CONTRIBUTIONS TO QUALITY IMPROVEMENT METHODOLOGIES AND COMPUTER EXPERIMENTS

A Thesis
Presented to
The Academic Faculty

by

Matthias Hwai Yong Tan

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
August 2013

CONTRIBUTIONS TO QUALITY IMPROVEMENT METHODOLOGIES AND COMPUTER EXPERIMENTS

Approved by:

Dr. C.F. Jeff Wu, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Jianjun Shi
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Roshan Joseph Vengazhiyil
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Santanu Dey
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Godfried Augenbroe
College of Architecture
Georgia Institute of Technology

Date Approved: 11 March 2013

To mom and dad,

for always being there for me.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Professor C.F. Jeff Wu for his support, guidance, and advice that helped me immensely throughout my doctoral studies. I am also thankful to Professor Jianjun Shi, Professor Roshan Joseph Vengazhiyil, Professor Santanu Dey, and Professor Godfried Augenbroe for serving on my dissertation committee, and providing valuable comments and suggestions. I thank all my friends for assisting me on various matters throughout the course my PhD studies. Finally, I would like to thank my parents and brothers for their unwavering support that helped me through difficult times.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This dissertation presents novel methodologies for five problem areas in modern quality improvement and computer experiments, i.e., selective assembly, robust design with computer experiments, multivariate quality control, model selection for split plot experiments, and construction of minimax designs.

Chapter 1 proposes generalizations of the selective assembly method to assemblies with any number of components. Selective assembly has traditionally been used to achieve tight specifications on the clearance of two mating parts. However, its applicability is not limited to this particular type of assembly. This chapter develops a generalized version of selective assembly, called GSA. It can be a powerful tool for improving the quality of assemblies of single units of different component types. Two variants of GSA are considered: direct selective assembly (DSA) and fixed bin selective assembly (FBSA). The former is selective assembly using information from measurements on component characteristics directly, whereas the latter is selective assembly of components sorted into bins. For each variant, the problem of matching the N components of each type to give N assemblies that minimize quality cost is formulated as a linear integer program. The component matching problem for DSA is an axial multi-index assignment problem, whereas for FBSA, it is an axial multi-index transportation problem. We use simulations to evaluate the performance of GSA and to find the optimal number of bins. Realistic examples are given to show that the proposed methods can significantly improve the quality of assemblies.

Chapter 2 proposes methods for robust design optimization with time consuming computer simulations that take into account uncertainty about the true function. Gaussian process models, which include the class of linear models, are widely employed for modeling responses as a function of control or noise factors. Using these models, the average loss at control factor settings can be estimated and compared. However, robust design optimization is often performed based on expected quadratic loss computed as if the posterior mean were the true response function. This can give very misleading results. We propose an expected quadratic loss criterion derived by taking expectation with respect to the noise factors and the posterior predictive process. Approximate but highly accurate credible intervals for average quadratic loss are constructed via numerical inversion of the Lugannani-Rice saddlepoint approximation. The coverage of the Lugannani-Rice intervals are compared with intervals constructed via moment-matching techniques on real data.

Chapter 3 proposes a Bayesian method for identifying mean shifts in multivariate normally distributed quality characteristics. Multivariate quality characteristics are often monitored using a single statistic or a few statistics. However, it is difficult to determine the causes of an out-of-control signal based on a few summary statistics. Therefore, if a control chart for the mean detects a change in the mean, the quality engineer needs to determine which means shifted and the directions of the shifts to facilitate identification of root causes. We propose a Bayesian approach that gives a direct answer to this question. For each mean, an indicator variable that indicates whether the mean shifted upwards, shifted downwards or remained unchanged is introduced. Prior distributions for the means and indicators capture prior knowledge about mean shifts and allow for

asymmetry in upward and downward shifts. The mode of the posterior distribution of the vector of indicators or the mode of the marginal posterior distribution of each indicator gives the most likely scenario for each mean. Evaluation of the posterior probabilities of all possible values of the indicators is avoided by employing Gibbs sampling. This renders the computational cost more affordable for high-dimension problems.

Chapter 4 proposes a Bayesian method for model selection in fractionated split plot experiments. We employ a Bayesian hierarchical model that takes into account the split plot error structure. Expressions for computing the posterior model probability and other important posterior quantities that require evaluation of at most two uni-dimensional integrals are derived. A novel algorithm called combined global and local search is proposed to find models with high posterior probabilities and to estimate posterior model probabilities. The proposed method is illustrated with the analysis of three real robust design experiments. Simulation studies demonstrate that the method has good performance.

The problem of choosing a design that is representative of a finite candidate set is an important problem in computer experiments. The minimax criterion measures the degree of representativeness because it is the maximum distance of a candidate point to the design. Chapter 5 proposes a method for finding minimax designs for finite design regions. We establish the relationship between minimax designs and the classical set covering location problem in operations research, which is a binary linear program. In particular, we prove that the set of minimax distances is the set of discontinuities of the function that maps the covering radius to the optimal objective function value. We show that solving the set covering location problem at the points of discontinuities, which can

be determined, gives minimax designs. These results are employed to design an efficient procedure for finding minimax designs for small sized candidate sets. A heuristic procedure is proposed to generate near-minimax designs for large candidate sets.

CHAPTER 1

GENERALIZED SELECTIVE ASSEMBLY

1.1 Introduction

Selective assembly is a method to achieve high-precision clearance between two mating parts. To illustrate the idea of selective assembly, consider a sleeve and a shaft assembly. The clearance Y is the difference between the inner-diameter of the sleeve X_1 and the outer diameter of the shaft X_2, i.e. $Y = X_1 - X_2$. In random assembly, a sleeve and a shaft are chosen randomly from the available supply. However, if the tolerance on Y is tight, we may match sleeves that have large inner diameters with shafts that have large outer diameters and sleeves that have small inner diameters with shafts that have small outer diameters. Our matching can be based on the measured inner diameters of the sleeves and the measured outer diameters of the shafts directly. This is what we call direct selective assembly (DSA). When there are many sleeves and shafts to be assembled, it is convenient to sort sleeves and shafts into bins and then assemble sleeves and shafts from matched bins. Assembly is performed by randomly choosing sleeves and shafts from matched bins. We call this fixed bin selective assembly (FBSA).

The literature on selective assembly has focused on sleeve-and-shaft type assemblies. Kwon et al. (1999) study the selective assembly of sleeve-and-shaft type assemblies, where X_1 and X_2 are assumed to have independent and identical normal distributions. They derive optimal partitioning of X_1 and X_2 for fixed number of bins under squared error loss and also propose a method to determine the optimal number of bins. Mease et al. (2004) generalize the optimal partitioning results of Kwon et al. (1999)

to the case where X_1 and X_2 have arbitrary distributions and also study optimal partitioning under absolute error loss. Matsura and Shinozaki (2007) study optimal partition limits under squared error loss in the presence of measurement error. These papers constrain the number of bins for sleeves and shafts to be the same, i.e., the bin of smallest shafts is matched with the bin of smallest sleeves, the bin of the second smallest shafts is matched with the bin of the second smallest sleeves and so on. In addition, the probability that X_1 falls within the limits of bin i for sleeves is constrained to be equal to the probability that X_2 falls within the limits of bin i for shafts. The focus of these papers is the minimization of the long-run loss of assemblies produced from each pair of matched bins weighted by the probabilities of the pairs of bins. No indication is given on how situations in which the number of sleeves and the number of shafts in a pair of matched bins are different can be handled. Obviously, these assumptions are not compatible with a batch production scenario. They seem to indicate that a mass production scenario, where the production rates of sleeves and shafts are equal, is being considered. However, there is a serious problem with this scenario. Even if we get the distributions of X_1 and X_2 perfectly right, each pair of matched bins would form an unstable queuing system. There is a tendency for the number of components in each bin to accumulate to excessive numbers since the number of components that go into each bin is random. This is different from random assembly, where the feeding of component parts into an assembly line can be adjusted so that there is no build-up of component inventories.

Coullard et al. (1998) study the problem of matching a given number of two component types. They present algorithms for solving the matching problem to maximize

yield and quality. Iwata et al. (1998) formulate the selective assembly problem as a bipartite network flow problem and present a fast algorithm to solve the problem. Other papers that considered selective assembly in the context of manufacturing quality include Kannan et al. (2008), Kannan et al. (2005), Kannan and Jayabalan (2002), Thesen and Jantayavichit (1999), Zhang and Fang (1999), Chan and Linn (1998), Fang and Zhang (1995), Pugh (1992) and Mansoor (1961).

Although selective assembly has traditionally been performed for sleeve-and-shaft and related assemblies, there is no reason to limit the applicability of selective assembly to only this particular type of assembly. It is the purpose of this chapter to develop selective assembly as a tool to improve the quality of assemblies of any number of components with any known form of assembly response function. We study the *selective assembly problem under a batch production environment*, where components are produced in batches and then assembled. Selective assembly seems to be more suitable for batch production than mass production; the added steps of selective assembly should be easily accommodated by a batch production facility due to the medium to low production rate requirements. Moreover, the queue instability problem of matched bins does not arise in batch production. The treatment of selective assembly in the statistical literature (e.g. Kwon et al. (1999) and Mease et al. (2004)) has focused on optimal binning designs for assemblies with two components and a linear response function. On the other hand, the treatment of selective assembly in the operations research literature (e.g. Coullard et al. (1998) and Iwata et al. (1998)) has concentrated on developing algorithms for matching components of two different types. In contrast, we treat the selective assembly of batches of N units of each of k different component types.

Moreover, our methods are applicable to assemblies with any assembly response function.

This chapter is organized as follows. Section 1.2 discusses DSA and gives two realistic examples to demonstrate the potential gains in quality cost. Section 1.3 discusses FBSA. Three realistic examples are given to demonstrate the potential gains in quality cost and to show how good bin designs can be found by a screening and ranking procedure. The problem of selecting bin designs that are robust to batch size and manufacturing process variations is addressed in Section 1.4. Concluding remarks are given in Section 1.5.

1.2 Direct Selective Assembly

Consider an assembly of one unit of each of k different types of components. Suppose the assemblies are produced in batches of size N. Then, in random assembly, N components of each type are supplied to the assembly line, where they are randomly matched and assembled. In direct selective assembly (DSA), the components are matched in a systematic manner based on measurements of the component characteristics. Let the assembly quality characteristic be denoted by $\boldsymbol{Y}$. Suppose that $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_k)$, where $\boldsymbol{f}$ is the known assembly response function and $\boldsymbol{X}_i$ is the vector of measured characteristics of component i. Let the characteristics of the N components of the ith type be denoted by $\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{iN}$. Then, if we assemble the $i_j th$ type j component for $j = 1, \dots, k$, we have a product with quality characteristic value

$$\boldsymbol{Y}_{i_1 i_2 \cdots i_k} = \boldsymbol{f}(\boldsymbol{X}_{1i_1}, \dots, \boldsymbol{X}_{ki_k}) \tag{1.1}$$

and quality cost

$$Q_{i_1 i_2 \cdots i_k} = L(\boldsymbol{Y}_{i_1 i_2 \cdots i_k}, \boldsymbol{T}), \tag{1.2}$$

where L is the quality loss function and $\boldsymbol{T}$ is the target for $\boldsymbol{Y}$. One common choice of the quality loss function L is

$$L(\boldsymbol{Y}_{i_1 i_2 \cdots i_k}, \boldsymbol{T}) = (\boldsymbol{Y}_{i_1 i_2 \cdots i_k} - \boldsymbol{T})^T \boldsymbol{C} (\boldsymbol{Y}_{i_1 i_2 \cdots i_k} - \boldsymbol{T}), \tag{1.3}$$

where $\boldsymbol{C}$ is a positive semidefinite matrix (Kapur and Cho, 1996).

There are a total of $(N!)^{k-1}$ possible ways to obtain N assemblies. However, we can find the combination of assemblies that minimizes quality cost by solving a binary linear program:

Program A

Direct Selective Assembly

$$\min QC = \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_k=1}^{N} Q_{i_1 i_2 \cdots i_k} D_{i_1 i_2 \cdots i_k}$$

subject to:

$$\sum_{i_2=1}^{N} \sum_{i_3=1}^{N} \cdots \sum_{i_k=1}^{N} D_{i_1 i_2 \cdots i_k} = 1 \ \forall i_1 = 1, \dots, N$$

$$\sum_{i_1=1}^{N} \sum_{i_3=1}^{N} \cdots \sum_{i_k=1}^{N} D_{i_1 i_2 \cdots i_k} = 1 \ \forall i_2 = 1, \dots, N$$

$$\vdots$$

$$\sum_{i_1=1}^{N} \cdots \sum_{i_{k-2}=1}^{N} \sum_{i_{k-1}=1}^{N} D_{i_1 i_2 \cdots i_k} = 1 \ \forall i_k = 1, \dots, N$$

$$D_{i_1 i_2 \cdots i_k} \in \mathbb{Z}, D_{i_1 i_2 \cdots i_k} \geq 0 \ \forall i_1, i_2, \dots, i_k \in \{1, \dots, N\}$$

The $D_{i_1 i_2 \cdots i_k}$'s are the decision variables. If $D_{i_1 i_2 \cdots i_k} = 1$, then one product is to be assembled from the $i_j th$ type j component for $j = 1, \dots, k$. The objective function of Program A is the total quality cost of the assemblies. The constraints of Program A say that each component of each type must belong to one assembly. Let $\boldsymbol{D} = (D_{11\cdots1}, \dots, D_{kk\cdots k})$ denote the vector of decision variables. Then, $\boldsymbol{D}$ is a feasible solution if N elements of $\boldsymbol{D}$ equal 1 and the remaining $N^k - N$ elements equal 0. The optimal solution $\boldsymbol{D}^*$ to the program gives the N assemblies that minimize quality cost.

Program A is a special case of the multidimensional assignment problem known

as the *axial multi-index assignment problem* (Bandelt et al., 2004; Queyranne and Spieksma, 1997; Gilbert and Hofstra, 1988; Pierskalla, 1968). The total number of decision variables is N^k and the program can be solved for a global optimal solution using a branch-and-bound algorithm (Vanderbei, 2001; Schrijver, 1986) in which lower bounds are obtained by solving linear relaxations. Note that if $k = 2$, then we have the usual assignment problem, which can be solved as a linear program (Dantzig and Thapa, 2003), i.e., we can drop the integrality constraints $D_{i_1 i_2 \cdots i_k} \in \mathbb{Z}\ \forall i_1, i_2, \ldots, i_k \in \{1, \ldots, N\}$ from Program A. Unfortunately, for $k > 2$, this cannot be done.

In order to implement DSA, all relevant characteristics of each component must be measured and the measured values for each component have to be labelled or tagged on the component. This can become extremely costly when the batch size is large. Thus, DSA would generally be most cost effective when the batch size N is small, as is usually seen in Just-In-Time production systems.

Note that some of the $Q_{i_1 i_2 \cdots i_k}$ values can be large, implying poor quality products would be obtained from certain combination of components. Thus, it might be more appropriate to scrap such assemblies rather than to produce it. Let c_j^S be the cost of scrapping one unit of component of type j. Then, if $Q_{i_1 i_2 \cdots i_k} > \sum_{j=1}^{k} c_j^S$, it is more cost effective to scrap the assembly. As such, we should replace $Q_{i_1 i_2 \cdots i_k}$ in Program A with $Q_{i_1 i_2 \cdots i_k} = \min\{L(\boldsymbol{Y}_{i_1 i_2 \cdots i_k}, \boldsymbol{T}), \sum_{j=1}^{k} c_j^S\}$. Note that we may also consider the possibility of reworking the components to get an acceptable assembly. In this case, $\sum_{j=1}^{k} c_j^S$ should be replaced with the rework cost.

Because the $\boldsymbol{X}_{ij}$'s are random, i.e. they vary from batch to batch, the $\boldsymbol{Y}_{i_1 i_2 \cdots i_k}$'s,

and $Q_{i_1 i_2 \cdots i_k}$'s are random quantities also. Therefore, the optimal solution and objective function value of Program A are random variables. Because the optimal objective value of Program A changes from batch to batch, we should assess the cost effectiveness of DSA by the expected quality cost per batch. We use the simulation procedure given in Appendix A.1 to estimate the expected quality cost for DSA. It is assumed that the $\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1N}, \boldsymbol{X}_{21}, \ldots, \boldsymbol{X}_{kN}$ are independently distributed and that for each i, $\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{iN}$ have common distribution $F_{\boldsymbol{X}_i}$.

In the following examples, we shall compare random assembly with selective assembly. For random assembly without inspection of assemblies, the expected quality cost per batch is

$$QC_{RA0} = NE[L(\boldsymbol{Y}, \boldsymbol{T})], \tag{1.4}$$

where $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$ and $\boldsymbol{X}_i \sim F_{\boldsymbol{X}_i}$. If 100% inspection of the assembled products is performed, then the expected quality cost is

$$QC_{RA} = NE\left[\min\left\{L(\boldsymbol{Y}, \boldsymbol{T}), \sum_{j=1}^{k} c_j^S\right\}\right], \tag{1.5}$$

assuming that products that have higher quality loss than the scrap cost are scrapped.

As shall be demonstrated in the examples, DSA is much better at improving the quality of assemblies than 100% inspection of the output of random assembly. Although selective assembly may be more costly than 100% inspection, the cost of 100% inspection can also be very high. Aside from an inspection cost for every batch, we also incur the additional cost of wasted effort due to assembling products that are ultimately scrapped. In cases where the assembly quality characteristics can only be measured via destructive testing, 100% inspection cannot be implemented. In contrast, direct selective assembly does not suffer from these disadvantages. Bad combinations of components are

scrapped even before they are assembled if it is optimal to do so. In addition, inspection of assembled products is not needed. Of course, these advantages can only be realized if $\boldsymbol{Y}$ can be predicted accurately based on $\boldsymbol{f}$. However, this is possible for many well studied assemblies, and also for mechanical assemblies for which $\boldsymbol{f}$ is derived from purely geometric considerations.

1.2.1 Example 1.1: Direct Selective Assembly of Bimetal thermostat

Consider a bimetal thermostat (see Appendix A.2) of unit width that is simply supported at both ends (Timoshenko, 1925). The amount of deflection at the midpoint upon heating of the thermostat from temperature T_0 to T is given by

$$Y = \frac{l^2}{8}\frac{6(\alpha_2-\alpha_1)(T-T_0)(1+m)^2}{h\left[3(1+m)^2+(1+mn)\left(m^2+\frac{1}{mn}\right)\right]}, \tag{1.6}$$

where

$l = distance\ between\ supports = length\ of\ thermostat\ at\ temperature\ T_0,$

$\alpha_1, \alpha_2 = coefficient\ of\ thermal\ expansion\ of\ the\ two\ metals,$

$n = E_1/E_2 = ratio\ of\ Young's\ moduli\ of\ the\ two\ metals,$

$m = t_1/t_2 = ratio\ of\ thicknesses\ of\ the\ two\ metals,$

$h = t_1 + t_2 = thickness\ of\ thermostat.$

Suppose we want to manufacture molybdenum/aluminium thermostats (Tierney and Eischen, 1997; Eischen, 1989) with a deflection of $Y = 1$mm when heated to 500 ℃ above room temperature. The material constants α_1, α_2, E_1, and E_2 are given by $4.9 \times 10^{-6}/℃$, $23 \times 10^{-6}/℃$, 32.5×10^{10}Pa, and 7×10^{10}Pa respectively, where the subscript 1 stands for molybdenum and the subscript 2 stands for aluminium.

The nominal dimensions for the thermostat are $l = 50mm$, $t_1 = 1mm$, and

$t_2 = 3mm$ and the thermostat is manufactured by joining molybdenum and aluminium strips. A manufactured molybdenum strip has thickness t_1 that follows a truncated normal distribution with support $[L_1, U_1] = [0.5,1.5]$, and mean and standard deviation before truncation $\mu_1 = 1$ and $\sigma_1 = 0.2$ respectively. On the other hand, a manufactured aluminium strip has thickness t_2 that follows a truncated normal distribution with support $[L_2, U_2] = [2.5,3.5]$, and mean and standard deviation before truncation $\mu_2 = 3$ and $\sigma_2 = 0.2$ respectively. We write

$t_1 \sim TN(1,0.2,0.5,1.5)$, and $t_2 \sim TN(3,0.2,2.5,3.5)$. (1.7)

Suppose that $L(Y,T) = 10(Y-T)^2$. If N molybdenum strips and N aluminium strips are randomly matched and joined to produce each batch of thermostats, then the expected quality cost of a batch is $QC_{RA0} = 10N\left\{E\left[(Y-1)^2\right]\right\}$. A simulation with 10,000 runs gives $\widehat{QC}_{RA0} = 0.05038N(0.000755N)$, where the quantity in brackets is the standard error. If $\sum_{j=1}^{2} c_1^S = 0.27$, then the quality cost is $QC_{RA} = 10N\{E[\min\{10(Y-1)^2, 0.27\}]\}$. A simulation with 10,000 runs gives $\widehat{QC}_{RA} = 0.04923N(0.000657N)$. For $N = 10, 15, 20$, we compute the quality cost when DSA is performed with and without the option of scrapping using simulations of 100 runs. Note that the simulation procedure is given in Appendix A.1. Table 1.1 gives estimates of the expected quality costs per batch of N thermostats for random assembly and DSA. We see that 50-60% reduction in quality cost is achieved with DSA (these percentages are obtained by comparing $\widehat{QC}_{RA0}$ with $\widehat{QC}_{SA0}$ and $\widehat{QC}_{RA}$ with $\widehat{QC}_{SA}$). This is a huge improvement and may more than offset the implementation cost of DSA. Moreover, DSA is seen to be far more effective at improving the quality cost than 100% inspection. Notice also that the percentage improvement is greater the larger the batch size. This is to be expected since

Table 1.1: Quality Costs for Random Assembly and Direct Selective Assembly (Bimetal Thermostat, Truncated Normally Distributed Strip Thickness)

N	$\widehat{QC}_{RA0}$ (no scrap)	$\widehat{QC}_{RA}$	$\widehat{QC}_{SA0}$ (no scrap)	% Improvement	$\widehat{QC}_{SA}$	% Improvement
10	0.504(0.0076)	0.492(0.0066)	0.258(0.0186)	49%	0.254(0.0158)	48%
15	0.756(0.0113)	0.738(0.0099)	0.362(0.0196)	52%	0.339(0.0197)	54%
20	1.008(0.0151)	0.985(0.0131)	0.454(0.0270)	55%	0.401(0.0186)	59%

Table 1.2: Quality Costs for Random Assembly and Direct Selective Assembly (Bimetal Thermostat, Uniformly Distributed Strip Thickness)

N	$\widehat{QC}_{RA0}$ (no scrap)	$\widehat{QC}_{RA}$	$\widehat{QC}_{SA0}$ (no scrap)	% Improvement	$\widehat{QC}_{SA}$	% Improvement
10	1.011(0.0134)	0.841(0.0088)	0.410(0.0256)	59%	0.377(0.0224)	55%
15	1.517(0.0201)	1.262(0.0132)	0.579(0.0371)	62%	0.497(0.0248)	61%
20	2.022(0.0268)	1.683(0.0176)	0.691(0.0341)	66%	0.669(0.0357)	60%

the number of choices of assemblies increases as N increases.

It is interesting to see the effect of changing the distribution of t_1 and t_2. Suppose that t_1 and t_2 are uniformly distributed over the same support as before, i.e.,

$t_1 \sim Unif(0.5,1.5)$, and $t_2 \sim Unif(2.5,3.5)$. (1.8)

Table 1.2 shows that for $N = 10,15,20$, we achieve 55-65% reduction in quality cost. The quality costs for random assembly and DSA are much larger when t_1 and t_2 are given by (1.8) than when they are given by (1.7). This is because (1.8) implies larger variability in t_1 and t_2. The absolute and percentage improvement achieved with DSA also tends to be larger when t_1 and t_2 are uniformly distributed. This suggests that greater cost reductions are attainable with DSA for processes with more variation.

1.2.2 Example 1.2: Direct Selective Assembly of Knuckle Joint Assembly

Unlike in Example 1.1, the current problem involves *multiple responses and*

multivariate component characteristics. The knuckle joint assembly (see Appendix A.2) shown in Singh et al. (2005) has two clearances Y_1 and Y_2 that must be maintained at target values. The clearances are functions of component dimensions given by $Y_1 = X_{2b} - X_1$, and $Y_2 = X_3 - (2X_{2a} + X_{2b})$, where X_1 is a dimension of component 1, X_{2a} and X_{2b} are dimensions of component 2, and X_3 is a dimension of component 3.

Suppose that $\boldsymbol{T} = (1,1)$ and that $L(\boldsymbol{Y}, \boldsymbol{T}) = 125(\boldsymbol{Y} - \boldsymbol{T})^T(\boldsymbol{Y} - \boldsymbol{T})$. Assume that X_1, X_{2a}, X_{2b}, and X_3 are independently distributed truncated normal random variables given by

$X_1 \sim TN(9,0.04,8.88,9.12)$, $X_{2a} \sim TN(5,0.04,4.88,5.12)$, $X_{2b} \sim TN(10,0.04,9.88,10.12)$, and $X_3 \sim TN(21,0.08,20.76,21.24)$. (1.9)

Simulations with $10{,}000$ runs give $\widehat{QC}_{RA0} = 2.1517N(0.02513N)$ and $\widehat{QC}_{RA} = 2.0973N(0.02268N)$, where we set $\sum_{j=1}^{3} c_3^S = 10$. We compute the quality cost for selective assembly for $N = 5{,}10$ using simulations of 300 runs. Table 1.3 gives estimates of the expected quality costs per batch of N knuckle joints for random assembly and DSA. With the implementation of DSA, we achieve 65-80% reduction in quality cost. As in Example 1.1, the improvement is larger for $N = 10$ than for $N = 5$.

Now, assume that X_1, X_{2a}, X_{2b}, and X_3 are uniformly distributed with the same support as given by (1.9). Table 1.4 gives estimates of the expected quality costs of

Table 1.3: Quality Costs for Random Assembly and Direct Selective Assembly (Knuckle Joint, Truncated Normally Distributed Component Dimensions)

N	$\widehat{QC}_{RA0}$ (no scrap)	$\widehat{QC}_{RA}$	$\widehat{QC}_{SA0}$ (no scrap)	% Improvement	$\widehat{QC}_{SA}$	% Improvement
5	10.759(0.1257)	10.487(0.1134)	3.868(0.1720)	64%	3.787(0.1593)	64%
10	21.517(0.2513)	20.973(0.2268)	4.184(0.1489)	81%	4.138(0.1504)	80%

Table 1.4: Quality Costs for Random Assembly and Direct Selective Assembly (Knuckle Joint, Uniformly Distributed Component Dimensions)

N	$\widehat{QC}_{RA0}$ (no scrap)	$\widehat{QC}_{RA}$	$\widehat{QC}_{SA0}$ (no scrap)	% Improvement	$\widehat{QC}_{SA}$	% Improvement
5	33.205(0.3393)	25.273(0.1788)	10.589(0.4675)	68%	9.848(0.3364)	61%
10	66.410(0.6786)	50.546(0.3577)	12.324(0.5204)	81%	12.084(0.4102)	76%

random assembly and DSA. With DSA, we achieve 60-80% reduction in quality cost. The quality costs for random and selective assembly are larger when t_1 and t_2 are uniformly distributed than when t_1 and t_2 are truncated normally distributed over the same support. In addition, the absolute improvement in quality cost achieved with the implementation of selective assembly also tends to be larger.

Examples 1.1 and 1.2 demonstrate that sizeable reductions in quality cost can be achieved with DSA. Such improvements can more than offset the implementation cost of DSA. Moreover, the substantial improvements in product quality can give the manufacturing organization a significant competitive advantage.

1.3 Fixed Bin Selective Assembly

It is seen in the previous section that selective assembly by matching components based on the measured values of their characteristics can result in considerable improvements in quality. However, this approach can be extremely costly when N is large. The cost of measuring component characteristics and keeping records on the measured values for each component can be prohibitive. Moreover, the number of decision variables and number of constraints for Program A increase as N increases, making the program more difficult to solve. To reduce these costs and still retain the benefits of component matching, selective assembly can be performed by sorting the N

components of each type into several bins. We call this fixed bin selective assembly (FBSA). In particular, we sort the N type i components into n_i bins for each $i = 1, \dots, k$, where a component of type i with measured characteristic $\boldsymbol{X}_i$ is sorted into the jth bin if $\boldsymbol{X}_i \in B_{ij}$. For each i, the bins $B_{i1}, \dots, B_{in_i}$ should be a partition of S_i, the sample space of $\boldsymbol{X}_i$, i.e.

$$\cup_{j=1}^{n_i} B_{ij} = S_i, \text{ and } B_{ij} \cap B_{il} = \phi \ \forall j \neq l. \tag{1.10}$$

Note that if $\boldsymbol{X}_i$ is a one-dimension random vector, then a common practice is to take each B_{ij} as an interval. If $\boldsymbol{X}_i$ has dimension greater than one, then it is convenient to take the B_{ij}'s as hyper-rectangles.

After sorting, the number of components in each bin is counted. Then, using this information, we determine the number of products to be assembled from bins $i_1, \dots, i_k$ of components of type $1, \dots, k$ respectively, for all $(i_1, \dots, i_k)$. Define

$$\boldsymbol{X}_{ij} = \boldsymbol{X}_i | \boldsymbol{X}_i \in B_{ij}. \tag{1.11}$$

If a product were assembled from a type j component randomly chosen from bin i_j, $j = 1, \dots, k$, then the quality characteristic of the product is described by a random variable given by

$$\boldsymbol{Y}_{i_1 i_2 \cdots i_k} = \boldsymbol{f}(\boldsymbol{X}_{1i_1}, \dots, \boldsymbol{X}_{ki_k}). \tag{1.12}$$

Hence, the expected quality cost of this product is

$$Q_{i_1 i_2 \cdots i_k} = E[L(\boldsymbol{Y}_{i_1 i_2 \cdots i_k}, \boldsymbol{T})]. \tag{1.13}$$

Let N_{ij} be the number of type i components in the jth bin for that component type and let n_i be the number of bins for type i components. Define a cell to be a combination of bins; we refer to cell $(i_1, \dots, i_k)$ to be the combination of bin i_j of type j components for $j = 1, \dots, k$. The component matching problem is to find the number of assemblies to be

produced from each cell such that the total quality cost is minimized. This problem can be formulated as an integer linear program:

Program B

Fixed Bin Selective Assembly

$$\min QC = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_k=1}^{n_k} Q_{i_1 i_2 \cdots i_k} D_{i_1 i_2 \cdots i_k}$$

subject to:

$$\sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} \cdots \sum_{i_k=1}^{n_k} D_{i_1 i_2 \cdots i_k} = N_{1 i_1} \ \forall i_1 = 1, \ldots, n_1$$

$$\sum_{i_1=1}^{n_1} \sum_{i_3=1}^{n_3} \cdots \sum_{i_k=1}^{n_k} D_{i_1 i_2 \cdots i_k} = N_{2 i_2} \ \forall i_2 = 1, \ldots, n_2$$

$$\vdots$$

$$\sum_{i_1=1}^{n_1} \cdots \sum_{i_{k-2}=1}^{n_{k-2}} \sum_{i_{k-1}=1}^{n_{k-1}} D_{i_1 i_2 \cdots i_k} = N_{k i_k} \ \forall i_k = 1, \ldots, n_k$$

$$D_{i_1 \cdots i_k} \in \mathbb{Z}, D_{i_1 \cdots i_k} \geq 0 \ \forall (i_1, \ldots, i_k) \in \{(j_1, \ldots, j_k) \in \mathbb{Z}^k : 1 \leq j_l \leq n_l, l = 1, \ldots, k\},$$

$$\text{where} \sum_{j=1}^{n_i} N_{ij} = N \ \forall i = 1, \ldots, k.$$

The $D_{i_1 i_2 \cdots i_k}$'s are the decision variables. If $D_{i_1 i_2 \cdots i_k} = x$, then x products are to be assembled from cell $(i_1, i_2, \ldots, i_k)$. The objective function of Program B is the expected total quality cost of the assemblies. The constraints of Program B say that all components in each bin must be utilized to build a total of N assemblies. Let $\boldsymbol{D} = \left(D_{11 \cdots 1}, \ldots, D_{n_1 n_2 \cdots n_k}\right)$ denote the vector of decision variables. Then, $\boldsymbol{D}$ is a feasible solution if the sum of the $\prod_{i=1}^{k} n_i$ elements of $\boldsymbol{D}$ equal N and all elements of $\boldsymbol{D}$ are nonnegative integers. The optimal solution $\boldsymbol{D}^*$ to the program gives the number of products to be assembled from each cell such that the expected total quality cost is minimized.

Program B is known as the *axial multi-index transportation problem* (Queyranne and Spieksma, 1997; Haley, 1963). It can be solved in exactly the same way as Program A, i.e. using a branch-and-bound algorithm with lower bounds obtained by solving linear

programming relaxations. For $k = 2$, the continuous relaxation of Program B becomes a transportation problem; thus, all basic solutions are integer points (Dantzig and Thapa, 2003) and the program must have an integer optimal solution. Unfortunately, for $k > 2$, this is not true.

If scrapping is an option, then we should replace $Q_{i_1 i_2 \cdots i_k}$ in (1.13) with

$$Q_{i_1 i_2 \cdots i_k} = \min\{E[L(\boldsymbol{Y}_{i_1 i_2 \cdots i_k}, \boldsymbol{T})], \sum_{j=1}^{k} c_j^S\}. \quad (1.14)$$

In essence, this leads us to a generalization of the notion of setting component tolerances. Instead of using specification limits to define hyper-rectangle acceptance regions, we now have acceptance regions that are unions of hyper-rectangles, as illustrated in Table 1.5 for $k = 2$, and $n_1 = n_2 = 4$. This can give an acceptance region that better approximates the shape of the contours of the assembly response function $\boldsymbol{f}$. Moreover, some undesirable combinations are scrapped only if it is optimal to do so according to Program B.

To evaluate the long-run performance of FBSA, we shall assume that all N components of the same type have identically distributed characteristics, and that all component characteristics are independently distributed. Moreover, we assume that the distribution for each component type remains the same from batch to batch. Given these

Table 1.5: Accept/ Reject Cells for Fixed Bin Selective Assembly

	B_{21}	B_{22}	B_{23}	B_{24}
B_{11}	Reject	Accept	Accept	Accept
B_{12}	Reject	Accept	Accept	Accept
B_{13}	Accept	Accept	Accept	Accept
B_{14}	Accept	Accept	Accept	Accept

assumptions, the $Q_{i_1 i_2 \cdots i_k}$'s in Program B are fixed quantities, unlike Program A. However, for all $i = 1, \ldots, k$, $(N_{i1}, \ldots, N_{in_i})$ is a multinomial random variable. Thus, the optimal solution and objective function value of Program B are random variables that changes from batch to batch.

For FBSA, it is desirable to determine the optimal number and formation of bins to minimize the expected total cost per batch of assemblies. The total cost is taken to be the sum of the quality cost and extra cost of FBSA over random assembly, which is a function of the number of bins $n_1, \ldots, n_k$ and also the batch size N. This latter cost includes the cost of sorting, and the increased cost of material handling due to segregation of components of the same type. Thus, we want to choose $\boldsymbol{n} = (n_1, \ldots, n_k)$ and $B_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, k$ such that the sum of the expected quality cost and the implementation cost of FBSA $C(\boldsymbol{n}, N)$ for a batch of products is minimized. Denote

$$\boldsymbol{\eta} = \left(N_{11}, \ldots, N_{1n_1}, N_{21}, \ldots, N_{2n_2}, \ldots, N_{k1}, \ldots, N_{kn_k}\right) \tag{1.15}$$

and let

$$\Omega = \left\{\boldsymbol{\eta} \in \mathbb{Z}^{\sum_{i=1}^{k} n_i} : \boldsymbol{\eta} \geq \boldsymbol{0}, \sum_{j=1}^{n_i} N_{ij} = N \ \forall i = 1, \ldots, k\right\} = \{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_m\}, \tag{1.16}$$

where $m = \prod_{i=1}^{k} \binom{N + n_i - 1}{n_i - 1}$. Let $P_{\boldsymbol{\eta}_i} = P(\boldsymbol{\eta} = \boldsymbol{\eta}_i)$, $p_{ij} = P(\boldsymbol{X}_i \in B_{ij})$, and $QC^*_{\boldsymbol{\eta}_i}$ be the minimum expected quality cost achieved when $\boldsymbol{\eta} = \boldsymbol{\eta}_i$. Then, the expected quality cost for a batch of size N is

$$E(QC) = \sum_{i=1}^{m} QC^*_{\boldsymbol{\eta}_i} P_{\boldsymbol{\eta}_i}, \tag{1.17}$$

where $P_{\boldsymbol{\eta}_i} = \prod_{j=1}^{k} \frac{N!}{N^i_{j1}! \cdots N^i_{jn_j}!} p_{j1}^{N^i_{j1}} \cdots p_{jn_j}^{N^i_{jn_j}}$, and

$$\boldsymbol{\eta}_i = \left(N^i_{11}, \ldots, N^i_{1n_1}, N^i_{21}, \ldots, N^i_{2n_2}, \ldots, N^i_{k1}, \ldots, N^i_{kn_k}\right).$$

We write the bin design problem formally as follows.

Program C

Bin Design for Fixed Bin Selective Assembly

$$\min_{n_i, B_{ij}} \left\{ TC = \sum_{i=1}^{m} QC^*_{\boldsymbol{\eta}_i} P_{\boldsymbol{\eta}_i} + C(\boldsymbol{n}, N) \right\}$$

subject to:

$$\bigcup_{j=1}^{n_i} B_{ij} = S_i, B_{ik} \cap B_{il} = \phi \ \forall k \neq l, \boldsymbol{n} \in \mathbb{N}^k.$$

In the following examples, we restrict attention to the case where each $\boldsymbol{X}_i$ is of a single dimension with sample space given by $S_i = [L_i, U_i]$, and we take

$$B_{ij} = \{x \in \mathbb{R}: b_{i,j-1} \leq x \leq b_{ij}\}, \tag{1.18}$$

where $b_{i0} = L_i$, $b_{in_i} = U_i$, $b_{i0} < b_{i1} < \cdots < b_{i,n_i-1} < b_{in_i}$; hence, the decision variables for Program C are $n_i, b_{i1}, \ldots, b_{i,n_i-1}, i = 1, \ldots, k$.

Program C is difficult to solve. Note that $QC^*_{\boldsymbol{\eta}_i}$ is a function of the decision variables $n_i, b_{i1}, \ldots, b_{i,n_i-1}, i = 1, \ldots, k$. Hence, to evaluate the objective function of Program C at a feasible point, Program B needs to be solved m times. Solving Program B involves computing the $Q_{i_1 i_2 \cdots i_k}$ values in addition to running a branch-and-bound algorithm. Moreover, because m can be a very large number even for small values of N and $n_1, \ldots, n_k$, computing the values of all the $QC^*_{\boldsymbol{\eta}_i}$ at a single point in the feasible region of Program C can be a daunting task in itself. The problem is further aggravated by the large number of decision variables.

In view of these problems, we restrict attention to ranking a finite number of alternative bin designs. Based on the limitations of the precision of measuring equipment, or on a comparison of $C(\boldsymbol{n}, N)$ to QC_{RA}, we can reduce the set of alternative $\boldsymbol{n}$'s from $\mathbb{N}^k$ to a finite set $\omega_{\boldsymbol{n}}$. Note that values of $\boldsymbol{n}$ such that $C(\boldsymbol{n}, N) > QC_{RA}$ need not be

considered. Since $C(\boldsymbol{n}, N)$ should be increasing in each of its argument, this allows us to narrow down the alternative choices of $\boldsymbol{n}$ to a finite set that contains $\boldsymbol{n}^*$, the optimal number of bins. However, it seems to be difficult to narrow down the infinite number of alternative b_{ij}'s. We shall compare the b_{ij}'s obtained using two heuristic methods (Mease et al., 2004), which we call bin-formation rules. These two rules are:

i. The equal-width rule: Choose the b_{ij}'s so that $b_{ij} - b_{i,j-1} = (U_i - L_i)/n_i$.

ii. The equal-area rule: Choose the b_{ij}'s so that $F_{X_i}(b_{ij}) - F_{X_i}(b_{i,j-1}) = 1/n_i$.

