularly greedy bidder, applying a too-high margin, would result in a non-competitive rate. Similarly, a particularly bargaining bidder, applying a too-low margin, would result in a better-than-competitive rate. Surely, some of the rate differences in the auction are determined by different margin strategies; however, it is very likely that most of the differences are determined by a different starting cost estimate, derived by different forecasting and prediction methods.

This aspect is a crucial feature of the problem of placing optimal bids. The feedback received on a given bid after the first-round submissions and the award received after the second-round submissions, depends, at least partially on the competitiveness of the rate

submitted (more variables are in reality considered by the customer, and this makes the reasoning on rate competitiveness noisy and only part of the bigger picture, which goes beyond the scope of this work). However, the rate submitted is the result of two separate components, cost and margin, that the customer has no way to discriminate.

As a very simple thought experiment, consider the two rates r_a and r_b submitted for a specific lane respectively by the two bidders identified by the letter subscripts a and b (thus the subscript here refers to the bidder, and the lane is the same). Each rate is the result of a similar process of adding a certain margin to a cost estimate, so that $r_a = c_a(1 + m_a)$ and $r_b = c_b(1 + m_b)$. Now, suppose that $r_a = r_b$. When looking at these rates from the perspective of the customer, they look the same. The two rates may very well be the result of equal cost estimates and equal margin strategies. However, it is also possible that the two cost estimates and margin strategies are different. If bidder a is trying to increase his chances of winning by lowering his margin, but he is applying this strategy to his cost estimates that happens to be higher than bidder b , then his margin strategy is lost in the confounding given by the different cost estimates, and in fact the two rates look the same in the eyes of the customer. Similarly, suppose that $r_a > r_b$. These rates suggest to the customer that bidder a may be applying a higher margin, and in any case, being a higher rate, the customer will prefer the bid of bidder b . This may very well be the case, that bidder a is applying a higher percent margin to the same cost estimate of bidder b ; but it can also be the case that the two bidders have different cost estimate approaches and they are, in truth, simply applying the

same margin, but their resulting rates are different because of the different cost estimate. Bidder b would get a competitive advantage here without it being explicitly his strategy to bid a lower rate.

If we consider that the full pool of bidders is comprised of several different cost estimation methods and margin strategies, we see that there are several combinations of reasons why the full set of rates for any given lane in any lane category may differ widely. The full set of rates submitted by all bidders in a given lane, will likely follow a mound-shaped distribution, centered around the competitive rate. For any bidder, placing a bid far above or far below that consensus rate is disadvantageous. On the one hand, placing a bid far above the other bids is clearly negative because it is very unlikely to receive positive feedback and to win the award. On the other hand, placing the bid far below the others can also be negative, because it can be seen as a very good bargain from the customer standpoint, and so it may receive a positive feedback and have higher chances to win the award, however it is going to give lower profits than necessary (a higher bid may have been just as competitive, or just as a bargain, without lowering the profits as much). Moreover, having an extremely lower rate than the consensus, without applying exceptionally low margins, may be a symptom of an excessively low cost estimate, which may turn harmful when the actual cost is realized.

Thus, in order to bid competitively and strategically, it would be beneficial to know where the general cost consensus lies. However, this type of information is clearly not an

easy one to obtain. The two-round auction mechanism, with feedback between rounds, allows a unique opportunity to learn more about the cost dynamic and strategic approach of the competitor bidders.

Suppose that the true underlying relationship between a margin point and the corresponding competitiveness of the bid was a deterministic step function:

$$f(x) = \begin{cases} 1, & \text{if } x < x_t. \\ 0, & \text{otherwise.} \end{cases} \quad (5.2)$$

where $f(x)$ is the function representing the relationship between the input x (in this context it can be a percent margin or a profit per load) and y (in this context it can be a feedback or award variable). In a deterministic context such a function would ideally be a step function as described in (5.2). In the real world of the auction context this function is more likely to resemble a logistic function.

In such a context, more than a switching point, it is of interest to find the value of the x (margin or profit per load) that is associated with an acceptable probability of the outcome variable. For example, if looking at the outcome of the first round of bidding in terms of green-yellow-red feedback, the x of interest, x_t , may be the value of percent margin that corresponds to a certain probability of obtaining green feedback. If looking at the outcome in terms of award of the second round, the x of interest, x_t may be the value

of percent margin that corresponds to a certain probability of obtaining an award. The two types of outcome may also be easily related if there is confidence in prior knowledge on the conversion rate. For example, if it is known that to obtain a 20% award, the amount of green feedback must be at least 30%, then the two outcomes could be used interchangeably accounting for the conversion rate.

5.3 Related literature

5.3.1 The Role of Feedback between Auction Stages

Haruvy and Jap (2013) investigate bidders behavior in a procurement auction settings in which the auction format allows for subsequent bidding over a certain bidding time frame and the bidders can observe at any moment the other submitted bids, even though they cannot observe the other bidders identity. The phenomenon that these authors highlight is that some bidders bid higher than the lowest current bid. In this setting the authors claim that the phenomenon is driven by a differential bidders quality and their ability to update their prior beliefs on the competitors' quality by observing their submitted bids. In this study, the authors find that high quality bidders bid more aggressively when bidding against what they perceive as other higher-quality competition, and less aggressively when bidding against potentially lower-quality competition.

Similarly to Haruvy and Jap (2013), we are concerned about the role of non-price re-

lated attributes that affect the auction dynamics and outcome. However, contrary to the multi-attribute case (Chen-Ritzo et al., 2005; Adomavicius et al., 2012), these non-price-related attributes are not explicitly stated and cannot formally be taken into account in devising a bidding strategy. Moreover, even if our auction mechanism is in rounds, and not continuous, and the bids of competitor bidders are not observable, the feedback mechanism between rounds allows the focal bidder to learn about the competitors positions, at least to the extent that the feedback details allow. In that sense, the focal bidder can use the feedback information as a strategic tool for learning and update his prior beliefs about his competition.

Armantier and Treich (2009) investigate biased probabilistic beliefs as a major driver for overbidding, and feedback as a significant mechanism for improving probability estimation and reduce overbidding. They show experimentally that subjects underestimate their probability of winning an auction, and overbid as a consequence of such biased belief. Then, when provided with a feedback on the precision of their predictions, subjects learn to make better predictions, and to curb overbidding significantly. The authors show that biased probabilistic beliefs are a driver for overbidding, and play a much more significant role than risk aversion. Other authors Neri (2015); Manski and Neri (2013) investigate the elicitation of subjective beliefs on continuous probability distributions and their effect on the choices made for pricing strategies in auctions. They also highlight heterogeneity in bidders beliefs.

The number of bidders affects the optimal bidding strategy in that it affects the Nash equilibrium and the behavior of the player (Füllbrunn and Neugebauer, 2013). However, the studies on this aspect of the auction usually assume the knowledge of the players about how many other bidders there are in the game. This is not the case in our two-round auction.

5.3.2 Regret Theory

A fully rational bidder should bid in order to maximize his own profit, given some cost evaluation. Upon winning, the realized profit would be the difference between revenues and costs. From a buyer's perspective, in the most simple procurement auction mechanism, first price sealed bid (FPSB), the lowest bid should win. In reality, on the one hand, the bidder places a bid according to not only their monetary evaluation of the lane auctioned, but also the strategic desirability of that lane. For example, it may be strategically necessary to be present on a lane even at a loss, for economies of scope. On the other hand, the shipper (customer, or auctioneer) usually does not award lanes based on a simple FPSB calculation. Other aspects to consider are diversification of their suppliers, existence of an established relationship with a carrier/broker, reputation, past service level.

According to regret theory, in FPSB auction, bidders tend to deviate from the rational Nash equilibrium bid for reasons that can not be explained by risk aversion. Engelbrecht-Wiggans (1989) proposed regret as an explanation of such non-rational bidding behavior. Specifically, upon winning, the winner's regret arises from the difference between the win-

ning bid and the second best, resulting in some 'money left of the table' for the winner. For the loser, regret arises when they miss the opportunity of winning at a favorable price, such as when the winning bid is above their cost evaluation. Both types of regret have been studied and empirically tested in Engelbrecht-Wiggans and Katok (2007, 2008, 2009). Regret has more influence on the bids when feedback information makes it more salient.

The above mentioned regret types are a behavioral effect related to the monetary value (lost profit) involved in the bidding. In literature, regret is presented at the level of the individual decision maker, however we can use the concepts of winner's and loser's regret at the strategic organizational level. Notice that we are not looking for using regret as a mechanism to explain underbidding or overbidding as in the past behavioral literature. We use these lenses to explain that the operator may bid sub-optimally because of an implicit utility function in which there is tension between profit seeking and probability of winning maximization. These two objectives can be weighted by calibrating the aggressiveness of the price discount strategy. This, in turn, depends on the lane desirability.

We can argue that, in addition to the expected profit, the anticipated regret is related to potentially lost profit increments, but also the lanes strategic value.

Moreover, knowing, or rather estimating, what non-monetary attributes are important for the shipper can influence the salience of such regrets. For example, knowing that a customer is fully price oriented, may drive the bidder to forgo the potential loss of money left on the table (related to winner's regret) in order to make sure he places the lowest

possible bid that may land the win. On the other end, for a service oriented customer that values a premium on high reliability and service level, the bidder may be more inclined to try to extract as much profit as possible, believing that a higher bid would be rewarded by the customer if associated with high service level.

We propose to apply experimental design plans to the decision making process of placing bids in the two-round auction. We argue that carefully designing an experiment to select margin levels and the proportion of lanes to bid at those margin level, can incorporate the double objective of estimating market behavior within a lane category in the first round of the auction, and maximize the expected profit resulting from price adjustments in the second round of the auction. With this approach to bidding, the operator can avoid behavioral biases due to anticipated regret, while still leveraging prior beliefs in the objective function used for design construction, as it discussed in the next section.

5.4 Constructing Bidding Designs

In this section, we investigate the construction of efficient bidding experiments. The goal is to define a method to construct a designed bidding experiment for the problem of bidding in two rounds auctions with uncertainty about the cost consensus among bidding competitors.