Thus, we suggest ranking the set ω_{alt} of alternative bin design that consists of all combinations of $\boldsymbol{n} \in \omega_{\boldsymbol{n}}$ and bin-formation rule. The procedure given in Appendix A.3 is used to estimate the expected total cost TC for each of the finite number of alternative bin designs suggested.

We use the two-stage screening and ranking procedure proposed by Nelson et al. (2001) to determine the best alternative in the set of alternative bin designs. Since this procedure requires the assumption of normality, we use as input data for the procedure the averages of total cost estimates $\widehat{TC}$ obtained from $r > 1$ runs of Procedure II (Schmeiser, 1982). The procedure is summarized in Appendix A.4. Steps 1-4 of Procedure III are steps for the screening of alternatives and steps 5-8 are steps for the ranking of alternatives.

In the following examples, we shall compare random assembly with FBSA. In all examples, the parameters α, r, and m_1 for Procedure III are fixed at $\alpha = 0.05$, $r = 10$, $m_1 = 10$. Note that $m_1 = 10$ replicates is the minimum recommended by Schmeiser (1982) and the average of $r = 10$ $\widehat{TC}$ values should be large enough for the normality assumption to be a good approximation. For Step 3 of Procedure II, we estimate each

$E\left[L\left(\boldsymbol{Y}_{i_1 i_2 \cdots i_k}, \boldsymbol{T}\right)\right]$ by a simulation with 10,000 run. In general, this would be convenient when either the loss function L is complicated, or the $\boldsymbol{Y}_{i_1 i_2 \cdots i_k}$'s are complicated functions of the $\boldsymbol{X}_{ji_j}$'s.

Assuming that products that have higher quality loss than the scrap cost are scrapped, the expected quality cost of random assembly per batch of assemblies is given by (1.5). Strictly speaking, the quality cost of random assembly with 100% inspection is not comparable to the total cost of FBSA. The cost of inspection and the cost of wasted effort due to assembling products that are scrapped after inspection must be added to the quality cost to obtain the total cost of random assembly. Note also that our specification of $C(\boldsymbol{n}, N)$ in the examples shall be somewhat arbitrary. This is unavoidable because $C(\boldsymbol{n}, N)$ depends on many production-system-specific factors such as the degree of automation, the material handling equipment etc. However, we point out that simulation models of the assembly process with and without implementation of FBSA can be used to determine $C(\boldsymbol{n}, N)$. For example, Activity Based Costing of manufacturing systems using discrete event simulation models is discussed by Spedding and Sun (1999).

1.3.1 Example 1.3: Fixed Bin Selective Assembly of Bimetal Thermostat

We investigate fixed bin selective assembly of the bimetal thermostat described in Example 1.1. Assume that $N = 100$, $C(n_1, n_2, 100) = 0.2(n_1 + n_2 - 2)$, and the distributions of t_1 and t_2 are given by (1.7). A simulation of 100,000 runs gives $\widehat{QC}_{RA} = 4.872(0.0208)$. This indicates that only values of (n_1, n_2) such that $n_1 + n_2 \leq 26$ need to be considered. However, the precision of the measuring equipment would likely place tighter upper bounds on the values of n_1 and n_2. We suppose that only a maximum of

four bins is allowed for each component. Thus, $\omega_{\boldsymbol{n}} = \left\{(n_1, n_2) \in \mathbb{N}^2 : 2 \leq n_1 \leq 4, 2 \leq n_2 \leq 4\right\}$. Note that if we set $n_1 = 1$ or $n_2 = 1$, we are essentially performing random assembly; if no scrapping of components or assemblies are allowed, then FBSA with $n_1 = 1$ or $n_2 = 1$ is equivalent to random assembly. Define

$$rule = \begin{cases} 1, if\ the\ equal\ width\ bin\ formation\ rule\ is\ used \\ 2, if\ the\ equal\ area\ bin\ formation\ rule\ is\ used \end{cases} \tag{1.19}$$

Then, our set of alternatives is given by

$$\omega_{alt} = \left\{(n_1, n_2, rule) \in \mathbb{N}^3 : 2 \leq n_1 \leq 4, 2 \leq n_2 \leq 4, 1 \leq rule \leq 2\right\}. \tag{1.20}$$

We employ Procedure III with $\delta = 0.1$ to screen and rank the alternatives in ω_{alt}. The results of the screening steps are shown in Table 1.6. Highlighted are the seven alternatives that remain after the screening phase. The column headed $\widehat{QC}_i$ gives estimates of the quality costs of the alternatives $\widehat{QC}_i = \overline{TC}_i^{(1)} - C(n_1, n_2, 100)$. It is interesting to see that the quality cost can increase with an increase in n_1 or n_2. However, the quality cost must decrease if the bins of the alternative with the larger values of n_1 and n_2 are nested within the bins of the alternative with the smaller values of n_1 and n_2.

The results of the ranking steps are presented in Table 1.7. We see that in order to rank the seven remaining alternatives, Procedure II had to be run $49, 36, 73, 70, 97, 44,$ and 56 times respectively with $M = 10$. Since $\overline{TC}_{(3,3,2)}^{(2)} = 3.523$ is the smallest, we declare $(3,3,2)$ as the best alternative. Recall that $\widehat{QC}_{RA} = 4.872(0.0208)$. Thus, the improvement in total cost, assuming that the total cost of random assembly equals the quality cost, is $\frac{4.872-3.523}{4.872} \approx 30\%$ and the improvement in quality cost is about $\frac{4.872-[3.523-0.2(3+3-2)]}{4.872} \approx 45\%$.

Table 1.6: Screening of Alternative Bin Designs (Bimetal Thermostat, Truncated Normally Distributed Strip Thickness)

n_1	n_2	$Rule$	s_i^2	$\overline{TC}_i^{(1)}$	$\widehat{QC}_i$
2	2	1	0.02094	3.659	3.259
3	2	1	0.01980	4.380	3.780
4	2	1	0.02339	3.894	3.094
2	3	1	0.01784	4.150	3.550
3	3	1	0.01635	3.668	2.868
4	3	1	0.02405	4.147	3.147
2	4	1	0.02708	3.842	3.042
3	4	1	0.01305	4.064	3.064
4	4	1	0.01850	3.782	2.582
2	2	2	0.02965	3.606	3.206
3	2	2	0.01922	3.924	3.324
4	2	2	0.02592	3.892	3.092
2	3	2	0.00861	3.757	3.157
3	3	2	0.02836	3.575	2.775
4	3	2	0.03822	3.581	2.581
2	4	2	0.01088	3.948	3.148
3	4	2	0.01912	3.630	2.630
4	4	2	0.02338	3.720	2.520

Table 1.7: Ranking of Alternative Bin Designs (Bimetal Thermostat, Truncated Normally Distributed Strip Thickness)

n_1	n_2	$Rule$	m_2	$\overline{TC}_i^{(2)}$	$std\ error$
2	2	1	59	3.671	0.0157
3	3	1	46	3.722	0.0248
2	2	2	83	3.589	0.0155
3	3	2	80	3.523	0.0176
4	3	2	107	3.678	0.0169
3	4	2	54	3.685	0.0169
4	4	2	66	3.670	0.0166

Table 1.8: Quality Cost $E[10(Y-1)^2]$ for Each Cell of (4,4,1) Alternative

		X_2			
		[2.5,2.75]	[2.75,3]	[3,3.25]	[3.25,3.5]
X_1	[0.5,0.75]	**0.3593**	0.0985	0.0085	0.0376
	[0.75,1]	**0.2841**	0.0812	0.0071	0.0277
	[1,1.25]	0.1505	0.0336	0.0052	0.0457
	[1.25,1.5]	0.0406	0.0052	0.0254	0.0928

For the alternative (4,4,1), we illustrate the bins and quality costs $E\left[10\left(Y_{i_1 i_2}-1\right)^2\right]$ in Table 1.8. The quality cost values are estimates obtained from simulations with 10,000 runs. We see that the two upper left bins have quality costs that exceed the scrapping cost $\sum_{j=1}^{2} c_j^S = 0.27$. Thus, $Q_{11} = Q_{21} = 0.27$. Table 1.8 also suggests that a rectangular tolerance region for (X_1, X_2) may not be a good choice since the union of the cells that would produce products of acceptable quality is not rectangular in shape.

In the case where the distributions of t_1 and t_2 are given by the uniform distributions (1.8), the equal-width and equal-area bin-formation rules lead to identical bins. Because of the larger variances, alternatives with up to five bins per component are considered and it is found that (4,4) is the best alternative. The improvement in total cost is about 40% and the improvement in quality cost is about 55%. Details are omitted for brevity.

Note that the choices of $\omega_{\boldsymbol{n}}$ in this and the next example were determined by trial and error. We incrementally increased the maximum number of bins for each component type and ran Procedure III. We stopped when further increases did not lead to a new best bin design. However, Examples 1.3-1.5 demonstrate that even by restricting attention to a very small subset of the set of all $\boldsymbol{n}$ with $C(\boldsymbol{n}, N) \leq QC_{RA}$, we can achieve substantial reductions in quality cost and total cost.

1.3.2 Example 1.4: Fixed Bin Selective Assembly of Fortini's Clutch

Consider Fortini's clutch (see Appendix A.2), which is an assembly of a cage, a hub, four ball bearings, and four springs as shown in the literature (Lee and Chen, 2007; Lee and Kwak, 2005; Forouraghi, 2002; Wu et al., 1998; Feng and Kusiak, 1997). The quality characteristic of interest is the contact angle Y given by

$Y = cos^{-1}[(X_1 + X_2)/(X_3 - X_2)]$,

where X_1 is a dimension of the hub, X_2 is the ball bearing diameter, and X_3 is the inner diameter of the cage. Note that there are four ball bearings per clutch, which means that there are four components of the same type per assembly and four contact angles. Program B is applicable to this assembly provided that we assume all ball bearings for each assembly have the same diameter. This assumption may be reasonable if the ball bearings are produced in groups of four, where each group of ball bearings is designated for one assembly, and the diameters of ball bearings in a group are highly correlated. In this case, one component of type 2 corresponds to a group of four ball bearings.

Suppose that

$X_1 \sim TN(55.29, 0.0793, 55.0521, 55.5279)$, $X_2 \sim TN(22.86, 0.0043, 22.8471, 22.8729)$,

$X_3 \sim TN(101.6, 0.0793, 101.3621, 101.8379)$. (1.21)

Let $N = 100$, $T = 0.12217\ rad = 7°$, and suppose that $L(Y,T) = 10^4(Y - T)^2$. Let the component scrap costs be given by $\sum_{j=1}^{3} c_j^S = 6$. Thus, the quality cost is such that when $Y < 0.09768\ rad = 5.6°$ or $Y > 0.14666\ rad = 8.4°$, then it is more cost effective to scrap the components. This is reasonable since Y must be between 5° and 9° (Feng and Kusiak, 1997; Wu et al., 1998; Forouraghi, 2002; Lee and Kwak, 2005). Based on a simulation with 10^5 runs, we obtain $\widehat{QC}_{RA} = 126.32(0.4957)$.

Let $C(\boldsymbol{n}, 100) = 5(n_1 + n_2 + n_3 - 3)$. Then, only those (n_1, n_2, n_3) that satisfy $n_1 + n_2 + n_3 \leq 28$ need to be included in $\omega_{\boldsymbol{n}}$. However, there are too many such (n_1, n_2, n_3) so that screening and ranking of all alternatives would take too much computing time. There are a total of $\sum_{j=3}^{28} \binom{j-1}{3-1} = 3276$ ways to choose strictly positive integers n_1, n_2, n_3 such that $3 \leq n_1 + n_2 + n_3 \leq 28$. Out of these 3276 triples, a total of $28 + 27 + 27 = 82$ contain a pair of 1's; hence, they are not allowable alternatives for bin design. Therefore, there are $3194\ (= 3276 - 82)$ allowable alternative values of $\boldsymbol{n}$. Assuming that both equal width and equal area bin formation rules are used, there are a total of $3194 \times 2 = 6388$ alternative bin designs. Since the total computation time needed to screen and rank the 80 alternatives given by (1.22) below is about 740 seconds on a MacBook Pro (2.4GHz Processor and 4GB RAM), we may need about 16 hours ($\approx 740 \times 6388/80$ seconds) computing time to find the best of the 6388 alternatives. To overcome this problem, we may place realistic upper bounds on the n_i's based on the variance of the X_i's, where the upper bound on n_i should be large if $var(X_i)$ is large. We may also screen alternatives sequentially using the group screening procedure proposed by Nelson et al. (2001), starting with alternatives with small n_i's and stopping when a further increase in each of the n_i's leads to negligible improvement in quality cost. But the statistical properties of such a procedure would be distorted and hard to evaluate if the decision to stop depends on what is observed in the screening process. For this example, we consider a much smaller set of alternative values of $\boldsymbol{n}$ given by $\omega_{\boldsymbol{n}} = \left\{(n_1, n_2, n_3) \in \mathbb{N}^3 : 1 \leq n_1 \leq 5, 1 \leq n_2 \leq 2, 1 \leq n_3 \leq 5, n_i + n_j > 2\ \forall i \neq j\right\}$. Note that we require that at least two of the three component types to have two or more bins; otherwise, we would essentially be performing random assembly. The set of

alternatives is given by

$$\omega_{alt} = \left\{(n_1, n_2, n_3, rule) \in \mathbb{N}^4 : (n_1, n_2, n_3) \in \omega_{\boldsymbol{n}}, 1 \leq rule \leq 2\right\}. \quad (1.22)$$

The alternatives $(3,1,3,2)$ and $(4,1,4,2)$ remain after the screening phase of Procedure III with δ set at 2. The results of the ranking steps are presented in Table 1.9. Since $\overline{TC}^{(2)}_{(4,1,4,2)} = 55.758 < 55.985 = \overline{TC}^{(2)}_{(3,1,3,2)}$, we declare $(4,1,4,2)$ as the best alternative. However, since the practically significant difference was set at $\delta = 2$ and the difference between the estimated total cost of the two alternatives is much less than 2, we probably cannot distinguish the two alternatives with a type I error of $\alpha = 0.05$. Note that the best alternative gives about 55% improvement in total cost and about 80% improvement in quality cost over random assembly.

Table 1.9: Ranking of Alternative Bin Designs (Fortini's Clutch, Truncated Normally Distributed Component Characteristics)

n_1	n_2	n_3	$Rule$	m_2	$\overline{TC}_i^{(2)}$	$std\ error$
3	1	3	2	19	55.985	0.2715
4	1	4	2	26	55.758	0.2445

For the case where X_1, X_2, and X_3 have uniform distributions with the same support as given in (1.21), we consider alternatives with up to seven bins for components 1 and 3, and up to three bins for component 2. The best alternative is $(6,2,5)$ and it gives about 70% improvement in total cost and 90% improvement in quality cost. Details are omitted.

1.3.3 Example 1.5: Fixed Bin Selective Assembly of Wheel Mounting Assembly

We now consider an example with multiple responses and a special form of

assembly response function which is such that certain bin combinations can be ruled out without performing any simulation to evaluate those combinations. This allows us to achieve some savings in simulation effort. The wheel mounting assembly described by Jeang and Chang (2002) has five components, where each component is of a different type. The two quality characteristics of interests are clearances given by $Y_1 = X_2 - X_4$, and $Y_2 = -X_1 - X_2 - X_3 + X_5$.

The component dimensions have truncated normal distributions given by $X_1 \sim TN(5, 0.045, 4.1, 5.09)$, $X_2 \sim TN(8.5, 0.03, 8.44, 8.56)$, $X_3 \sim TN(4, 0.045, 3.91, 4.09)$, $X_4 \sim TN(8.36, 0.04, 8.28, 8.44)$, $X_5 \sim TN(17.7, 0.035, 17.63, 17.77)$.

Let $N = 100$, $\boldsymbol{T} = (0.14, 0.2)^T$, $L(\boldsymbol{Y}, \boldsymbol{T}) = 3000(Y_1 - T_1)^2 + 12000(Y_2 - T_2)^2$, $\sum_{j=1}^{5} c_j^S = 270$, and $C(\boldsymbol{n}, 100) = 40(n_1 + n_2 + n_3 + n_4 + n_5 - 5)^2$. A simulation with 10^4 runs yields $\widehat{QC}_{RA} = 6122.1(0.68)$. This suggests that we should compare all $(n_1, n_2, n_3, n_4, n_5)$ such that $n_1 + n_2 + n_3 + n_4 + n_5 \leq 17$. However, this would give rise to a very large $\omega_{\boldsymbol{n}}$ set; moreover, Program B contains many decision variables even for small values of n_1, n_2, n_3, n_4, and n_5. Therefore, we consider the set of values of $(n_1, n_2, n_3, n_4, n_5)$ given by

$$\omega_{\boldsymbol{n}}' = \left\{(n_1, n_2, n_3, n_4, n_5) \in \mathbb{N}^5 : 1 \leq n_1, n_3 \leq 3, 1 \leq n_2, n_4, n_5 \leq 2\right\}.$$

A few of the elements of $\omega_{\boldsymbol{n}}'$ give rise to FBSA designs that are essentially equivalent to random assembly. These are elements where four or five of the n_i's equal 1, or where only n_4 and exactly one of n_1, n_3, or n_5 is greater than 1. With these bin combinations, component matching does not contribute to the improvement of assembly quality. However, quality might be improved by rejecting assemblies from certain cells. We remove these values of $\boldsymbol{n}$ from $\omega_{\boldsymbol{n}}'$ to get $\omega_{\boldsymbol{n}}$. For simplicity, we shall only consider the

equal width rule. Thus, there are a total of 59 alternatives.

The single alternative $(n_1, n_2, n_3, n_4, n_5) = (2,2,2,1,2)$ remains after screening of the alternatives using Procedure III with $\delta = 100$. It has a total cost of $\overline{TC}^{(1)}_{(2,2,2,1,2)} = 1887.0(15.02)$ and a quality cost of $\widehat{QC}_{(2,2,2,1,2)} = 1247.0(15.02)$. Thus, we achieve a 70% improvement in total cost and an 80% improvement in quality cost.

Remark: In Examples 1.3-1.5, the cost function $C(\boldsymbol{n}, N)$ is somewhat arbitrary. However, the main purpose of these examples is to demonstrate that a substantial improvement in quality cost can be achieved with FBSA, which suggests that implementation of FBSA may be justified in many cases. Another purpose of these examples is to show that good bin designs can be found by restricting attention to only a small set of values of $\boldsymbol{n}$ with small values of n_i that satisfy $C(\boldsymbol{n}, N) \leq QC_{RA}$.

1.3.4 Summary of Examples and Comparison

Examples 1.1 and 1.2 show that DSA is good at achieving significant cost reductions of 50-80% even for small batch sizes ($N = 5$ to $N = 20$). Cost reduction increases with batch size. Note that Example 1.1 considers an assembly with univariate response and univariate component characteristics, whereas Example 1.2 considers an assembly with multivariate response and characteristics.

Examples 1.3-1.5 show that FBSA with optimal bin designs can achieve significant cost reductions and that large batch sizes can be handled with FBSA. They demonstrate that good bin designs can be found by restricting consideration to alternatives with small number of bins for each component type and finding the optimal bin design among the alternatives using a screening and ranking procedure. Reductions in total cost range from 30% for the bimetal thermostat to 70% for Fortini's clutch and the

wheel mounting assembly; reductions in quality cost range from 45% for the bimetal thermostat to 90% for Fortini's clutch. Reductions tend to increase when the variances of the distributions of component characteristics increase.

In the case where we have a moderately large batch size, both DSA and FBSA may be considered. Clearly, DSA will be more costly than FBSA. However, we can expect less improvement with FBSA than DSA (as may be seen by comparing Examples 1.1 and 1.3) since FBSA uses less information than DSA in matching components.

1.4 Robustness of FBSA Design

1.4.1 Robustness to Variation in Batch Size

In practice, the batch size N is determined by customer orders or production planning methods. Thus, we should consider robust bin design with N treated as a noise factor. Note that for an arbitrary batch with number of components in each bin given by the vector $\boldsymbol{\eta}$ defined in (1.15), the total cost is $TC_{\boldsymbol{\eta}}(\boldsymbol{n}, rule, N) = QC^{*}_{\boldsymbol{\eta}}(\boldsymbol{n}, rule, N) + C(\boldsymbol{n}, rule, N)$. This makes explicit the fact that the two sources of variation of the total cost is $\boldsymbol{\eta}$ and N. In other words, we can split the variation in total cost of a batch into two components: variation between different batches due to variation in component manufacturing, and variation due to changes in batch sizes. Thus, the mean and variance of the total cost per batch are given by

$$\mu_{TC}(\boldsymbol{n}, rule) = E\left(TC_{\boldsymbol{\eta}}(\boldsymbol{n}, rule, N)\right) = E\left[E\left(TC_{\boldsymbol{\eta}}(\boldsymbol{n}, rule, N)\middle|N\right)\right], \tag{1.23}$$

and

$$\sigma^2_{TC}(\boldsymbol{n}, rule) = var\left(TC_{\boldsymbol{\eta}}(\boldsymbol{n}, rule, N)\right)$$
$$= E\left[var\left(TC_{\boldsymbol{\eta}}(\boldsymbol{n}, rule, N)\middle|N\right)\right] + var\left[E\left(TC_{\boldsymbol{\eta}}(\boldsymbol{n}, rule, N)\middle|N\right)\right]. \tag{1.24}$$

Equations (1.23) and (1.24) suggest that we may choose the bin design $(\boldsymbol{n}, rule)$ to minimize $\mu_{TC}(\boldsymbol{n}, rule)$ and $\sigma^2_{TC}(\boldsymbol{n}, rule)$ or a composite of both criteria. It may also be of interest to consider the effect of the choice of bin design on $\mu_{QC}(\boldsymbol{n}, rule)$ and $\sigma^2_{QC}(\boldsymbol{n}, rule)$.

In the following examples, we find that the correlation between $\mu_{TC}(\boldsymbol{n}, rule)$ and $\sigma^2_{TC}(\boldsymbol{n}, rule)$ is very high. Thus, it would be appropriate to focus on finding a bin design that minimizes $\mu_{TC}(\boldsymbol{n}, rule)$. Note that optimization of $\mu_{TC}(\boldsymbol{n}, rule)$ can be achieved using Procedure III.

1.4.2 Example 1.6: Mean and Variance of Total Cost, Bimetal Thermostat and Fortini's Clutch

Consider the bimetal thermostat of Example 1.3. Assume $p(N = 50) = p(N = 100) = p(N = 150) = 1/3$. Since the cost of implementing FBSA should be higher for larger batch sizes, we set $C(n_1, n_2, N) = 0.002N(n_1 + n_2 - 2)$, and let ω_{alt} be given by (1.20).

We can estimate $\mu_{TC}(\boldsymbol{n}, rule)$ and $\sigma^2_{TC}(\boldsymbol{n}, rule)$ using simulation data obtained with procedure II. For $N = 50, 100, 150$, and $(\boldsymbol{n}, rule) \in \omega_{alt}$, we run procedure II with $M = 100$ and compute the sample mean and variance. Then, we estimate $\mu_{TC}(\boldsymbol{n}, rule)$ and $\sigma^2_{TC}(\boldsymbol{n}, rule)$ using sample analogues of (1.23) and (1.24).

Assuming that the thicknesses are distributed as in (1.7), we find that for the set of alternatives ω_{alt} considered, the correlation between the estimates of $\mu_{TC}(\boldsymbol{n}, rule)$ and $\sigma^2_{TC}(\boldsymbol{n}, rule)$ is 0.96. Moreover, using the same data, we find that the correlation between the estimates of $\mu_{QC}(\boldsymbol{n}, rule)$ and $\sigma^2_{QC}(\boldsymbol{n}, rule)$ is 0.97. On the other hand, if the

thicknesses are distributed as in (1.8) and ω_{alt} is given by (1.21), the correlation between the estimates of $\mu_{TC}(\boldsymbol{n}, rule)$ and $\sigma^2_{TC}(\boldsymbol{n}, rule)$ is 0.97, and the correlation between the estimates of $\mu_{QC}(\boldsymbol{n}, rule)$ and $\sigma^2_{QC}(\boldsymbol{n}, rule)$ is 0.98. These observations suggests that we need only focus on finding the bin design that minimizes the mean total cost $\mu_{TC}(\boldsymbol{n}, rule)$.

Lastly, we mention that similar results are obtained with the Fortini's Clutch of Example 1.4 for the case where $p(N = 50) = p(N = 100) = p(N = 150) = 1/3$, and $C(n_1, n_2, n_3, N) = 0.05N(n_1 + n_2 + n_3 - 3)$.

1.4.3 Robustness to Misspecification of Distributions

In this section, we consider the robustness of the optimal bin design to misspecification of the distributions of component characteristics. If the distributions of some characteristics are misspecified, the calculated values of the $Q_{i_1 i_2 \cdots i_k}$'s would be in error. Moreover, we would generate the $\boldsymbol{\eta}$ vector in Step 4 of Procedure II from wrong multinomial distributions. Thus, our estimate of the total cost of each alternative and our choice of the optimal bin design would likely be erroneous. However, we think that robustness of bin design to misspecification of distributions is not a significant practical problem. In implementing FBSA, every batch of each component type will be sorted into bins. Data on the total number of components sorted into each bin of a component type can be used to determine whether the specified distribution for that component type is correct (e.g., Pearson goodness-of-fit test). In addition, measurements on a characteristic can be recorded and used to estimate the distribution. After FBSA has been implemented for a long time, there would be an abundance of data to accurately determine the correct distributions. Nevertheless, we present the results of a numerical study of the robustness

of optimal bin designs in Example 1.7.

1.4.4 Example 1.7: Robustness of Optimal Bin Designs for Fortini's Clutch and Bimetal Thermostat

We shall study how cost savings over random assembly with 100% inspection are affected when the optimal bin design is determined from misspecified component distributions. We consider the Fortini's clutch problem discussed in Example 1.4. The distributions of X_1, X_2, and X_3 after component inspection are given by

$X_1 \sim TN(\mu_1, \sigma_1, 55.0521, 55.5279)$, $X_2 \sim TN(\mu_2, \sigma_2, 22.8471, 22.8729)$, and

$$X_3 \sim TN(\mu_3, \sigma_3, 101.3621, 101.8379) \tag{1.25}$$

where $\mu_1, \sigma_1, \mu_2, \sigma_2, \mu_3, \sigma_3$ are unknown. The bounds on $[55.0521, 55.5279]$, $[22.8471, 22.8729]$, and $[101.3621, 101.8379]$ on X_1, X_2, and X_3 are known since these are the specification limits. Assume that to estimate the unknown parameters, we measure the dimensions of $n = 16$ pieces of each component type (hub, ball bearing, and cage) randomly sampled from batches that have not been inspected. Then, Bayesian inferences with the usual noninformative priors give

$$\frac{\mu - \bar{x}}{s/\sqrt{n}} \Big| \, data \sim t_{n-1} \text{ and } \frac{\sigma}{s\sqrt{n-1}} \Big| \, data \sim 1/\sqrt{\chi^2_{n-1}}, \tag{1.26}$$

where $\bar{x}$ is the sample mean and s is the sample standard deviation. Thus, we have

$$0.991 = P\left(-3 \le \frac{\mu - \bar{x}}{s/\sqrt{16}} \le 3\right) = P\left(-0.75 \le \frac{\mu - \bar{x}}{s} \le 0.75\right), \tag{1.27}$$

$$0.991 = P\left(\left(\chi^2_{15,0.9955}\right)^{-1/2} \le \sigma/\left(s\sqrt{15}\right) \le \left(\chi^2_{15,0.0045}\right)^{-1/2}\right) = P(0.673 \le \sigma/s \le 1.823). \tag{1.28}$$

We use the values $\bar{x}_1 = 55.29, s_1 = 0.0793, \bar{x}_2 = 22.86, s_2 = 0.0043, \bar{x}_3 = 101.6, s_3 = 0.0793$ to determine the optimal bin design, which in Example 1.4, was found to be

$(4,1,4,2)$. However, since these estimates are subject to error, we should quantitatively study the robustness of the design.

It is of interest to determine the performance of the optimal bin design at the endpoints and midpoints of the credible intervals given by (1.27) and (1.28). This would require running a 3^6 design. Fortunately, it is possible to reduce the number of runs and still be able to observe the best and worst performance of the optimal design.
The effect of increasing standard deviation of each characteristic is to increase the standard deviation of Y. Thus, we can form a compound factor F_V consisting of the three levels $(0.673,1,1.823)$, where $F_V = 0.673,1,1.823$ correspond to setting the standard deviations of all X_i's to their lower credible limits $(0.673s)$, nominal values (s), and upper credible limits $(1.823s)$ respectively.

It can be shown that Y is a decreasing function of X_1 and X_2, and an increasing function of X_3. Hence, the mean of Y decreases when the means of X_1 and X_2 increase, and the mean of X_3 decrease. Thus, we can form a compound factor F_M consisting of the three levels $(-1,0,1)$, where $F_M = -1,0,1$ correspond to setting the mean of (X_1, X_2, X_3) at $(\bar{x}_1 - 0.75s_1, \bar{x}_2 - 0.75s_2, \bar{x}_3 + 0.75s_3)$, $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$, and $(\bar{x}_1 + 0.75s_1, \bar{x}_2 + 0.75s_2, \bar{x}_3 - 0.75s_3)$ respectively. The levels of F_M and F_V are shown in Table 1.10. Thus, introducing compound factors reduce our experiment from a 3^6 factorial to a 3^2 factorial.

The results of the 3^2 factorial experiment are shown in Table 1.11. The column labelled $\widehat{TC}$ gives an estimate of the total cost of FBSA with bin design $(4,1,4,2)$ and the column labelled $\hat{\sigma}(\widehat{TC})$ gives its standard error. Each estimate $\widehat{TC}$ is obtained by running procedure II 50 times with $M = 10$. The column labelled $\widehat{QC}_{RA}$ gives an estimate of the

Table 1.10: Levels of the Compound Factors F_M and F_V

Level	-1	0	+1
F_M	$\mu_1 = 55.2305$ $\mu_2 = 22.8568$ $\mu_3 = 101.6595$	$\mu_1 = 55.29$ $\mu_2 = 22.86$ $\mu_3 = 101.6$	$\mu_1 = 55.3495$ $\mu_2 = 22.8632$ $\mu_3 = 101.5405$
Level	0.673	1	1.823
F_V	$\sigma_1 = 0.0534$ $\sigma_2 = 0.0029$ $\sigma_3 = 0.0534$	$\sigma_1 = 0.0793$ $\sigma_2 = 0.0043$ $\sigma_3 = 0.0793$	$\sigma_1 = 0.1446$ $\sigma_2 = 0.0078$ $\sigma_3 = 0.1446$

Table 1.11: Estimates of Total Cost of FBSA with Bin Design (4,1,4,2), Cost of Random Assembly, and Percentage Savings of FBSA Over Random Assembly for the 3^2 Experiment

F_M	F_V	$\widehat{TC}$	$\hat{\sigma}(\widehat{TC})$	$\widehat{QC}_{RA}$	$\hat{\sigma}(\widehat{QC}_{RA})$	%Savings
-1	0.673	198.33	0.70	198.13	1.65	-0.1
0	0.673	42.35	0.09	60.91	0.85	30.5
1	0.673	225.54	0.91	235.34	2.01	4.2
-1	1	196.44	0.89	220.13	2.01	10.8
0	1	55.57	0.18	123.81	1.55	55.1
1	1	217.62	0.94	247.80	2.26	12.2
-1	1.823	132.46	0.98	243.91	2.23	45.7
0	1.823	73.86	0.42	223.83	2.18	67.0
1	1.823	140.94	1.03	261.01	2.34	46.0

quality cost of random assembly (100% inspection of output) obtained from 10,000 runs, and the column labelled $\hat{\sigma}(\widehat{QC}_{RA})$ gives its standard error. Finally, the column labeled %Savings gives the quantity $(\widehat{QC}_{RA} - \widehat{TC})/\widehat{QC}_{RA} \times 100$. Comparison of the F_V column and the %Savings column reveals that F_V has a strong linear effect (if we code F_V so that its levels are $-1,0,1$). Thus, underestimation of the component standard deviations is much less serious than overestimation. If the true standard deviations are larger than their estimates (i.e., $F_V = 1.823$ gives the true standard deviations while $F_V = 1$ represent the nominal standard deviations used to derive the optimal bin design), the percentage cost

reduction would be larger than predicted. On the other hand, if the true standard deviations are smaller than estimated (i.e., $F_V = 0.673$ versus $F_V = 1$), the percentage cost reduction would be less than predicted. Comparison of the F_M column with the %Savings column reveals that F_M has a strong quadratic effect. If errors in estimation of the component means act in the same direction to maximize the error in estimating the response mean, the percentage cost reduction can be much less than predicted (i.e., %Savings for $F_M = -1$ and $F_M = 1$ are much less than %Savings for $F_M = 0$). These eyeball analyses can be confirmed by a formal analysis (using a half-normal plot or Lenth's test) of linear and quadratic contrasts obtained by decomposing the two main effects and two two-factor interaction components (Wu and Hamada, 2009). We point out that the high and low levels of F_M and F_V represent worst case scenarios. Overall, Table 1.11 indicates that FBSA would very likely achieve cost reductions over random assembly.

In Table 1.12, we give estimates of the total cost for the bin designs $(4,1,4,2)$ and $(6,2,5,1)$ when X_1, X_2, and X_3 have distributions given by (1.21) and when they have uniform distributions with the same support as in (1.21). Note that for the latter case, we obtained $(6,2,5,1)$ as the optimal bin design, as mentioned at the end of Example 1.4.

Table 1.12: Total Cost Estimates of Bin Designs Optimal for Truncated Normally Distributed and Uniformly Distributed Component Characteristics

	Truncated Normal			Uniform	
	$\widehat{TC}$	$\hat{\sigma}(\widehat{TC})$		$\widehat{TC}$	$\hat{\sigma}(\widehat{TC})$
(4,1,4,2)	55.764	0.056	(4,1,4,2)	101.040	0.207
(6,2,5,1)	75.032	0.025	(6,2,5,1)	77.127	0.100
$\widehat{QC}_{RA}$	126.320	0.496	$\widehat{QC}_{RA}$	278.950	0.754

Table 1.12 shows that, although there is a non-negligible increase in total cost over the minimum attainable cost if the component distributions are misspecified, significant cost savings are still achieved with bins optimal with respect to misspecified distributions. We may use the results of Table 1.11 to predict the cost savings achieved with bin design $(4,1,4,2)$ when the component distributions are uniform. The standard deviations of the uniformly distributed component characteristics are given by $F_V = 1.732$ and the means are given by $F_M = 0$. If we fit a constant mean Kriging model with exponential correlation function (Santner et al., 2003) in the input space (F_M, F_V) to the %Savings data in Table 1.11, we obtain the prediction 65.5. This is very close to the observed %Savings, which is 63.8 $(= (278.95 - 101.04)/278.95 \times 100)$. This suggests that the %Savings depend strongly on the first and second moments of the component distributions.

Finally, we point out that for the bimetal thermostat assembly problem described in Example 1.3, Y is not monotonic in t_1. However, over the sample space $[0.5,1.5] \times [2.5,3.5]$, Y achieves its maximum near $(t_1, t_2) = (0.5,2.5)$ and its minimum at $(t_1, t_2) = (1.5,3.5)$. Therefore, we can use a compound factor F_M for the mean with levels $(\bar{x}_1 - 0.75s_1, \bar{x}_2 - 0.75s_2)$, $(\bar{x}_1, \bar{x}_2)$,and $(\bar{x}_1 + 0.75s_1, \bar{x}_2 + 0.75s_2)$. Results for the bimetal thermostat are similar to the results for Fortini's clutch given above and are omitted.

1.5 Conclusions

The industrial revolution was brought about by the idea of interchangeable parts, which Eli Whitney demonstrated with muskets. This led to the move away from craftsmen to mass production. Each product made by craftsmen is unique in the sense that some components of a product cannot be replaced with the components of another

product of the same type. In contrast, interchangeable parts are parts that can be randomly matched to build an acceptable assembly. Thus, it might seem that selective assembly is a step back towards the days of the craftsmen, and an antithesis of the concept of interchangeable manufacturing. However, we take the position that selective assembly is a valuable quality improvement tool. Selective assembly improves quality by introducing a component sorting and matching step to improve the assembly of interchangeable parts. It does not call for the replacement of interchangeable manufacturing, which is necessary to ensure easy maintenance and good uniformity in performance of products. Rather, selective assembly is a strategy to gain a competitive advantage in product quality, which can be crucial for profit and growth in a business environment that demands perfection in quality.

Implementation of selective assembly requires added steps in the flow of material through the manufacturing facility. Although these extra steps may cause congestion in a mass production facility, they should be easily accommodated in batch production facilities or job shops since production rate requirements are not high in these facilities. Moreover, the numerical examples presented in this chapter also show that larger reductions in quality cost are possible with GSA than with 100 percent inspection and rework.

We developed selective assembly as a tool that can be used to improve the quality of products that are assemblies of one unit of each of a finite number of component types provided that the assembly response function is known. We studied two versions of selective assembly: direct selective assembly and fixed bin selective assembly. For each version, we formulated the problem of matching components to give a batch of

assemblies that minimize expected quality cost as a linear integer program. Realistic examples were given to demonstrate that significant reductions in average quality cost can be achieved.

A few problems require further research. Firstly, extending the selective assembly methods proposed in this chapter to assemblies that consist of multiple components of the same type is one important area for further research. Secondly, for fixed bin selective assembly, methods are needed to determine the optimal bin formation for given number of bins. Thirdly, although it is theoretically possible to consider more than one measurement per component type for fixed bin selective assembly, it seems practically impossible to handle this case since we can have many ways to partition the more-than-one-dimension sample space of the characteristics into sub-regions that define the bins for selective assembly. Thus, approaches to reduce the number of partitions that have to be considered are needed.

1.6 References

Bandelt, H.J., Maas, A., and Spieksma, F.C.R. (2004). "Local Search Heuristics for Multi-Index Assignment Problems with Decomposable Costs," *Journal of the Operational Research Society*, 55, 694-704.

Chan, K.C. and Linn, R.J. (1998). "A Grouping Method for Selective Assembly of Parts of Dissimilar Distributions," *Quality Engineering*, 11(2), 221-234.

Coullard, C.R., Gamble, A.B., and Jones, P.C. (1998). "Matching Problems in Selective Assembly Operations," *Annals of Operations Research*, 76, 95-107.

Dantzig, G.B. and Thapa, M.N. (2003). *Linear Programming 2: Theory and Extensions*, New York: Springer-Verlag.

Eischen, J.W. (1989). "Geometric Nonlinearity in a Bimaterial Strip," *Proceedings of the International Congress of Applied Mechanics*, Beijing, P.R.C., 1-7.

Fang, X.D. and Zhang, Y. (1995). "A New Algorithm for Minimizing the Surplus Parts in Selective Assembly," *Computers and Industrial Engineering*, 28(2), 341-350.

Feng, C. and Kusiak, A. (1997). "Robust Tolerance Design with the Integer Programming Approach," *Journal of Manufacturing Science and Engineering*, 119, 603-610.