In the following subsections we introduce the notation needed for describing our design construction approach, followed by the description of a compound optimization criterion, its construction algorithm, and a numerical example.

In order to make strategic choices in terms of bidding design, we consider two main purposes driving the design construction: the first is the ability to estimate the true functional relationship between control variables (factors) that we can manipulate and outcome variables resulting from the auction; the second is profit maximization. In particular we can consider as control variables the margin levels or the intended profit per load in any given bid.

5.4.1 Design notation

Consider the classical settings of a logistic regression model in which the mean binary response Y is distributed according to a logistic model with logit link, for which the probability mass function can be written as follows:

$$p_i(\mathbf{x}_i, \boldsymbol{\beta}) = \frac{e^{\mathbf{f}^T(\mathbf{x}_i)\boldsymbol{\beta}}}{1 + e^{\mathbf{f}^T(\mathbf{x}_i)\boldsymbol{\beta}}} = \frac{1}{1 + e^{-\mathbf{f}^T(\mathbf{x}_i)\boldsymbol{\beta}}}, \quad (5.3)$$

where \mathbf{x}_i is the $n \times 1$ vector of factor settings for the i -th run, $\mathbf{f}^T(\mathbf{x}_i)$ is the $1 \times p$ row vector of model terms for the i -th run of the experiment, and $\boldsymbol{\beta}$ is the $p \times 1$ vector of unknown parameters.

In the context of the bidding experiment there are multiple factors to take into account as factors or covariates. For simplicity, we will construct designs for which we consider one control factor X in a polynomial model with four parameters. For the third order model:

$$\mathbf{f}^T(\mathbf{x}_i)\boldsymbol{\beta} = \beta_0 + \sum_{j=1}^p \beta_j x_i^j \quad i = 1, \dots, n. \quad (5.4)$$

where $p = 3$.

The other covariates to consider would be the lane and load characteristics. These features are basically used to create categories of lanes with similar cost dynamics, as described in a previous section. For this reason, we will consider that our designs will be implemented within a category, so that those characteristics are the same for every lane in the category. Moreover, in our design construction we will discuss the choices related to the percent margins to apply to the lanes in a category, so that the starting cost estimate that is unique to every lane stays outside of this picture. Were we to reason on rates, the unique cost estimates would have confounded our design choices. Thus we will consider our design problem as the problem of choosing, within a lane category, the percent margin levels x_i and the percent of lanes in that category to bid at those margin levels, l_i .

In experimental design jargon, this problem is the construction of an approximate design (Fedorov, 1972).

Let \mathbf{d} denote the $n \times 2$ design matrix

$$\mathbf{d} = \begin{bmatrix} x_1 & l_1 \\ \vdots & \vdots \\ x_n & l_n \end{bmatrix}. \quad (5.5)$$

where $\sum_{i=1}^n l_i = 1$. For a design \mathbf{d} , the Fisher information matrix corresponding to a model \mathbf{f} is:

$$\mathbf{I}(\mathbf{d}, \boldsymbol{\beta}) = \mathbf{X}^T \mathbf{W}(\mathbf{d}, \boldsymbol{\beta}) \mathbf{X}, \quad (5.6)$$

where $\mathbf{W}(\mathbf{d}, \boldsymbol{\beta}) = \text{diag}[\mathbf{P}(\mathbf{d}, \boldsymbol{\beta}) - \mathbf{p}(\mathbf{d}, \boldsymbol{\beta})\mathbf{p}^T(\mathbf{d}, \boldsymbol{\beta})]$, with $\mathbf{P}(\mathbf{d}, \boldsymbol{\beta}) = \text{diag}(p_1, \dots, p_n)$, $\mathbf{p}(\mathbf{d}, \boldsymbol{\beta}) = (p_1, \dots, p_n)^T$, and $p_i = p(\mathbf{x}_i, \boldsymbol{\beta})$ is as given in (5.3). For a fixed $\boldsymbol{\beta}$, a design \mathbf{d}^* is D-optimal if it maximizes the logarithm of the determinant of the information matrix in (5.6):

$$\mathbf{d}^* = \underset{\mathbf{d}}{\text{argmax}} \log(|\mathbf{I}(\mathbf{d}, \boldsymbol{\beta})|). \quad (5.7)$$

As noted above, since $\boldsymbol{\beta}$ is unknown, the information matrix in (5.6) cannot be used as a direct measure of information provided by the design, unless assumptions are made regarding the parameters in the model. Let $\boldsymbol{\beta}^0$ denote a prior estimate of the parameter vector that can be used to compute the information matrix in (5.6).

The design criterion in (5.7), like most design criteria for non-linear regression models, depends on the model's unknown coefficients, which means we must make assumptions about the values of such coefficients in order to compute the criterion for a given design. As discussed in Chapter 4, several approaches have been proposed in the literature to face this conundrum and each approach has its advantages and disadvantages. *Locally optimal designs* can be constructed employing a prior guess of the values of the parameters. This approach is simple, but it can be inefficient if the true parameters differ appreciably from the prior values. The other approach is that of *Bayesian optimal designs*, which accounts for the uncertainty of the parameters by integrating the design criterion over a prior distribution of such parameters. This approach can be computationally cumbersome, but it constructs designs that may be more efficient for a larger space of potential model parameter values.

5.4.2 Design construction

We implement a compound optimization approach Cook and Wong (1994) that considers both the D-efficiency of the third-order model and a measure of profit efficiency. We explore the tradeoff between design efficiency and profitability in order to select the best compound design for the purpose of this paper.

Notice that henceforth we refer to the D-optimal design as the design that is constructed by the sole criterion of D-efficiency. Such a design may not be unique, and it is dependent on the prior knowledge of the coefficients. We will explore the degree of such dependency.

Our compound optimization approach is similar to that given by Jones and Nachtsheim (2011b) and Errore et al. (2017a). Let \mathbf{D} denote a D-optimal design. Then, the D-efficiency of a design \mathbf{d} is given by

$$\phi_D(\mathbf{d}, \boldsymbol{\beta}) = \left[\frac{|\mathbf{X}'_1(\mathbf{d}, \boldsymbol{\beta})\mathbf{W}\mathbf{X}_1(\mathbf{d}, \boldsymbol{\beta})|}{|\mathbf{X}'_1(\mathbf{D}, \boldsymbol{\beta})\mathbf{W}\mathbf{X}_1(\mathbf{D}, \boldsymbol{\beta})|} \right]^{1/p}. \quad (5.8)$$

For $0 \leq \alpha \leq 1$, let:

$$\phi_\alpha(\mathbf{d}, \boldsymbol{\beta}) = (1 - \alpha)\phi_D(\mathbf{d}, \boldsymbol{\beta}) + \alpha\phi_P(\mathbf{d}, x_t^0), \quad (5.9)$$

where $\phi_P(\mathbf{d}, x_t^0)$ is a scaled measure of the degree of profitability, which we define as follows. Let x_t be the initial guess on the margin level that converts to the desired percent outcome. To compute a measure of profitability notice that any bid placed at any given margin level x_t has a profitability proportional to that percent margin. The total profit given by any set of bids, hence any design \mathbf{d} , is bound to be proportional to the sum of the margins multiplied by the fraction of lanes bid at those margin levels. However, one must consider that margins above the desired conversion threshold would not result in sufficient profit, and they should be somewhat penalized in this profit function. On the other hand, the two round dynamic allows to estimate the conversion function (the logistic regression model of main interest) and eventually to re-allocate the lanes initially placed at too high margins to

the maximum margin of desired conversion. This margin may be estimated with the results of the first round submissions and feedback, and it can be the same as the initial guess or be quite different. Thus, with a conservative approach, one can compute the profitability value of the first round design as if any bid above the guessed threshold margin was bid at the threshold margin instead.

In other words, we compute the profit value of a design \mathbf{d} as follows:

$$\phi_P(\mathbf{d}, x_t^0) = \sum_i x_i l_i + x_t^0 \sum_j l_j \quad i = 1, \dots, t-1, \quad j = t, \dots, n. \quad (5.10)$$

where the design points lower than the threshold margin x_t^0 are valued proportionally to their weights l_i , and the cumulative weights of the design points above the threshold margin are valued as if they were equal to the threshold margin. This approach to compute the profit criterion could be conceived in other ways. The choice here is quite conservative but it carries the meaning of the two-round auction. The idea is that when, after the first round, the threshold margin is estimated, the lanes that were bid above that level can be lowered to such margin and hence have the potential profit of that level.

The scaled value of this criterion is computed in relative terms to the maximum value, given by placing the whole weight on the threshold margin x_t^0 .

Clearly, a design with $\phi_P(\mathbf{d}, x_t^0) = x_t^0$ would be a design of this kind, with only one design point for maximum value of the profit criterion. Such design would not be a good

design in terms of estimability, or D-efficiency. Thus we will limit our exploration of the compound designs to designs that give at most an α weight near to but not equal to 1.

Our compound construction algorithm thus performs the following steps:

1. Construct the D-optimal design. This is also the design that maximizes (5.9) when $\alpha = 0$.
2. For evenly spaced values of $\alpha \in (0, 1)$, construct a design that maximizes (5.9). We refer to these designs as compound-optimal designs.
3. Omit the construction of the design that maximizes the profit criterion for the reason explained above, with $\alpha = 1$.

To apply this approach, designs can be computed using a prior based on the internal predictive model that estimates the relationship between margin and percent award. With marginal effects of margin on percent award we can fit a logistic regression model with a constant and linear term in the predictor variable margin. This coefficients can be used as priors to construct locally optimal designs.

By construction, the compound objective function, will shift weights from the D-optimal design points to the threshold point; the more so as α increases.

5.4.3 Numerical example

To illustrate our proposed design construction method, consider the following numerical example.

Historical data and predictive models can be used to estimate the functional relationship between a response of interest and a factor under study. This information can be used to formulate initial beliefs about the threshold margin and the parameters to input into the design construction algorithm. For example, within a lane category with certain common load and lane characteristics, a model that represents the predicted percent of award (*%Award*) for a given margin (*Margin*) can be fit using a logistic regression model.