Forouraghi, B. (2002). "Worst-Case Tolerance Design and Quality Assurance via Genetic Algorithms," *Journal of Optimization Theory and Applications*, 113(2), 251-268.

Gilbert, K.C. and Hofstra, R.B. (1988). "Multidimensional Assignment Problems," *Decision Sciences*, 19(2), 306-321.

Haley, K.B. (1963). "The Multi-Index Problem," *Operations Research*, 11(3), 368-379.

Iwata,S., Matsui, T., and McCormick, S.T. (1998). "A Fast Bipartite Network Flow Algorithm for Selective Assembly," *Operations Research Letters*, 22, 137-143.

Jeang, A. and Chang C.L. (2002). "Combined Robust Parameter and Tolerance Design Using Orthogonal Arrays," *International Journal of Advanced Manufacturing Technology*, 19, 442-447.

Kannan, S.M., Asha, A., and Jayabalan, V. (2005). "A New Method in Selective Assembly to Minimize Clearance Variation for a Radial Assembly Using Genetic Algorithm," *Quality Engineering*, 17(4), 595-607.

Kannan, S.M. and Jayabalan, V. (2002). "A New Grouping Method for Minimizing the Surplus Parts in Selective Assembly," *Quality Engineering*, 14(1), 67-75.

Kannan, S.M., Jeevanantham, A.K., and Jayabalan, V. (2008). "Modeling and Analysis of Selective Assembly Using Taguchi's Loss Function," *International Journal of Production Research*, 46(15), 4309-4330.

Kapur, K.C., and Cho, B. (1996). "Economic Design of the Specification Region for Multiple Quality Characteristics," *IIE Transactions*, 28(3) 237-248.

Kwon, H., Kim, K., and Chandra, M.J. (1999). "An Economic Selective Assembly Procedure for Two Mating Components with Equal Variance," *Naval Research Logistics*, 46, 809-821.

Lee, S.H. and Kwak, B.M. (2005). "Response Surface Augmented Moment Method for Efficient Reliability Analysis," *Structural Safety*, 28, 261-272.

Lee, S.H. and Chen, W. (2007). "A Comparative Study of Uncertainty Propagation Methods for Black-Box Type Functions," *Proceedings in IDETC/CIE*, Las Vegas, Nevada, 1-10.

Mansoor, E.M. (1961). "Selective Assembly: Its Analysis and Applications," *International Journal of Production Research*, 1(1), 13-24.

Matsura, S. and Shinozaki, N. (2007). "Optimal Binning Strategies under Squared Error Loss in Selective Assembly with Measurement Error," *Communications in Statistics – Theory and Methods*, 36(16), 2863-2876.

Mease, D., Sudjianto, A., and Nair, V.N. (2004). "Selective Assembly in Manufacturing: Statistical Issues and Optimal Binning Strategies," *Technometrics*, 46(2). 165-175.

Nelson, B.L., Swann, J., Goldsman, D., and Song, W. (2001). "Simple Procedures for Selecting the Best Simulated System When the Number of Alternatives is Large," *Operations Research*, 49(6), 950-963.

Pierskalla, W.P. (1968). "The Multidimensional Assignment Problem," *Operations Research*, 16(2), 422-431.

Pugh, G.A. (1992). "Selective Assembly with Components of Dissimilar Variance," *Computers and Industrial Engineering*, 23(4), 487-491.

Queyranne, M. and Spieksma, F.C.R. (1997). "Approximation Algorithms for Multi-Index Transportation Problems with Decomposable Costs," *Discrete Applied Mathematics*, 76, 239-253.

Rinott, Y. (1978). "On Two-Stage Selection Procedures and Related Probability-Inequalities," *Communications in Statistics – Theory and Methods*, **7**(8), 799-811.

Santner, T.J., Williams, B.J., and Notz, W.I. (2003). *The Design and Analysis of Computer Experiments*, New York: Springer-Verlag.

Schmeiser, B. (1982). "Batch Size Effects in the Analysis of Simulation Output," *Operations Research*, 30(3), 556-568.

Schrijver, A. (1986). *Theory of Linear and Integer Programming*, New York: Wiley.

Singh, P.K., Jain, S.C., and Jain, P.K. (2005). "Advanced Optimal Tolerance Design of Mechanical Assemblies with Interrelated Dimension Chains and Process Precision Limits," *Computers in Industry*, 56(2), 179-194.

Spedding, T.A. and Sun, G.Q. (1999). "Application of Discrete Event Simulation to the Activity Based Costing of Manufacturing Systems," *International Journal of Production Economics*, 58, 289-301.

Thesen, A. and Jantayavichit, A. (1999). "Design and Evaluation of a Selective Assembly Station for High Precision Scroll Compressor Shells," *Proceedings of the 1999 Winter Simulation Conference*. Phoenix, Arizona, 694–700.

Tierney, J.W. and Eischen, J.W. (1997). "Residual Stress Analysis of Bimaterial Strips Under Multiple Thermal Loading," *Journal of Electronic Packaging*, 119, 281-287.

Timoshenko, S. (1925). "Analysis of Bi-Metal Thermostats," *Journal of the Optical Society of America*, 11(3), 223-233.

Vanderbei, R.J. (2001). *Linear Programming: Foundations and Extensions* (2nd ed.), New York: Springer-Verlag.

Wu, C., Chen, Z., and Tang, G. (1998). "Component Tolerance Design for Minimum Quality Loss and Manufacturing Cost," *Computers in Industry*, 35, 223-232.

Wu, C.F.J. and Hamada, M.S. (2009). *Experiments: Planning, Analysis, and Optimization* (2nd ed.), New York: Wiley.

Zhang, Y. and Fang, X.D. (1999). “Predict and Assure the Matchable Degree in Selective Assembly via PCI-based Tolerance,” *Journal of Manufacturing Science and Engineering*, 121(3), 494-500.

CHAPTER 2

ROBUST DESIGN OPTIMIZATION WITH QUADRATIC LOSS DERIVED FROM GAUSSIAN PROCESS MODELS

2.1 Introduction

Robust parameter design is a quality improvement methodology for designing products and processes to be insensitive to variation in a set of factors, called noise factors. Noise factors can be controlled during experimentation but not during process operation or product use. The objective of robust parameter design is to find settings of the control factors so that the response is maintained as close to the target as possible under noise variations. As such, widely used average loss criteria for robust design optimization include the mean squared error and variance.

Most of the statistical literature on robust parameter design has focused on methods developed for physical experimentation. However, due to the rapid increase in computing power, computer simulations have become an important tool for product development. Because these simulations can be extremely time-consuming, metamodels are needed to facilitate exploration and optimization of simulators. To construct a metamodel, a training data set is acquired by running a computer experiment. In many cases, Gaussian process models are utilized in statistical metamodel building (Sacks et al. 1989). A Gaussian process model can be interpreted as a prior for the unknown function that a simulator represents. Parameters of a Gaussian process prior are estimated using training data and the prior process is updated using the same data to yield a posterior process. Statistical inferences on the simulator are then made using the posterior process.

There are many techniques for searching for robust settings based on linear models of the response (e.g., Koksoy and Doganaksoy 2003), including ones that take into account model estimation uncertainty (Miró-Quesada and Del Castillo 2004). Methods for constructing confidence regions for optimal robust settings of linear models have also been developed (Myers et al. 1997; Ginsburg and Ben-Gal 2006). Peterson and Kuhn (2005) give a method for constructing simultaneous confidence intervals for the minimum mean squared error loss at various distances from the design center. Myers et al. (1997) give methods to construct approximate prediction and tolerance limits on the response.

However, most of the techniques developed for linear models cannot be applied to Gaussian process models, and many are only applicable to linear models with specific forms. Hence, approaches for finding robust design solutions and quantifying uncertainty in the estimated average loss are needed for Gaussian process models. Apley and Chen (2006) propose an approximate method for constructing confidence intervals for a loss function suitable for smaller-the-better quality characteristics. Williams et al. (2000) and Lehman et al. (2004) introduce expected improvement (EI) criteria for finding robust settings. Chang et al. (1999) and Chang et al. (2001) describe the robust design of a femoral component for total hip arthroplasty. Their criteria for robust design are derived by treating the posterior mean as if it were the true response function. Despite the vast literature on robust design, it seems that methods for constructing credible intervals for quadratic loss derived from either linear models or Gaussian process models are lacking. Moreover, uncertainty about the true response function is often ignored in robust design. Optimization is typically based on the average quadratic loss derived as if the fitted

model/ posterior mean were the true function. This chapter demonstrates that such a practice can be deficient and introduces better methods for robust design optimization.

Although computer experiments and physical experiments are very different in nature, linear regression and Gaussian process modeling are actually closely related. The posterior predictive process for Gaussian process models contains the predictive process for linear models as a special case. We propose the use of a general expected quadratic loss criterion, where the expectation is taken with respect to the noise factors and the posterior process. We shall develop saddlepoint-approximation-based methods to construct credible intervals for the loss that are applicable to Gaussian process and linear models. The criterion and its credible interval allow us to properly take into account uncertainty about the true response function in performing robust design optimization. In examples involving real data, we demonstrate the superiority of the proposed criterion over the commonly used criterion that ignores response function uncertainty. We also demonstrate the impressive accuracy of the Lugannani-Rice saddlepoint approximation.

The rest of the chapter is organized as follows. Section 2.2 reviews the Gaussian process modeling framework for computer experiments. Section 2.3 derives the proposed expected quadratic loss criterion. Sections 2.4 and 2.5 present methods for constructing credible intervals for quadratic loss derived from posterior normal and t processes respectively. Section 2.6 discusses a method for partially accounting for uncertainty in the correlation parameters and sequential design using quantiles. Section 2.7 gives three examples. Concluding remarks are given in Section 2.8.

2.2 Gaussian Process Modeling

In this section, we briefly review the main ideas of Gaussian process modeling of

simulators. It is assumed that the output of a simulator $y(\cdot)$ can be modeled by

$$Y(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta} + G(\boldsymbol{x}), \tag{2.1}$$

where $\boldsymbol{x} \in \chi \subset \mathbb{R}^m$, χ is the design region, $\boldsymbol{f}(\boldsymbol{x}) = \left(1, f_1(\boldsymbol{x}), \dots, f_{p-1}(\boldsymbol{x})\right)^T$, $\boldsymbol{\beta} = \left(\beta_0, \beta_1, \dots, \beta_{p-1}\right)^T$, and $G(\boldsymbol{x})$ is a zero mean stationary Gaussian process. Given any two points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, the covariance of $Y(\boldsymbol{x}_i)$ and $Y(\boldsymbol{x}_j)$ is given by $cov\left[Y(\boldsymbol{x}_i), Y(\boldsymbol{x}_j)\right] = cov\left[G(\boldsymbol{x}_i), G(\boldsymbol{x}_j)\right] = \sigma^2 R(\boldsymbol{x}_i, \boldsymbol{x}_j)$, where $R(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is the correlation function. The most commonly used correlation function is the Gaussian correlation function

$$R(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp\left[-\sum_{u=1}^{m} \theta_u \left(x_{iu} - x_{ju}\right)^2\right], \tag{2.2}$$

where x_{iu} is the uth element of $\boldsymbol{x}_i$, and $\theta_u > 0, u = 1, \dots, m$.

In a computer experiment, the output is observed at N values of inputs given by the rows of $\boldsymbol{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_N)^T$. This yields a vector $\boldsymbol{Y}$ of observed outputs. The matrix $\boldsymbol{X}$ is called the design, and the choice of each $\boldsymbol{x}_i$ is restricted to χ. Based on data from the experiment, a best linear unbiased estimator of the output can be constructed and is given by

$$\hat{y}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^T\widehat{\boldsymbol{\beta}} + \boldsymbol{r}(\boldsymbol{x})^T\boldsymbol{R}^{-1}(\boldsymbol{Y} - \boldsymbol{F}\widehat{\boldsymbol{\beta}}), \tag{2.3}$$

where $\boldsymbol{r}(\boldsymbol{x}) = \left(R(\boldsymbol{x}, \boldsymbol{x}_1), \dots, R(\boldsymbol{x}, \boldsymbol{x}_N)\right)^T$, $\boldsymbol{F} = \left(\boldsymbol{f}(\boldsymbol{x}_1), \dots, \boldsymbol{f}(\boldsymbol{x}_N)\right)^T$, $\boldsymbol{R} = \left(R(\boldsymbol{x}_i, \boldsymbol{x}_j)\right)_{N\times N}$, and

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{Y}. \tag{2.4}$$

In (2.3), $\boldsymbol{R}$, $\boldsymbol{r}$, and $\widehat{\boldsymbol{\beta}}$ depend on the correlation function $R(\cdot)$, which in turn, depends on the correlation parameters $\boldsymbol{\theta}$. One approach to estimating $\boldsymbol{\theta}$ is the maximum likelihood method, i.e., to maximize the likelihood

$$\psi(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2) = [(2\pi\sigma^2)^N|\boldsymbol{R}|]^{-1/2}\exp[-(\boldsymbol{Y} - \boldsymbol{F}\boldsymbol{\beta})^T\boldsymbol{R}^{-1}(\boldsymbol{Y} - \boldsymbol{F}\boldsymbol{\beta})/2\sigma^2].$$

It turns out that given $\boldsymbol{\theta}$, $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}}$ is given in (2.4), and $\sigma^2 = \hat{\sigma}^2 = \left(\boldsymbol{Y} - \boldsymbol{F}\widehat{\boldsymbol{\beta}}\right)^T \boldsymbol{R}^{-1}(\boldsymbol{Y} - \boldsymbol{F}\widehat{\boldsymbol{\beta}})/N$ maximizes $\psi(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2)$. Thus, the maximum likelihood estimate (MLE) $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained by maximizing $\psi\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\beta}}, \hat{\sigma}^2\right)$, which is equivalent to minimizing

$$S(\boldsymbol{\theta}) = Nlog(\hat{\sigma}^2) + log(|\boldsymbol{R}|). \tag{2.5}$$

Because the output of deterministic simulations lacks random error, it seems more natural to view model (2.1) as representing the prior distribution for the simulator. Using a weak prior for $\boldsymbol{\beta}|\boldsymbol{\theta}, \sigma^2$, $p(\boldsymbol{\beta}|\boldsymbol{\theta}, \sigma^2) \propto 1$, it can be shown that (Handcock and Stein 1993; O'Hagan 1994)

$$Y(\cdot)|\boldsymbol{Y}, \boldsymbol{\theta}, \sigma^2 \sim GP\left(\mu(\cdot|\boldsymbol{\theta}), V(\cdot,\cdot|\boldsymbol{\theta}, \sigma^2)\right), \tag{2.6}$$

where $GP(\mu(\cdot|\boldsymbol{\theta}), V(\cdot,\cdot|\boldsymbol{\theta}, \sigma^2))$ denotes a Gaussian process with mean function $\mu(\cdot|\boldsymbol{\theta})$ and covariance function $V(\cdot,\cdot|\boldsymbol{\theta}, \sigma^2)$; $\mu(\cdot|\boldsymbol{\theta}) = \hat{y}(\boldsymbol{x})$, and

$$V(\cdot,\cdot|\boldsymbol{\theta}, \sigma^2) =$$

$$\sigma^2 \left\{R\left(\boldsymbol{x}^i, \boldsymbol{x}^j\right) - \boldsymbol{r}\left(\boldsymbol{x}^i\right)^T \boldsymbol{R}^{-1}\boldsymbol{r}\left(\boldsymbol{x}^j\right) + \left[\boldsymbol{f}\left(\boldsymbol{x}^i\right) - \boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{r}\left(\boldsymbol{x}^i\right)\right]^T (\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}\left[\boldsymbol{f}\left(\boldsymbol{x}^j\right) - \boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{r}\left(\boldsymbol{x}^j\right)\right]\right\}. \tag{2.7}$$

It is common to base inference on (2.6) using $\widehat{\boldsymbol{\theta}}$ in place of $\boldsymbol{\theta}$ and $\tilde{\sigma}^2 = \frac{N}{N-p}\hat{\sigma}^2$ in place of σ^2, i.e.,

$$Y(\cdot)|\boldsymbol{Y}, \widehat{\boldsymbol{\theta}}, \tilde{\sigma}^2 \sim GP\left(\mu\left(\cdot|\widehat{\boldsymbol{\theta}}\right), V\left(\cdot,\cdot|\widehat{\boldsymbol{\theta}}, \tilde{\sigma}^2\right)\right). \tag{2.8}$$

A fully Bayesian approach would place a prior on $(\boldsymbol{\theta}, \sigma^2)$ and integrate out $(\boldsymbol{\theta}, \sigma^2)$ from the product of the prior and $Y(\cdot), \boldsymbol{Y}|\boldsymbol{\theta}, \sigma^2$. If we use the prior $p(\sigma^2|\boldsymbol{\theta}) \propto \sigma^{-2}$ and integrate out σ^2 from (2.6), we obtain a t-process. In particular, we obtain (Handcock and Stein 1993; O'Hagan 1994)

$$Y(\cdot)|\boldsymbol{Y},\boldsymbol{\theta}\sim st\left(\mu(\cdot|\boldsymbol{\theta}),\tilde{V}(\cdot,\cdot|\boldsymbol{\theta})\right), \tag{2.9}$$

where $st\left(\mu(\cdot|\boldsymbol{\theta}),\tilde{V}(\cdot,\cdot|\boldsymbol{\theta})\right)$ denotes a Student t process with mean function $\mu(\cdot|\boldsymbol{\theta})$, and covariance function $\frac{N-p}{N-p-2}\tilde{V}(\cdot,\cdot|\boldsymbol{\theta}) = \frac{N-p}{N-p-2}V(\cdot,\cdot|\boldsymbol{\theta},\tilde{\sigma}^2)$. Inferences can be made using (2.9) if we replace $\boldsymbol{\theta}$ with $\widehat{\boldsymbol{\theta}}$, i.e.,

$$Y(\cdot)|\boldsymbol{Y},\widehat{\boldsymbol{\theta}}\sim st\left(\mu\left(\cdot|\widehat{\boldsymbol{\theta}}\right),\tilde{V}\left(\cdot,\cdot|\widehat{\boldsymbol{\theta}}\right) = V\left(\cdot,\cdot|\widehat{\boldsymbol{\theta}},\tilde{\sigma}^2\right)\right). \tag{2.10}$$

Finally, let $p(\boldsymbol{\theta})$ denote the prior distribution of $\boldsymbol{\theta}$. Then, it can be shown that the posterior distribution of $\boldsymbol{\theta}$ is given by

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) \propto |\boldsymbol{R}|^{-1/2}|\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F}|^{-1/2}[\boldsymbol{Y}^T(\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{F}(\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{R}^{-1})\boldsymbol{Y}]^{-(n-p)/2}p(\boldsymbol{\theta}). \tag{2.11}$$

While integrating out $\boldsymbol{\theta}$ from (2.9) is analytically intractable, it is possible to integrate out $\boldsymbol{\theta}$ numerically.

Note that by setting $R\left(\boldsymbol{x}^i,\boldsymbol{x}^i\right) = 1$, $R\left(\boldsymbol{x}^i,\boldsymbol{x}^j\right) = 0$ $(i \neq j)$, $\boldsymbol{r}\left(\boldsymbol{x}^i\right) = \boldsymbol{0}$, $\boldsymbol{R} = \boldsymbol{I}$, (2.9) gives the posterior predictive surface for the usual linear model with independent and identically distributed normal errors and prior $p(\boldsymbol{\beta},\sigma^2) \propto \sigma^{-2}$ (Zellner and Chetty 1965). Thus, the Bayesian predictive process for linear models is a special case of the predictive process for Gaussian process models.

2.3 Quadratic Loss

In robust parameter design applications (Myers and Montgomery 2002; Wu and Hamada 2009), the simulator output $y(\cdot)$ represents a quality characteristic of interest and $\boldsymbol{x} = (\boldsymbol{x}_c^T,\boldsymbol{x}_e^T)^T$, where $\boldsymbol{x}_c$ and $\boldsymbol{x}_e$ are the vector of control and noise factor settings respectively. Let the quality cost of a system with quality characteristic $y(\boldsymbol{x})$ be given by

$L[y(\boldsymbol{x})]$. Then, the average quality cost is given by

$$Q_{y(\cdot)}(\boldsymbol{x}_c) = E_{\boldsymbol{X}_e}\left\{L\left[y\begin{pmatrix}\boldsymbol{x}_c\\ \boldsymbol{X}_e\end{pmatrix}\right]\right\}, \tag{2.12}$$

where the expectation is taken with respect to the distribution of the noise factors $\boldsymbol{X}_e$. Our objective is to find an $\boldsymbol{x}_c$ such that $Q_{y(\cdot)}(\boldsymbol{x}_c)$ is minimized. A commonly used approach to this problem would be to replace the unknown function $y(\cdot)$ by the posterior mean function $\mu(\cdot)$ (for simplicity of notation, we omit reference to the MLE $\widehat{\boldsymbol{\theta}}$) and then optimize $Q_{\mu(\cdot)}(\boldsymbol{x}_c)$. However, this method ignores the uncertainty about the true function $y(\cdot)$, which is captured by the posterior process (2.10). Each realization of (2.10) is a function on χ and the true function $y(\cdot)$ can be any one of these realizations. Hence, the uncertainty about the average loss at $\boldsymbol{x}_c$ is represented by a random variable given by

$$Q_{Y(\cdot)}(\boldsymbol{x}_c)\left|\left(\boldsymbol{Y}, \widehat{\boldsymbol{\theta}}\right) = E_{\boldsymbol{X}_e}\left\{L\left[Y\begin{pmatrix}\boldsymbol{x}_c\\ \boldsymbol{X}_e\end{pmatrix}\right]\right| Y(\cdot)\right\}\left|\left(\boldsymbol{Y}, \widehat{\boldsymbol{\theta}}\right).\right. \tag{2.13}$$

It follows that we should base our inference on the expected quality cost obtained by taking the expectation of (2.13) with respect to (2.10). This yields

$$\bar{Q}(\boldsymbol{x}_c) = E\left(E_{\boldsymbol{X}_e}\left\{L\left[Y\begin{pmatrix}\boldsymbol{x}_c\\ \boldsymbol{X}_e\end{pmatrix}\right]\right| Y(\cdot)\right\}\left|\boldsymbol{Y}, \widehat{\boldsymbol{\theta}}\right.\right), \tag{2.14}$$

which is a criterion that appropriately accounts for the uncertainty about $y(\cdot)$.

The focus of this chapter shall be on the widely used quadratic loss $L(y) = \delta(y - T)^2$, where δ is a constant and T is the target. This loss function arises naturally with *nominal-the-best* quality characteristics (Wu and Hamada 2009). However, it can also be used with *smaller-the-better* and *larger-the-better* characteristics. Suitable targets for these two classes of responses are $T = \min_{\boldsymbol{x}\in\chi}[\hat{y}(\boldsymbol{x})]$ and $T = \max_{\boldsymbol{x}\in\chi}[\hat{y}(\boldsymbol{x})]$ respectively (Peterson and Kuhn 2005). Moreover, some positive-valued smaller-the-better characteristics have a natural target of 0.

We shall assume in the rest of the chapter that the distribution of the noise factors is discrete, or can be discretized, with mass function $\left(p(\boldsymbol{x}_{e1}), \ldots, p(\boldsymbol{x}_{el})\right)$. For a univariate continuous distribution, we can find an l-point discrete approximation that has the same first $2l-1$ moments as the continuous distribution (Miller and Rice 1983). On the other hand, for an M-dimensional continuous distribution that is a product of independent univariate marginals, we can discretize the distribution in the following manner. First, discretize each of the marginals using the method proposed by Miller and Rice (1983) to obtain $\left(p(\tau_{j1}), \ldots, p\left(\tau_{jl_j}\right)\right), j = 1, \ldots, M$. Then, form the M-dimensional discrete distribution with support $\times_{j=1}^{M} \left\{\tau_{j1}, \ldots, \tau_{jl_j}\right\}$, where $\times$ denotes Cartesian product, and probability mass $p\left[\left(\tau_{1i_1}, \ldots, \tau_{Mi_M}\right)\right] = p\left(\tau_{1i_1}\right) \cdots p\left(\tau_{Mi_M}\right)$. This guarantees that the first $2l_j - 1$ moments of the jth marginal of the discrete multivariate distribution match the corresponding moments of the jth marginal of the continuous multivariate distribution. Moreover, independence is preserved with this discrete approximation.

Let $\boldsymbol{Y}(\boldsymbol{x}_c)^T = \left[Y\begin{pmatrix}\boldsymbol{x}_c \\ \boldsymbol{x}_{e1}\end{pmatrix}, \ldots, Y\begin{pmatrix}\boldsymbol{x}_c \\ \boldsymbol{x}_{el}\end{pmatrix}\right]$, $\boldsymbol{\mu}(\boldsymbol{x}_c)^T = \left\{\mu\left[\begin{pmatrix}\boldsymbol{x}_c \\ \boldsymbol{x}_{e1}\end{pmatrix}\middle|\, \widehat{\boldsymbol{\theta}}\right], , \ldots, \mu\left[\begin{pmatrix}\boldsymbol{x}_c \\ \boldsymbol{x}_{el}\end{pmatrix}\middle|\, \widehat{\boldsymbol{\theta}}\right]\right\}$,

and suppose that we can write

$$Q_{Y(\cdot)}(\boldsymbol{x}_c) = E_{\boldsymbol{X}_e}\{L[\boldsymbol{Y}(\boldsymbol{x}_c)]|Y(\cdot)\} = [\boldsymbol{Y}(\boldsymbol{x}_c) - \boldsymbol{T}]^T \boldsymbol{H}[\boldsymbol{Y}(\boldsymbol{x}_c) - \boldsymbol{T}], \tag{2.15}$$

where $\boldsymbol{H}$ is some $l \times l$ symmetric matrix. This implies that

$$Q_{\mu(\cdot)}(\boldsymbol{x}_c) = [\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]^T \boldsymbol{H}[\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]. \tag{2.16}$$

For the loss function $L(Y) = (Y - T)^2$, we substitute $\boldsymbol{H} = diag\left(p(\boldsymbol{x}_{e1}), \ldots, p(\boldsymbol{x}_{el})\right)$, and $\boldsymbol{T} = T\boldsymbol{1}_l$, where $\boldsymbol{1}_l$ is a $l \times 1$ matrix of 1's in (2.15). On the other hand, for cases where only the variance is of interest, we have

$$\boldsymbol{H} = (\boldsymbol{I} - \boldsymbol{1}_l \boldsymbol{p}^T)^T diag(\boldsymbol{p})(\boldsymbol{I} - \boldsymbol{1}_l \boldsymbol{p}^T) \tag{2.17}$$

and $\boldsymbol{T} = \boldsymbol{0}$, where $\boldsymbol{I}$ is the $l \times l$ identity matrix, and $\boldsymbol{p} = \left(p(\boldsymbol{x}_{e1}), \ldots, p(\boldsymbol{x}_{el})\right)^T$.

The loss functions introduced in the previous paragraph give rise to positive semidefinite $\boldsymbol{H}$. However, for any smooth loss function, the average loss at $\boldsymbol{x}_c$ for a given realization $Y(\cdot)|\boldsymbol{Y}, \widehat{\boldsymbol{\theta}}$ of the posterior process can be approximated by a (possibly non-positive semidefinite) quadratic form in degenerate normal or multivariate t random variables obtained by a Taylor series expansion of the loss function. Although this chapter focuses on $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ given by (2.15), the methods developed herein can be modified in a straightforward fashion to accommodate this more general case.

If $v = N - p$ is large, (2.8) is an approximation of (2.10). Taking expectation of (2.15) with respect to (2.8), we obtain the *robust design criterion*

$$\bar{Q}_1(\boldsymbol{x}_c) = E\left\{[\boldsymbol{Y}(\boldsymbol{x}_c) - \boldsymbol{T}]^T \boldsymbol{H}[\boldsymbol{Y}(\boldsymbol{x}_c) - \boldsymbol{T}] | \boldsymbol{Y}, \widehat{\boldsymbol{\theta}}, \tilde{\sigma}^2\right\}$$

$$= [\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]^T \boldsymbol{H}[\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}] + trace[\boldsymbol{H}\boldsymbol{V}(\boldsymbol{x}_c)], \quad (2.18)$$

where $\left(\boldsymbol{V}(\boldsymbol{x}_c)\right)_{ij} = V\left(\begin{pmatrix}\boldsymbol{x}_c \\ \boldsymbol{x}_{ei}\end{pmatrix}, \begin{pmatrix}\boldsymbol{x}_c \\ \boldsymbol{x}_{ej}\end{pmatrix} \middle| \widehat{\boldsymbol{\theta}}, \tilde{\sigma}^2\right)$, where $V\left(\cdot,\cdot|\widehat{\boldsymbol{\theta}}, \tilde{\sigma}^2\right)$ is given by (2.7).

On the other hand, if $v = N - p > 2$ but is small, inferences should be based on (2.10). It follows from Theorem B.2 of Appendix B.1 that

$$\bar{Q}_2(\boldsymbol{x}_c) = [\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]^T \boldsymbol{H}[\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}] + \frac{1}{1-2/v} trace[\boldsymbol{H}\boldsymbol{V}(\boldsymbol{x}_c)]. \quad (2.19)$$

Note that if $\boldsymbol{H}$ is positive semidefinite (and $\boldsymbol{H} \neq \boldsymbol{0}$), then $\bar{Q}_2(\boldsymbol{x}_c) - Q_{\mu(\cdot)}(\boldsymbol{x}_c) > 0$ since $trace[\boldsymbol{H}\boldsymbol{V}(\boldsymbol{x}_c)] > 0$. Thus, the $Q_{\mu(\cdot)}$ criterion, which ignores uncertainty about the true response function, always gives an *over-optimistic* estimate of the true loss. In most cases, (2.18) and (2.19) are complicated nonlinear functions of $\boldsymbol{x}_c$. We search for optimal solutions with respect to these two criteria using the pattern search algorithm proposed by Lewis and Torczon (1999) (this algorithm is available in Matlab; see Appendix B.4 for

parameter values used to control the search). We also find the MLE $\widehat{\boldsymbol{\theta}}$ using this algorithm because it is stable and does not require derivative information. However, since (2.18) and (2.19) are not convex in general, multiple starting points are used.

2.4 Construction of Credible Intervals when $v = \infty$ with the Lugannani-rice saddlepoint approximation

To construct credible intervals for $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ given in (2.15), we need its cumulative distribution function (cdf). However, this is hard to obtain. This section introduces approximations of the distribution of $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ for the case where (2.8) is used for inference. The *Lugannani-Rice approximation* (Lugannani and Rice 1980; Butler 2007) and its numerical inversion are discussed. We shall develop this approach in complete generality, i.e., *the method is applicable to any positive semidefinite* $\boldsymbol{V}(\boldsymbol{x}_c)$ *and any symmetric* $\boldsymbol{H}$. We shall also briefly discuss the chi-square and lognormal approximations, which are obtained by matching moments. These two approximations are appropriate when $\boldsymbol{H}$ is positive semidefinite.

For positive definite $\boldsymbol{V}(\boldsymbol{x}_c)$, the cumulant-generating function (cgf), which is the key ingredient for saddlepoint approximations, is well-known. However, the special case of singular $\boldsymbol{V}(\boldsymbol{x}_c)$ deserves attention. For example, when $\begin{pmatrix}\boldsymbol{x}_c \\ \boldsymbol{x}_{ei}\end{pmatrix}$ appears in the design matrix, we would have $Y\begin{pmatrix}\boldsymbol{x}_c \\ \boldsymbol{x}_{ei}\end{pmatrix} = \mu\begin{pmatrix}\boldsymbol{x}_c \\ \boldsymbol{x}_{ei}\end{pmatrix}$ with probability 1. Thus, when some of the noise factors are qualitative, $\boldsymbol{V}(\boldsymbol{x}_c)$ would be singular for all $\boldsymbol{x}_c$'s that appear in the design. Moreover, to improve computational stability, it would be desirable to utilize an expression for the cgf that does not involve the inversion of $\boldsymbol{V}(\boldsymbol{x}_c)$. Hence, we derive an expression for the cgf of $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ that holds for any positive semidefinite $\boldsymbol{V}(\boldsymbol{x}_c)$ and any

symmetric $\boldsymbol{H}$ in Appendix B.1.

Let $\boldsymbol{\Gamma\Lambda\Gamma}^T$ be the spectral decomposition of $\boldsymbol{V}(\boldsymbol{x}_c)^{1/2}\boldsymbol{H}\boldsymbol{V}(\boldsymbol{x}_c)^{1/2}$, where $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues $\lambda_1 \leq \cdots \leq \lambda_l$. It follows from Theorem B.1 in Appendix B.1 that the cumulant-generating function of $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ is given by

$$K(t) = t[\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]^T\boldsymbol{H}[\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}] - \frac{1}{2}\sum_{i=1}^{L}\ln(1 - 2t\lambda_i) + \sum_{i=1}^{L}\frac{2t^2u_i^2}{1-2t\lambda_i}, \tag{2.20}$$

where u_i is the ith element of the vector $\boldsymbol{\Gamma}^T\boldsymbol{V}(\boldsymbol{x}_c)^{1/2}\boldsymbol{H}[\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]$. Note that if $\boldsymbol{V}(\boldsymbol{x}_c)$ is singular, the support of $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ would not be $(0, \infty)$ in general, but rather (ε, ∞), where $\varepsilon \geq 0$ if $\boldsymbol{H}$ is positive semidefinite (and $\boldsymbol{H} \neq \boldsymbol{0}$).

A credible interval for $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ can be constructed based on its cgf (2.20) using the Lugannani-Rice saddlepoint approximation for the cdf. The Lugannani-Rice approximation is given by

$$\hat{F}(x) = \Phi(w) + \phi(w)\left(\frac{1}{w} - \frac{1}{u}\right) \tag{2.21}$$

for $x \in (\varepsilon, \infty) \setminus K_1$, where $K_1 = \bar{Q}_1(\boldsymbol{x}_c)$ (from (2.18)) is the first cumulant of $Q_{Y(\cdot)}(\boldsymbol{x}_c)$, $\Phi(x)$ and $\phi(x)$ are the cdf and probability density function (pdf) of the standard normal distribution,

$$w = sign(t_*)\sqrt{2[t_*x - K(t_*)]}, \tag{2.22}$$

$$u = t_*\sqrt{K''(t_*)}, \tag{2.23}$$

and t_* is the saddlepoint, which is defined implicitly by

$$K'(t_*) = x, t_* \in I = (-|2\min(\lambda_1, 0)|^{-1}, [2\max(\lambda_l, 0)]^{-1}). \tag{2.24}$$

Expressions for $K'(t)$ and $K''(t)$ can be obtained from (B.2) and (B.3) by replacing $\boldsymbol{\mu}$ with $\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}$ and $\boldsymbol{V}$ with $\boldsymbol{V}(\boldsymbol{x}_c)$. An approximate $100(1-\alpha)\%$ credible interval for $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ is given by $[LCL, UCL]$, where LCL and UCL satisfy the equations

$\hat{F}(LCL) = \alpha/2$, and $\hat{F}(UCL) = 1 - \alpha/2$. (2.25)

Equation (2.25) can be solved using numerical methods such as Newton's method. In particular, we can find the roots of the equation $\hat{F}(x) = \alpha$ by iterating according to

$$x_{i+1} = x_i - \left[\hat{F}(x_i) - \alpha\right]/\hat{f}(x_i), \tag{2.26}$$

where

$$\hat{f}(x) = \hat{F}'(x) = \frac{1}{\sqrt{2\pi K''(t_*)}} \exp\left(-\frac{w^2}{2}\right)\left\{1 + \frac{1}{u^2} + \frac{K'''(t_*)}{2u[K''(t_*)]^{1.5}} - \frac{u}{w^3}\right\}, \tag{2.27}$$

and $K'''(t)$ is given by (B.4) in Appendix B.1. Note that for each iteration given by (2.26), we need to solve (2.24) for t_*. This can be done by iterating according to

$$t_{i+1} = t_i - [K'(t_i) - x]/K''(t_i). \tag{2.28}$$

A good starting point x_0 is essential for the fast convergence of (2.26). Credible limits constructed using moment-matching methods developed below provide good starting points. A starting point is also needed for (2.28). Since $K'(\cdot)$ is a strictly increasing function from I onto (ε, ∞), and $K'(0) = K_1$, we can choose t_0 from $I^- = (-|2\min(\lambda_1, 0)|^{-1}, 0)$ or $I^+ = (0, [2\max(\lambda_l, 0)]^{-1})$ according to whether $x < K_1$ or $x > K_1$.

When $\boldsymbol{V}(\boldsymbol{x}_c)$ is positive definite, $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ is distributed as a linear combination of noncentral chi-squares. Let $\lambda_1, \ldots, \lambda_s$ be the distinct eigenvalues of $\boldsymbol{HV}(\boldsymbol{x}_c)$, and $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_s$ be $l \times l$ positive semidefinite matrices of rank r_i such that $\boldsymbol{E}_i\boldsymbol{E}_j = \boldsymbol{0}$ for all $j \neq i$, $\boldsymbol{E}_i^2 = \boldsymbol{E}_i$, $\sum_{i=1}^{s} r_i = l$, and $\boldsymbol{HV}(\boldsymbol{x}_c) = \sum_{i=1}^{s} \lambda_j \boldsymbol{E}_j$. Baldessari (1967) proves that

$$Q_{Y(\cdot)}(\boldsymbol{x}_c) = [\boldsymbol{Y}(\boldsymbol{x}_c) - \boldsymbol{T}]^T \boldsymbol{H}[\boldsymbol{Y}(\boldsymbol{x}_c) - \boldsymbol{T}] \sim \sum_{j=1}^{s} \lambda_j \chi^2_{r_j}(\delta_j), \tag{2.29}$$

where $\chi^2_{r_j}(\delta_j)$ is a noncentral chi-square variate with r_j degrees of freedom and

noncentrality parameter $\delta_j = [\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]^T \boldsymbol{E}_j \boldsymbol{V}(\boldsymbol{x}_c)^{-1} [\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]$. For positive semidefinite $\boldsymbol{H}$, the eigenvalues of $\boldsymbol{HV}(\boldsymbol{x}_c)$ are nonnegative. Thus, we can approximate $\sum_{j=1}^{s} \lambda_j \chi_{r_j}^2(\delta_j)$ by $g\chi_h^2$, where g and h are chosen so that the first two moments of $g\chi_h^2$ match those of $\sum_{j=1}^{s} \lambda_j \chi_{r_j}^2(\delta_j)$. The first moment (first cumulant) of (2.29) is $K_1 = \bar{Q}_1(\boldsymbol{x}_c)$ given in (2.18) and it follows from Corollary B.1 of Appendix B.1 that the second central moment (second cumulant) is

$$K_2 = 2trace([\boldsymbol{HV}(\boldsymbol{x}_c)]^2) + 4[\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]^T \boldsymbol{HV}(\boldsymbol{x}_c)\boldsymbol{H}[\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]. \quad (2.30)$$

Since $E(g\chi_h^2) = gh$ and $var(g\chi_h^2) = 2g^2h$, we set $g = K_2/2K_1$ and $h = 2K_1^2/K_2$. Thus, we arrive at the approximation

$$Q_{Y(\cdot)}(\boldsymbol{x}_c) \sim^a (K_2/2K_1)\chi^2_{2K_1^2/K_2}, \quad (2.31)$$

where $\sim^a$ denotes approximately equal in distribution.