Using one such set of data, we estimated values of the parameters of the logistic regression model that can be used as prior values to construct a locally D-optimal design. Appropriately scaling both the dependent and independent variable, we obtained a significant fit of a logistic regression model with one linear term in *Margin* and an intercept; consequently we used the estimated values of such parameters as priors for β_0 and β_1 , respectively $\beta_0^0 = 3.893$ and $\beta_1^0 = -9.265$, and we kept the two other parameters' priors equal to zero, $\beta_2^0 = \beta_3^0 = 0$. Moreover, we use this data to identify a threshold margin that corresponds to a desirable conversion rate. For example, within the specific lane category, a margin of $x_t = .1032$ is associated with an award percent conversion of 20%, a margin of $x_t = .0633$ is associated with an award percent conversion of 30%, and a margin of $x_t = .0144$ is associated with an award percent conversion of 40% (these values will be

Table 5.1: D-optimal design

x_i	l_i
-0.15	0.25
0.28	0.25
0.56	0.25
1	0.25

transformed in the scaled axis before computing the designs). Thus, for example a margin of $x_i^0 = .063$ (re-scaled as 0.3763 for the axis transformation) can be used as a prior for the threshold margin to refer to in the profit optimization criterion in (5.10).

A locally D-optimal approximate design constructed with (5.9) and $\alpha = 0$ identifies the optimal design points listed in Table 5.1, and each point is assigned a weight of 0.25. This design has a profit efficiency of $\phi_p^s(\mathbf{d}, x_i^0) = .88$, which is an 88% efficiency in respect to a design that puts all the weight on the threshold margin x_i^0 .

Computing a sequence of compound designs with the optimization criterion (5.9) for α increments of .1, we obtain the efficiency trace in Figure 5.1. It is apparent that a mini-max approach to choosing a compound design in this example would select the compound design constructed with $\alpha = .7$. This design is shown in Table 5.2, and it has a 94.34% D-efficiency and a 94% profit-efficiency.

As expected, the compound design distributes weights on the design points in a manner that favors the margin points greater than or equal to the threshold margin, while keeping the four design points necessary for the estimation of the logistic regression model of the

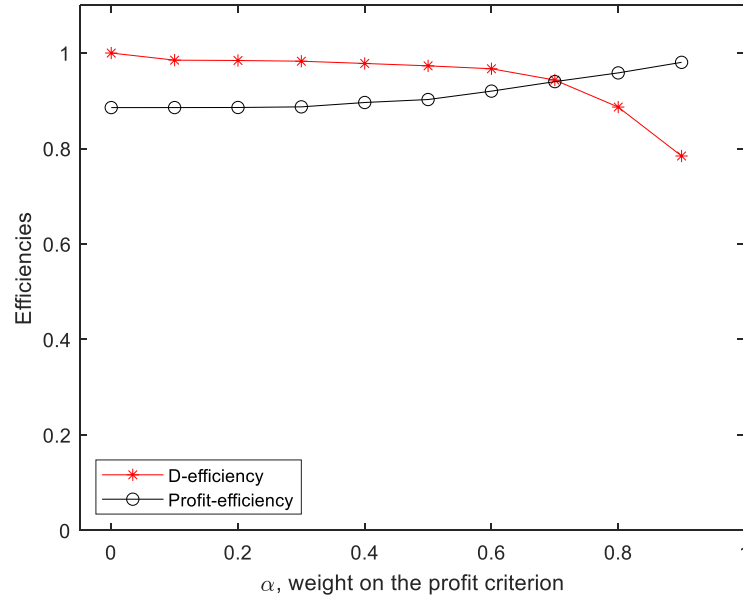


Figure 5.1: Efficiency trace: D-efficiency and Profit-efficiency vs α in (5.9)

third order. Incorporating the profit criterion in the design construction through the compound criterion (5.9) and the appropriate α weight on the profit component impacts the distribution of the lanes on the margin points. The percentage of lanes bid at higher margin levels are available for exploitation via decreasing margins in the second round, while the percentage of lanes bid at lower margin levels are “sacrificed” for exploration, in the sense of being used for the purpose of estimating the parameters of the functional relationship between margins and awards.

Table 5.2: Compound design with $\alpha = .7$

x_i	l_i
-0.06	0.15
0.33	0.27
0.58	0.28
1	0.30

5.5 Implementation strategies

To implement our bidding experiment, the following general procedure should be applied. For a given large customer bid package, the first task is to cluster the lanes of the package into categories that group together lanes with similar cost dynamics. Within a lane category, the steps to follow are:

1. Identify a reasonable margin range that can be considered as the design space for the variable X to use in the construction algorithm. This step can be performed with a combination of historical data and expert knowledge of the pricing operator.
2. Apply the internal predictive model that uses historical data to predict the expected outcome (say the *PercentAward*) associated with the margins within the identified range.
3. Fit the predicted outcome values versus the percent margin with a logistic regression model in two parameters (a constant and a linear term in X) in order to obtain a

prior on the first two coefficients of the model that will be assumed in the design construction algorithm.

4. Select a desired conversion rate in the predicted model and identify the corresponding threshold margin to be used as a profit optimization reference for the profit criterion of the design construction routine.
5. Use the construction algorithm for compound designs described in the previous section and identify the minimax design that best accomplishes the multi-objective of estimation and profit maximization.
6. Use the selected design, margins and weights, to price the group of lanes of the selected category. Notice that the individual lane should be priced considering a cost base component that is unique to the lane, as in (5.1). To this base cost, the margin applied should be chosen according to the created design. Assignment of lanes to margin levels should be done randomly.
7. After feedback, the realized outcome should be used to fit a model with up to the third order term. Such model can be used to identify post hoc the threshold margin hypothesized initially.
8. For the second round of submissions, with the objective to maximize the probability of conversion in awards, re-position all the lanes initially priced at margins higher than the new estimated threshold at that margin level.

The construction method and implementation strategy discussed here is under further refinement and it is the focus of future research to be completed in the near future with the goal of publication of this work.

References

- Abdelbasit, K. M. and Plackett, R. L. (1983). Experimental design for binary data. *Journal of the American Statistical Association*, 78(381):90–98.
- Adomavicius, G., Gupta, A., and Sanyal, P. (2012). Effect of information feedback on the outcomes and dynamics of multisourcing multiattribute procurement auctions. *Journal of Management Information Systems*, 28(4):199–230.
- Ai, M., Xu, X., and Wu, C. F. J. (2010). Optimal blocking and foldover plans for regular two-level designs. *Statistica Sinica*, 20(1):183–207.
- Akaike, H. (1992). Information theory and an extension of the maximum likelihood principle. Technical report.
- Armantier, O. and Treich, N. (2009). Subjective probabilities in games: An application to the overbidding puzzle. *International Economic Review*, 50(4):1079–1102.
- Beattie, S. D., Fong, D. K., and Lin, D. K. (2002). A two-stage Bayesian model selection strategy for supersaturated designs. *Technometrics*, 44(1):55–63.
- Box, G. E. P. and Lucas, H. L. (1959). Design of experiments in non-linear situations. *Biometrika*, 46(1-2):77–90.
- Box, G. E. P. and Meyer, R. D. (1986). An analysis for unreplicated fractional factorials. *Technometrics*, 28(1):11–18.

- Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n . *Annals of Statistics*, 35(6):2313–2351.
- Chaloner, K. (1993). A note on optimal Bayesian design for nonlinear problems. *Journal of Statistical Planning and Inference*, 37:229.
- Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21:191–208.
- Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304.
- Chen-Ritzo, C. H., Harrison, T. P., Kwasnica, A. M., and Thomas, D. J. (2005). Better, faster, cheaper: An experimental analysis of a multiattribute reverse auction mechanism with restricted information feedback. *Management Science*, 51(12):1753–1762.
- Chernoff, H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, 24(4):586–602.
- Chipman, H., Hamada, M., and Wu, C. F. (1997). A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics*, 39(4):372–381.
- Choi, N. H., Li, W., and Zhu, J. (2010). Variable selection with the strong heredity constraint and its oracle property. *Journal of the American Statistical Association*, 105(489):354–364.
- Cook, R. D. and Wong, W. K. (1994). On the Equivalence of Constrained and Compound Optimal Designs. *Journal of the American Statistical Association*, 89(426):687–692.
- Cox, D. R. (1988). A note on design when response has an exponential family distribution. *Biometrika*, 75(1):161–164.

- Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 49(1):1–18.
- Dette, H., Melas, V. B., and Pepelyshev, A. (2004). Optimal designs for a class of nonlinear regression models. *Annals of Statistics*, 32(5):2142–2167.
- Dette, H. and Sahm, M. (1998). Minimax optimal designs in nonlinear regression models. *Statistica Sinica*, 8(4):1249–1264.
- Diamond, N. T. (1991). The use of a class of foldover designs as search designs. *Austral. J. Statist.*, 33(2):159–166.
- Diamond, N. T. (1995). Some properties of a foldover design. *Australian Journal of Statistics*, 37(3):345–352.
- Draguljić, D., Woods, D. C., Dean, A. M., Lewis, S. M., and Vine, A. J. E. (2014). Screening strategies in the presence of interactions. *Technometrics*, 56(1):1.
- DuMouchel, W. and Jones, B. (1994). A simple bayesian modification of d-optimal designs to reduce dependence on an assumed model. *Technometrics*, 36(1):37–47.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2):407–499.
- Elsawah, A. M. and Fang, K. T. (2019). A catalog of optimal foldover plans for constructing U-uniform minimum aberration four-level combined designs. *Journal of Applied Statistics*, 46(7):1288–1322.
- Elsawah, A. M. and Qin, H. (2015). A new strategy for optimal foldover two-level designs. *Statistics and Probability Letters*, 103:116–126.