Approximation (2.31) can be viewed as an application of Patnaik's and Satterthwaite's approximations. We first approximate of $\chi_{r_j}^2(\delta_j)$ by $\rho_j \chi_{v_j}^2$ (Patnaik 1949), where ρ_j and v_j are chosen so that the first two moments of $\chi_{r_j}^2(\delta_j)$ and $\rho_j \chi_{v_j}^2$ match. Then, we approximate $\sum_{j=1}^{s} \lambda_j \rho_j \chi_{v_j}^2$ by $g\chi_h^2$ (Box 1954), where g and h are chosen so that the first two moments of $\sum_{j=1}^{s} \lambda_j \rho_j \chi_{v_j}^2$ and $g\chi_h^2$ match. Note that even if $\boldsymbol{V}(\boldsymbol{x}_c)$ is positive semidefinite, the first two cumulants of $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ are still given by (2.18) and (2.30), as shown in Corollary B.1 in Appendix B.1. Thus, it seems reasonable to use (2.31) even when $\boldsymbol{V}(\boldsymbol{x}_c)$ is singular. However, it should be noted that when $\boldsymbol{V}(\boldsymbol{x}_c)$ is singular, $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ has support on (ε, ∞), whereas the chi-square approximation has support on $(0, \infty)$.

The saddlepoint approximation is expected to be more accurate than the chi-

square approximation. However, inversion of (2.21) is more time-consuming than computation of the quantiles of (2.31). Nevertheless, development of (2.31) serves an important purpose: we shall use the credible limits given by (2.31) as starting points for (2.26) to find LCL and UCL given in (2.25).

Finally, as an alternative to the saddlepoint and chi-square approximations discussed above, we shall also consider using three other approximations. The first is a lognormal approximation obtained by matching the first three moments of the lognormal distribution with the first three moments of $Q_{Y(\cdot)}(\boldsymbol{x}_c)$. This approximation is given by

$$[\boldsymbol{Y}(\boldsymbol{x}_c) - \boldsymbol{T}]^T \boldsymbol{H} [\boldsymbol{Y}(\boldsymbol{x}_c) - \boldsymbol{T}] \sim^a \ln[N(\gamma, \tau^2)] + \varepsilon', \tag{2.32}$$

where $\tau^2 = \ln(s_*)$; $s_* > 1$ is the unique positive solution of the equation $s^3 + 3s^2 - 4 - K_3^2/K_2^3 = 0$; K_r is the rth cumulant of $Q_{Y(\cdot)}(\boldsymbol{x}_c)$, which can be obtained from (B.5) in Appendix B.1; $\gamma = 0.5\left[\ln\left(\frac{K_2}{\exp(\tau^2)-1}\right) - \tau^2\right]$; and $\varepsilon' = K_1 - \exp(\gamma + \tau^2/2)$. Note that $\varepsilon' \neq \varepsilon$ in general. The second approximation is the noncentral chi-square approximation introduced by Liu et al. (2009). However, in our simulations, we have found that it often reduces to Pearson's three-moment approximation (Imhof 1961), where a shift is introduced into (2.31). The third approximation is introduced by Solomon and Stephens (1977). It is obtained by matching the first three moments of $g(\chi_h^2)^e$ and $Q_{Y(\cdot)}(\boldsymbol{x}_c)$.

2.5 Construction of Credible Intervals when v is Small

When (2.10) is used for inference, the first two central moments of $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ are given by (see Theorem B.2 in Appendix B.1) $M_1 = c_1 a_1 + b_1, v > 2$, and

$$M_2 = (c_2 - c_1^2)a_1^2 + c_2 a_2 + c_1 b_2, v > 4, \tag{2.33}$$

$$a_r = 2^{r-1}(r-1)!\, trace\{[\boldsymbol{H}\boldsymbol{V}(\boldsymbol{x}_c)]^r\},$$

$b_r = 2^{r-1}r!\,[\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]^T[\boldsymbol{H}\boldsymbol{V}(\boldsymbol{x}_c)]^{r-1}\boldsymbol{H}[\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]$, and $c_r = [(1-2/v)(1-4/v)\cdots(1-2r/v)]^{-1}$. Note that $M_1 = \bar{Q}_2(\boldsymbol{x}_c)$, where $\bar{Q}_2(\boldsymbol{x}_c)$ is given by (2.19).

It is evident that $M_1 > K_1$ and $M_2 > K_2$, where K_1 and K_2 are given by (2.18) and (2.30) respectively. Thus, the first two central moments of $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ are inflated by uncertainty about σ^2. As such, for small v, credible intervals constructed based on (2.8) would be too narrow and centered too low.

We propose the following saddlepoint-based approximation to overcome this problem. When (2.10) holds, $\boldsymbol{Y}(\boldsymbol{x}_c) =^d \boldsymbol{Z}/\sqrt{\chi_v^2/v} + \boldsymbol{\mu}(\boldsymbol{x}_c)$, where $=^d$ denotes equality in distribution, $\boldsymbol{Z}\sim N\big(\boldsymbol{0}, \boldsymbol{V}(\boldsymbol{x}_c)\big)$, χ_v^2 is a chi-squared random variable with v degrees of freedom, and $\boldsymbol{Z}$ and χ_v^2 are independent. Thus,

$$Q_{Y(\cdot)}(\boldsymbol{x}_c) = [\boldsymbol{Y}(\boldsymbol{x}_c) - \boldsymbol{T}]^T\boldsymbol{H}[\boldsymbol{Y}(\boldsymbol{x}_c) - \boldsymbol{T}]$$

$$=^d \left[\boldsymbol{Z}/\sqrt{\Xi} + \boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}\right]^T\boldsymbol{H}\left[\boldsymbol{Z}/\sqrt{\Xi} + \boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}\right], \tag{2.34}$$

$\Xi = \chi_v^2/v$ and

$$\left\{\left[\boldsymbol{Z}/\sqrt{\Xi} + \boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}\right]^T\boldsymbol{H}\left[\boldsymbol{Z}/\sqrt{\Xi} + \boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}\right]\middle|\, \Xi = \xi\right\} =^d (\boldsymbol{Z}_*)^T\boldsymbol{H}\boldsymbol{Z}_*, \tag{2.35}$$

where $\boldsymbol{Z}_*\sim N\big(\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}, \xi^{-1}\boldsymbol{V}(\boldsymbol{x}_c)\big)$. The Lugannani-Rice approximation to the cdf of (2.35) is

$$\hat{F}(x|\Xi = \xi) = \Phi(w) + \phi(w)\left(\frac{1}{w} - \frac{1}{u}\right) \tag{2.36}$$

for $x \in (\varepsilon, \infty) \setminus K_1$, where $K_1 = \xi^{-1}trace[\boldsymbol{H}\boldsymbol{V}(\boldsymbol{x}_c)] + [\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]^T\boldsymbol{H}[\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}]$; w, u, t_* are given by (2.22)-(2.24); I is given by (2.24) with λ_1 and λ_l replaced by $\xi^{-1}\lambda_1$ and $\xi^{-1}\lambda_l$; and $K(t)$, $K'(t)$, and $K''(t)$ are given by (B.1)-(B.3) with $\boldsymbol{\mu}$ replaced by $\boldsymbol{\mu}(\boldsymbol{x}_c) - \boldsymbol{T}$, and $\boldsymbol{V}$ replaced by $\xi^{-1}\boldsymbol{V}(\boldsymbol{x}_c)$, the latter implying that each λ_i should be replaced by $\xi^{-1}\lambda_i$, and each u_i^2 should be replaced by $\xi^{-1}u_i^2$. Now, multiplying $\hat{F}(x|\Xi = \xi)$ by the

pdf $g(\xi)$ of Ξ and integrating over all ξ gives an approximation

$$\hat{F}(x) = \int_0^\infty \hat{F}(x|\Xi = \xi) g(\xi)\, d\xi \tag{2.37}$$

of the cdf of (2.34). However, to compute the integral, we need to utilize a numerical method or a simulation method. The use of simulation is not practical since computing $\hat{F}(x|\Xi = \xi)$ is difficult and we want to minimize the number of such computations. It seems reasonable to approximate (2.37) using the expectation of $\hat{F}(x|\Xi = \xi)$ with respect to an n-point discrete approximation of Ξ, which is essentially the Gaussian quadrature method of integration (Stoer and Bulirsch 1993). By choosing the n support points $\xi_1, \dots, \xi_n$ and the probability masses $p_1, \dots, p_n$ using the method described in Miller and Rice (1983), we can match the first $2n - 1$ moments of the discrete approximation with the corresponding moments of the distribution of Ξ. Thus, if $\hat{F}(x|\Xi = \xi)$ is closely approximated by a polynomial with degree $2n - 1$ in ξ, the approximation

$$\hat{F}(x) \approx \hat{F}_n(x) = \sum_{i=1}^n p_i \hat{F}(x|\Xi = \xi_i) \tag{2.38}$$

would be good. Since we want to minimize the number of evaluations of $\hat{F}(x|\Xi = \xi)$ and at the same time, to approximate (2.37) with good accuracy, it seems reasonable to use a 3-point or 4-point approximation. For the 4-point approximation, it can be shown that $\xi_1, \xi_2, \xi_3, \xi_4$ are the (distinct and positive) roots of the equation

$$v^3 x^4 - 4(v+6)v^2 x^3 + 6(v+6)(v+4)vx^2 + (v+6)(v+4)(v+2)(1-4x) = 0, \tag{2.39}$$

whereas p_1, p_2, p_3, p_4 are given by

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ \xi_1 & \xi_2 & \xi_3 & \xi_4 \\ \xi_1^2 & \xi_2^2 & \xi_3^2 & \xi_4^2 \\ \xi_1^3 & \xi_2^3 & \xi_3^3 & \xi_4^3 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 + 2/v \\ (1 + 2/v)(1 + 4/v) \end{pmatrix}. \tag{2.40}$$

Note that the values $\xi_1, \xi_2, \xi_3, \xi_4$, and p_1, p_2, p_3, p_4 need to be computed only once for a given value of v, and they can be used repeatedly to construct credible intervals for $Q_{Y(\cdot)}(\boldsymbol{x}_c)$ for various values of $\boldsymbol{x}_c$. In the examples in this chapter, we shall use the 4-point approximation. Equation (2.28) and a modified version of (2.26) given by

$$x_{i+1} = x_i - \left[\sum_{i=1}^{n} p_i \hat{F}(x|\Xi = \xi_i) - \alpha\right] / \sum_{i=1}^{n} p_i \hat{f}(x|\Xi = \xi_i), \tag{2.41}$$

where $\hat{f}(x|\Xi = \xi_i)$ is given by (2.27) with $\boldsymbol{V}(\boldsymbol{x}_c)$ replaced by $\xi_i^{-1}\boldsymbol{V}(\boldsymbol{x}_c)$, can be used to numerically solve $\hat{F}_n(x) = \alpha$ for x.

For comparison with the saddlepoint-based approximation, we shall also construct credible intervals using the chi-square and lognormal approximations obtained from (2.31) and (2.32) with M_1, M_2, and M_3^2/M_2^3 (see Theorem B.2 in Appendix B.1) replacing K_1, K_2, and K_3^2/K_2^3. The approximations introduced by Liu et al. (2009) and Solomon and Stephens (1977) shall be modified in a similar manner. The confidence limits obtained from the chi-square approximation can be used as starting points in (2.41).

2.6 Uncertainty in Correlation Parameters, Sequential Design, and Multiple Responses

This section introduces a method to partially take into account uncertainty in the correlation parameters $\boldsymbol{\theta}$ and discuss the problems of sequential design and multiple responses.

It is possible to account for uncertainty in $\boldsymbol{\theta}$ in the expected loss criterion and the construction of credible intervals by simply taking expectation of (2.19) and (2.38) with respect to the density given by (2.11). In doing this, numerical methods need to be used to find the constant of proportionality in (2.11) and to integrate the product of (2.11) and

(2.19) and also the product of (2.11) and (2.38). This is cumbersome for multidimensional $\boldsymbol{\theta}$. However, we may partially account for uncertainty in $\boldsymbol{\theta}$ by supposing that the correlation function is known to be

$$R(\boldsymbol{x}_i, \boldsymbol{x}_j) = \theta_0^{\left[\sum_{u=1}^{m} \hat{\theta}_u (x_{iu} - x_{ju})^2\right]}, \tag{2.42}$$

where $\theta_0 \in (0,1)$ and $\hat{\theta}_u$ is the MLE of θ_u conditioned on $\theta_0 = e^{-1}$. This is more general than the form in (2.2). Thus, the only uncertain parameter in the correlation function is θ_0, which has the posterior density (2.11). Since there is only a single uncertain parameter, we can easily perform the numerical integrations mentioned above. To perform this task, we first compute the first $2n_0 - 1$ moments of the posterior distribution of θ_0 using numerical integration and then construct an n_0-point discrete approximation $\{(\theta_0^1, p_0^1), \ldots, (\theta_0^{n_0}, p_0^{n_0})\}$ of the distribution of θ_0 (Miller and Rice 1983). This would allow the efficient computation of the $\bar{Q}_3$ criterion

$$\bar{Q}_3(\boldsymbol{x}_c) = \sum_{i=1}^{n_0} p_0^i \left[\bar{Q}_2(\boldsymbol{x}_c) | \theta_0^i\right] \tag{2.43}$$

and the saddlepoint-based approximation

$$\hat{F}_{n \times n_0}(x) = \sum_{i=1}^{n_0} p_0^i \sum_{j=1}^{n} p_j \hat{F}\left(x | \Xi = \xi_j, \theta_0^i\right), \tag{2.44}$$

where $\hat{F}\left(x | \Xi = \xi_j, \theta_0^i\right)$ is computed using (2.21). In the examples, we shall set $n_0 = 5$ and assume that θ_0 is *a priori* uniformly distributed on $[0.001, 0.99]$. This is to ensure that the posterior distribution is proper. While (2.43) and (2.44) does not fully take into account uncertainty in the correlation parameters, they do give an idea of the amount of uncertainty present. Obvious modifications of (2.43) and (2.44) can be used to fully account for uncertainty in $\boldsymbol{\theta}$ if we have a sample from the posterior distribution of $\boldsymbol{\theta}$.

Credible intervals are valuable in the sequential design of computer experiments. For instance, they can be used to determine whether the average loss is poorly estimated

and where the next design point should be added. In the latter application, instead of using an expected improvement criterion as in Williams et al. (2000) and Lehman et al. (2004), we propose the Q_α criterion, which is "to minimize the α quantile Q_α of $Q_{Y(\cdot)}$". Using this criterion, the next design point $\boldsymbol{x}_{N+1} = \begin{pmatrix} \boldsymbol{x}_{c,N+1} \\ \boldsymbol{x}_{e,N+1} \end{pmatrix}$ is chosen by setting $\boldsymbol{x}_{c,N+1}$ equal to the point with the minimum α quantile Q_α of $Q_{Y(\cdot)}$. For this application, the quickness and deterministic nature of the proposed method for constructing credible intervals make it more appealing than Monte Carlo simulation. The Q_α criterion takes into account both location and dispersion of the distribution of the average loss and is an optimistic decision rule. Note that optimism is good in the sequential design context since we do not want to miss potentially good control factor settings. The choice of $\boldsymbol{x}_{e,N+1}$ can be determined by maximizing the minimum pairwise distance between design points (which is the approach taken in Example 2.2) or some other space-filling or optimal design criterion. We shall demonstrate that this simple criterion is quite effective in finding the true optimum.

The proposed robust design criterion and method for constructing credible intervals can be applied to cases where there are multiple responses, the loss is a quadratic form in the responses, and the modelling method of Conti et al. (2009) is employed. Conditioned on the covariances between responses at a fixed $\boldsymbol{x}$ and the spatial correlation parameters $\boldsymbol{\theta}$, the posterior distribution of $\boldsymbol{Y}^q(\boldsymbol{x}_c) = \left(\boldsymbol{Y}_1(\boldsymbol{x}_c), \ldots, \boldsymbol{Y}_q(\boldsymbol{x}_c)\right)$ is a matrix-variate normal distribution. In this case, the saddlepoint approximation is directly applicable. On the other hand, conditioned on only $\boldsymbol{\theta}$, $\boldsymbol{Y}^q(\boldsymbol{x}_c)$ has a matric-variate t distribution (Box and Tiao 1973), which is a mixture of a matrix-variate normal and an inverse wishart distribution. The conditioning method (2.38) for approximating the cdf

can be applied in conjunction with samples from the inverse Wishart distribution.

2.7 Examples

2.7.1 Example 2.1: Analysis of Springback Data

A finite-element-analysis (FEA) simulation experiment was run by Chen and Koc (2007) to study the effects of *blankholder force* (d_1), *friction* (d_2), and *material* (d_3) on springback variation of a drawing process. A fourth factor, *part thickness* (d_4), is a noise factor whose effect on the variation of springback is of interest. Two of the responses are the *springback of wall opening angle* Y_1 and the *springback of the flange angle* Y_2. The 30-run design matrix and observed output for this experiment are given in Appendix B.2. We shall analyze Y_1 and Y_2 separately. In our analysis, we use the coded levels $x_j = \left(d_j - \min_{1\le i\le N} d_{ji}\right)/\left(\max_{1\le i\le N}\ d_{ji} - \min_{1\le i\le N} d_{ji}\right)$, where d_{ji} is the value of d_j for the ith experiment run. Thus, the design region is $\chi = [0,1]^4$. We assume the distribution of *part thickness* is discrete with support on the coded levels $\{0, 1/9,\ 2/9, \dots, 1\}$ and probability mass of $1/10$ at each support point. For both responses, the correlation function (2.2) is used and the prior mean is a constant β_0.

2.7.1.1 Response Y_1

Minimizing $S(\boldsymbol{\theta})$ in (2.5) gives $\widehat{\boldsymbol{\theta}} = (2.88, 0.34748, 0.48417, 13.476)$, $\hat{\beta}_0 = 15.426$, and $\hat{\sigma}^2 = 10.041$. We shall consider loss functions given by $\boldsymbol{H} = 0.1(\boldsymbol{I} - 0.1\boldsymbol{J})$, $T = 0$ (variance loss) and $\boldsymbol{H} = 0.1\boldsymbol{I}$, $T = 15$ (target loss) and draw inferences using (2.8), (2.10), (2.43) and (2.44). We denote the minimizer of $Q_{\mu(\cdot)}(\boldsymbol{x}_c)$ by $\boldsymbol{x}_c^0$, and the minimizer of $\bar{Q}_i(\boldsymbol{x}_c)$ by $\boldsymbol{x}_c^i$, $i = 1,2,3$. Table 2.1 gives $\boldsymbol{x}_c^0$ and $\boldsymbol{x}_c^1$, the expected loss $\bar{Q}_1$ at these two control factor settings, and the 90% credible intervals constructed by inverting

Table 2.1: Optimal Robust Settings, Expected Loss, and 90% Credible Intervals Based on the Posterior Normal Process for Response Y_1

Criterion		x_1	x_2	x_3	$\bar{Q}_1$	LCL	UCL
Target=15	$\boldsymbol{x}_c^0$	0.6010	0.1120	0.1067	2.0407	0.3908	5.5559
	$\boldsymbol{x}_c^1$	0.5273	0.4111	0.2594	1.5739	0.4427	3.8815
Variance	$\boldsymbol{x}_c^0$	1	0.5177	0.1858	1.8352	0.2617	5.299
	$\boldsymbol{x}_c^1$	0.5292	0.5172	0.5347	0.9321	0.1910	2.2028

Table 2.2: Optimal Robust Settings, Expected Loss, and 90% Credible Intervals for Response Y_1

Criterion		x_1	x_2	x_3	$\bar{Q}_3(\bar{Q}_2)$	LCL	UCL
Target=15	$\boldsymbol{x}_c^0$	0.60102	0.11204	0.10675	2.3638 (2.179)	0.3896 (0.3859)	6.6856 (6.12)
	$\boldsymbol{x}_c^2$	0.52492	0.43018	0.26924	1.7499 (1.6521)	0.478 (0.4647)	4.4182 (4.1264)
	$\boldsymbol{x}_c^3$	0.51983	0.48428	0.28989	1.7426	0.5137	4.2313
Variance	$\boldsymbol{x}_c^0$	1	0.51778	0.18581	2.0788 (1.9638)	0.26638 (0.2588)	6.1997 (5.8245)
	$\boldsymbol{x}_c^2$	0.52536	0.53568	0.55338	0.97177 (0.96526)	0.19368 (0.1992)	2.3699 (2.3062)
	$\boldsymbol{x}_c^3$	0.51379	0.6058	0.6115	0.95963	0.20794	2.2652

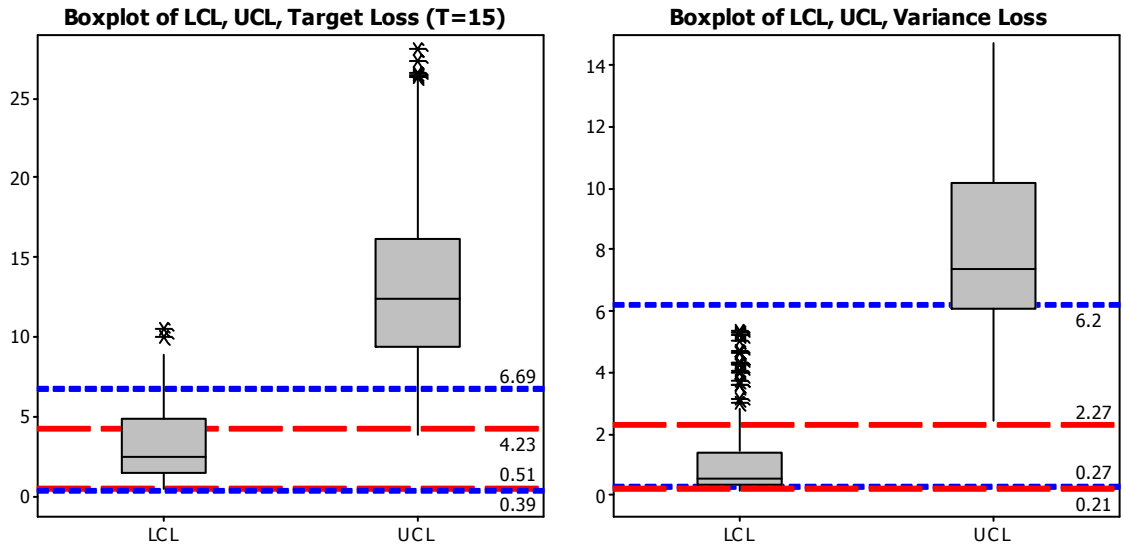

Figure 2.1: Boxplots of LCL and UCL for 90% Credible Intervals for Target Loss (Left) and Variance Loss (Right), Response Y_1

(2.21). Table 2.2 gives $\boldsymbol{x}_c^0$, $\boldsymbol{x}_c^2$, $\boldsymbol{x}_c^3$, the $\bar{Q}_2$ (in parentheses) and $\bar{Q}_3$values, and 90%

credible intervals constructed using (2.38) (in parentheses) and (2.44) respectively. We see that the optimal robust settings obtained from $Q_{\mu(\cdot)}$ and $\bar{Q}_i, i = 1,2,3$ are very different. The credible intervals for the average loss at $\boldsymbol{x}_c^1, \boldsymbol{x}_c^2$, and $\boldsymbol{x}_c^3$ are much narrower than the credible intervals for the average loss at $\boldsymbol{x}_c^0$. In fact, $Q_{\mu(\cdot)}(\boldsymbol{x}_c^0) = 0.1738$ for target loss and $Q_{\mu(\cdot)}(\boldsymbol{x}_c^0) = 0.0995$ for variance loss, which are below the lower credible limits. Since $Q_{\mu(\cdot)}(\boldsymbol{x}_c)$ ignores the uncertainty about the simulator output, using this criterion without confirmation experiments can be dangerous. In contrast, the optimal settings and credible intervals for the average loss do not change much when (2.10) is used instead of (2.8) (uncertainty in σ^2 is taken into account) or when (2.43) and (2.44) is used instead of (2.10) (uncertainty in $\boldsymbol{\theta}$ is partially accounted for).

To determine how good $\boldsymbol{x}_c^0$ and $\boldsymbol{x}_c^3$ are compared to other control factor settings, we compare the 90% credible intervals for the loss at each of these points to the 90% credible intervals for points on a grid of 125 points. Figure 2.1 display box plots of the credible limits for the 125 grid points and each of the two loss functions, where the credible limits are computed based on (2.44). The long dashed lines plot the credible limits for the loss at $\boldsymbol{x}_c^3$ ((0.51,4.23) for target loss and (0.21,2.27) for variance loss), and the short dashed lines plot the credible limits for the loss at $\boldsymbol{x}_c^0$ ((0.39,6.69) for target loss and (0.27,6.20) for variance loss). We see that for the target loss, the upper credible limit (i.e., 4.23) of the loss at $\boldsymbol{x}_c^3$ is higher than the lower credible limit of the loss at more than 50% (middle line of the boxplot of LCL) of the 125 grid points. For the variance loss, the upper credible limit (i.e., 2.27) of the loss at $\boldsymbol{x}_c^3$ is higher than the lower credible limit of the loss at more than 75% (upper edge of the box in the boxplot of LCL) of the 125 grid points. Because the interval for $\boldsymbol{x}_c^3$ overlaps with the intervals for many of the

125 points, the true average loss is not well estimated by $\bar{Q}_3$. However, in this situation, the optimal setting identified using the $\bar{Q}_3$ criterion considerably outperforms the optimal setting found using the $Q_{\mu(\cdot)}$ criterion.

As a general rule, we can consider any point $\boldsymbol{x}_c$ with credible intervals that contain the optimal value $\bar{Q}_3^*$ of $\bar{Q}_3$ as a potential optimal robust setting. Thus, if among points on a grid, there are many with credible interval that contains $\bar{Q}_3^*$, the true average loss is not sufficiently well estimated by $\bar{Q}_3$ to allow us to distinguish between many of the $\boldsymbol{x}_c$'s and $\boldsymbol{x}_c^3$. However, $\bar{Q}_3$ is still preferable as a criterion to $Q_{\mu(\cdot)}$ since uncertainty about the true function penalizes $\bar{Q}_3$ but not $Q_{\mu(\cdot)}$. Another application of credible intervals computed for many $\boldsymbol{x}_c$'s is that if $\boldsymbol{x}_c^3$ is deemed infeasible, we may choose from among the feasible $\boldsymbol{x}_c$ the one with the tightest credible interval that contains the optimal value of $\bar{Q}_3$.

2.7.1.2 Approximate Credible Intervals for Average Quadratic Loss, Response Y_1

We now assess the accuracy of the credible intervals constructed from the Lugannani-Rice, chi-square, lognormal, power chi-square (Solomon and Stephens 1977), and noncentral chi-square (Liu et al. 2009) approximations for the posterior t process. We construct 90% credible intervals of the loss on a grid of 125 points on the control factor space $[0,1]^3$ and the points $\boldsymbol{x}_c^0$ and $\boldsymbol{x}_c^2$ for the loss functions $\boldsymbol{H} = 0.1(\boldsymbol{I} - 0.1\boldsymbol{J})$, $T = 0$ and $\boldsymbol{H} = 0.1\boldsymbol{I}$, $T = 15$. For each loss function and point in the control factor space, a total of 20,000 loss values are simulated and the *fraction of loss values* that fall into each of the five approximate credible intervals are recorded. In the simulations, we generate $\boldsymbol{Y}(\boldsymbol{x}_c)^T = \left[Y\begin{pmatrix}\boldsymbol{x}_c\\ \boldsymbol{x}_{e1}\end{pmatrix}, \dots, Y\begin{pmatrix}\boldsymbol{x}_c\\ \boldsymbol{x}_{el}\end{pmatrix}\right]$ from either (2.8) or (2.10) and compute

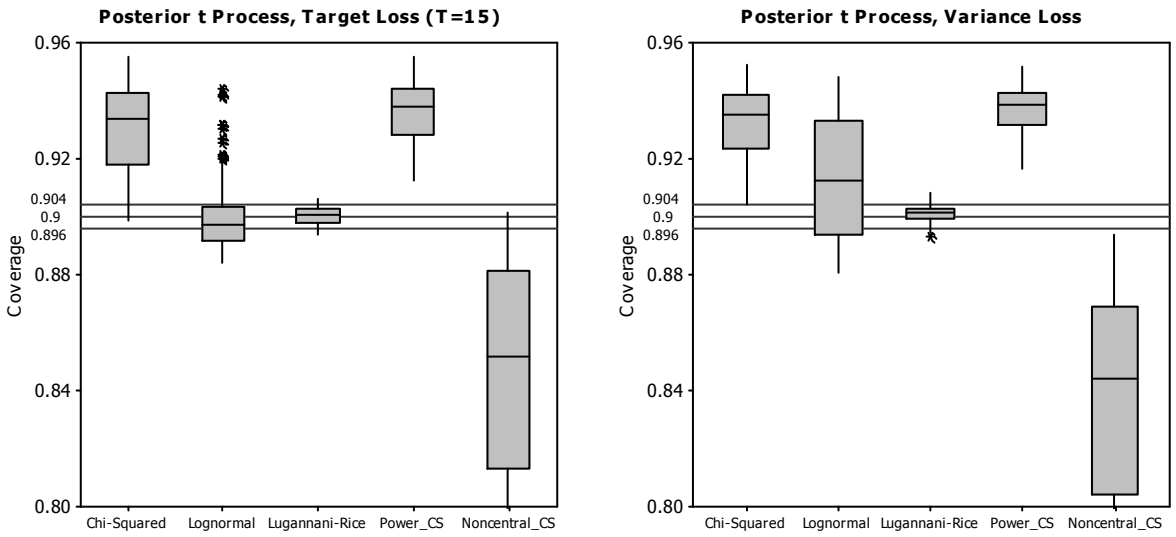

Figure 2.2: Empirical Coverage of 90% Credible Intervals for Response Y_1, Posterior t Process

$Q_{Y(\cdot)}(\boldsymbol{x}_c) = [\boldsymbol{Y}(\boldsymbol{x}_c) - \boldsymbol{T}]^T \boldsymbol{H} [\boldsymbol{Y}(\boldsymbol{x}_c) - \boldsymbol{T}]$. For simplicity, we will refer to the fraction of loss values that fall into an interval as the "empirical coverage".

Figure 2.2 shows boxplots of the empirical coverage at the 127 points for the posterior t process with 90% being the nominal level. In the figure, horizontal lines are drawn at two standard errors of empirical coverage from the nominal level, where the standard errors are computed assuming that the true coverage equals the nominal level. We see that the Lugannani-Rice credible intervals outperform all other approximations. The noncentral chi-square approximation (Liu et al. 2009) performs much worse than the others (note that the lower whisker actually extends beyond the figure window). The chi-square and power chi-square intervals are conservative, whereas the lognormal intervals have highly variable coverage. Boxplots for the posterior normal process are given in the Appendix B.3. For a nominal level of 95%, similar conclusions can be drawn from boxplots (not shown in this chapter) of the empirical coverages. Thus, this simulation indicates that Lugannani-Rice credible intervals are superior to credible intervals

obtained via moment matching techniques.

Note that comparison of credible intervals based on (2.38), and (2.44) allow us to determine how uncertainty in θ_0 affects the credible interval size. We have found that for all loss functions and grid points considered above, the credible intervals constructed using (2.44) are 0 to 10 percent longer than the credible intervals constructed using (2.38), with a median of about 5 percent. Thus, we may miss some potentially promising $\boldsymbol{x}_c$, i.e., those with credible intervals that contain the optimal $\bar{Q}_3$ value if we use (2.38) instead of (2.44) to construct credible intervals. However, Table 2.2 suggests that we do not lose much by using the $\bar{Q}_2$ criterion in place of the $\bar{Q}_3$ criterion (since $\boldsymbol{x}_c^2$ and $\boldsymbol{x}_c^3$ give similar $\bar{Q}_3$ values and the credible intervals for the average loss at these points are similar).

It is difficult to compare the cost of computing credible intervals using Monte Carlo simulation with the cost of computing credible intervals using the saddlepoint-based approximations because it is hard to determine the number of Monte Carlo runs needed to construct intervals that achieve the same accuracy as the saddlepoint method. However, it is computationally cheap to construct credible intervals using the saddlepoint-based approximations. We observed that each interval constructed using (2.38) takes a fraction of a second (we stop the iterations on the x_i's when the change is less than 10^{-4} and we stop the iterations on the t_i's when the change is less than 10^{-6}). This is similar to the cost of a few hundred simulation runs. Moreover, it is known that the Lugannani-Rice approximation maintains high *relative accuracy* at the tails (Daniels 1987). On the other hand, the performance of Monte Carlo simulation in estimating quantiles tends to deteriorate as α approaches 0 or 1.

2.7.1.3 Response Y_2

For this response, we consider two loss functions: $\boldsymbol{H} = 0.1(\boldsymbol{I} - 0.1\boldsymbol{J})$, $T = 0$ and $\boldsymbol{H} = 0.1\boldsymbol{I}$, $T = 10$. Results similar to Y_1 are obtained. For target loss, the values of $\boldsymbol{x}_c^2$ and $\boldsymbol{x}_c^3$ are very different. However, they appear to be equally desirable (in terms of $\bar{Q}_3$ and credible intervals). The optimal settings are $\boldsymbol{x}_c^0 = (0,0.854,0.204)$, $\boldsymbol{x}_c^2 = (0.286,0.351,0.045)$, and $\boldsymbol{x}_c^3 = (0,0.078,0.556)$. The $\bar{Q}_3$ values at $\boldsymbol{x}_c^0$, $\boldsymbol{x}_c^2$, and $\boldsymbol{x}_c^3$ are 0.5439, 0.3204, and 0.3197 respectively. Lastly, the 90% credible intervals (constructed based on (2.44)) for the average loss at $\boldsymbol{x}_c^0$, $\boldsymbol{x}_c^2$ and $\boldsymbol{x}_c^3$ are $(0.082,1.607)$, $(0.060,0.858)$ and $(0.045,0.773)$ respectively.

2.7.2 Example 2.2: Sequential Robust Parameter Design

This example illustrates the use of quantiles computed based on (2.44) for sequential robust design optimization as described in Section 2.6. We assume that the true function is

$y(x_1, x_2) = [4 - 2.1(-1 + 2x_1)^2 + (-1 + 2x_2)^3/3](-1 + 2x_1)^2 + (-1 + 2x_1)(-1 + 2x_2) + [-4 + 4(-1 + 2x_2)^2](-1 + 2x_2)^2$.

The initial design is shown in Figure 2.3. It is assumed that x_2 is a noise factor with a continuous uniform distribution on $[0,1]$. A 6-point discrete approximation of this continuous distribution is employed. We attempt to find the optimum value of x_1 with respect to target loss with target 0.5. The true average loss function has two local minima, one at $x_1 = 0.229$ with average loss $AL = 0.202$ and the other at $x_1 = 0.764$ with $AL = 0.242$.

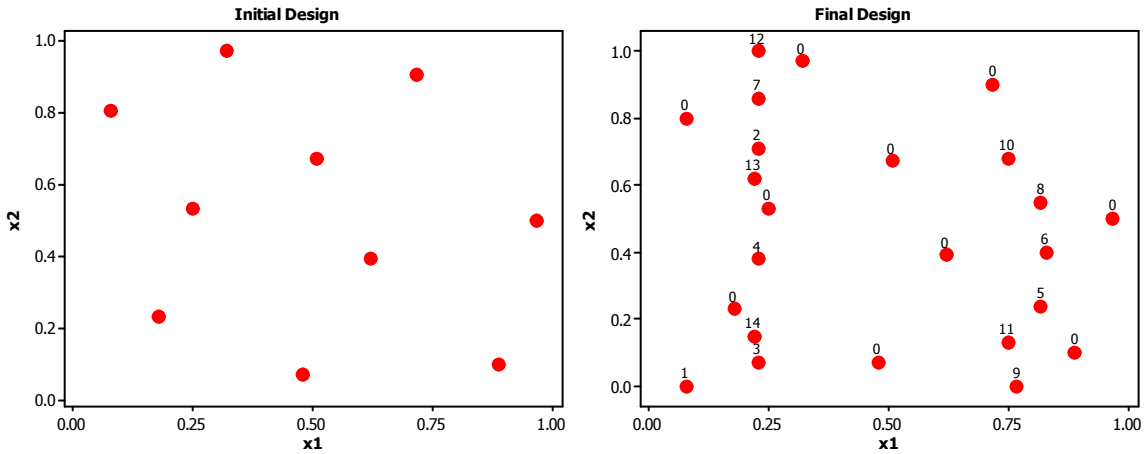

Figure 2.3: Initial Design and Final Design (order in which Points are added is also shown), Example 2.2

A sequential design scheme is employed. The next design point is chosen to be the point that minimizes the 2.5% quantile $Q_{0.025}$ of $Q_{Y(\cdot)}(x_c)$ (note that $x_c = x_1$), which is approximated with (2.44). This sequential scheme is terminated after the *potential improvement* $PI = \bar{Q}_3(x_c^3) - Q_{0.025}(x_c^4) = \min_{x_c \in [0,1]} \bar{Q}_3(x_c) - \min_{x_c \in [0,1]} Q_{0.025}(x_c)$, where $x_c^4 = \text{argmin}_{x_c \in [0,1]} Q_{0.025}(x_c)$, is less than 0.01. Figure 2.3 shows the final set of points and the order in which they are added. Notice that except for the first added point,

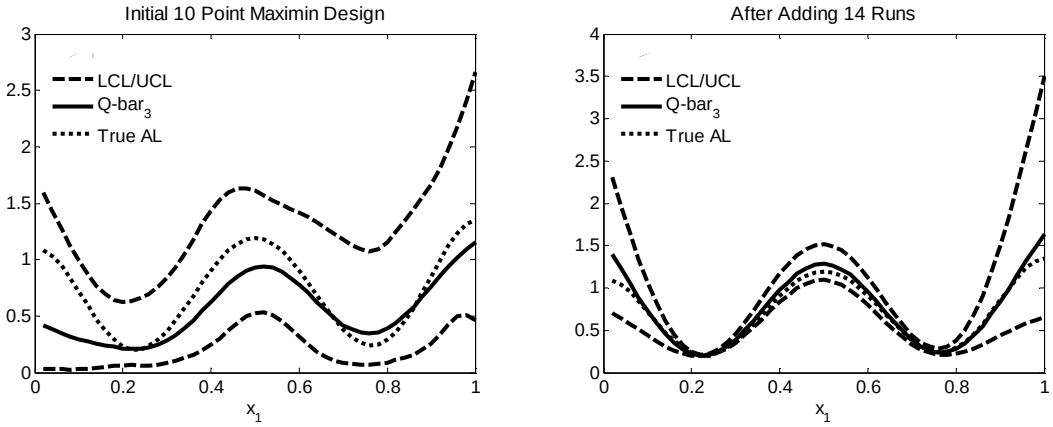

Figure 2.4: Plot of $\bar{Q}_3$ (Solid Line), True Average Loss (Dotted Line), and Upper and Lower Credible Limits (Dashed Line) Using Data from Initial Design (Left) and Final Design (Right), Example 2.2

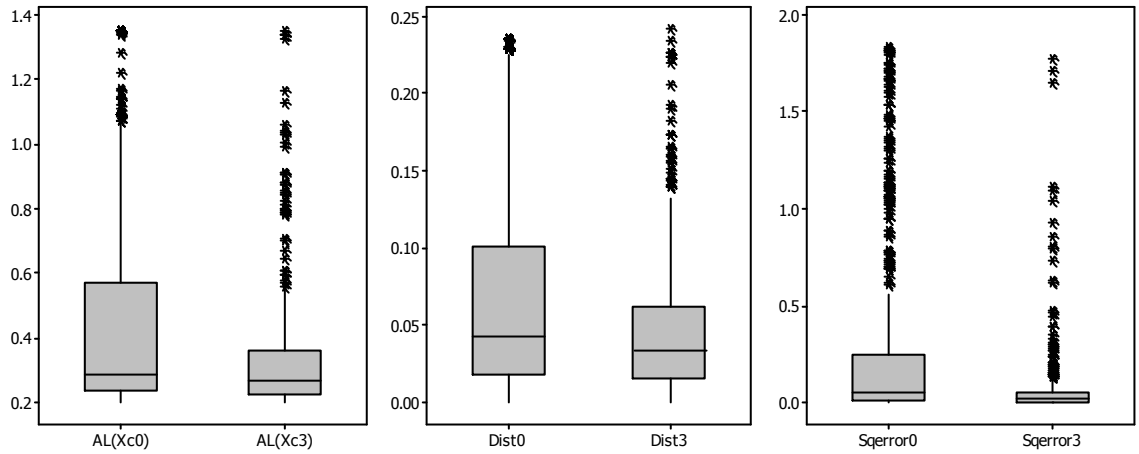

Figure 2.5: Boxplot of Three Measures of Performance of the $Q_{\mu(\cdot)}$ and $\bar{Q}_3$ Criteria Computed for 500 Randomly Sampled Latin Hypercube Designs, Example 2.2

all the others concentrate around 0.23 and 0.76. In our simulation, some of the computations are performed with low accuracy to save time (e.g., stopping the Newton's iterations for computing the quantiles $Q_{0.025}$ when the change is less than 10^{-3}). With data from the initial design, $x_c^3 = 0.225$ and $\bar{Q}_3(x_c^3) = 0.210$, which is very close to the true global minimum $AL = 0.202$ at $x_c = 0.229$. In contrast, $x_c^0 = 0.051$, which is far from any local minimum, and $Q_{\mu(\cdot)}(x_c^0) = 0.007$, which is far too optimistic. At the final iteration, both $Q_{\mu(\cdot)}$ and $\bar{Q}_3$ criteria give optimal robust settings and estimates of the minimum average loss that are close to the true values. Figure 2.4 plots $\bar{Q}_3(x_c)$ and 95% credible intervals obtained with the initial and final designs together with the true average loss $AL(x_c)$. The average losses at the two local minima are estimated with a high degree of accuracy.

We have also applied the $Q_{0.025}$ criterion to several other functions and obtained similar successes. However, we have found examples in which this sequential design scheme leaves some local minima unexplored.

The implementation of the $Q_{0.025}$ sequential design criterion would be difficult if

Monte Carlo simulation were used to compute $Q_{0.025}$. In addition to time-consuming simulations, minimization of $Q_{0.025}$ would be difficult due to the stochastic nature of the estimates. These difficulties are overcome with the saddlepoint-based approximation (2.44).

Note that the assumption that the correlation function is given by (2.42) implies that we are placing a different prior on the original correlation parameters $\boldsymbol{\theta}$ at each iteration of the sequential design scheme because the MLE of $\boldsymbol{\theta}$ is updated after every iteration. However, this does represent an improvement over completely ignoring uncertainty in $\boldsymbol{\theta}$. Moreover, as demonstrated, the $Q_{0.025}$ criterion yields good results for sequential robust design.

Finally, we give some simulation results to demonstrate the effectiveness of $\bar{Q}_3$ over $Q_{\mu(\cdot)}$ for randomly sampled Latin Hypercube designs. A total of 500 Latin Hypercube designs of size 10 are randomly sampled and the quantities $AL(x_c^0)$, $AL(x_c^3)$, $dist0 = min\{\|x_c^0 - 0.229\|, \|x_c^0 - 0.764\|\}$, $dist3 = min\{\|x_c^3 - 0.229\|, \|x_c^3 - 0.764\|\}$, $Sqerror0 = \left[Q_{\mu(\cdot)}(x_c^0) - AL(x_c^0)\right]^2$, $Sqerror3 = [\bar{Q}_3(x_c^3) - AL(x_c^3)]^2$ are computed. The first two quantities measure the goodness of x_c^0 and x_c^3 as optimal robust settings respectively; the third and fourth quantities measure the distance of x_c^0 and x_c^3 from the true local minima respectively; and the fifth and sixth quantities measure the accuracy of $Q_{\mu(\cdot)}$ and $\bar{Q}_3$ as estimators of the true average loss at x_c^0 and x_c^3 respectively. Boxplots of these performance measures are given in Figure 2.5. The figure suggests that $\bar{Q}_3$ is superior to $Q_{\mu(\cdot)}$ as a robust design criterion. In this simulation setup, the MLE of the correlation parameters is highly unstable. So, it seems important to take into account this source of uncertainty. Similar evidence of the advantage of $\bar{Q}_3$ over $Q_{\mu(\cdot)}$ was

obtained with several other response functions.

2.7.3 Example 2.3: Robust Design with Linear Model

In this example, we shall apply the proposed methodology to perform robust design optimization with a linear model. Data for a 27-run finite element experiment described by Bawaneh (2007) is given in Appendix B.2. The response Y is chip thickness; the factors A, B, C, n, m are material constants, which we assume are control factors. For the sake of illustration, we assume that the friction coefficient FC is a noise factor. In the following, we work with the coded factor levels $x_A, x_B, x_C, x_n, x_m, x_{FC}$ corresponding to A, B, C, n, m, FC respectively. Each factor is coded such that the minimum and maximum coded levels in the design matrix are 0 and 1 respectively.

The full second order model is taken as the complete model. Stepwise regression suggests the model $\hat{y} = 312 + 218x_n - 114x_A + 158x_nx_{FC} + 149x_mx_{FC} - 132x_Cx_n - 129x_B + 42.8x_m^2 + 103x_Ax_B + 86.4x_Bx_n$. Common practice in robust design with linear models (Myers and Montgomery 2002) is to find robust settings based on $\hat{y}$, which is equal to the posterior mean $\mu(\boldsymbol{x})$. This is equivalent to robust design using the $Q_{\mu(\cdot)}$ criterion. In the following, we use (2.10) for inference but for simplicity, we ignore model uncertainty (see Chipman 1998). Note that, although the data is from a computer experiment, the linear model given above fits the data very well ($R^2 = 98.2$). Thus, there is no need for using a Gaussian process model with a Gaussian correlation function.

It is assumed that the noise factor has discrete support {0,0.2,0.4,0.6,0.8,1} with equal probability mass on each point. For the target loss given by $\boldsymbol{H} = \boldsymbol{I}/6$, $T = 700$, the optimal control factor setting based on the $\bar{Q}_2$ criterion is $\boldsymbol{x}_c^2 = (x_A, x_B, x_C, x_n, x_m) = (0,0,0,1,0.57248)$. We generate a 200-point Latin hypercube design on the control factor

space and compute a 90% credible interval for each point. Plotted in Figure 2.6 (left) are boxplots of LCL and UCL for the 200 points, and also $LCL = 4264$ and $UCL = 14483$ for the loss at $\boldsymbol{x}_c^2$. It is seen that $\boldsymbol{x}_c^2$ is clearly superior to most other points since the upper credible limit for the loss at $\boldsymbol{x}_c^2$ is below the lower credible limit of the loss at most of the 200 points. However, $\boldsymbol{x}_c^0$ is nearly identical to $\boldsymbol{x}_c^2$. Shown in Figure 2.6 (center) is a boxplot of the estimated coverages (based on simulations with 20,000 runs) of the credible intervals at the 200 points and $\boldsymbol{x}_c^2$. It shows that the saddlepoint approximation is highly accurate.

For the variance loss given by $\boldsymbol{H} = (\boldsymbol{I} - \boldsymbol{J}/6)/6$, $T = 0$, the set of multiple optimal solutions for the criterion $\bar{Q}_2$ is given by $\omega^2 = \{(x_A, x_B, x_C, 0,0): 0 \leq x_A, x_B, x_C \leq 1\}$. For each of the points in the set ω^2, $\boldsymbol{HV}(\boldsymbol{x}_c)$ does not depend on (x_A, x_B, x_C) and $\boldsymbol{H\mu}(\boldsymbol{x}_c) = \boldsymbol{0}$. This implies that the loss at each point in ω^2 has the same posterior distribution. Thus, we have the very desirable situation of a wide region of equivalent optimal control factor settings. Note that for linear models and $\boldsymbol{H}$ given by (2.17), it is true in general that $\boldsymbol{HV}(\boldsymbol{x}_c)$ does not depend on $\boldsymbol{x}_c$. Hence, in this special

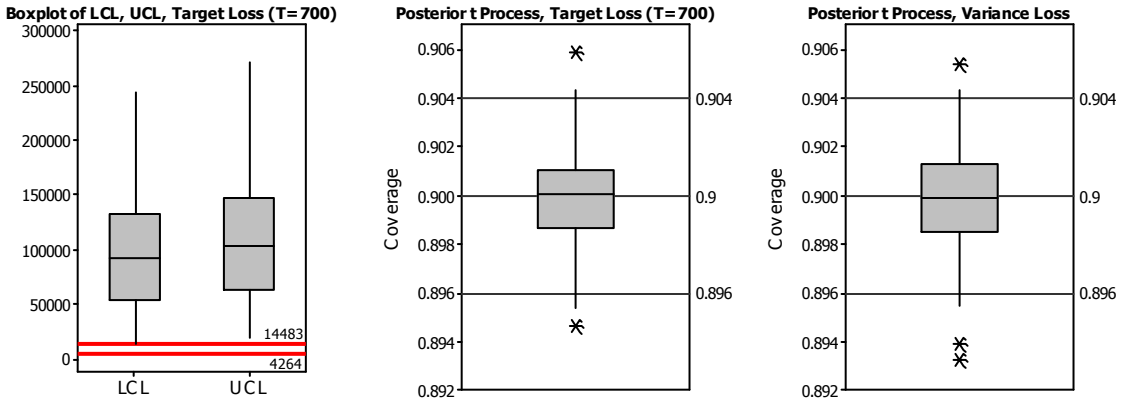

Figure 2.6: Boxplots of LCL and UCL for Target Loss (left), Estimated Coverage of Credible Intervals for Target Loss (center), and Variance Loss (right), Example 2.3

case, optimal settings obtained with the $Q_{\mu(\cdot)}$ and $\bar{Q}_2$ criteria coincide. Shown in Figure 2.6 (right) is a boxplot of the estimated coverage of credible intervals for the loss at 200 Latin hypercube points and $\boldsymbol{x}_c^2 = (1,1,1,0,0)$. Again, the Lugannani-Rice approximation is highly accurate.

In this example, we have worked with a linear model of the form commonly used for robust design, i.e., one that is linear in the noise factor. However, we point out that the $\bar{Q}_2$ criterion and our method of constructing credible intervals can be applied to linear models of any form. This is an important advantage over traditional mean variance modelling in response surface methodology (Myers and Montgomery 2002), which is applicable only to models linear in the noise factors.

2.8 Conclusions

We have proposed a general Bayesian framework for robust design optimization. We derive general quadratic loss criteria that accounts for uncertainty about the true response function. We also show how highly accurate credible intervals for the loss can be constructed using numerical inversion of the Lugannani-Rice approximation to the posterior distribution of the loss.

Two real data sets are analyzed. The examples demonstrate a significant advantage of using the $\bar{Q}_i, i = 1,2,3$ criteria over the $Q_{\mu(\cdot)}$ criterion, i.e., the former criteria favor settings with small response function uncertainty. The examples also show the usefulness of credible intervals for comparing alternative control factor settings. Moreover, it is demonstrated that the Lugannani-Rice approximation of the cdf of quadratic loss derived from Gaussian process models is considerably better than chi-square and lognormal approximations derived by matching moments.

The methodology presented in this chapter is general. Our criteria include as special case any loss function that can be written as a quadratic form in a set of random observations on the posterior process. Our method for constructing credible intervals can be applied to any quadratic form in non-degenerate or degenerate normal or multivariate-t random variables. The criteria and method for constructing credible intervals can be applied to any Gaussian process model or linear model. However, our methodology does not fully take uncertainty incurred in estimating the correlation parameters $\boldsymbol{\theta}$ into account. In addition, we have not addressed the question of quadratic loss criteria derived from multiple responses. Both are important areas for further research.

2.9 References

Apley, D.W., Liu, J., and Chen, W. (2006). "Understanding the Effects of Model Uncertainty in Robust Design with Computer Experiments," *Journal of Mechanical Design*, 128, 945-958.

Baldessari, B. (1967). "The Distribution of a Quadratic Form of Normal Random Variables," *The Annals of Mathematical Statistics*, 38(6), 1700-1704.

Bawaneh, M.A. (2007). "*Determination of Material Constitutive Models Using Orthogonal Machining Tests*," Ph.D. Dissertation, Wichita State University, Dept. of Industrial and Manufacturing Engineering.

Box, G.E.P. (1954). "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification," *The Annals of Mathematical Statistics*, 25(2), 290-302.

Box, G.E.P., and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*, Reading, MA: Addison-Wesley.

Butler, R.W. (2007). *Saddlepoint Approximations with Applications*, Cambridge, U.K.: Cambridge University Press.

Chang, P.B., Williams, B.J., Bhalla, K.S.B., Belknap, T.W., Santner, T.J., Notz, W.I., and Bartel, D.L. (2001). "Design and Analysis of Robust Total Joint Replacements: Finite Element Model Experiments with Environmental Variables," *Journal of Biomechanical Engineering*, 123, 239-246.

Chang, P.B., Williams, B.J., Notz, W.I., Santner, T.J., and Bartel, D.L. (1999). "Robust Optimization of Total Joint Replacements Incorporating Environmental Variables," *Journal of Biomechanical Engineering*, 121, 304-310.

Chen, P., and Koc, M. (2007). "Simulation of Springback Variation in Forming of Advanced High Strength Steels," *Journal of Materials Processing Technology*, 190, 189-198.

Chipman, H. (1998). "Handling Uncertainty in Analysis of Robust Design Experiments," *Journal of Quality Technology*, 30(1), 11-17.

Conti, S., Gosling, J.P., Oakley, J.E., and O'Hagan, A. (2009). "Gaussian Process Emulation of Dynamic Computer Codes," *Biometrika*, 96(3), 663-676.

Daniels, H.E. (1987). "Tail Probability Approximations," *International Statistical Review*, 55(1), 37-48.

Feuerverger, A. and A.C.M. Wong. (2000). "Computation of Value-at-Risk for Nonlinear Portfolios," *Journal of Risk*, Vol. 3, No. 1, pp. 37-55.

Ginsburg, H., and Ben-Gal, I. (2006). "Designing Experiments for Robust-Optimization Problems: the V_S-Optimality Criterion," *IIE Transactions*, 38, 445-461.

Handcock, M.S., and Stein, M.L. (1993). "A Bayesian Analysis of Kriging," *Technometrics*, 35(4), 403-410.

Imhof, J.P. (1961). "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48(3/4), 419-426.

Khuri, A.I. (2009). *Linear Model Methodology*, Boca Raton: Chapman & Hall.

Koksoy, O., and Doganaksoy, N. (2003). "Joint Optimization of Mean and Standard Deviation Using Response Surface Methods," *Journal of Quality Technology*, 35(3), 239-252.

Lehman, J.S., Santner, T.J., and Notz, W.I. (2004). "Designing Computer Experiments to Determine Robust Control Variables," *Statistica Sinica*, 14, 571-590.

Lewis, R.M., and Torczon, V. (1999). "Pattern Search Algorithms for Bound Constrained Minimization," *SIAM Journal of Optimization*, 9(4), 1082-1099.

Liu, H., Tang, Y., and Zhang, H.H. (2009). "A New Chi-Square Approximation to the Distribution of Non-Negative Definite Quadratic Forms in Non-Central Normal Variables," *Computational Statistics and Data Analysis*, 53, 853-856.

Lugannani, R., and Rice, S. (1980). "Saddle Point Approximation for the Distribution of the Sum of Independent Random Variables," *Advances in Applied Probability*, 12, 475-490.

Miller, A.C., and Rice, T.R. (1983). "Discrete Approximations of Probability Distributions," *Management Science*, 29(3), 352-362.

Miro-Quesada, G., and Del Castillo, E. (2004). "A Bayesian Approach for Multiple Response Surface Optimization in the Presence of Noise Variables," *Journal of Applied Statistics*, 31(3), pp. 251-270.

Myers, R.H., Kim, Y., and Griffiths, K.L. (1997). "Response Surface Methods and the Use of Noise Variables," *Journal of Quality Technology*, 29(4), 429-440.

Myers, R.H., and Montgomery, D.C. (2002). *Response Surface Methodology* (2nd ed.), New York: Wiley.

O'Hagan, A. (1994). *Kendalls Advanced Theory of Statistics: Volume 2B, Bayesian Analysis*, London, U.K.: Edwards Arnold.

Patnaik, P.B. (1949). "The Non-Central χ^2 and F Distributions and Their Applications," *Biometrika*, 36, 202-232.

Peterson, J.J., and Kuhn, A.M. (2005). “Ridge Analysis with Noise Variables,” *Technometrics*, 47(3), 274-283.

Sacks, J., Welch, W.J., Mitchell, T.J., and Wynn, H.P. (1989). “Design and Analysis of Computer Experiments,” *Statistical Science*, 4(4), 409-435.

Solomon, H., and Stephens, M.A. (1977). “Distribution of a Sum of Weighted Chi-Square Variables,” *Journal of the American Statistical Association*, 72(360), 881-885.

Stoer, J., and Bulirsch, R. (1993). *Introduction to Numerical Analysis* (2nd ed.), New York: Springer-Verlag.

Williams, B.J., Santner, T.J., and Notz, W.I. (2000). “Sequential Design of Computer Experiments to Minimize Integrated Response Functions,” *Statistica Sinica*, 10, 1133-1152.

Wu, C.F.J., and Hamada, M.S. (2009). *Experiments: Planning, Analysis, and Optimization* (2nd ed.), New York: Wiley.

Zellner, A., and Chetty, V.K. (1965). “Prediction and Decision Problems in Regression Models from the Bayesian Point of View,” *Journal of the American Statistical Association*, 60(310), 608-616.

CHAPTER 3

A BAYESIAN APPROACH FOR INTERPRETING MEAN SHIFTS IN MULTIVARIATE QUALITY CONTROL

3.1 Introduction

The performance or quality of a process is often characterized by multiple variables. Thus, effective control of a process can only be achieved by jointly monitoring all relevant variables. Yeh et al. (2006) and Bersimis et al. (2007) review the literature on control charts for the covariance matrix. Control charts for the mean include the well-known Hotelling T^2 charts and multivariate CUSUM and EWMA charts (Bersimis et al., 2007). Monitoring multivariate quality characteristics using a small number of summary statistics is a common practice. However, these statistics does not effectively support the diagnosis task, which is to determine the cause of the out-of-control signal.

Identification of mean shifts among a large number of quality characteristics often provides important information for the diagnosis task. For example, the quality control of fruit juice can be performed by comparing randomly selected samples against base samples using a multivariate control chart for amino acids and other constituents. If the control chart indicates differences between samples, information on the identity of the variables whose mean shifts and the shift directions can help to determine the source of adulteration of the juice. For instance, Zhang et al. (2009) give criteria for identifying the various types adulterations of pomegranate juice (e.g., amino acid proline>25mg/L is indicative of added grape products).

This chapter proposes a Bayesian method for identifying the means that shifted

and the directions of the shifts. The method provides a new diagnostic tool for Phase II monitoring. For p quality characteristics, there are 3^p possible scenarios for the means since each mean can increase, decrease, or remain in-control. In the Bayesian approach, each scenario corresponds to a value of an indicator vector and the most probable of the 3^p scenarios is found by sampling from the posterior distribution of the indicator via Gibbs sampling.

The problem of interpreting an out-of-control signal from a T^2 control chart has been widely studied in the literature. Many of the proposed approaches attempt to identify a subset of variables that has the most significant contribution to the large observed T^2 value by decomposing T^2. This is the basic idea underpinning Murphy (1987), Doganaksoy et al. (1991), Runger et al. (1996), and Mason et al. (1995, 1997). Mason et al. (1995, 1997) propose decomposing the T^2 statistic into independent components. However, there are several problems with the T^2 decomposition method. Firstly, for p variables, there are $p!$ possible decompositions; this makes the T^2 decomposition impractical for high-dimension problems. Secondly, there are no clear cut rules for jointly interpreting the components of the decomposition. Thirdly, results are sensitive to the significance levels used. Li et al. (2008) introduce a solution to the first two problems, which is to use a Bayesian causal network that describes the causal relationship between variables.

Hawkins (1991) proposes a procedure that is based on the likelihood-ratio tests of a shift in each mean. A comparison of the approaches proposed by Murphy (1986), Doganaksoy et al. (1991), Hawkins (1991), and Mason et al. (1995) is given by Das and Prakash (2008).

Recently, Wang and Jiang (2009) propose a penalized likelihood variable selection method to identify variables with shifted means. Zou et al. (2011) propose a heuristic Bayesian Information Criterion (BIC) for shift detection; the search for a model that minimizes the BIC criterion is restricted to those models that are optimal with respect to an adaptive-LASSO-type penalized likelihood. Their method can be applied to diagnose both covariance and mean shifts. Hereafter, we abbreviate the methods proposed by Wang and Jiang (2009) and Zou et al. (2011) as WJPLM and LEB respectively. Capizzi and Masarotto (2011) and Zou and Qiu (2009) propose EWMA-type control charts that employ test statistics based on the least angle regression algorithm and the adaptive-LASSO-type penalized likelihood function respectively. Diagnostic information is a byproduct of the charting statistics.

Unlike most of the available techniques in the literature, the Bayesian approach described in this chapter gives more specific and direct information about shifts in the mean, i.e., it gives the means that shifted upwards, the means that shifted downwards, and those that remained in-control. It allows prior knowledge to be incorporated in a statistical framework. We believe that this is an advantage rather than a disadvantage since in all cases, engineering knowledge must be used to identify assignable causes whenever a control chart signals.

Our approach is inspired by George and McCulloch's (1993) Bayesian Stochastic Search Variable Selection (SSVS) approach. However, it is *different* from SSVS. Our approach is developed for solving the problem of comparing the means of two different populations, i.e., phase I and phase II process means. In contrast, SSVS is developed for solving the problem of variable selection in regression. SSVS is limited to independent

and identically distributed residual error settings, and there are only two decisions per variable (whether the coefficient is zero or nonzero). On the other hand, our approach is for general multivariate normal variables and there are three decisions per variable. Furthermore, the priors we use can capture prior information such as most likely shift magnitudes that have values different from zero, and different ranges and probabilities of upward and downward shifts. These features cannot be modeled by SSVS. Finally, we also develop an empirical Bayes method for specifying some of the prior parameters.

Unlike our approach, WJPLM and LEB can only incorporate prior information via assumptions (such as assumptions on the number of variables that shifted and allowable shift directions), which have strong effects on results. Moreover, these approaches ignore uncertainty incurred in estimating the in-control means and covariance matrix (the phase I estimates are assumed to be equal to the population parameters) whereas our approach takes this uncertainty into account. Ignoring estimation uncertainty may not be justifiable when the phase I sample size is small. However, the proposed approach incurs a high computation cost, except when compared to the best subset variant of WJPLM. Note that our approach cannot be compared with the methods proposed by Capizzi and Masarotto (2011) and Zou and Qiu (2009) as those methods are for simultaneous monitoring and diagnosis. Hence, in this chapter, we shall only compare our approach with WJPLM and LEB.

The chapter is organized as follows: In Section 3.2, we briefly describe the mean diagnostic problem, state our assumptions, and review the T^2 chart and WJPLM. Section 3.3 gives our proposed Bayesian hierarchical model. Guidelines for specifying prior distributions are given in Section 3.4. Section 3.5 gives a Gibbs sampling procedure for

sampling from the posterior distribution of the parameters, and decision rules for mean shifts. Section 3.6 presents four examples and Section 3.7 concludes the chapter.

3.2 Multivariate Quality Control

This section describes the diagnostic problem addressed by this chapter and states the assumptions and notations used throughout the chapter. We also review the Hotelling T^2 control chart, the t-test and WJPLM. While the proposed approach can be used independently of any control chart, an interesting application of the approach is the diagnosis of an out-of-control signal given by a T^2 chart. Discussion of the T^2 chart also provides a nice context for introducing the mean shift diagnosis problem.

Throughout this chapter, we assume that the variables of interest $\boldsymbol{x} = (x_1, \dots, x_p)^T$ have a multivariate normal distribution with in-control mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{x} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$. We let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$ denote the N in-control phase I observations, and n denote the sample size of the phase II sample *suspected* to have been drawn from a common normal distribution different from $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$. All samples are assumed independent. Furthermore, in this chapter, we assume that the covariance matrix remains in-control so that $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in phase II. The objective of the mean shift diagnosis problem is to identify which components of $\boldsymbol{\mu}$ are different from $\boldsymbol{\mu}_0$. We let $\bar{\boldsymbol{x}} = \sum_{i=1}^{N} \boldsymbol{x}_i / N$ and $\boldsymbol{S} = \sum_{i=1}^{N} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T$ denote the sample mean and sample dispersion matrix of the phase I sample respectively. The sample covariance matrix of the phase I sample is denoted by $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{S}/(N-1)$. Similarly, we let $\bar{\boldsymbol{x}}_f$ and $\boldsymbol{S}_f$ denote the phase II sample mean and dispersion matrix respectively (the subscript f stands for future).

The standard control chart for monitoring $\boldsymbol{x}$ is the T^2 chart with statistic $T^2 =$

$n(\bar{\boldsymbol{x}}_f - \bar{\boldsymbol{x}})^T \hat{\boldsymbol{\Sigma}}^{-1}(\bar{\boldsymbol{x}}_f - \bar{\boldsymbol{x}})$ plotted with an upper control limit. It can be shown that if each phase I observation is independently distributed as $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$, and $\bar{\boldsymbol{x}}_f \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}/n)$, then $T^2 \sim \frac{(N+n)(N-1)p}{N(N-p)} F(p, N-p)$, where $F(p, N-p)$ is the F-distribution with p and $N-p$ degrees of freedom. Thus, to control the type I error at α, the upper control limit UCL should be determined using the upper 100α percentile of $F(p, N-p)$.

When $T^2 > UCL$, the control chart indicates that either the mean has shifted, i.e., $\boldsymbol{\mu} = E(\bar{\boldsymbol{x}}_f) \neq \boldsymbol{\mu}_0$, or the covariance matrix has changed, or both. This chapter assumes that the covariance matrix remains in-control. Whether the assumption of in-control covariance matrix is reasonable can be decided through the use of control charts for the covariance matrix (see Yeh et al. (2006)) or through tests of the hypothesis $H_0: \boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_1$, where $\boldsymbol{\Sigma}_0$ is the covariance matrix for the phase I samples and $\boldsymbol{\Sigma}_1$ is the covariance matrix for the phase II sample that triggered the out-of-control signal. A standard test for H_0 is the likelihood ratio test (Timm, 2002).

If it is determined that the mean shifted, supplementary information can be provided by computing t-statistics for testing each of the hypotheses $H_0: \mu_{0i} = \mu_i$, where μ_{0i} is the ith component of $\boldsymbol{\mu}_0$, and μ_i is the ith component of $\boldsymbol{\mu}$. The t-statistic for testing $H_0: \mu_{0i} = \mu_i$ is given by

$$t_i = \left(\bar{x}_{fi} - \bar{x}_i\right) / \sqrt{\tilde{\sigma}_i^2 \left(\frac{1}{n} + \frac{1}{N}\right)}, \tag{3.1}$$

where $\tilde{\sigma}_i^2$ is the *pooled* sample variance for the ith variable (the sample variance of the combined phase I and phase II samples), and $\bar{x}_{fi}$ and $\bar{x}_i$ are the ith components of $\bar{\boldsymbol{x}}_f$ and $\bar{\boldsymbol{x}}$ respectively. We interpret the t_i's using the following decision rule:

If $t_i < -\gamma$, then $\mu_i < \mu_{0i}$; if $t_i > \gamma$, then $\mu_i > \mu_{0i}$; if $-\gamma \leq t_i \leq \gamma$, then $\mu_i = \mu_{0i}$. (3.2)

The positive scalar γ is some suitably chosen cut-off.

A more modern approach to variable selection is WJPLM. This approach applies the forward selection method or the best subset method to the variable selection problem where the response is given by the column vector $\boldsymbol{R}(\bar{\boldsymbol{x}}_f - \bar{\boldsymbol{x}})$, the regressors are given by the columns of $\boldsymbol{R}$, and $\boldsymbol{R}$ is defined by the Cholesky decomposition $\left(\hat{\boldsymbol{\Sigma}}/n\right)^{-1} = \boldsymbol{R}^T\boldsymbol{R}$. The forward selection algorithm is terminated when the model size is L and the best subset method chooses the best model of size L, where L is specified by the engineer.

3.3 Bayesian Hierarchical Model

In this section, we shall develop a Bayesian hierarchical model that can be used to determine the means that shifted and the directions of the shifts when given a suspected out-of-control phase II sample of size n. Prior information is obtained from phase I data, which is then combined with the likelihood for phase II to yield the desired posterior inference.

It can be shown that $\bar{\boldsymbol{x}} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}/N)$, $\boldsymbol{S} \sim W(\boldsymbol{\Sigma}, N-1)$, where $W(\boldsymbol{\Sigma}, N-1)$ denotes a Wishart distribution with scale matrix $\boldsymbol{\Sigma}$ and $N-1$ degrees of freedom, $\bar{\boldsymbol{x}}$ and $\boldsymbol{S}$ are independent, and $(\bar{\boldsymbol{x}}, \boldsymbol{S})$ is a sufficient statistic for $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$. Thus, if we use noninformative priors for $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}$ given by $p(\boldsymbol{\mu}_0) \propto 1, p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{p+1}{2}}$, the resulting posterior distributions for $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}^{-1}$ are (see, e.g., Chapter 8 of Box and Tiao (1973))

$$p(\boldsymbol{\mu}_0|\bar{\boldsymbol{x}}, \boldsymbol{S}) \propto |1 + N(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)^T\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)|^{-\frac{N}{2}}, \tag{3.3}$$

$$p(\boldsymbol{\Sigma}^{-1}|\bar{\boldsymbol{x}}, \boldsymbol{S}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{N-p-2}{2}} exp\left\{-\frac{1}{2}trace(\boldsymbol{\Sigma}^{-1}\boldsymbol{S})\right\}. \tag{3.4}$$

For a phase II sample with sample mean $\bar{\boldsymbol{x}}_f \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$ and sample dispersion matrix $\boldsymbol{S}_f \sim W(\boldsymbol{\Sigma}, n-1)$, the likelihood is

$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1} | \bar{\boldsymbol{x}}_f, \boldsymbol{S}_f) \propto$

$$|\boldsymbol{\Sigma}|^{-\frac{1}{2}} exp\left[-\frac{1}{2} n\left(\bar{\boldsymbol{x}}_f - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\bar{\boldsymbol{x}}_f - \boldsymbol{\mu}\right)\right] |\boldsymbol{\Sigma}|^{-\frac{n-1}{2}} exp\left[-\frac{1}{2} trace\left(\boldsymbol{\Sigma}^{-1} \boldsymbol{S}_f\right)\right]. \quad (3.5)$$

Note that if $n = 1$, we simply set $\boldsymbol{S}_f = \boldsymbol{0}$. It is of interest to determine which components of $\boldsymbol{\mu} - \boldsymbol{\mu}_0$ are nonzero and the signs of the nonzero components. A similar problem arises in the area of Bayesian variable selection (George and McCulloch, 1993, 1997) in which the objective is to determine the variables with nonzero coefficients in a linear model. The key idea in Bayesian variable selection is to model each regression coefficient as a mixture of two distributions, where coefficients drawn from one distribution tend to be close to zero, and coefficients drawn from the other distribution tend to be large. Indicator variables are introduced to indicate the distribution from which a coefficient is drawn.

In a similar vein, we introduce indicator variables $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^T$ so that $\delta_i = -1$ indicates that μ_i has decreased, $\delta_i = 0$ indicates that μ_i has remained unchanged, and $\delta_i = 1$ indicates that μ_i has increased. The reason that we let δ_i take on three levels rather than two (one indicating that the ith mean shifted and the other indicating that it remained in-control) is that this would allow us to handle cases where the most likely shift magnitudes are known to be large (instead of close to zero) and cases where we have asymmetrical prior information about upward and downward shifts.

The parameters of the Bayesian model *for phase II* are $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, and $\boldsymbol{\Sigma}^{-1}$, and we need to specify the prior distribution $p(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1})$. Note that it is not justifiable to use the posterior of $(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1})$ from phase I as the prior for phase II. The phase II sample mean is expected to be different, and the proposed method is intended to unravel the differences. We assume $\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}^{-1}$ are independent and $p(\boldsymbol{\mu}|\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) = p(\boldsymbol{\mu}|\boldsymbol{\delta})$ so that $p(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) =$

$p(\boldsymbol{\mu}|\boldsymbol{\delta})p(\boldsymbol{\delta})p(\boldsymbol{\Sigma}^{-1})$. The assumption that $\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}^{-1}$ are independent is reasonable because in many cases, prior knowledge suggests that mean shifts does not depend on the inverse covariance matrix. The dependence structure captured by the covariance matrix is due to *common cause* variation; on the other hand, mean shifts are due to *special cause* variation. For the sake of mathematical tractability, we further assume that $\boldsymbol{\mu}|\boldsymbol{\delta}$ has a multivariate normal distribution, that is

$$p(\boldsymbol{\mu}|\boldsymbol{\delta}) \propto |\boldsymbol{\psi}_{\boldsymbol{\delta}}|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\theta}_{\boldsymbol{\delta}})^T \boldsymbol{\psi}_{\boldsymbol{\delta}}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta}_{\boldsymbol{\delta}})\right\}, \tag{3.6}$$

where $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ is the mean and $\boldsymbol{\psi}_{\boldsymbol{\delta}}$ is the covariance matrix. The subscript $\boldsymbol{\delta}$ of $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ and $\boldsymbol{\psi}_{\boldsymbol{\delta}}$ indicates that the mean and covariance matrix depend on $\boldsymbol{\delta}$.

Because we assume that $p(\boldsymbol{\mu}|\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) = p(\boldsymbol{\mu}|\boldsymbol{\delta})$ and $p(\boldsymbol{\mu}|\boldsymbol{\delta})$ is given by (3.6), the joint distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^{-1}$ given $\boldsymbol{\delta} = \mathbf{0}$ does not match the phase I posterior for $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}^{-1})$ (see the remark at the end of Section 3.4.2). However, we adopt (3.6) because it provides a prior that is easy to interpret and tune to capture prior information. Moreover, the prior distribution $p(\boldsymbol{\mu}|\boldsymbol{\delta} = \mathbf{0})$ with some reasonable values for $\boldsymbol{\theta}_{\mathbf{0}}$ and $\boldsymbol{\psi}_{\mathbf{0}}$ (which shall be given in Section 3.4.2) works very well**.**

Detailed specification of the parameters of $p(\boldsymbol{\mu}|\boldsymbol{\delta})$ and the prior for $\boldsymbol{\delta}$, $p(\boldsymbol{\delta})$ shall be discussed in the next section. Here, we shall discuss specification of the prior $p(\boldsymbol{\Sigma}^{-1})$ for $\boldsymbol{\Sigma}^{-1}$ since this is a simpler problem. It is obvious that we should set $p(\boldsymbol{\Sigma}^{-1})$ equal to the distribution given by (3.4) if the covariance matrix is assumed to remain unchanged. Thus,

$$p(\boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}^{-1}|^t exp\left[-\frac{1}{2} trace(\boldsymbol{\Sigma}^{-1}\boldsymbol{S})\right], \tag{3.7}$$

where $t = (N - p - 2)/2$ as indicated in (3.4). Note that $N \geq p + 1$ is sufficient for the Wishart distribution (3.7) to be nondegenerate. This condition is almost always met in

practice.

Our proposed Bayesian hierarchical model consists of $l(\bar{\boldsymbol{x}}_f, \boldsymbol{S}_f | \boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1})$, $p(\boldsymbol{\mu}|\boldsymbol{\delta})$, $p(\boldsymbol{\Sigma}^{-1})$, and $p(\boldsymbol{\delta})$. The quantities of primary interest are the posterior probabilities of $\boldsymbol{\delta}$, i.e., $p(\boldsymbol{\delta}|\bar{\boldsymbol{x}}_f, \boldsymbol{S}_f)$. The posterior probability of $\boldsymbol{\delta} = (\delta_1, \dots, \delta_p)^T$ is the probability that the state of the ith mean is given by $\delta_i, i = 1, \dots, p$ in light of prior knowledge, phase I data, and the phase II data $(\bar{\boldsymbol{x}}_f, \boldsymbol{S}_f)$. A comparison of the posterior probabilities of all possible values of $\boldsymbol{\delta}$ would yield information about which combination of out-of-control means is more likely. In addition, the marginal posterior distribution of δ_i can also be useful for deciding whether δ_i shifted upwards, downwards, or remained in-control. However, direct computation of the posterior distribution of $\boldsymbol{\delta}$ is clearly infeasible for practical problems. Section 3.5 provides a Gibbs sampling procedure that solves this computational problem. Before describing that procedure, we discuss the specification of prior distributions for $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$ in the next section.

3.4 Specification of Prior Distributions

3.4.1 Prior Distribution for Indicator Variables

A straightforward choice for $p(\boldsymbol{\delta})$ is the independence prior

$$p(\boldsymbol{\delta}) = \prod_{i=1}^{p} p_{1i}^{I(\delta_i=-1)} p_{2i}^{I(\delta_i=0)} p_{3i}^{I(\delta_i=1)}, \tag{3.8}$$

where p_{1i}, p_{2i}, and $p_{3i} = 1 - p_{1i} - p_{2i}$ are the prior probabilities that the ith mean shifted downwards, remained in-control, and shifted upwards respectively. Adopting the independence prior for $p(\boldsymbol{\delta})$ is equivalent to assuming that the δ_i's are a priori independent, which is justified in cases where there is no prior knowledge about the causal relationships between the variables. We propose that the *default choice* of

$\boldsymbol{p}_i = (p_{1i}, p_{2i}, p_{3i})$ be taken as

$$\boldsymbol{p}_i = (0.25, 0.5, 0.25). \tag{3.9}$$

This choice is justified if the engineer is ignorant about whether the $\boldsymbol{ith}$ mean changed or not, and if it changed, whether it increased or decreased. If we specify $\boldsymbol{p_i}$ as given in (3.9) for all $\boldsymbol{i = 1, \dots, p}$, the expected number of shifted means is $\boldsymbol{p/2}$; this quantity provides one way to check whether it is reasonable to specify $\boldsymbol{p_i}$ according to (3.9). Other choices of $\boldsymbol{p_i}$ might be considered if there is some information about shifts in the $\boldsymbol{ith}$ mean. For instance, if the $\boldsymbol{ith}$ variable is a smaller-the-better quality characteristic, a downward shift in the $\boldsymbol{ith}$ mean may be unlikely and $\boldsymbol{p_{1i}}$ should be small.