- Engelbrecht-Wiggans, R. (1989). The effect of regret on optimal bidding in auctions. *35(6):685–692*.
- Engelbrecht-Wiggans, R. and Katok, E. (2007). Regret in auctions: Theory and evidence. *Economic Theory*, 33(1):81–101.
- Engelbrecht-Wiggans, R. and Katok, E. (2008). Regret and feedback information in first-price sealed-bid auctions. *Management Science*, 54(4):808–819.
- Engelbrecht-Wiggans, R. and Katok, E. (2009). A direct test of risk aversion and regret in first price sealed-bid auctions. *Decision Analysis*, 6(2):75–86.
- Errore, A., Jones, B., Li, W., and Nachtsheim, C. J. (2017a). Benefits and fast construction of efficient two-level foldover designs. *Technometrics*, 59(1):48–57.
- Errore, A., Jones, B., Li, W., and Nachtsheim, C. J. (2017b). Using definitive screening designs to identify active first-and second-order factor effects. *Journal of Quality Technology*, 49(3):244–264.
- Fedorov, V. V. (1972). *Theory of optimal experiments*. Academic Press New York and London.
- Ford, I., Titterington, D. M., and Kitsos, C. P. (1989). Recent advances in nonlinear experimental design. *Technometrics*, 31(1):49–60.
- Füllbrunn, S. and Neugebauer, T. (2013). Varying the number of bidders in the first-price sealed-bid auction: Experimental evidence for the one-shot game. *Theory and Decision*, 75(3):421–447.
- Gotwalt, C. M. (2010). Addendum to "Fast computation of designs robust to parameter uncertainty for nonlinear settings". *Technometrics*, 52(1):137.

- Gotwalt, C. M., Jones, B. A., and Steinberg, D. M. (2009). Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics*, 51(1):88–95.
- Haines, L. M., Kabera, G. M., and Ndlovu, P. (2018). D-optimal designs for the two-variable binary logistic regression model with interaction. *Journal of Statistical Planning and Inference*, 193:136–150.
- Haruvy, E. and Jap, S. D. (2013). Differentiated bidders and bidding behavior in procurement auctions. *Journal of Marketing Research*, 50(2):241–258.
- Heise, M. A. and Myers, R. H. (1996). Optimal Designs for Bivariate Logistic Regression. 52(2):613–624.
- Holcomb, D. R., Montgomery, D. C., and Carlyle, W. M. (2003). Analysis of supersaturated designs. *Journal of Quality Technology*, 35(1):13–27.
- Holland-Letz, T. (2017). On the combination of c- and D-optimal designs: General approaches and applications in dose–response studies. *Biometrics*, 73(1):206–213.
- Hotelling, H. (1944). Some improvements in weighing and other experimental techniques. *Source: The Annals of Mathematical Statistics*, 15(3):297–306.
- Jones, B. (2013). Comment: Enhancing the search for compromise designs. *Technometrics*, 55(3):278–280.
- Jones, B. and Nachtsheim, C. J. (2011a). A class of three-level designs for definitive screening in the presence of second-order effects. *Journal of Quality Technology*, 43(1):1–15.
- Jones, B. and Nachtsheim, C. J. (2011b). Efficient designs with minimal aliasing. *Technometrics*, 53(1):62–71.
- Jones, B. and Nachtsheim, C. J. (2013). Definitive screening designs with added two-level categorical factors. *Journal of Quality Technology*, 45(2):121–129.

- Jones, B. and Nachtsheim, C. J. (2016). Blocking schemes for definitive screening designs. *Technometrics*, 58(1):74–83.
- Kabera, M. G., Haines, L. M., and Ndlovu, P. (2012). A note on the construction of locally D- and Ds-optimal designs for the binary logistic model with several explanatory variables. *Statistics and Probability Letters*, 82(5):865–870.
- Khuri, A. I., Mukherjee, B., Sinha, B. K., and Ghosh, M. (2006). Design issues for generalized linear models: A review. *Statistical Science*, 21(3):376–399.
- Li, G. and Majumdar, D. (2008). D-optimal designs for logistic models with three and four parameters. *Journal of Statistical Planning and Inference*, 138(7):1950–1959.
- Li, R. and Lin, D. K. J. (2002). Data analysis in supersaturated designs. *Statistics & Probability Letters*, 59:135–144.
- Li, W., Nachtsheim, C. J., Wang, K., Reul, R., and Albrecht, M. (2013). Conjoint analysis and discrete choice experiments for quality improvement. *Journal of Quality Technology*, 45(1):74–99.
- Li, W. and Zhu, J. (2014). Comment: Model selection with strong and weak heredity constraints. *Technometrics*, 56(1):21–22.
- Li, X., Sudarsanam, N., and Frey, D. D. (2006). Regularities in data from factorial experiments. *Complexity*, 11(5):32–45.
- Lin, C. D., Miller, A., and Sitter, R. R. (2008). Folded over non-orthogonal designs. *Journal of Statistical Planning and Inference*, 138(10):3107–3124.
- Lin, D. K. J. (1993). A new class of supersaturated designs. *Technometrics*, 35(1):28–31.
- Lu, X. and Wu, X. (2004). A strategy of searching active factors in supersaturated screening experiments. *Journal of Quality Technology*, 36(4):392–399.