3.4.2 Prior Distribution for Mean

We now discuss the specification of $p(\boldsymbol{\mu}|\boldsymbol{\delta})$, which we have assumed to be a normal distribution with mean $\boldsymbol{\theta_\delta}$ and covariance matrix $\boldsymbol{\psi_\delta}$. We set

$$\boldsymbol{\theta_\delta} = \left(\bar{x}_1 - I(\delta_1 = -1)c_{1d} + I(\delta_1 = 1)c_{1u}, \dots, \bar{x}_p - I(\delta_p = -1)c_{pd} + I(\delta_p = 1)c_{pu}\right)^T \tag{3.10}$$

$$\boldsymbol{\psi_\delta} = diag\left\{\left[a_{1d}^{2I(\delta_1=-1)} a_{1u}^{2I(\delta_1=1)}\right]\hat{\sigma}_1^2/N, \dots, \left[a_{pd}^{2I(\delta_p=-1)} a_{pu}^{2I(\delta_p=1)}\right]\hat{\sigma}_p^2/N\right\}, \tag{3.11}$$

where $\hat{\sigma}_i$ is the sample standard deviation of the phase I data for the ith variable, i.e., the square root of the ith diagonal element of $\widehat{\boldsymbol{\Sigma}}$. It follows from (3.10) and (3.11) that $\mu_i|\boldsymbol{\delta} = \mu_i|\delta_i$ and $\mu_1|\delta_1, \dots, \mu_p|\delta_p$ are independently distributed. We also obtain

$$\mu_i|(\delta_i = 0) \sim N(\bar{x}_i, \hat{\sigma}_i^2/N),\ \mu_i|(\delta_i = -1) \sim N(\bar{x}_i - c_{id}, a_{id}^2\hat{\sigma}_i^2/N),$$

$$\mu_i|(\delta_i = +1) \sim N(\bar{x}_i + c_{iu}, a_{iu}^2\hat{\sigma}_i^2/N). \tag{3.12}$$

Figure 3.1 illustrates a standardized version of these priors, which will be

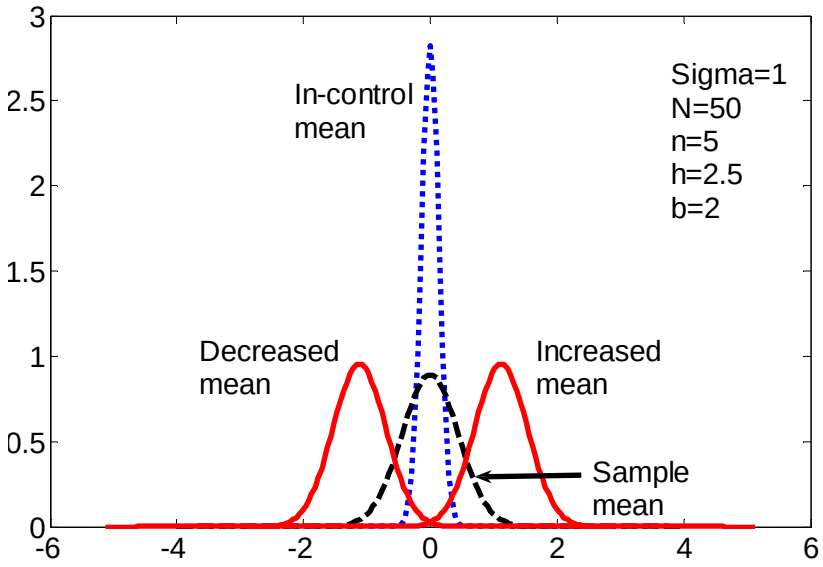

Figure 3.1: Density Functions of $(\mu_i - \bar{x}_i)/\hat{\sigma}_i|(\delta_i = -1)$ (Standardized Prior for Decreased Mean), $(\mu_i - \bar{x}_i)/\hat{\sigma}_i|(\delta_i = 0)$ (Standardized Prior for In-Control Mean), $(\mu_i - \bar{x}_i)/\hat{\sigma}_i|(\delta_i = 1)$ (Standardized Prior for Increased Mean) and $\left(\bar{x}_{fi} - \bar{x}_i\right)/\hat{\sigma}_i\big|(\mu_i = \bar{x}_i, \sigma_i = \hat{\sigma}_i)$ (Distribution of Standardized Phase II Sample Mean)

discussed later in Section 3.4.4. Note that $\mu_i|(\delta_i = 0) \sim N(\bar{x}_i, \hat{\sigma}_i^2/N)$ is entirely determined from phase I data and it is a good approximation of the posterior distribution of μ_{0i} derived from (3.3) when N is large. On the other hand, the distributions of $\mu_i|(\delta_i = -1)$ and $\mu_i|(\delta_i = +1)$ are each controlled by two parameters. The choices of these parameters are crucial; thus, this subject shall be discussed in the remainder of the section.

Note that $p(\boldsymbol{\mu}|\boldsymbol{\delta})$ and $p(\boldsymbol{\delta})$ should be required to jointly satisfy certain restrictions. Suppose that $[\bar{x}_i + \varepsilon_i, \bar{x}_i + \zeta_i]$ and $[\bar{x}_i - \zeta_i, \bar{x}_i - \varepsilon_i]$, where ε_i is small and ζ_i is large, contain all possible upward and downward shifts respectively. Then, it is logical to require that $p(\delta_i = 1|\mu_i)/p(\delta_i = -1|\mu_i)$ be larger than one in the interval $[\bar{x}_i + \varepsilon_i, \bar{x}_i + \zeta_i]$ and smaller than one in $[\bar{x}_i - \zeta_i, \bar{x}_i - \varepsilon_i]$. This is equivalent to requiring that

$$o(\mu_i) = log\left[\frac{p(\mu_i|\delta_i=1)p(\delta_i=1)}{p(\mu_i|\delta_i=-1)p(\delta_i=-1)}\right] \tag{3.13}$$

satisfies

$$o(\mu_i) > 0 \ \forall \mu_i \in [\bar{x}_i + \varepsilon_i, \bar{x}_i + \zeta_i] \text{ and } o(\mu_i) < 0 \ \forall \mu_i \in [\bar{x}_i - \zeta_i, \bar{x}_i - \varepsilon_i]. \quad (3.14)$$

Equation (3.14) will be satisfied whenever $p(\delta_i = 1) = p(\delta_i = -1)$ ($p_{1i} = p_{3i}$ when $p(\boldsymbol{\delta})$ is given by (3.8)), $c_{id} = c_{iu}$, and $a_{id} = a_{iu}$. It will also be satisfied when $a_{id} = a_{iu}$ and $p(\delta_i = 1)/p(\delta_i = -1) = exp[(c_{iu}^2 - c_{id}^2)/(2a_i^2\hat{\sigma}_i^2/N)]$. Otherwise, (3.14) should be checked.

In sections 3.4.3 and 3.4.4, we shall discuss specification of $p(\boldsymbol{\mu}|\boldsymbol{\delta})$ when prior information about mean shifts is available and when little prior information is available.

Remark: Due to the assumption that $p(\boldsymbol{\mu}|\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) = p(\boldsymbol{\mu}|\boldsymbol{\delta})$, $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}|\boldsymbol{\delta} = \mathbf{0})$ is not the same as $p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}^{-1}|\bar{\boldsymbol{x}}, \boldsymbol{S})$, where $p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}^{-1}|\bar{\boldsymbol{x}}, \boldsymbol{S})$ is the posterior distribution $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}^{-1})$ given phase I data. To satisfy the requirement that $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}|\boldsymbol{\delta} = \mathbf{0})$ equals $p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}^{-1}|\bar{\boldsymbol{x}}, \boldsymbol{S})$, we can set

$$p(\boldsymbol{\mu}|\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{D}_{\boldsymbol{\delta}}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\delta}}|^{\frac{1}{2}} exp\left\{-\frac{1}{2}N(\boldsymbol{\mu} - \boldsymbol{\theta}_{\boldsymbol{\delta}})^T \boldsymbol{D}_{\boldsymbol{\delta}}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\delta}}(\boldsymbol{\mu} - \boldsymbol{\theta}_{\boldsymbol{\delta}})\right\}. \quad (3.15)$$

In this case, we can have $p(\boldsymbol{\mu}|\boldsymbol{\delta} = \mathbf{0}, \boldsymbol{\Sigma}^{-1}) = N(\bar{\boldsymbol{x}}, \boldsymbol{\Sigma}/N)$ by setting $\boldsymbol{\theta}_{\mathbf{0}} = \bar{\boldsymbol{x}}$ and $\boldsymbol{D}_{\mathbf{0}} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. Together with $p(\boldsymbol{\Sigma}^{-1})$ given in (3.7), $p(\boldsymbol{\mu}|\boldsymbol{\delta} = \mathbf{0}, \boldsymbol{\Sigma}^{-1})p(\boldsymbol{\Sigma}^{-1})$ would now be the same as $p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}^{-1}|\bar{\boldsymbol{x}}, \boldsymbol{S})$. However, this prior has the disadvantage that the represented prior knowledge is hard to understand because the distribution has a complicated form. For interpretability, we may want $\boldsymbol{D}_{\boldsymbol{\delta}}$ to be diagonal so that the variances of $\mu_i|(\delta_i = -1)$ and $\mu_i|(\delta_i = 1)$ depends only on the ith diagonal element of $\boldsymbol{\Sigma}$ and not on the other elements. However, the existence of correlations between the μ_i's given $(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta})$ for $\boldsymbol{\delta} \neq \mathbf{0}$ is hard to interpret. For example, why would μ_1 and μ_2 be correlated when μ_1 shifts upwards and μ_2 shifts downwards and why would they have the same correlation as x_1 and x_2? Certainly, we can let the entire covariance

matrix of $p(\boldsymbol{\mu}|\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1})$ depend on $\boldsymbol{\delta}$ and set all the correlations involving μ_i to zero whenever $\delta_i \neq 0$. But, it would then be difficult to sample $\boldsymbol{\Sigma}^{-1}$ from its full conditional distribution as the distribution would not be a Wishart distribution. Compared to (3.15), a prior for the mean given by (3.12) is arguably easier to interpret. For this reason, we prefer (3.12) over (3.15) despite the fact that (3.15) give the correct prior for the case where $\boldsymbol{\delta} = \mathbf{0}$.

3.4.3 Case 1: Prior Information about Mean Shifts is Available

In cases where the quality engineer has a good idea about mean shifts, specification of the prior parameters in (3.10) and (3.11) is a rather straightforward exercise. The prior information that needs to be elicited for upward and downward shifts in each mean is the *most likely value and range*.

Clearly, c_{id} and c_{iu} should be set equal to most likely magnitudes of downward and upward shifts of the ith mean respectively. Now, if the range of upward shifts in the ith mean is $[\varepsilon_{iu}, \zeta_{iu}]$, then we should set a_{iu} so that $c_{iu} + 2a_{iu}\hat{\sigma}_i/\sqrt{N} \geq \zeta_{iu}$ and $c_{iu} - 2a_{iu}\hat{\sigma}_i/\sqrt{N} \leq \varepsilon_{iu}$. If the range of magnitudes of downward shifts is $[\varepsilon_{id}, \zeta_{id}]$, then we should set a_{id} so that $-c_{id} - 2a_{id}\hat{\sigma}_i/\sqrt{N} \leq -\zeta_{id}$ and $-c_{id} + 2a_{id}\hat{\sigma}_i/\sqrt{N} \geq -\varepsilon_{id}$. Note that ± 2 constants are used because a normal random variable has a high (roughly 95%) probability of being within two standard deviations of its mean. These considerations give

$$a_{iu} = \frac{\sqrt{N}}{2\hat{\sigma}_i} max\{\zeta_{iu} - c_{iu}, c_{iu} - \varepsilon_{iu}\},\ a_{id} = \frac{\sqrt{N}}{2\hat{\sigma}_i} max\{\zeta_{id} - c_{id}, c_{id} - \varepsilon_{id}\}. \tag{3.16}$$

Three forms of prior knowledge can be captured by the proposed priors $p(\boldsymbol{\mu}|\boldsymbol{\delta})$ and $p(\boldsymbol{\delta})$. The first is the most likely magnitudes of a shift, which is often large.

Assignable causes often produce large shifts; for instance, in thermocompression processes, a marked decrease in strength of gold-gold bonds is observed when surface contamination is present (Jellison, 1975). Moreover, large shifts are more critical and more easily detected than small shifts. Secondly, it is often the case that information about mean shifts is asymmetrical, i.e., an increase in a mean is likely to be within a certain range and a decrease is likely to be in another range and the most likely upward and downward shifts are different. For instance, if x_i is a smaller-the-better quantity (page 268 of Wu and Hamada (2009)), we should set $c_{iu} > c_{id}$ whereas for larger-the-better quantity, we should set $c_{iu} < c_{id}$. This is because process improvements are often smaller than process deteriorations. Thirdly, based on knowledge of the process, the engineer may have reason to believe that a particular mean is more likely to have increased than decreased and vice versa. These three forms of prior knowledge cannot be incorporated by modeling each mean as a mixture of two normal distributions (one for the in-control mean and one for the out-of-control mean).

3.4.4 Case 2: Little or No Prior Information about Mean Shifts is Available

In cases where prior information about mean shifts is hard to obtain, we consider using a symmetric prior for the mean with *only two parameters*, i.e., $c_{id} = c_{iu} = c_i = h\hat{\sigma}_i/\sqrt{n}$ and $a_{id} = a_{iu} = a$. This implies that $\mu_i|\delta_i \sim N(\bar{x}_i + \delta_i h\hat{\sigma}_i/\sqrt{n}, a^{2|\delta_i|}\hat{\sigma}_i^2/N)$, which gives

$$(\mu_i - \bar{x}_i)/\hat{\sigma}_i|\delta_i \sim N(\delta_i h/\sqrt{n}, a^{2|\delta_i|}/N). \quad (3.17)$$

Figure 3.1 illustrates this prior. It can be seen that the large reduction in the number of parameters is achieved through *standardization* of all variables (variable i is standardized by $\bar{x}_i$ and $\hat{\sigma}_i$), and assuming that the magnitudes of upward and downward shifts of the

standardized variables can be approximately modeled by the same normal distribution $N(h/\sqrt{n}, a^2/N)$. Our recommendations for specifying the prior distributions for the mean and indicators are summarized in Figure 3.2. Detailed discussions about these choices are given below.

In a nutshell, our suggestion concerning the choices of the parameters h and a is based on the rationale that $\mu_i|(\delta_i = 1)$ should be centered at the upper tail of $\bar{x}_{fi}\big|(\mu_i = \bar{x}_i, \sigma_i = \hat{\sigma}_i)$, (this determines h) and the densities of $\mu_i|(\delta_i = 0)$ and $\mu_i|(\delta_i = 1)$ should have at least a slight overlap (this determines a). Figure 3.1 illustrates this idea. Note that

Simplified Priors for the Mean and Indicators

1. ith element of $\boldsymbol{\theta_\delta}$: $\bar{x}_i + \delta_i c_i = \bar{x}_i + \delta_i h\hat{\sigma}_i/\sqrt{n}$.
2. $\boldsymbol{\psi_\delta}$: Diagonal matrix with ith diagonal element $\left(a^{2|\delta_i|}\right)\hat{\sigma}_i^2/N$.
3. Prior for indicators: $p(\boldsymbol{\delta}) = \prod_{i=1}^{p} p_{1i}^{I(\delta_i=-1)} p_{2i}^{I(\delta_i=0)}(1-p_{1i}-p_{2i})^{I(\delta_i=1)}$.
   Default choice for p_{1i} and p_{2i}: $p_{1i} = 0.25$, $p_{2i} = 0.5$.

Choice of Hyperparameters h and a:

Suggestion 1: Compare results obtained from several reasonable choices.
Set $a = max\left\{\frac{h}{b}\sqrt{\frac{N}{n}} - 1{,}1\right\}$. Try several values of $h \in [2{,}8]$ and $b \in [0.5{,}2]$. Make sure that $\left[-h/\sqrt{n} - 2a/\sqrt{N}, h/\sqrt{n} + 2a/\sqrt{N}\right]$ contains all $\left(\bar{x}_{fi} - \bar{x}_i\right)/\hat{\sigma}_i$.

Suggestion 2: Estimate prior parameters via the empirical Bayes method.
Find $I = \left\{i: \sqrt{n}\left|\bar{x}_{fi} - \bar{x}_i\right|/\hat{\sigma}_i > 2\right\}$. Set

$$h = h^{EB} = \begin{cases} \sqrt{n}\sum_{i\in I}\frac{|\bar{x}_{fi}-\bar{x}_i|}{\hat{\sigma}_i}, if\ I\ is\ nonempty \\ 2, \qquad\qquad if\ I\ is\ empty \end{cases} \text{ and}$$

$$a = a^{EB} = \begin{cases} max\left\{\sqrt{N}stdev\left\{\frac{|\bar{x}_{fi}-\bar{x}_i|}{\hat{\sigma}_i}, i \in I\right\}, \frac{h^{EB}}{2}\sqrt{\frac{N}{n}} - 1{,}1\right\}, if\ |I| \geq 2 \\ max\left\{\frac{h^{EB}}{2}\sqrt{\frac{N}{n}} - 1{,}1\right\}, \qquad\qquad if\ |I| \leq 1 \end{cases}.$$

Alternatively, set $h = h^{EB}$, $a = max\left\{\frac{h^{EB}}{b}\sqrt{\frac{N}{n}} - 1{,}1\right\}$, and try several values of b. We call this the partially empirical Bayes method.

Figure 3.2: Summary of Recommended Prior Parameter Choices When Little or No Prior Information about Mean Shifts is Available

if we plot $(\mu_i - \bar{x}_i)/\hat{\sigma}_i|\delta_i$, the same plot would be obtained for all i. Thus, we need to use *only one* plot to check whether the prior specification is reasonable.

In the design of control charts, n is chosen so that $\hat{\sigma}_i/\sqrt{n}$ is small compared to C_i to get good power for detecting a shift of size C_i (page 247 of Montgomery (2009)). Similarly, for the proposed approach, the relative sizes of $\hat{\sigma}_i/\sqrt{n}$ and c_i determine the type I and type II error rates, where the type I error rate is defined as the average number of in-control means declared out-of-control and the type II error rate is defined as the average number of out-of-control means whose shift directions are misidentified. Thus, it is convenient to measure c_i in units of $\hat{\sigma}_i/\sqrt{n}$, i.e., $c_i = h\hat{\sigma}_i/\sqrt{n}$. We recommend that h be at least 2 to achieve reasonable type I error. The reason for this choice is as follows: When the ith mean is in-control, $\bar{x}_{fi}\big|(\mu_i = \mu_{0i}, \sigma_i) \sim N(\mu_{0i}, \sigma_i^2/n)$ and so, $\bar{x}_{fi}$ falls within the interval $\left(\bar{x}_i - 2\hat{\sigma}_i/\sqrt{n}, \bar{x}_i + 2\hat{\sigma}_i/\sqrt{n}\right)$ most of the time (see Figure 3.1). Thus, if we specify h to be less than 2, the type I error rate would be high. On the other hand, if h is too large ($h\hat{\sigma}_i/\sqrt{n} > 3a\hat{\sigma}_i/\sqrt{N}$), the type II error rate would be high for small shifts. This is because for fixed a, a large h implies that large shifts are likely and small shifts are unlikely. The problem can be mitigated somewhat by choosing a large a. However, prior information is diluted because the priors for $\mu_i|(\delta_i = 1)$ and $\mu_i|(\delta_i = -1)$ would overlap somewhat and also allow very large shifts.

We set

$$a = max\left\{\frac{c_i}{b\hat{\sigma}_i/\sqrt{N}} - 1{,}1\right\} = max\left\{\frac{h}{b}\sqrt{\frac{N}{n}} - 1{,}1\right\}, \tag{3.18}$$

where $b > 0$. This arises from the following consideration. For reasons explained in the next paragraph, we would like to choose the constant a so that there is reasonable overlap

between the priors for the increased mean, in-control mean and decreased mean. Thus, we set $\bar{x}_i + c_i - ba\hat{\sigma}_i/\sqrt{N} = \bar{x}_i + b\hat{\sigma}_i/\sqrt{N}$. This means that b standard deviation units above the center of $\mu_i|(\delta_i = 0)$ should be b standard deviation units below the center of $\mu_i|(\delta_i = 1)$. By symmetry, setting $\bar{x}_i - c_i + ba\hat{\sigma}_i/\sqrt{N} = \bar{x}_i - b\hat{\sigma}_i/\sqrt{N}$ yields the same result. Since there is typically less information about shifted means than in-control means, we should set a to be 1 or larger. These considerations lead to (3.18).

If $b \leq 2$, there is at least a slight overlap between the density functions of $\mu_i|(\delta_i = 0)$ and $\mu_i|(\delta_i = 1)$, and between the density functions of $\mu_i|(\delta_i = 0)$ and $\mu_i|(\delta_i = -1)$. This ensures that the Gibbs sampler do not get stuck in one of the conditional distributions $\mu_i|\delta_i$. Moreover, too large values for b imply strong prior knowledge on the mean shifts. This can give rise to large type I error rates because $\delta_i = 1$ and $\delta_j = 0$ for all $j \neq i$ cannot explain the data well if $\bar{x}_{fi} \gg \bar{x}_i + c_i + 2a\hat{\sigma}_i/\sqrt{N}$ and $\bar{x}_{fj} = \bar{x}_j$ for all $j \neq i$. The data may be better explained by $(\delta_i, \delta_l) = (1,1)$ and $\delta_j = 0$ for all $j \notin \{i, l\}$. The smaller the value of b, the larger the variances of $\mu_i|(\delta_i = -1)$ and $\mu_i|(\delta_i = 1)$, and so, the prior contains less information about the mean shifts. This reduces shift detection power since for $\mu_i > \bar{x}_i$, $p(\delta_i = 1|\mu_i)/p(\delta_i = 0|\mu_i)$ is reduced. Thus, we recommend that b be at least 0.5.

Note that a small b (which gives a large a) produces considerable overlap between the densities of $\mu_i|(\delta_i = 1)$ and $\mu_i|(\delta_i = -1)$, which may seem unreasonable because if a mean shifted downwards, it cannot be much larger than $\bar{x}_i$. However, because $o(\mu_i)$ is a straight line with positive slope through $\bar{x}_i$ regardless of the value of b, the prior distribution makes sense regardless of the value of b. Nonetheless, the slope of $o(\mu_i)$ is $2Nc_i/a^2\hat{\sigma}_i^2$, which decreases with an increase in a. As a consequence, we may

observe that if the posterior probability that $\delta_i = 1$ is largest, the posterior probability that $\delta_i = -1$ can be a close second and vice versa. However, this does not affect decisions about the mean shifts as the decisions are based on the value of δ_i that gives the largest posterior probability (see Section 3.5).

For the analysis of a real dataset, we suggest that several values of h and b be tried and the resulting decisions be compared. The values of h and b should be chosen so that the interval $\left[-h/\sqrt{n} - 2a/\sqrt{N}, h/\sqrt{n} + 2a/\sqrt{N}\right]$ contains all $\left(\bar{x}_{fi} - \bar{x}_i\right)/\hat{\sigma}_i$ (e.g., see Figure 3.9). Otherwise, if $\left(\bar{x}_{fi} - \bar{x}_i\right)/\hat{\sigma}_i > h/\sqrt{n} + 2a/\sqrt{N}$, $\delta_i = 1$ alone cannot explain this large deviation well and if $\left(\bar{x}_{fi} - \bar{x}_i\right)/\hat{\sigma}_i < h/\sqrt{n} - 2a/\sqrt{N}$, $\delta_i = -1$ alone cannot explain this large deviation well. We have found that decisions based on Decision Rule 1 or Decision Rule 2 (see Section 3.5) are often *robust* to changes in values of b and h. However, trying several values give confidence to the decisions obtained from the proposed approach.

We can also choose h and b using an *empirical Bayesian (EB) approach*. Because $(\mu_i - \bar{x}_i)/\hat{\sigma}_i|(\delta_i = -1) \sim N\left(-h/\sqrt{n}, a^2/N\right)$, and $(\mu_i - \bar{x}_i)/\hat{\sigma}_i|(\delta_i = 1) \sim N\left(h/\sqrt{n}, a^2/N\right)$, the quantity $(|\mu_i - \bar{x}_i|/\hat{\sigma}_i)|(\delta_i \neq 0)$ would be approximately $N\left(h/\sqrt{n}, a^2/N\right)$ if $h/\sqrt{n}$ is large compared to $a/\sqrt{N}$. Since $\mu_1|\delta_1, \ldots, \mu_p|\delta_p$ are independently distributed, $(|\mu_1 - \bar{x}_1|/\hat{\sigma}_1)|(\delta_1 \neq 0), \ldots, \left(|\mu_p - \bar{x}_p|/\hat{\sigma}_p\right)|\left(\delta_p \neq 0\right)$ are independent and identically distributed with approximate common distribution $N\left(h/\sqrt{n}, a^2/N\right)$. Based on this observation, we can estimate h and a as follows. Let $I = \left\{i: \sqrt{n}\left|\bar{x}_{fi} - \bar{x}_i\right|/\hat{\sigma}_i > 2\right\}$, i.e., I is the set of indices associated with large discrepancy between phase I and phase II means. Then, we can estimate h by

$$h^{EB} = \begin{cases} \sqrt{n}\sum_{i\in I}\frac{|\bar{x}_{fi}-\bar{x}_i|}{\hat{\sigma}_i}, if\ I\ is\ nonempty \\ 2, \qquad\quad if\ I\ is\ empty \end{cases}. \tag{3.19}$$

We propose two ways to choose a. The first is to simply use (3.18) and try various values of b; the second is to estimate a by

$$a^{EB} = \begin{cases} max\left\{\sqrt{N}stdev\left\{\frac{|\bar{x}_{fi}-\bar{x}_i|}{\hat{\sigma}_i}, i\in I\right\}, \frac{h^{EB}}{2}\sqrt{\frac{N}{n}}-1{,}1\right\}, if\ |I|\geq 2 \\ max\left\{\frac{h^{EB}}{2}\sqrt{\frac{N}{n}}-1{,}1\right\}, \qquad\qquad if\ |I|\leq 1 \end{cases}, \tag{3.20}$$

where $stdev\{|\bar{x}_{fi}-\bar{x}_i|/\hat{\sigma}_i, i\in I\}$ denotes the standard deviation of the set of values $\{|\bar{x}_{fi}-\bar{x}_i|/\hat{\sigma}_i, i\in I\}$ and $|I|$ denotes the number of elements in I. We call the approach of setting $h = h^{EB}$ and trying several values of b the partially empirical Bayes (PEB) method and we call the approach of setting $h = h^{EB}$ and $a = a^{EB}$ the empirical Bayes (EB) method.

The reason we do not set a^{EB} equal to $\sqrt{N}stdev\{|\bar{x}_{fi}-\bar{x}_i|/\hat{\sigma}_i, i\in I\}$ is that we want to ensure that a^{EB} is not smaller than (3.18) with $b = 2$. Too small a value for a^{EB} is undesirable for reasons discussed previously. In addition, we want to ensure that the prior does not rule out shifts in variables whose indices are not in I.

Before we end this section, we point out that the EB method determines the prior for the mean entirely from data. Hence, it is very convenient in practice.

3.5 Gibbs Sampling and Decision Rules for Mean Shifts

This section gives a Gibbs sampling procedure for sampling from the posterior distribution of $\boldsymbol{\delta}$ and decision rules for identifying mean shifts.

Since the sample space of $\boldsymbol{\delta}$ consists of 3^p points, direct calculation of all 3^p probabilities is infeasible even when p is moderately large. To give an idea of the growth

of 3^p with p, note that $3^{10} = 59049$ and $3^{15} = 14348907$. However, not all values of $\boldsymbol{\delta}$ are equally of interest. We are primarily interested in those values of $\boldsymbol{\delta}$ with large probabilities, especially the posterior mode. This is adequate information for making inference about the mean shifts and the directions of the shifts. Gibbs sampling is a tool that allows us to discover the most probable values of $\boldsymbol{\delta}$ (Gelfand and Smith, 1990; George and McCulloch, 1993, 1997).

From the joint posterior distribution

$$p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta} | \bar{\boldsymbol{x}}_f, \boldsymbol{S}_f\right) \propto l(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1} | \bar{\boldsymbol{x}}_f, \boldsymbol{S}_f) p(\boldsymbol{\mu}|\boldsymbol{\delta}) p(\boldsymbol{\Sigma}^{-1}) p(\boldsymbol{\delta}), \tag{3.21}$$

we can easily obtain the full conditional distributions

$$p(\boldsymbol{\Sigma}^{-1} | \bar{\boldsymbol{x}}_f, \boldsymbol{S}_f, \boldsymbol{\delta}, \boldsymbol{\mu}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{n}{2}+t} exp\left\{-\frac{1}{2} trace\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{S}_f + \boldsymbol{S} + n(\bar{\boldsymbol{x}}_f - \boldsymbol{\mu})(\bar{\boldsymbol{x}}_f - \boldsymbol{\mu})^T\right)\right]\right\}, \tag{3.22}$$

$$p(\boldsymbol{\mu} | \bar{\boldsymbol{x}}_f, \boldsymbol{S}_f, \boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) \propto \left|\boldsymbol{V}_{\boldsymbol{\delta},\boldsymbol{\Sigma}}\right|^{1/2} exp\left\{-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{g}_{\boldsymbol{\delta},\boldsymbol{\Sigma}})^T \boldsymbol{V}_{\boldsymbol{\delta},\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu} - \boldsymbol{g}_{\boldsymbol{\delta},\boldsymbol{\Sigma}})\right\}, \tag{3.23}$$

$$p(\boldsymbol{\delta} | \bar{\boldsymbol{x}}_f, \boldsymbol{S}_f, \boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\psi}_{\boldsymbol{\delta}}|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\theta}_{\boldsymbol{\delta}})^T \boldsymbol{\psi}_{\boldsymbol{\delta}}^{-1}(\boldsymbol{\mu} - \boldsymbol{\theta}_{\boldsymbol{\delta}})\right\} p(\boldsymbol{\delta}), \tag{3.24}$$

where $\boldsymbol{g}_{\boldsymbol{\delta},\boldsymbol{\Sigma}} = (\boldsymbol{\psi}_{\boldsymbol{\delta}}^{-1} + n\boldsymbol{\Sigma}^{-1})^{-1}\left(\boldsymbol{\psi}_{\boldsymbol{\delta}}^{-1}\boldsymbol{\theta}_{\boldsymbol{\delta}} + n\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{x}}_f\right)$ and $\boldsymbol{V}_{\boldsymbol{\delta},\boldsymbol{\Sigma}} = (\boldsymbol{\psi}_{\boldsymbol{\delta}}^{-1} + n\boldsymbol{\Sigma}^{-1})^{-1}$. The conditional distributions (3.22)-(3.24) enable us to use a Gibbs sampling algorithm (Figure 3.3) for sampling from the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}, \boldsymbol{\delta})$. The algorithm given in Figure 3.3 holds for general $p(\boldsymbol{\mu}|\boldsymbol{\delta})$ given by (3.6). For the simplified prior given by (3.12), Step 4 of the Gibbs sampling algorithm given in Figure 3.3 simplifies to Step 4' given in Figure 3.4. The Gibbs sampling algorithm creates a Markov chain

$$(\boldsymbol{\Sigma}^{-1})^1, \boldsymbol{\mu}^1, \delta_1^1, \ldots, \delta_p^1, (\boldsymbol{\Sigma}^{-1})^2, \boldsymbol{\mu}^2, \delta_1^2, \ldots, \delta_p^2, \ldots., (\boldsymbol{\Sigma}^{-1})^i, \boldsymbol{\mu}^i, \delta_1^i, \ldots, \delta_p^i, \ldots \tag{3.25}$$

which has the property that the values of $\boldsymbol{\delta}^i = \left(\delta_1^i, \ldots, \delta_p^i\right)^T$ that appear in the simulation would be a sample from $p(\boldsymbol{\delta}|\bar{\boldsymbol{x}}_f, \boldsymbol{S}_f)$ for i large enough. Here, the reason for working

directly with $\boldsymbol{\Sigma}^{-1}$ instead of $\boldsymbol{\Sigma}$ becomes evident. Direct sampling of $\boldsymbol{\Sigma}^{-1}$ avoids the need to invert $\boldsymbol{\Sigma}$ in the computation of $\boldsymbol{g}_{\boldsymbol{\delta},\boldsymbol{\Sigma}}$ and $\boldsymbol{V}_{\boldsymbol{\delta},\boldsymbol{\Sigma}}$. In addition, $\boldsymbol{\Sigma}$ has an inverse Wishart distribution and we are not aware of a direct method for generating random matrices from this distribution; the most common method is to take the inverse of a draw from the corresponding Wishart distribution.

1. Start with $\boldsymbol{\mu}^0 = \bar{\boldsymbol{x}}$ and $\boldsymbol{\delta}^0 = \left(\delta_1^0, \dots, \delta_p^0\right)^T$. Set $i = 1$.

2. Sample $(\boldsymbol{\Sigma}^{-1})^i$ from the Wishart distribution with scale matrix $\left(\boldsymbol{S}_f + \boldsymbol{S} + n\left(\bar{\boldsymbol{x}}_f - \boldsymbol{\mu}^{i-1}\right)\left(\bar{\boldsymbol{x}}_f - \boldsymbol{\mu}^{i-1}\right)^T\right)^{-1}$ and $v = n + 2t + p + 1 = N + n - 1$ degrees of freedom.

3. Sample $\boldsymbol{\mu}^i$ from a normal distribution with mean $\boldsymbol{g}_{\boldsymbol{\delta}^{i-1},\boldsymbol{\Sigma}^i} = \left(\boldsymbol{\psi}_{\boldsymbol{\delta}^{i-1}}^{-1} + n(\boldsymbol{\Sigma}^{-1})^i\right)^{-1}\left(\boldsymbol{\psi}_{\boldsymbol{\delta}^{i-1}}^{-1}\boldsymbol{\theta}_{\boldsymbol{\delta}^{i-1}} + n(\boldsymbol{\Sigma}^{-1})^i\bar{\boldsymbol{x}}_f\right)$, and covariance matrix $\boldsymbol{V}_{\boldsymbol{\delta}^{i-1},\boldsymbol{\Sigma}^i} = \left(\boldsymbol{\psi}_{\boldsymbol{\delta}^{i-1}}^{-1} + n(\boldsymbol{\Sigma}^{-1})^i\right)^{-1}$.

4. Define $\boldsymbol{\delta}_j^i(r) = (\delta_1^i, \dots, \delta_{j-1}^i, r, \delta_{j+1}^{i-1}, \dots, \delta_p^{i-1})^T$. For $j = 1, \dots, p$, sample δ_j^i from the discrete distribution:

$$p(r) = \frac{\left|\boldsymbol{\psi}_{\boldsymbol{\delta}_j^i(r)}\right|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}^i - \boldsymbol{\theta}_{\boldsymbol{\delta}_j^i(r)}\right)^T \boldsymbol{\psi}_{\boldsymbol{\delta}_j^i(r)}^{-1}\left(\boldsymbol{\mu}^i - \boldsymbol{\theta}_{\boldsymbol{\delta}_j^i(r)}\right)\right\} p(\boldsymbol{\delta}_j^i(r))}{\sum_{q=-1}^{1}\left|\boldsymbol{\psi}_{\boldsymbol{\delta}_j^i(q)}\right|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}^i - \boldsymbol{\theta}_{\boldsymbol{\delta}_j^i(q)}\right)^T \boldsymbol{\psi}_{\boldsymbol{\delta}_j^i(q)}^{-1}\left(\boldsymbol{\mu}^i - \boldsymbol{\theta}_{\boldsymbol{\delta}_j^i(q)}\right)\right\} p\left(\boldsymbol{\delta}_j^i(q)\right)},$$

where $r \in \{-1,0,1\}$.

Figure 3.3: Gibbs Sampling Algorithm

4'. Define $\boldsymbol{\delta}_j^i(r) = (\delta_1^i, \dots, \delta_{j-1}^i, r, \delta_{j+1}^{i-1}, \dots, \delta_p^{i-1})^T$. For $j = 1, \dots, p$, sample δ_j^i from the discrete distribution:

$$p(r) = \frac{\left(\phi_j(r)\right)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left[\mu_j^i - \theta_j(r)\right]^2\left(\phi_j(r)\hat{\sigma}_j^2/N\right)^{-1}\right\} p(\boldsymbol{\delta}_j^i(r))}{\sum_{q=-1}^{1}\left(\phi_j(q)\right)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left[\mu_j^i - \theta_j(q)\right]^2\left(\phi_j(q)\hat{\sigma}_j^2/N\right)^{-1}\right\} p\left(\boldsymbol{\delta}_j^i(q)\right)}, r \in \{-1,0,1\},$$

where $\theta_j(r) = \left(\bar{x}_j - I(r=-1)c_{jd} + I(r=1)c_{ju}\right)$ and $\phi_j(r) = a_{jd}^{2I(r=-1)}a_{ju}^{2I(r=1)}$.

Figure 3.4: Simplification of Step 4 of Gibbs Sampling Algorithm Given in Figure 3.3

Decision Rule 1: For each j, choose the value d_j of δ_j that appears most frequently in (3.25) for $i > \tau$ and make decision $(d_1, \ldots, d_p)$.
Decision Rule 2: Choose the value $\boldsymbol{d}$ of $\boldsymbol{\delta}$ that appears most frequently in (3.25) for $i > \tau$ and make decision $\boldsymbol{d}$.

Figure 3.5: Decision Rules for Identifying Mean Shifts Using Steady State Gibbs Sampler Output $\{\boldsymbol{\delta}^i : i > \tau\}$

Mean shift directions would be identified from $\boldsymbol{\delta}$ values that have high posterior probability or δ_i values that have high marginal posterior probability. We give two decision rules in Figure 3.5 to determine mean shift directions. The idea underlying the first decision rule is that of making a decision based on the mode of the marginal posterior distribution of each δ_j while the idea underlying the second decision rule is that of making a decision based on the posterior mode of $\boldsymbol{\delta}$. We estimate these modes using the close-to-steady-state portion ($i > \tau$) of (3.25), where τ is the burn-in period. This gives Decision Rules 1 and 2.

Before we end this section, we point out that while the focus in this chapter is on shift directions, the shift magnitudes can be easily estimated from the components of $E\left(\boldsymbol{\mu}|\bar{\boldsymbol{x}}_f, \boldsymbol{S}_f\right) - \bar{\boldsymbol{x}}$. This latter quantity can be estimated from the samples generated by (3.25).

3.6 Examples

We shall give four examples to demonstrate the effectiveness of our proposed methodology. In the first example, we show that the proposed approach is superior to the t-test and is as good as WJPLM over wide ranges of recommended prior parameter values. Note that Wang and Jiang (2009) propose the use of the forward selection algorithm for model selection. In contrast, we shall use the best subset method. The latter performs better than the former but it incurs a high computation cost. Even with the

use of fast algorithms like those proposed Furnival and Wilson (1974), the computation can be too costly for high dimensional problems (e.g., $p > 30$). This is primarily due to the exponential increase in number of possible models with the dimension of the problem. In comparison, the most computationally intensive parts of the proposed approach are Steps 2 and Step 3 in Figure 3.3 (assuming Step 4' in Figure 3.4 is used), which require $O(p^3)$ operations. We conclude that the proposed approach requires $O(N_{iter}p^3)$ operations, where N_{iter} is the number of iterations of the Markov chain (3.25). This means that the proposed approach can be more affordable than the best subset variant of WJPLM when p is large. In the second example, we show that proper choices of h and b can yield performance superior to WJPLM with correct L, and that asymmetric choices of $p(\mu_i)$ and $p(\delta_i)$ that correctly reflect the true state of nature can give even better results.

In the third example, we compare our approach with the LEB diagnostic procedure proposed by Zou et al. (2011). The example demonstrates that our proposed method combined with the likelihood ratio test for equality of covariance matrices yields a powerful method for diagnosing shifts. The example also includes an analysis of a real dataset; for this dataset, the proposed method yields the same shift decision reached by LEB for wide ranges of prior parameters. The fourth example analyzes an interesting set of process monitoring data for a fruit juice process originally given by Fuchs and Kenett (1998) and reproduced in Beltran (2006). Matlab code for implementing the empirical Bayes version of the proposed approach is given in Appendix C.3.

3.6.1 Example 3.1: Performance Comparison I

We consider a problem where $p = 12$, $N = 90$, $n = 6$ (12 variables, 90 phase I

samples, 6 out-of-control phase II samples). The in-control mean is $\boldsymbol{\mu}_0 = \mathbf{0}$ and the population covariance matrix $\boldsymbol{\Sigma}_0$ (given in Appendix C.1) is generated randomly from the inverse Wishart distribution with scale matrix equal to the identity matrix $\boldsymbol{I}$, and $p + 1$ degrees of freedom. This distribution yields a random correlation matrix that has marginally uniformly distributed correlations (Barnard et al., 2000). The shifted mean is $\boldsymbol{\mu}_1 = (\Delta\sigma_1, -\Delta\sigma_2, \Delta\sigma_3, -\Delta\sigma_4, 0, \ldots, 0)$. The prior distributions $p(\boldsymbol{\mu}|\boldsymbol{\delta})$ and $p(\boldsymbol{\delta})$ are specified according to Figure 3.2. We change Δ, h, and b according to the experimental design given in Table 3.1 (the empirical Bayes method is denoted by EB) and we replicate each run in the design 100 times. For each replicate, we simulate $\bar{\boldsymbol{x}}$, $\boldsymbol{S}$, $\bar{\boldsymbol{x}}_f$, and $\boldsymbol{S}_f$ independently from $N(\mathbf{0}, \boldsymbol{\Sigma}_0/N)$, $W(\boldsymbol{\Sigma}_0, N-1)$, $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_0/n)$, and $W(\boldsymbol{\Sigma}_0, n-1)$ respectively. Given $\bar{\boldsymbol{x}}$, $\boldsymbol{S}$, $\bar{\boldsymbol{x}}_f$, and $\boldsymbol{S}_f$, we obtain samples from $p(\boldsymbol{\delta}|\bar{\boldsymbol{x}}_f, \boldsymbol{S}_f)$ using the Gibbs sampling algorithm in Figure 3.3 with Step 4 replaced by Step 4' given in Figure 3.4 and we apply the decision rules given in Figure 3.5. Two performance measures are calculated:

1. Type I Error Rate=(number of in-control means incorrectly identified as out-of-control)/(total number of in-control means).
2. Type II Error Rate=(number of out-of-control means incorrectly identified as in-control or whose shift-direction is incorrectly identified)/(total number of out-of-control means).

Adding a constant to both $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ will not change the performance of our procedure. It has the desirable *invariance property* that shifting or rescaling all observations by the same amount does not change the posterior distribution of $\boldsymbol{\delta}$ (if the priors are specified according to Figure 3.2).

For all simulation runs, we terminate sampling from the Markov chain (3.25) after $N_{iter} = 3000$ iterations and we set the burn-in period at $\tau = 1000$. These choices are found to work well in many trial runs. Table 3.1 presents the results of the simulation. The error rates for Decision Rule 1 are given in the columns labeled D1 and the error rates for Decision Rule 2 are given in the columns labeled D2.

It can be seen from Table 3.1 that except for a few cases, Decision Rules 1 and 2 have comparable performance. Table 3.1 also suggests that a smaller b tends to give smaller type I error rates and a larger h ($h/\sqrt{n} > \Delta$) tends to give larger type II error rates. Note that for $\Delta = 1.4$, $b = 0.5$ and $b = 2$ give comparable type II error rates but $b = 0.5$ gives smaller type I error rates. Thus, for larger shifts, a smaller b may be better. For all shift sizes, the EB method gives good performance; its superior performance when the shift size is small, i.e., $\Delta = 0.6$, is noteworthy.

Table 3.1: Factor Settings and Results for Simulation Experiment

Shift Δ	h	b	Type II Error Rate		Type I Error Rate	
			D1	D2	D1	D2
0.6	$\sqrt{6}$	0.5	0.445	0.433	0.011	0.030
	$1.8\sqrt{6}$	0.5	0.525	0.528	0.009	0.016
	$\sqrt{6}$	2	0.340	0.373	0.054	0.083
	$1.8\sqrt{6}$	2	0.505	0.503	0.014	0.031
	EB		0.330	0.335	0.053	0.086
1	$\sqrt{6}$	0.5	0.080	0.068	0.014	0.011
	$1.8\sqrt{6}$	0.5	0.098	0.110	0.009	0.013
	$\sqrt{6}$	2	0.028	0.035	0.040	0.045
	$1.8\sqrt{6}$	2	0.073	0.068	0.016	0.019
	EB		0.043	0.053	0.035	0.044
1.4	$\sqrt{6}$	0.5	0.000	0.000	0.006	0.005
	$1.8\sqrt{6}$	0.5	0.008	0.008	0.005	0.010
	$\sqrt{6}$	2	0.000	0.000	0.025	0.026
	$1.8\sqrt{6}$	2	0.008	0.008	0.015	0.018
	EB		0.003	0.000	0.014	0.013

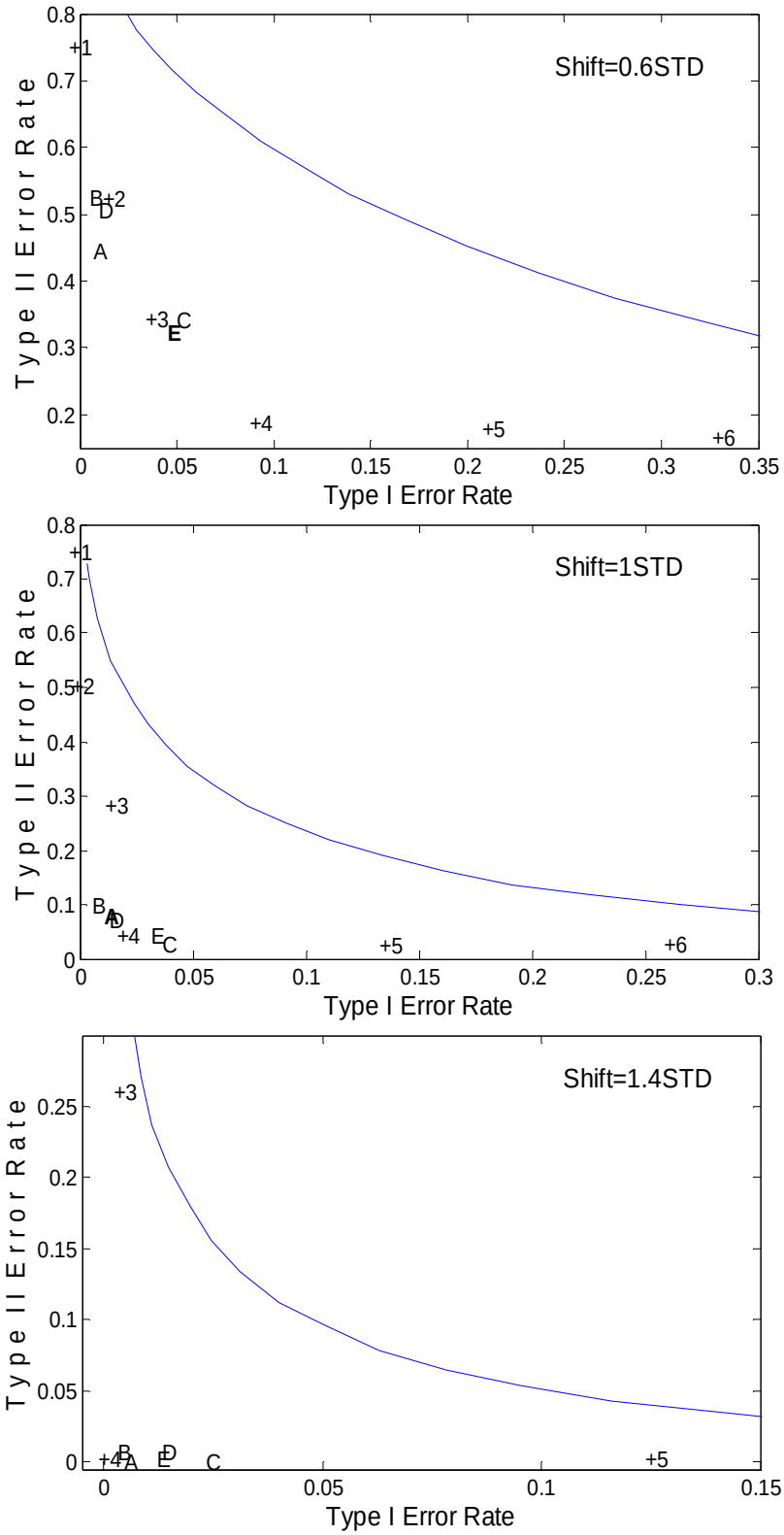

Figure 3.6: Type II Error Rate versus Type I Error Rate for Proposed Approach, WJPLM, and t-test. Top: $\Delta = 0.6$, Middle: $\Delta = 1$, Bottom: $\Delta = 1.4$. The solid line is for the t-test. The symbol $+L$ is for WJPLM with model size L. Symbols A-D are for prior distributions in Table 3.1. Symbol E is for the empirical Bayes approach.

We plot the type II error rate versus the type I error rate for the five different priors (results for Decision Rule 1 are used), the t-test, and WJPLM in Figure 3.6. Symbols A, B, C, D denote the priors given by $(h, b) = (\sqrt{6}, 0.5), (1.8\sqrt{6}, 0.5), (\sqrt{6}, 2), (1.8\sqrt{6}, 2)$ respectively and symbol E denotes the EB method. The continuous curve for the t-test is obtained by changing γ in (3.2) over small steps (the type I and II error rates are both functions of γ). WJPLM with model size (specified number of mean shifts) L is plotted as $+L$ in the figure. Figure 3.6 clearly indicate that our approach and WJPLM are superior to the t-test. When $\Delta = 0.6$, our approach performs similarly to WJPLM with $L = 2$ and $L = 3$ (specified number of mean shifts less than four, which is the correct number of mean shifts). However, as the shift size increases to $\Delta = 1$ and $\Delta = 1.4$, our approach has performance comparable to WJPLM with correctly specified number of mean shifts, i.e., $L = 4$. In particular, for $\Delta = 0.6$, the EB method performs like WJPLM with $L = 3$ and for $\Delta = 1$ and $\Delta = 1.4$, the EB method performs like WJPLM with $L = 4$. Note that selecting the model of correct size with minimum residual sum of squares, which is WJPLM with $L = 4$, is probably the best performance we can expect from a frequentist method. Since in practice, the actual number of means that shifted cannot be known, the performance of the EB method is impressive.

It can also be concluded from an examination of Table 3.1 and Figure 3.6 that the proposed approach is *robust.* The type I and type II error rates for the proposed approach change only slightly (compared to changes for WJPLM with different L) even though h and b are changed over wide ranges.

We have also investigated the effect of n on performance. When $\Delta = 0.6$ and n is increased to 12, the type II and type I error rates of the EB method decrease to 0.098 and

0.024 respectively for Decision Rule 1 and 0.075 and 0.023 respectively for Decision Rule 2 (compare with Table 3.1). For $\Delta = 1$ and $n = 12$, the estimated type II error rates are 0.005 (for both decision rules) and the estimated type I error rates are 0.0175 and 0.02. Thus, huge improvements can be attained with an increase in n.

We shall now illustrate a case of poor prior specification. The prior given by $h = 0.6\sqrt{6}$ and $b = 2$ is a poor choice for mean shifts $\Delta = 1,1.4$. This is illustrated in Figure 3.7 (left), which plots the distribution of $(\mu_i - \bar{x}_i)/\hat{\sigma}_i^2 \big| \delta_i$ together with the estimated in-control distribution of the standardized sample mean $\left(\bar{x}_{fi} - \bar{x}_i\right)/\hat{\sigma}_i \big| (\mu_i = \bar{x}_i, \sigma_i = \hat{\sigma}_i)$. It can be seen that shifts of sizes $\hat{\sigma}_i$ and $1.4\hat{\sigma}_i$ (vertical lines) are unlikely under this prior since the two vertical lines fall far out in the tail of the prior for the increased mean. Hence, performance of these choices is poor when $\Delta = 1$ or $\Delta = 1.4$, as shown in Table 3.2. In particular, we see that the type I error rates are very large compared to the values in Table 3.1. An example of a good prior specification ($h = \sqrt{6}$ and $b = 0.5$) is given in the right of Figure 3.7. In this case, shifts of sizes $\Delta = 1,1.4$ are in regions of concentration of the prior density for the increased/ decreased mean.

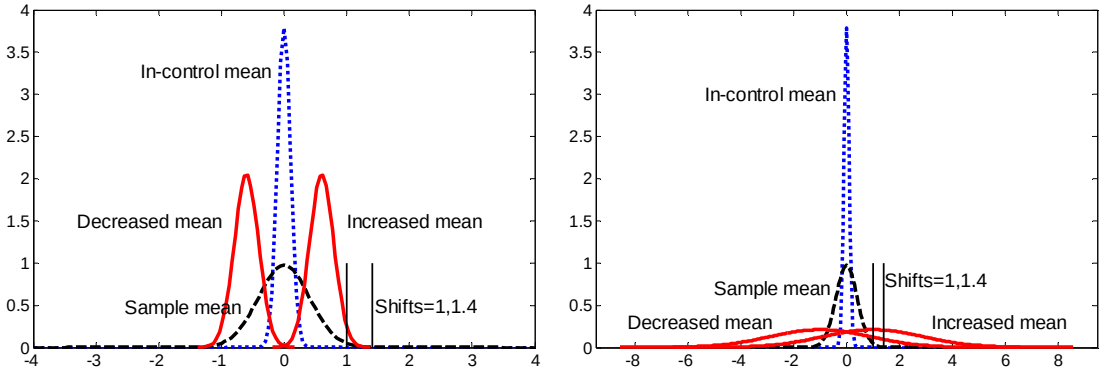

Figure 3.7: Poor Prior Specification (left), Good Prior Specification (right)

Table 3.2: Type I and Type II Error Rates for Poor Prior Distribution for Mean

Shift	h	b	Type II Error Rate		Type I Error Rate	
			D1	D2	D1	D2
1	$0.6\sqrt{6}$	2	0.000	0.003	0.173	0.191
1.4	$0.6\sqrt{6}$	2	0.000	0.000	0.375	0.394

3.6.2 Example 3.2: Performance Comparison II

This example demonstrates that proper choices of h and b can yield performance superior to WJPLM with correct L. It also investigates the performance of the PEB method and the effect of using asymmetric priors for the mean. In this example, $p = 6$, $N = 45$ and $n = 3$. We work with two different covariance matrices $\boldsymbol{\Sigma}_0$ (see Appendix B), each generated randomly from the inverse Wishart distribution with scale matrix $\boldsymbol{I}$, and $p + 1$ degrees of freedom. The mean shift is $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0 = (00{,}0{,}0{,}2\sigma_5/\sqrt{3}, -2\sigma_6/\sqrt{3})$. Six different choices of h and b and the PED with two values of b are investigated (see Table 3.3). Table 3.3 gives the results for the first covariance matrix. Figure 3.8 plots the type II error rate versus the type I error rate obtained with Decision Rule 1 for the eight priors using the symbols given in Table 3.3. It also plots the error rates of WJPLM, and the t-test. The figure shows that the prior given by $(h, b) = (1.5{,}1.6)$ can outperform WJPLM with $L = 2$ in terms of type II error rates (by quite a large margin for the second covariance matrix). For the first covariance matrix, the PEB method and the other priors (except B) are somewhat inferior to WJPLM with $L = 2$. For the second covariance matrix, the PEB with $b = 0.9$ (symbol G) and the prior $(h, b) = (2{,}0.9)$ (symbol C) are as good as WJPLM with $L = 2$ while the priors $(h, b) = (1.5{,}0.9), (2{,}1.6)$ (symbols A,D) have smaller type II error rates but larger type I error rates. For both covariance matrices, the t-test is dominated by the proposed approach.

Table 3.3: Results for Simulation Experiment and Labels for Figure 3.8

h	b	Type II Error Rate		Type I Error Rate		Symbol
		D1	D2	D1	D2	
1.5	0.9	0.290	0.290	0.120	0.113	A
1.5	1.6	0.165	0.160	0.168	0.155	B
2	0.9	0.300	0.230	0.083	0.078	C
2	1.6	0.275	0.255	0.113	0.128	D
2.5	0.9	0.365	0.325	0.065	0.073	E
2.5	1.6	0.380	0.325	0.095	0.103	F
PEB	0.9	0.370	0.340	0.083	0.093	G
PEB	1.6	0.280	0.225	0.100	0.093	H
Asymmetric Prior 1		0.140	0.150	0.023	0.023	-
Asymmetric Prior 2		0.050	0.060	0.013	0.013	-

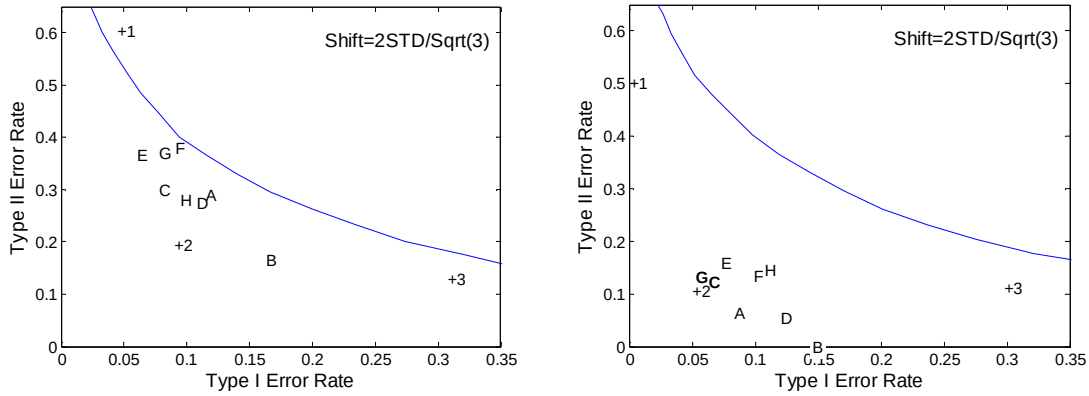

Figure 3.8: Type II Error Rate versus Type I Error Rate for Proposed Approach, WJPLM, and t-test for First (Left) and Second (Right) Covariance Matrices

We also consider two choices of asymmetric priors for the mean:

Asymmetric Prior 1: For $i = 1, \ldots, 4$, $p(\mu_i|\delta_i)$ and $p(\delta_i)$ are specified according to Figure 3.4 with $h = 3$ and $b = 0.9$; $(c_{5d}, c_{5u}) = (4,1.5)\hat{\sigma}_i/\sqrt{n}$, $(c_{6d}, c_{6u}) = (1.5,4)\hat{\sigma}_i/\sqrt{n}$; for $i = 5,6$, $a_{id} = max\left\{\frac{c_{id}}{b_{id}\hat{\sigma}_i/\sqrt{N}} - 1,1\right\}$, $a_{iu} = max\left\{\frac{c_{iu}}{b_{iu}\hat{\sigma}_i/\sqrt{N}} - 1,1\right\}$, $b_{5d} = b_{6u} = 0.9$, $b_{5u} = b_{6d} = 1.5$, and default $\boldsymbol{p}_i$ for all $i = 5,6$.

Asymmetric Prior 2: same as Asymmetric Prior 1 except that $\boldsymbol{p}_5 = (0.2,0.4,0.4)$ and $\boldsymbol{p}_6 = (0.4,0.4,0.2)$.

For Asymmetric Prior 1, $o(\mu_5) > 0$ if and only if $\mu_5 \in [-0.03\hat{\sigma}_5, 1.97\hat{\sigma}_5]$ and for Asymmetric Prior 2, $o(\mu_5) > 0$ if and only if $\mu_5 \in [-0.15\hat{\sigma}_5, 2.09\hat{\sigma}_5]$. This says that Asymmetric Prior 1 is good for upward shifts in μ_5 of sizes less than about $1.97\hat{\sigma}_5$ and downward shifts of size more than $0.03\hat{\sigma}_5$. A similar statement can be made for Asymmetric Prior 2. Table 3.3 presents type I and II error rates for the two asymmetric priors (the results are for the first covariance matrix). Compared with all other priors, the asymmetric priors give better results since the priors incorporate "correct" prior information. The priors $p(\mu_5|\delta_5 = 1)$ and $p(\mu_6|\delta_6 = -1)$ are concentrated around the upward and downward shifts of size $2\sigma_5/\sqrt{3}$ in the fifth and sixth mean respectively. Asymmetric Prior 2 gives better results than Asymmetric Prior 1 because for Asymmetric Prior 2, $\boldsymbol{p}_5$ and $\boldsymbol{p}_6$ give more probability to the true state of nature.

3.6.3 Example 3.3: Comparison with LEB

In this example, we compare our method with the LASSO based diagnostic procedure LEB introduced by Zou et al. (2011). In Section 4.1 of their paper, they consider a few simulated cases that involve only mean shifts. In these cases $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0 = (1,1,0,\dots,0)^T$ and $\boldsymbol{\Sigma}_0 = \left(0.5^{|i-j|}\right)$. We shall compare our approach with LEB for four cases of (N, n, p): $(50,25,4)$, $(1000,25,4)$, $(100,50,6)$, $(1000,50,6)$.

Note that Zou et al. (2011) apply their method to diagnose changes in the elements of the covariance matrix as well as the means. Thus, to ensure fair comparison, our method is combined with the likelihood ratio test of equality of two covariance matrices (Timm, 2002). We assume that if the null hypothesis is rejected, then the means are declared in-control but the covariance matrix is declared out-of-control. For comparison with the results in Zou et al. (2011), we measure the performance of the

combined method with two quantities: the relative frequency with which the combined method yields perfectly correct decisions on both means and covariance matrix (C), and the expected number of errors in mean shift decisions ($ENEM$). Note that Zou et al. (2011) employ the expected number of errors in the decisions on all parameters ENE as a performance measure instead of $ENEM$. However, this performance measure cannot be adopted here because we would need a diagnostic method for identifying the elements in the covariance matrix that shifted. Nevertheless, if the combined method is enhanced to include a diagnostic procedure for the covariance matrix, we will have $ENE \in [ENEM, ENEM + \alpha p(p+1)/2]$, where α is the significance level of the test for the covariance matrix and $p(p+1)/2$ is the number of elements in the covariance matrix.

Table 3.4 summarizes the results for four different choices of (h, b), i.e., (3,0.5),(3,1),(6,0.5), (6,1), and the PEB method with $b = 0.5$ and $b = 1$. The table also reproduces the performance estimates for LEB given by Zou et al. (2011). The number of simulations for each prior is 200 and we set $N_{iter} = 3000$ and $\tau = 1000$. For $n = 25$, the shift sizes are $5\sigma_i/\sqrt{n}, i = 1,2$ and for $n = 50$, the shift sizes are $7.07\sigma_i/\sqrt{n}, i = 1,2$. Thus, $h = 3$ and $h = 6$ can be thought of as poor and moderately good guesses of the mean shifts respectively. The significance level of the covariance matrix test is fixed at 0.05 and the cut-off point is obtained from a chi-squared approximation. Simulation suggests that this approximation is accurate. Thus, for our approach, $ENE \in [ENEM, ENEM + 0.5]$ for $p = 4$ and $ENE \in [ENEM, ENEM + 1.05]$ for $p = 6$. For the purpose of comparison, we set $ENE = ENEM + 0.25$ for $p = 4$ and $ENE = ENEM + 0.525$ for $p = 6$, which are perhaps worse-case estimates.

Table 3.4: Estimates of Performance Measures C, $ENEM$, and ENE for Six Prior Distribution Choices and LEB

h	b	(N,n,p)							
		(50,25,4)		(1000,25,4)		(100,50,6)		(1000,50,6)	
		C	$ENEM$ (ENE)	C	$ENEM$ (ENE)	C	$ENEM$ (ENE)	C	$ENEM$ (ENE)
3	0.5	0.81	0.28 (0.53)	0.91	0.15 (0.40)	0.88	0.16 (0.69)	0.84	0.25 (0.78)
3	1	0.80	0.30 (0.55)	0.86	0.20 (0.45)	0.69	0.42 (0.94)	0.80	0.25 (0.78)
6	0.5	0.86	0.22 (0.47)	0.90	0.19 (0.44)	0.90	0.17 (0.69)	0.95	0.10 (0.63)
6	1	0.88	0.19 (0.44)	0.89	0.18 (0.43)	0.83	0.27 (0.79)	0.90	0.15 (0.67)
PEB	0.5	0.88	0.17 (0.42)	0.92	0.13 (0.38)	0.92	0.16 (0.68)	0.93	0.12 (0.65)
PEB	1	0.82	0.24 (0.49)	0.87	0.18 (0.43)	0.87	0.21 (0.73)	0.90	0.16 (0.68)
LEB		C	ENE	C	ENE	C	ENE	C	ENE
		0.37	1.04	0.36	1.26	0.51	0.85	0.50	0.94

We see that except for the combination of $(N,n,p) = (100,50,6)$ and $(h,b) = (3,1)$, the performances of the six priors are significantly better than LEB (larger values of C and smaller values of ENE are preferred). Some of the best results are obtained with the PEB method. The marked decrease in performance when $(N,n,p) = (100,50,6)$ and $(h,b) = (3,1)$ is due to the fact that $(h,b) = (3,1)$ yields a somewhat informative prior centered far from the true shift, and $N = 100$ is not large enough for accurately estimating the in-control parameters. It can be concluded from the results in Table 3.4 that for processes that frequently suffer from mean shifts, it is better to use the combined approach than to use the LEB procedure to perform a simultaneous check on all parameters for a shift.

We now apply the proposed approach (without the covariance matrix test) to the wine quality control example discussed by Zou et al. (2011), which is based on a real

dataset. This problem involves 11 variables that are measurements from various physicochemical tests. Zou et al. (2011) show that the correct change point (which is known) can be found with a change point method. We take all observations after the change point as the out-of-control sample. Thus, we have $N = 870$ Phase I samples, and $n = 11$ out-of-control samples. We compute $\lambda_i = \sqrt{n}(\bar{x}_{fi} - \bar{x}_i)/\hat{\sigma}_i$ for $i = 1, \ldots, 11$ and found that the four largest λ_i are for $i = 4,5,8,11$ with values $1.95, 3.32, 3.88, -4.38$ respectively. The fifth largest λ_i only have magnitude 1.45. Eight priors given by $(h, b) \in \{2,3,5,7\} \times \{1,1.5\}$ and the EB method are tried. For the EB method, we set $I = \{4,5,8,11\}$. All nine prior distributions result in appreciable probability for $(\mu_i - \bar{x}_i)/\hat{\sigma}_i | (\delta_i = 1)$ over the interval $[1.95/\sqrt{11}, 4.38/\sqrt{11}]$ (see Figure 3.9 for a plot of the priors obtained via the EB method). The results obtained for all priors are similar. All nine posterior modes of $\boldsymbol{\delta}$ are identical, with a 1 in the fifth and eight positions, a -1 in the eleventh position, and zeros elsewhere. The results are obtained with $N_{iter} = 20000$

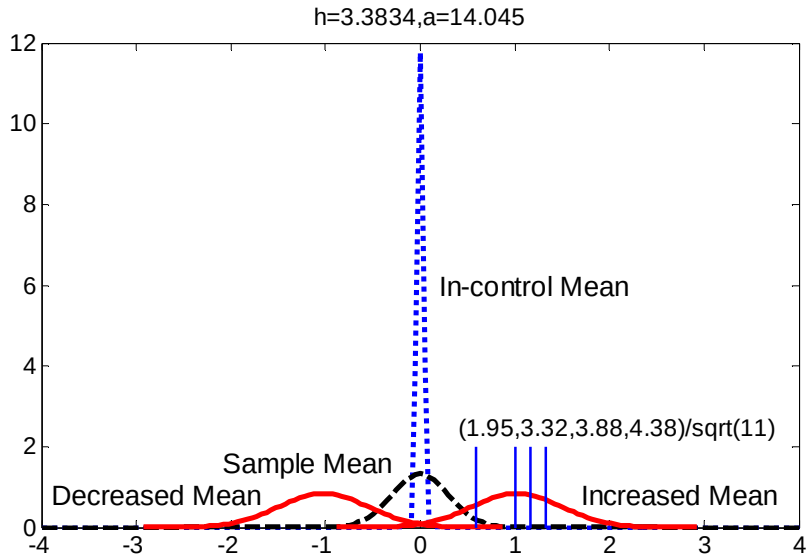

Figure 3.9: Plot of Prior Distributions Obtained via the EB Method and Density of $(\bar{x}_{fi} - \bar{x}_i)/\hat{\sigma}_i | (\mu_i = \bar{x}_i, \sigma_i = \hat{\sigma}_i)$

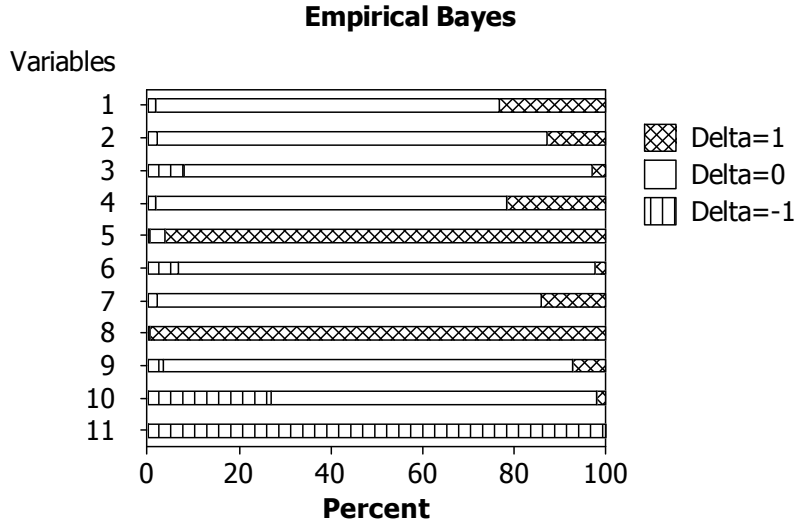

Figure 3.10: Marginal Posterior Distribution of Each Indicator for EB Method, Wine Data

and $\tau = 10000$. For the EB method, the marginal posterior distributions of $\delta_1, \ldots, \delta_{11}$ are plotted in Figure 3.10. The results clearly indicate that the fifth and eight mean shifted upwards and the eleventh mean shifted downwards. This agrees with the conclusions reached by Zou et al. (2011) using LEB.

Finally, note that although we set $I = \{4,5,8,11\}$, the EB method identifies a shift in μ_5, μ_8, and μ_{11} but not in μ_4. Thus, the choice of I does not determine the means that will be identified as shifted.

3.6.4 Example 3.4: Fruit Juice Data

We analyze process monitoring data collected from a fruit juice process given in Table 4.3a of Beltran (2006). The data, which is in Appendix C.1, consists of the concentrations in microgram per standard volume of $p = 11$ amino acids (names of the amino acids are given in the appendix). We take the phase I observations as the first 25 observations and the phase II observations as the remaining $n = 11$ observations.

However, there are some phase I observations that plot outside the 95% control limit of a T^2 chart for the phase I observations (see Bersimis et al. (2007) for a formula for the control limit). We removed the four observations (rows 16, 20, 22, and 25 of the data in Appendix C.1) that fall above the control limit, reconstructed the control chart, and further removed two observations (rows 11 and 12) that fall above the revised control limit. The remaining observations all appear to be in-control. Thus, we have $N = 19$ in-control phase I samples. A T^2 chart for the fruit juice process (without the T^2 value for observations that were removed) is plotted in Figure 3.11. We see that four of the phase II samples plot above the 95% upper control limit.

Ten of the values of $\lambda_i = \sqrt{n}\left(\bar{x}_{fi} - \bar{x}_i\right)/\hat{\sigma}_i$ for $i = 1, \dots, 11$ exceed two in absolute value; only $|\lambda_3| < 2$. Taking $I = \{1,2,4, \dots, 10\}$, we find that $h^{EB} = 3.58$ and $a^{EB} = 2.01 = \sqrt{N}stdev\left\{\left(\bar{x}_{fi} - \bar{x}_i\right)/\hat{\sigma}_i, i \in I\right\} > (h^{EB}/2)\sqrt{(N/n)} - 1 = 1.35$. The posterior mode of $\boldsymbol{\delta}$ ($N_{iter} = 20000$ and $\tau = 10000$) is $(0,0,0,1,0,0,0,1, -1, -1,0)$ and it has probability about 0.3. The marginal posterior distributions of the δ_i's, which are plotted in Figure 3.12, give the same information. This indicates that means 4 and 8 shifted upwards whereas means 9 and 10 shifted downwards. Although $\lambda_3 = 0.141$ is smallest in absolute value, the posterior probability that $\delta_3 = -1$ is quite high (about 0.3). This can be explained as follows. Variable 3 has positive correlations with variables 4 and 8 but negative correlations with variables 9 and 10. Thus, if it is assumed that one or more of means 4, 8, 9, and 10 did not shift so that the unusually large or small values of $\bar{x}_{fi}, i = 4,8,9,10$ are explained by natural variation, we would expect λ_3 to be quite large and positive. Since λ_3 is small, mean 3 could have shifted downwards. We also tried $(h, b) = (2.5,0.8), (5,1)$. For these priors, the shift decisions given by Decision Rules 1 and 2 are

identical to those obtained with the EB method.

Appendix C.2 gives a plot of the observations for each variable. The figure gives some indication that the means of variables 4, 9 and 10 shifted. However, it seems hard to tell by a visual inspection of the figure whether the mean of variable 8 shifted.

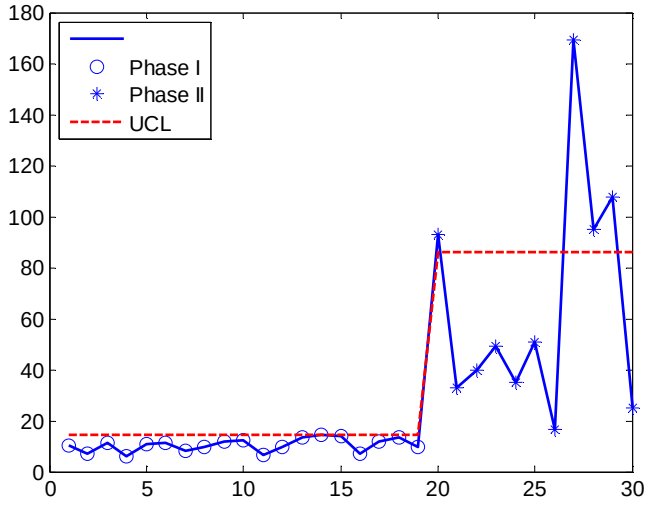

Figure 3.11: T^2 Chart for Fruit Juice Data

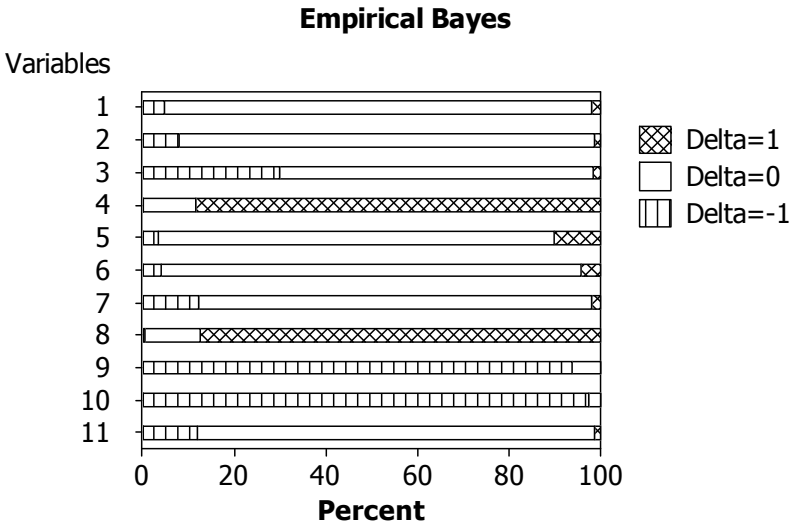

Figure 3.12: Marginal Posterior Distribution of Each Indicator for EB Method, Fruit Juice Data

3.7 Conclusions

We have proposed a Bayesian approach to identify the means that shifted and the direction of the shifts when a control chart for the mean of normal variables signals but the data indicate that the covariance matrix remain in-control. We introduce an indicator variable for each mean whose values $-1,0,1$ indicate whether the mean shifted downwards, remained in-control, or shifted upwards. The prior for each mean conditioned on its indicator captures prior information about the in-control and shifted states of the mean. The Bayesian hierarchical model is specified by prior distributions for the shifts, the indicators of the shift directions, and the covariance matrix. The prior distributions for the in-control mean and the inverse covariance matrix are derived from phase I data. Assumptions are made to simplify the prior distributions and guidelines are given to choose the prior parameters effectively. A Gibbs sampling algorithm for sampling from the posterior distribution of the vector of indicators is given.

We propose two decision rules to identify the most probable state of each mean. The first is to pick the value of each indicator with the highest marginal posterior probability. The second is to pick the posterior mode of the vector of indicators. These quantities can easily be estimated by their sample counterparts. Monte Carlo simulation shows that the proposed Bayesian approach always performs better than the t-test and can have performance comparable to the best subset variant of WJPLM with correctly specified number of mean shifts. Moreover, it can also outperform the LEB approach when shifts in the mean are more common than shifts in the covariance matrix. In all examples, the EB or PEB method for specifying the prior for the mean gives good results. The EB and PEB methods are attractive because the former is entirely data driven and the

latter only requires the user to specify one prior parameter. For these reasons, we recommend the EB and PEB methods for practical use.

We have considered modeling the mean with a mixture of three normal distributions. However, it is straightforward to extend our work to include modeling the increased and decreased mean with truncated normal distributions. Truncated prior distributions may yield better results. If we let $p(\boldsymbol{\mu}|\boldsymbol{\delta}) \propto |\boldsymbol{\psi}_{\boldsymbol{\delta}}|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\theta}_{\boldsymbol{\delta}})^T \boldsymbol{\psi}_{\boldsymbol{\delta}}^{-1}(\boldsymbol{\mu} - \boldsymbol{\theta}_{\boldsymbol{\delta}})\right\} I_{\Xi_{\boldsymbol{\delta}}}(\boldsymbol{\mu})$, where $I_{\Xi_{\boldsymbol{\delta}}}(\boldsymbol{\mu}) = 1$ if $\boldsymbol{\mu} \in \boldsymbol{\Xi}_{\boldsymbol{\delta}}$ and $I_{\Xi_{\boldsymbol{\delta}}}(\boldsymbol{\mu}) = 0$ otherwise, and $\boldsymbol{\Xi}_{\boldsymbol{\delta}}$ is a hyperrectangular set that depends on $\boldsymbol{\delta}$, then the full conditional distribution of each μ_i is truncated normal. Hence, we can sample from the posterior distribution of $\boldsymbol{\delta}$ using the Gibbs sampling algorithm given in Figure 3.3 with a modified Step 3. The modified Step 3 is to sample from $\mu_1, \ldots, \mu_p$ one at a time, where each μ_i is sampled from a truncated normal distribution.

Finally, we mention that we are researching the topic of incorporating prior knowledge about causal relationships between variables through the prior $p(\boldsymbol{\delta})$.

3.8 References

Barnard, J., R. McCulloch, and X. Meng. (2000). "Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage," *Statistica Sinica*, 10, 1281-1311.

Beltran, L.A. (2006). "Nonparametric Multivariate Statistical Process Control Using Principal Component Analysis and Simplicial Depth," PhD Dissertation, Department of Industrial and Management Systems, University of Central Florida.

Bersimis, S., S. Psarakis, and J. Panaretos. (2007). "Multivariate Statistical Process Control Charts: An Overview," *Quality and Reliability Engineering International*, 23, 517-543.

Box, G.E.P., and G.C. Tiao. (1973). *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.

Capizzi, G. and Masarotto, G. (2011). "A Least Angle Regression Control Chart for Multidimensional Data," *Technometrics*, 53(2), 285-296.

Das, N., and V. Prakash. (2008). "Interpreting the Out-of-Control Signal in Multivariate Control Chart – a Comparative Study," *International Journal of Advanced Manufacturing Technology*, 37, 966-979.