- Manski, C. F. and Neri, C. (2013). First- and second-order subjective expectations in strategic decision-making: Experimental evidence. *Games and Economic Behavior*, 81(1):232–254.
- Margolin, B. H. (1969). Results on factorial designs of resolution IV for the $2n$ and $2n3m$ series. *Technometrics*, 11(3):431–444.
- Marley, C. J. and Woods, D. C. (2010). A comparison of design and model selection methods for supersaturated experiments. *Computational Statistics and Data Analysis*, 54(12):3158–3167.
- Mathew, T. and Kumar Sinha, B. (2001). Optimal designs for binary data under logistic regression. *Journal of Statistical Planning and Inference*, 93:295–307.
- McLeod, R. G. and Brewster, J. F. (2008). Optimal foldover plans for two-level fractional factorial split-plot designs. *Journal of Quality Technology*, 40(2):227–240.
- Meyer, R. K. and Nachtsheim, C. J. (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, 37(1):60–69.
- Miller, A. (2002). *Subset selection in regression*. Boca Raton, FL: CRC Press.
- Miller, A. and Sitter, R. R. (2001). Using the folded-over 12-run Plackett-Burman design to consider interactions. *Technometrics*, 43(1):44–55.
- Miller, A. and Sitter, R. R. (2005). Using folded-over nonorthogonal designs. *Technometrics*, 47(4):502–513.
- Monahan, J. and Genz, A. (1997). Spherical-Radial Integration Rules for Bayesian Computation. *Journal of the American Statistical Association*, 92(438):664–674.
- Neri, C. (2015). Eliciting beliefs in continuous-choice games: a double auction experiment. *Experimental Economics*, 18(4):569–608.

- Phoa, F. K., Pan, Y. H., and Xu, H. (2009). Analysis of supersaturated designs via the Dantzig selector. *Journal of Statistical Planning and Inference*, 139(7):2362–2372.
- Ponce de Leon, A. C. and Atkinson, A. C. (1991). Optimum experimental design for discriminating between two rival models in the presence of prior information. *Biometrika*, 78(3):601–608.
- Sebastiani, P. and Settimi, R. (1997). A note on D-optimal designs for a logistic regression model. *Journal of Statistical Planning and Inference*, 59(2):359–368.
- Sebastiani, P. and Settimi, R. (1998). First-order optimal designs for non-linear models. *Journal of Statistical Planning and Inference*, 74(1):177–192.
- Sitter, R. and Wu, C. F. J. (1993). Optimal designs for binary response experiments: Fieller, D, and A criteria. *Scandinavian Journal of Statistics*, 20(4):329–341.
- Sitter, R. R. and Torsney, B. (1995). Optimal designs for binary response experiments with two design variables. *Statistica Sinica*, 5(2):405–419.
- Street, D. J. and Burgess, L. (2004). Optimal and near-optimal pairs for the estimation of effects in 2-level choice experiments. *Journal of Statistical Planning and Inference*, 118:185–199.
- Sun, D. X., Li, W., and Ye, K. Q. (2008). Algorithmic construction of catalogs of non-isomorphic two-level orthogonal designs for economic run sizes. *Statistics and Applications*, 6:141–155.
- Webb, S. (1968). Non-orthogonal designs of even resolution. *Technometrics*, 10(2):291–299.
- Westfall, P. H., Young, S. S., and Lin, D. K. J. (1998). Forward selection error control in the analysis of supersaturated designs. *Statistica Sinica*, 8(1):101–117.

- Woods, D. C. and Lewis, S. M. (2011). Continuous optimal designs for generalized linear models under model uncertainty. *Journal of Statistical Theory and Practice* 2011 5:1, 5(1):137–145.
- Woods, D. C., Lewis, S. M., Eccleston, J. A., and Russell, K. G. (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics*, 48(2):284–292.
- Wu, C. F. J. (1993). Construction of supersaturated designs through partially aliased interactions. *Biometrika*, 80(3):661–670.
- Wu, C. F. J. and Hamada, M. J. (2008). *Experiments: Planning, Analysis, and Optimization*. New York NY Wiley.
- Xiao, L., Lin, D. K., and Bai, F. (2012). Constructing definitive screening designs using conference matrices. *Journal of Quality Technology*, 44(1):2–8.
- Yang, J. and Mandal, A. (2015). D-optimal factorial designs under generalized linear models. *Communications in Statistics: Simulation and Computation*, 44(9):2264–2277.
- Yang, M. and Huang, S. (2011). Optimal designs for generalized linear models with multiple design variables. *Statistica Sinica*, 21(3):1415–1430.
- Yang, P. and Li, W. (2019). Some properties of foldover designs with column permutations. *Metrika*, 82(6):705–717.
- Yu, J., Goos, P., and Vandebroek, M. (2008). Model-robust design of conjoint choice experiments. *Communications in Statistics: Simulation and Computation*, 37(8):1603–1621.
- Yuan, M., Joseph, V. R., and Lin, Y. (2007). An efficient variable selection approach for analyzing designed experiments. *Technometrics*, 49(4):430–439.

- Zhang, Q. Z., Zhang, R. C., and Liu, M. Q. (2007). A method for screening active effects in supersaturated designs. *Journal of Statistical Planning and Inference*, 137(6):2068–2079.