Doganaksoy, N., F.W. Faltin, W.T. Tucker. (1991). "Identification of Out-of-Control Quality Characteristics in a Multivariate Manufacturing Environment," *Communications in Statistics – Theory and Methods*, 20(9), 2775-2790.

Fuchs, C. and Kenett, R. (1998) *Multivariate Quality Control*, New York: Marcel Dekker.

Furnival, G.M. and Wilson, R.W. (1974). "Regressions by Leaps and Bounds," *Technometrics*, 16(1), 499-511.

Gelfand, A.E. and A.F.M. Smith. (1990). "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85(410), pp. 398-409.

George, E.I., and R.E. McCulloch. (1993). "Variable Selection Via Gibbs Sampling," *Journal of the American Statistical Association*, 88(423), pp. 881-889.

George, E.I., and R.E. McCulloch. (1997). "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7, 339-373.

Hawkins, D.M. (1991). "Multivariate Quality Control Based on Regression-Adjusted Variables," *Technometrics*, 33(1), pp.61-75.

Jellison, J.L. (1975). "Effect of Surface Contamination on the Thermocompression Bondability of Gold," *IEEE Transactions on Parts, Hybrids, and Packaging*, PHP-11(2), 206-211.

Li, J., J. Jin, and J. Shi. (2008). "Causation-Based T^2 Decomposition for Multivariate Process Monitoring and Diagnosis," *Journal of Quality Technology*, 40(1), pp. 46-58.

Mason, R.L., N.D. Tracy, and J.C. Young. (1995). "Decomposition of T^2 for Multivariate Control Chart Interpretation," *Journal of Quality Technology*, 27(1), pp. 99-108.

Mason, R.L., N.D. Tracy, and J.C. Young. (1997). "A Practical Approach for Interpreting Multivariate T^2 Control Chart Signals," *Journal of Quality Technology*, 29(1), 396-406.

Montgomery, D.C. (2009). *Introduction to Statistical Quality Control*. 6th Edition, New York: Wiley.

Murphy, B.J. (1987). "Selecting Out of Control Variables with the T^2 Multivariate Quality Control Procedure," *The Statistician*, 36(2), pp. 571-581.

Runger, G.C., F.B. Alt, and D.C. Montgomery. (1996). "Contributors to a Multivariate Statistical Process Control Chart Signal," *Communications in Statistics – Theory and Methods*, 25(6), pp. 2203-2213.

Timm, N.H. (2002). *Applied Multivariate Analysis*. Springer-Verlag, New York, NY.

Wang, K. and Jiang, W. (2009). "High-Dimensional Process Monitoring and Fault Isolation via Variable Selection," *Journal of Quality Technology*, 41(2), pp. 247-258.

Wu, C.F.J. and Hamada, M.S. (2009). *Experiments: Planning, Analysis, and Optimization*. 2nd Edition. New York: Wiley.

Yeh, A.B., D.K.J. Lin, and McGrath, R.N. (2006). "Multivariate Control Charts for Monitoring Covariance Matrix: A Review," *Quality Technology and Quantitative Management*, 3(1), 415-436.

Zhang, Y., Krueger, D., Durst, R., Lee, R., Wang, D., Seeram, N., and Heber, D. (2009). "International Multidimensional Authenticity Specification (IMAS) Algorithm for Detection of Commercial Pomegranate Juice Adulteration," *Journal of Agricultural and Food Chemistry*, 57, 2550-2557.

Zou, C., Jiang, W., and Tsung, F. (2011). "A LASSO-Based Diagnostic Framework for Multivariate Statistical Process Control," *Technometrics*, 53(2), 297-309.

Zou, C. and Qiu, P. (2009). "Multivariate Statistical Process Control Using LASSO," *Journal of the American Statistical Association*, 104(488), 1586-1596.

CHAPTER 4

A BAYESIAN APPROACH FOR MODEL SELECTION IN FRACTIONATED SPLIT PLOT EXPERIMENTS WITH APPLICATIONS IN ROBUST PARAMETER DESIGN

4.1 Introduction

Split plot designs are widely used in industrial experimentation. In these experiments, the wholeplot and subplot treatments are *separately randomized*. Due to the restriction in randomization, two sources of errors are present. A wholeplot error is incurred when a wholeplot treatment is applied to a wholeplot unit whereas a subplot error is incurred when a subplot treatment is applied to a subplot unit. As a consequence of this error structure, effects in split plot experiments can be divided into wholeplot and subplot effects. Subplot effects tend to be estimated with smaller variance than wholeplot effects. As such, analyzing a split plot experiment as if it were completely randomized can be misleading. Spurious wholeplot effects may be found significant and real subplot effects may be missed.

Traditional methods for analyzing split plot experiments are the ANOVA (Giesbrecht and Gumpertz, 2004; Wu and Hamada, 2009) and generalized least squares/ likelihood (Letsinger, 1996; Goos et al., 2006; Næs et al., 2007; Jones and Nachtsheim, 2009) methods. Although these methods are useful for analyzing many split plot experiments, there are many experiments in robust parameter design (RPD) where they cannot be applied because there are insufficient degrees of freedom for simultaneously estimating all wholeplot and subplot effects, and the wholeplot and subplot error

variances. In these cases, a Bayesian approach may be taken. Some early work on Bayesian analysis of variance components model are Hill (1965), and Tiao and Tan (1965). Gilmour and Goos (2009) and Wolfinger and Kass (2000) analyze split plot experiments using Bayesian methods; the latter paper does not discuss selection of significant effects. The analysis in Gilmour and Goos (2009) is based on the marginal posterior distribution of each effect, although they employ mixture priors for the coefficients and note the utility of such priors for distinguishing active and inactive effects. Other recent works include Vounatsou and Smith (1997) and Sun et al. (1996). Recently, Bingham and Goh (2012) extend the stochastic search variable selection (SSVS) approach of George and McCulloch (1993) and Chipman et al. (1997) for split plot experiments.

Highly fractionated orthogonal array designs that are run with restrictions in randomizations are often employed in RPD. These designs can provide independent estimates of main effects, but if two-factor interactions are also taken into account, a pattern of complex aliasing emerges (Chapter 9 of Wu and Hamada, 2009). A special type of design used in RPD is the *crossed array* (or inner-outer array in Taguchi's terminology), which is built from the Cartesian product of two orthogonal arrays. In Taguchi's applications, control factors are assigned to one of the arrays, called the control array, whereas noise and signal factors are assigned to the other array, which we call the *signal-noise array*. Two methods have been proposed to analyze a crossed array: the performance measure modeling and the response modeling approaches (Chapter 11 of Wu and Hamada, 2009). In the response modeling approach, crossed array designs are often analyzed as if they were completely randomized. However, in reality, many crossed

arrays are run as split plot experiments to reduce cost of experimentation (Box and Jones, 2001). In some cases, the control factors are the wholeplot factors because they are more difficult to change whereas in other cases, the signal and noise factors are the wholeplot factors. Split-plot designs obtained by crossing regular two-level fractional factorial designs for the wholeplot and subplot factors can be analyzed using half-normal plots (Box and Jones, 2001). However, this simple analysis method has limited applicability because crossed arrays employed in practice often consists of highly fractionated control and signal-noise arrays with more than two-levels for some or all of the factors.

In view of the high cost of experimentation and practical restrictions, the possibility of running a split plot experiment with fewer wholeplots than wholeplot effects and a small number of subplot runs is of important practical interest. Designs that are proposed to satisfy this need include fractional factorial split plot designs (Bingham and Sitter, 2003) and Taguchi's split unit designs (Taguchi, 1987). Split unit designs are highly fractionated mixed-level designs derived from orthogonal arrays by grouping rows according to columns designated for wholeplot factors. These designs are difficult to analyze with most existing methods if two-factor interactions in addition to main effects are entertained.

To address the above problems, we propose a Bayesian model selection methodology for analyzing *any balanced* split plot experiment, i.e., the number of subplots in each wholeplot is the same. In particular, it can be applied even if the design is nonorthogonal and the full model cannot be fitted due to insufficient number of runs. We enhance the conjugate hierarchical model for SSVS (George and McCulloch, 1997) to account for the split plot error structure. We derive an expression for the posterior

probability of a model that requires *computation of at most two uni-dimensional integrals,* and employ this quantity for model selection. We use forward selection (FS) to find good models in regular two-level fractional factorial experiments and propose a new algorithm, called combined global and local search (GLS), for more complex designs. GLS is an algorithm that searches for models with posterior probability above a threshold using a large and diverse set of starting points. While it explores the model space by changing the indicator components one at a time as in Gibbs sampling, it differs from Gibbs sampling because it uses large number of starting points and avoids unnecessary revisiting of the same models. Another attractive feature of GLS is that despite the fact that it is not a Markov Chain Monte Carlo (MCMC) method, it can be used for estimating posterior model probabilities using a simple procedure that we propose. To allow the proposed method to be automated and routinely used by practitioners, we propose default prior choices and show that they give good frequentist properties. Our approach is different from Bingham and Goh's (2012) approach in a few ways. First, we use a different parameterization of the covariance matrix and a different prior for the model coefficients. Second, we give explicit expressions for the posterior model probability and other posterior quantities, and we propose novel algorithms for finding good models. Third, we apply the proposed method to real problems in RPD. Finally, we give more emphasis on default prior choices and frequentist properties.

This chapter is organized as follows. Section 4.2 introduces the sampling model for balanced split plot experiments. Section 4.3 gives the Bayesian hierarchical model and expressions for some posterior quantities. Section 4.4 discusses prior specification. Section 4.5 discusses model search algorithms and computational issues. Simulation

studies of the proposed method and analysis of three real split plot robust design experiments are given in Section 4.6. Concluding remarks are given in Section 4.7.

4.2 Split Plot Design and Sampling Model

In this chapter, we assume that the split plot design is of the balanced type, i.e., there are N wholeplots and r subplots per wholeplot. This gives a total of $n = Nr$ observed response values. The wholeplot factor levels are denoted by $\boldsymbol{d}_1^w, \dots, \boldsymbol{d}_N^w$ and the subplot factor levels for the ith wholeplot are denoted by $\boldsymbol{d}_{i1}^s, \dots, \boldsymbol{d}_{ir}^s$. The corresponding values for the wholeplot and subplot effects levels are denoted by $\boldsymbol{x}_1^w, \dots, \boldsymbol{x}_N^w$ and $\boldsymbol{x}_{i1}^s, \dots, \boldsymbol{x}_{ir}^s$ respectively, where $\boldsymbol{x}_i^w = \boldsymbol{f}(\boldsymbol{d}_i^w)$ and $\boldsymbol{x}_{ij}^s = \boldsymbol{g}(\boldsymbol{d}_i^w, \boldsymbol{d}_{ij}^s)$. We write the design matrix and vector of observations as

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{d}_1^w, \boldsymbol{d}_{11}^s \\ \vdots \\ \boldsymbol{d}_1^w, \boldsymbol{d}_{1r}^s \\ \boldsymbol{d}_2^w, \boldsymbol{d}_{21}^s \\ \vdots \\ \boldsymbol{d}_N^w, \boldsymbol{d}_{Nr}^s \end{pmatrix}, \boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \vdots \\ \boldsymbol{Y}_N \end{pmatrix}, \boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{ir})^T, \tag{4.1}$$

where Y_{ij} is the response for the jth subplot of the ith wholeplot. The corresponding model matrix for the full model, excluding the column of ones, is given by

$$\boldsymbol{X} = \begin{pmatrix} \widetilde{\boldsymbol{X}}_1 \\ \vdots \\ \widetilde{\boldsymbol{X}}_N \end{pmatrix}, \text{ where } \widetilde{\boldsymbol{X}}_i = \begin{pmatrix} (\boldsymbol{x}_i^w, \boldsymbol{x}_{i1}^s) \\ \vdots \\ (\boldsymbol{x}_i^w, \boldsymbol{x}_{ir}^s) \end{pmatrix}. \tag{4.2}$$

We assume that the columns of $\boldsymbol{X}$ are centered and standardized so that the sum of squares of each of its p columns is n. For crossed array designs, the subplot treatment combinations are the same for all wholeplots. One of the arrays consists of the rows $\boldsymbol{d}_1^w, \dots, \boldsymbol{d}_N^w$, and the other consists of the rows $\boldsymbol{d}_1^s, \dots, \boldsymbol{d}_r^s$. The design $\boldsymbol{D}$ is the Cartesian product of the two arrays.

The sampling model for a balanced split plot experiment is

$$Y_{ij} = \beta_0 + \boldsymbol{x}_{ij}\boldsymbol{\beta} + \varepsilon_i + e_{ij}, i = 1, \dots, N, j = 1, \dots, r, \quad (4.3)$$

where $\boldsymbol{x}_{ij} = \left(\boldsymbol{x}_i^w, \boldsymbol{x}_{ij}^s\right)$; $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression coefficients; $\varepsilon_1, \dots, \varepsilon_N$ are the wholeplot errors, which are independent and identically distributed (iid) with common distribution $N(0, \sigma_w^2)$; $e_{11}, \dots, e_{Nr}$ are the subplot errors, which are iid $N(0, \sigma_s^2)$ and independent of the wholeplot errors.

Assuming model (4.3) holds, the variance of an individual observation is $\sigma^2 = var(Y_{ij}) = \sigma_w^2 + \sigma_s^2$ and the correlation between subplot observations in the same wholeplot is $\varphi = corr(Y_{ij}, Y_{il}) = \sigma_w^2/(\sigma_w^2 + \sigma_s^2)$. Thus, observations from different subplots in the same wholeplot are *positively correlated*. It is easy to see that observations from different wholeplots are independent.

Variable selection can be an important problem for fractionated split plot designs that are used to study a large number of factors. Due to the large number of effects, it is desirable and perhaps necessary to identify a sparse model rather than to include all effects in (4.3). However, due to the error structure of (4.3), many variable selection techniques cannot be applied to this problem.

4.3 Bayesian Hierarchical Model for Variable Selection

In this section, we present the proposed Bayesian hierarchical model, and derive expressions for the posterior probability of a model and some other posterior quantities. In the following two sections, we discuss prior specification and model search. It is clear from the discussion in Section 4.2 that $\boldsymbol{Y}|\beta_0, \boldsymbol{\beta}, \sigma^2, \varphi \sim N\left(\beta_0\mathbf{1} + \boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{R}_\varphi\right)$, where $\mathbf{1}$ is a vector of ones,

$$\boldsymbol{R}_{\varphi} = \boldsymbol{I}_N \otimes \begin{pmatrix} 1 & \varphi & \cdots & \varphi \\ \varphi & 1 & \cdots & \varphi \\ \vdots & \vdots & \ddots & \vdots \\ \varphi & \varphi & \cdots & 1 \end{pmatrix} \tag{4.4}$$

is the $n \times n$ correlation matrix and $\boldsymbol{I}_N$ is the $N \times N$ identity matrix.

The main inference problem is the identification of important effects, i.e., variable selection. Thus, as in George and McCulloch (1993), we introduce an indicator vector $\boldsymbol{\delta}$, where a 1 in position i indicates that effect i is large and a 0 in position i indicates that effect i is small. In other words, when $\delta_i = 1$, effect i is included in the model and when $\delta_i = 0$, it is not. Note that there are alternative Bayesian variable selection approaches that do not use indicators (see the review paper by O'Hara and Sillanpää (2009)). Following the conjugate hierarchical setup of George and McCulloch (1997), we let

$$\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}, c \sim N\left(\mathbf{0}, \sigma^2 \boldsymbol{S}_{\boldsymbol{\delta},c}\right), \tag{4.5}$$

where $\boldsymbol{S}_{\boldsymbol{\delta},c}$ is a diagonal matrix given by $\boldsymbol{S}_{\boldsymbol{\delta},c} = diag\{cI(\delta_i = 1) + dI(\delta_i = 0)\}$, c is given a prior $p(c)$ concentrated on values larger than d, and d is a small nonnegative number. Thus, conditioned on $\delta_i = 1$, c and σ^2, β_i will have variance $c\sigma^2$ and conditioned on $\delta_i = 0$ and σ^2, β_i will have variance $d\sigma^2$. As in SSVS, we let $p(\beta_0) \propto 1$,

$$p(\sigma^2) \propto (\sigma^2)^{-\left(\frac{\upsilon}{2}+1\right)} exp\left(-\frac{\upsilon\lambda}{2\sigma^2}\right), \tag{4.6}$$

$$p(\boldsymbol{\delta}) = \rho^{\|\boldsymbol{\delta}\|_1}(1-\rho)^{p-\|\boldsymbol{\delta}\|_1}. \tag{4.7}$$

Finally, we set the joint prior for the parameters $\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}, c, \varphi$ as

$$p(\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}, c, \varphi) = p(\beta_0)p(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}, c)p(\sigma^2)p(\boldsymbol{\delta})p(\varphi)p(c), \tag{4.8}$$

where $p(\varphi)$ and $p(c)$ are the priors of φ and c. Choices of $p(\boldsymbol{\delta})$, $p(c)$, $p(\varphi)$ and d, υ, λ shall be discussed in the next section. Note that unlike the parameter c, the parameter d is a fixed value.

The primary objective of this chapter is to address the problem of model selection for a split plot experiment with many effects of interest. This objective can be achieved by finding one or more models that have high posterior probability $p(\boldsymbol{\delta}|\boldsymbol{Y})$. We shall search for models with high posterior probability using the FS and GLS algorithms to be discussed in Section 4.5.

However, identification of good models is not sufficient for achieving the goals of robust design experiments. The engineer also needs to be given point estimates and credible intervals of effects. Inference on $\boldsymbol{\beta}$ based on $p(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\delta}_1^*)$, where $\boldsymbol{\delta}_1^*$ is the *highest posterior probability* (HPP) model, is the natural step after model selection if there is not too much model uncertainty. If model uncertainty is high, then inference on $\boldsymbol{\beta}$ should be based on $p(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\delta} \in \mathcal{M})$, where $\mathcal{M}$ is a set of high probability models. It is also useful to perform conditional inferences on $\boldsymbol{\beta}$ based on $p(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\delta})$ for each $\boldsymbol{\delta} \in \mathcal{M}$ in cases where the intention is to interpret effects, predict, or optimize based on each model separately to compare the results. In this chapter, we provide an expression for computing $E(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\delta})$ and show that $p(\beta_i|\boldsymbol{Y}, \boldsymbol{\delta})$ is a mixture of t-distributions. The latter fact is used to construct credible intervals for β_i. Since we can compute $p(\beta_i|\boldsymbol{Y}, \boldsymbol{\delta})$ exactly and $p(\boldsymbol{\delta}|\boldsymbol{Y})$ up to a proportionality constant, we can also compute $p(\beta_i|\boldsymbol{Y}, \boldsymbol{\delta} \in \mathcal{M})$ using the relation $p(\beta_i|\boldsymbol{Y}, \boldsymbol{\delta} \in \mathcal{M}) = \sum_{\boldsymbol{\delta}\in\mathcal{M}} p(\beta_i|\boldsymbol{Y}, \boldsymbol{\delta})p(\boldsymbol{\delta}|\boldsymbol{Y}) \,/\, \sum_{\boldsymbol{\delta}\in\mathcal{M}} p(\boldsymbol{\delta}|\boldsymbol{Y})$.

Aside from model selection and effect estimation, the experimenter may be interested in estimating the quantity φ also. This is important for the optimal design of follow-up runs (Goos, 2002), which may be needed when there is substantial model or parameter uncertainty. Point and interval estimates can be obtained with $p(\varphi|\boldsymbol{Y}, \boldsymbol{\delta} \in \mathcal{M}) = \sum_{\boldsymbol{\delta}\in\mathcal{M}} p(\varphi, \boldsymbol{\delta}|\boldsymbol{Y}) \,/\, \sum_{\boldsymbol{\delta}\in\mathcal{M}} p(\boldsymbol{\delta}|\boldsymbol{Y})$.

The problem of computing the posterior distribution of the indicator and other parameters is computationally challenging. A commonly used method is MCMC as in SSVS. However, we are able to integrate out $(\boldsymbol{\beta}, \sigma^2)$ from the joint posterior distribution of all parameters because we use the normal-inverse gamma prior for these parameters. This allows us to derive expressions for the posterior model probability (see (4.10)) and other posterior quantities of interest (see (4.11)-(4.14)) that require evaluation of at most two uni-dimensional integrals. Consequently, the use of MCMC is not necessary. Integrals are computed with Gaussian quadrature, and simple but reliable search algorithms are employed to find models with high posterior probabilities. This approach eliminates the need for convergence diagnostics, and allows the method to be automated and routinely used by unsophisticated practitioners.

In the following, we shall state results about various posterior quantities of interest. We give the joint posterior distribution of all parameters $p(\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}, c, \varphi|\boldsymbol{Y})$ in Proposition 4.3.1. This can be used to derive all other posterior distributions of interest. Proposition 4.3.2 gives the posterior distribution of the indicator $p(\boldsymbol{\delta}|\boldsymbol{Y})$. The posterior distribution of the effects given the indicator $p(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\delta})$ and the corresponding marginal distribution $p(\beta_i|\boldsymbol{Y}, \boldsymbol{\delta})$ are given in Propositions 4.3.3 and 4.3.4. Lastly, the posterior distribution of the correlation parameter $p(\varphi|\boldsymbol{Y}, \boldsymbol{\delta} \in \mathcal{M})$ is given in Proposition 4.3.5. These propositions give important quantities for the analysis of a split plot experiment. Proofs are outlined in Section D.2 of Appendix D. Note that several other quantities such as the posterior distribution of $\beta_0 + \boldsymbol{x}\boldsymbol{\beta}$ can be derived from the results given below.

Proposition 4.3.1: The joint posterior distribution of $(\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}, c, \varphi)$ is

$$p(\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}, c, \varphi|\boldsymbol{Y}) \propto L(\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}, \varphi|\boldsymbol{Y})p(\beta_0)p(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}, c)p(\sigma^2)p(\boldsymbol{\delta})p(\varphi)p(c)$$

$$\propto (\sigma^2)^{-(n+p+v+2)/2}\left|\boldsymbol{R}_{\varphi}\right|^{-1/2} exp\left(-\frac{v\lambda+RSS_{\boldsymbol{\delta},c,\varphi}}{2\sigma^2}\right) exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta}-\boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}\right)^T \boldsymbol{G}_{\boldsymbol{\delta},c,\varphi}^{-1}\left(\boldsymbol{\beta}-\right.\right.$$

$$\left.\left.\boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}\right)\right\} exp\left(-\frac{(\beta_0-\bar{Y})^2}{2\sigma^2\left(\mathbf{1}^T\boldsymbol{R}_{\varphi}^{-1}\mathbf{1}\right)^{-1}}\right)\left|\boldsymbol{S}_{\boldsymbol{\delta},c}\right|^{-1/2} p(\boldsymbol{\delta})p(\varphi)p(c), \tag{4.9}$$

where $L(\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}, \varphi|\boldsymbol{Y})$ is the likelihood,

$\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi} = \boldsymbol{S}_{\boldsymbol{\delta},c} - \boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T\left(\boldsymbol{R}_{\varphi} + \boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T\right)^{-1}\boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}$, $\boldsymbol{m}_{\boldsymbol{\delta},c,\varphi} = \boldsymbol{G}_{\boldsymbol{\delta},c,\varphi}\boldsymbol{X}^T\boldsymbol{R}_{\varphi}^{-1}(\boldsymbol{Y}-\overline{\boldsymbol{Y}})$, $\overline{\boldsymbol{Y}} = \bar{Y}\mathbf{1}$,

$\bar{Y} = \sum_{i=1}^{N}\sum_{j=1}^{r} Y_{ij}\,/n$, $RSS_{\boldsymbol{\delta},c,\varphi} = (\boldsymbol{Y}-\overline{\boldsymbol{Y}})^T\left(\boldsymbol{R}_{\varphi} + \boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T\right)^{-1}(\boldsymbol{Y}-\overline{\boldsymbol{Y}})$, and $\mathbf{1}^T\boldsymbol{R}_{\varphi}^{-1}\mathbf{1} = \frac{n}{1+(r-1)\varphi}$.

Proposition 4.3.2: The posterior distribution of $\boldsymbol{\delta}$ is given by

$$p(\boldsymbol{\delta}|\boldsymbol{Y}) = \int_0^1\int_0^{\infty} p(\boldsymbol{\delta}, c, \varphi|\boldsymbol{Y})\,dcd\varphi \propto \int_0^1\int_0^{\infty}\left[v\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\right]^{-(n-1+v)/2}\left|\boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T + \right.$$

$$\left.\boldsymbol{R}_{\varphi}\right|^{-1/2}\left(\mathbf{1}^T\boldsymbol{R}_{\varphi}^{-1}\mathbf{1}\right)^{-1/2} p(\boldsymbol{\delta})p(\varphi)p(c)dc\,d\varphi. \tag{4.10}$$

Proposition 4.3.3: The posterior distribution of $\boldsymbol{\beta}$ conditional on $\boldsymbol{\delta}$ is given by

$$p(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\delta}) = K\int_0^1\int_0^{\infty} t\left(\boldsymbol{\beta}; \boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}, (n+v-1)^{-1}\left(v\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\right)\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi}, n+v-1\right)$$

$$\left(v\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\right)^{-(n+v-1)/2}\left|\boldsymbol{R}_{\varphi} + \boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T\right|^{-1/2}\left(\mathbf{1}^T\boldsymbol{R}_{\varphi}^{-1}\mathbf{1}\right)^{-1/2} p(\varphi)p(c)dcd\varphi, \tag{4.11}$$

where $t\left(\boldsymbol{\beta}; \tilde{\boldsymbol{m}}, \tilde{\boldsymbol{G}}, \tilde{v}\right) = \frac{\Gamma\left(\frac{\tilde{v}+p}{2}\right)}{\Gamma\left(\frac{\tilde{v}}{2}\right)(\tilde{v})^{p/2}\pi^{p/2}}\left|\tilde{\boldsymbol{G}}\right|^{-1/2}\left[1 + \frac{1}{n+v-1}(\boldsymbol{\beta}-\tilde{\boldsymbol{m}})^T\tilde{\boldsymbol{G}}^{-1}(\boldsymbol{\beta}-\tilde{\boldsymbol{m}})\right]^{-(\tilde{v}+p)/2}$

is the multivariate t probability density function with mean $\tilde{\boldsymbol{m}}$, scale matrix $\tilde{\boldsymbol{G}}$ and $\tilde{v}$ degrees of freedom, and

$K =$

$$\left[\int_0^1\int_0^{\infty}\left(v\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\right)^{-(n+v-1)/2}\left|\boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T + \boldsymbol{R}_{\varphi}\right|^{-1/2}\left(\mathbf{1}^T\boldsymbol{R}_{\varphi}^{-1}\mathbf{1}\right)^{-1/2} p(\varphi)p(c)\,dcd\varphi\right]^{-1}.$$

Remark: The distribution $\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\delta}$ is a mixture of multivariate t distributions with weight function

$$\omega(c,\varphi) = K\left(\upsilon\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\right)^{-(n+v-1)/2}\left|\boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T + \boldsymbol{R}_\varphi\right|^{-1/2}\left(\mathbf{1}^T\boldsymbol{R}_\varphi^{-1}\mathbf{1}\right)^{-1/2}p(\varphi)p(c).$$

Corollary 4.3.1: The posterior mean of $\boldsymbol{\beta}$ conditional on $\boldsymbol{\delta}$ is given by

$$E(\boldsymbol{\beta}|\boldsymbol{Y},\boldsymbol{\delta}) = \int_0^1\int_0^\infty \boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}\,\omega(c,\varphi)dcd\varphi. \tag{4.12}$$

Proposition 4.3.4: Let m_i denote the ith component of $\boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}$, θ_i denote the ith diagonal element of $(n+v-1)^{-1}\left(\upsilon\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\right)\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi}$. Then, $p(\beta_i|\boldsymbol{Y},\boldsymbol{\delta}) = \int_0^\infty\int_0^1 \omega(c,\varphi)t(\beta_i;m_i,\theta_i,n+v-1)\,d\varphi dc$, where $t(\beta_i;m_i,\theta_i,n+v-1)$ is the univariate t-distribution with mean m_i, scale θ_i, and $n+v-1$ degrees of freedom. Thus,

$$P(\beta_i \le x|\boldsymbol{Y},\boldsymbol{\delta}) = \int_0^\infty\int_0^1 \omega(c,\varphi)P\left(t_{n+v-1} \le (x-m_i)/\sqrt{\theta_i}\right)d\varphi dc, \tag{4.13}$$

where t_{n+v-1} is a t random variable with mean 0, scale 1, and $n+v-1$ degrees of freedom.

Proposition 4.3.5: The posterior distribution of correlation parameter φ conditional on $\boldsymbol{\delta} \in \mathcal{M}$ is given by

$$p(\varphi|\boldsymbol{Y},\boldsymbol{\delta}\in\mathcal{M}) \propto \sum_{\boldsymbol{\delta}\in\mathcal{M}}\int_0^\infty\left[\upsilon\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\right]^{-(n-1+v)/2}\left|\boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T + \boldsymbol{R}_\varphi\right|^{-1/2}$$
$$\left(\mathbf{1}^T\boldsymbol{R}_\varphi^{-1}\mathbf{1}\right)^{-1/2}p(c)dcp(\boldsymbol{\delta})p(\varphi). \tag{4.14}$$

4.4 Prior Specification

This section discusses the problem of prior specification.

i. Choice of $p(\varphi)$: The key feature of the proposed Bayesian hierarchical model that distinguishes it from SSVS is the correlation parameter φ. An obvious choice for $p(\varphi)$ is the beta prior, i.e., $(\varphi) \propto \varphi^{a-1}(1-\varphi)^{b-1}$. In this chapter, we set $a = b = 2$ so that $p(\varphi)$ is symmetric with a mode of 0.5. This has the effect of pulling the posterior mode

of φ towards 0.5 when the data are scarce. An alternative choice of (a, b) is $a = b = 1$, which yields a uniform prior for φ. This choice tends to yield a posterior mode of φ closer to 0 or 1 than the previous choice. Large values of a and b are not recommended since the beta distribution becomes more concentrated as a and b get large.

If φ has a prior concentrated near 0, models with large ratios of number of wholeplot effects to subplot effects are favored since the wholeplot error variance is small compared to the subplot error variance. Conversely, a prior for φ concentrated near 1 will favor models with relatively fewer wholeplot effects. If we set $\varphi = 0$ or $\varphi = 0.99$ (if $\varphi = 1$, the covariance matrix for $\boldsymbol{Y}$ would be singular), we would observe two opposite and extreme behaviors in the model selection process. With little data available, a prior for φ with mode at 0.5 such as the beta prior with $a = b = 2$ is recommended. This can help avoid extreme behavior in the model selection process.

ii. Choice of $p(c)$: We use a discrete uniform prior for c with support

$$\{(1/2)^2, (3/4)^2, 1^2, 2^2, 3^2, 4^2, 5^2\}. \quad (4.15)$$

We have found that results obtained with a fixed value of c can be sensitive to the choice of c. If one fixed value of c is used, a large value of c (so that $\sqrt{c}\sigma$ is large relative to most nonzero β_i's) tends to favor (usually sparse) models with large effects. In this case, small effects can be missed. On the other hand, small values of c tend to favor less sparse models. However, a very small value of c would again favor sparse models possibly because the Bayesian model does not support the hypothesis of a true model with many effects that are considerably larger than $\sqrt{c}\sigma$. The support (4.15) of the proposed prior for c covers both small and large values of c. Thus, for the discrete uniform prior proposed above, results obtained are averaged over small and large values of c.

iii. Choice of d, v and λ: In this chapter, we set $d = v = \lambda = 0$. Note that $v = \lambda = 0$ yields the common noninformative prior for σ^2 while $d = 0$ implies that we set $\beta_i = 0$ if $\delta_i = 0$. These choices are also employed by Box and Meyer (1993). The results given in Section 4.3 are derived assuming that $\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}, c$ has a density, i.e., $d > 0$. However, the expressions for $p(\boldsymbol{\delta}|\boldsymbol{Y})$, $p(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\delta})$, $E(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\delta})$, $p(\beta_i|\boldsymbol{Y}, \boldsymbol{\delta})$, and $p(\varphi|\boldsymbol{Y}, \mathcal{M})$ given in Propositions 4.3.2-4.3.5 and Corollary 4.3.1 are well defined for $d = 0$. The choice $d = 0$ in each of these expressions can be interpreted as the limiting value as $d \to 0$. In addition, these results can also be obtained by employing a degenerate normal prior for $\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}, c$, where conditional on $\sigma^2, \boldsymbol{\delta}$, and c, $\beta_i = 0$ if $\delta_i = 0$ and the other β_i's are iid normal with mean 0 and variance $c\sigma^2$.

iv. Choice of ρ in (4.7): Following the recommendation of Box and Meyer (1993), we set $\rho = 0.25$. We have found that this choice works well. However, other values of ρ can be tried. For example, to avoid missing important effects, one may want to try $\rho = 0.5$. Although this chapter employs iid Bernoulli priors for the δ_i's, alternative forms of $p(\boldsymbol{\delta})$ such as the heredity prior proposed by Chipman et al. (1997) and a beta-binomial prior can also be employed.

Remark: Some papers (e.g., Gelfand et al. (1990)) on Bayesian variance components models employ independent priors for σ_w^2 and σ_s^2. However, it seems that in many experiments, if σ_w^2 is large, then σ_s^2 should also be large and vice versa. For example, suppose a wholeplot is a batch of material and a subplot is a smaller quantity taken from a batch. Then, if the batches are manufactured by an unstable process so that variation between batches is large (σ_w^2 is large), within batch variation would likely be large also (σ_s^2 is large). The proposed prior $p(\varphi, \sigma^2) \propto \varphi^{a-1}(1-\varphi)^{b-1}/\sigma^2$ reflects this

relationship between σ_w^2 and σ_s^2. It can be shown that $E(\sigma_w^2|\sigma_s^2) = \sigma_s^2 a/(b-2)$ if $b > 2$ and the mode of $p(\sigma_w^2|\sigma_s^2)$ is $m = max\{\sigma_s^2(a-1)/b, 0\}$. Thus, $E(\sigma_w^2|\sigma_s^2)$ is proportional to σ_s^2 for $b > 2$ and the mode is proportional to σ_s^2 for $a > 1$. This implies that a large σ_s^2 tends to give rise to a large σ_w^2 and vice versa.

4.5 Computation Strategies

This section proposes two algorithms for model search: FS and GLS. Some issues in computing the posterior quantities given in Section 4.3 are also discussed.

4.5.1 Forward Selection Algorithm

For regular two-level fractional factorial designs, we propose the use of a FS algorithm, i.e., a greedy optimization algorithm, to find models that have a high posterior probability $p(\boldsymbol{\delta}|\boldsymbol{Y})$. Because the normalization constant for $p(\boldsymbol{\delta}|\boldsymbol{Y})$ given by (4.10) is not known, we cannot compute $p(\boldsymbol{\delta}|\boldsymbol{Y})$. Our implementation of FS maximizes the *log posterior odds ratio* $o(\boldsymbol{\delta}) = \ln[p(\boldsymbol{\delta}|\boldsymbol{Y})/p(\mathbf{0}|\boldsymbol{Y})]$, which is equivalent to maximization of $p(\boldsymbol{\delta}|\boldsymbol{Y})$. The simple procedure is given below:

FS Algorithm

1. Start with $\boldsymbol{\delta} = \mathbf{0}$, the null model. Set $I = \{\}$ and $J = \{1, \ldots, p\}$.
2. Find the term $i^* \in J$ that, when added to the current model, yields the largest posterior odds ratio, i.e., $i^* = argmax_{i \in J}\{o(\boldsymbol{\delta}_{I \cup \{i\}})\}$. Set $I = I \cup \{i^*\}$ and $J = J \backslash \{i^*\}$.
3. Repeat Step 2 a total of p times. This generates a set of p models (one model of each size).
4. Choose from the p models the one with the highest *positive* posterior odds ratio

$o(\boldsymbol{\delta})$. If there are no models with positive $o(\boldsymbol{\delta})$, choose the null model.

Note that if ρ, the prior probability that an effect is active, is increased, larger models tend to be favored and the HPP model may change. However, the sequence of effects entered into the model will remain the same if the value of ρ is changed. This is because for a given model $\boldsymbol{\delta}$, a change in ρ from ρ_1 and ρ_2 merely increase $o(\boldsymbol{\delta})$ by $\|\boldsymbol{\delta}\|_1\{\ln[\rho_2/(1-\rho_2)] - \ln[\rho_1/(1-\rho_1)]\}$.

4.5.2 Global and Local Search Algorithms

For designs that do not give independent effect estimates, FS often yields poor models. In such cases, we use GLS, which consists of a global search (GS) algorithm and a local search (LS) algorithm:

Global Search (GS) Algorithm

1. Start with an indicator $\boldsymbol{\delta}^c = \boldsymbol{\delta}^0 \in \boldsymbol{D}_{\boldsymbol{\delta}^0}$. Set $\zeta = -\infty$.
2. For $i = 1, \ldots, p$, obtain $\boldsymbol{\delta}^i$ from $\boldsymbol{\delta}^c$ by switching the value of δ_i^c (from 0 to 1 or 1 to 0).
3. Find $i^* = argmax_i\{o(\boldsymbol{\delta}^i)\}$.
4. If $o(\boldsymbol{\delta}^{i^*}) \leq \zeta$, **stop** and return $\boldsymbol{\delta}^{opt} = \boldsymbol{\delta}^c$. Otherwise, set $\boldsymbol{\delta}^c = \boldsymbol{\delta}^{i^*}$, $\zeta = o(\boldsymbol{\delta}^c)$ and go to Step 2.
5. Repeat Steps 1-4 for each $\boldsymbol{\delta}^0 \in \boldsymbol{D}_{\boldsymbol{\delta}^0}$. In this chapter, we take $\boldsymbol{D}_{\boldsymbol{\delta}^0}$ to be a 100-run maximin design on the model space $\{0,1\}^p$.

Local Search (LS) Algorithm

1. Specify t and compute the *minimum acceptable criterion value* $MACV = o(\boldsymbol{\delta}^*_{GS}) - \ln t$, where $\boldsymbol{\delta}^*_{GS}$ is the HPP model found by GS.
2. Start with $I = \{\boldsymbol{\delta}^{opt}\}$ and $J = \{\boldsymbol{\delta}^{opt}\}$, where $\boldsymbol{\delta}^{opt}$ is a local optimal solution found

with GS such that $o(\boldsymbol{\delta}^{opt}) \geq MACV$. Set $\boldsymbol{\delta}^0 = \boldsymbol{\delta}^{opt}$.

3. For $i = 1, \ldots, p$, obtain $\boldsymbol{\delta}^i$ from $\boldsymbol{\delta}^0$ by switching the value of δ_i^0 (from 0 to 1 or 1 to 0). If $\boldsymbol{\delta}^i \notin I$ and $o(\boldsymbol{\delta}^i) \geq MACV$, set $I = I \cup \{\boldsymbol{\delta}^i\}$ and $J = J \cup \{\boldsymbol{\delta}^i\}$.
4. Set $J = J \backslash \{\boldsymbol{\delta}^0\}$. If J is empty or if $|I| \geq 10^4$, **stop** and return I. Otherwise, set $\boldsymbol{\delta}^0$ equal to the first element in J and return to Step 3.
5. Repeat Steps 2-4 for each $\boldsymbol{\delta}^{opt}$ that satisfies $o(\boldsymbol{\delta}^{opt}) \geq MACV$.

Note that GS returns a local optimal solution $\boldsymbol{\delta}^{opt}$ for each given starting indicator $\boldsymbol{\delta}^0$. Given the current indicator $\boldsymbol{\delta}^c$ (which is initialized as $\boldsymbol{\delta}^0$), it finds an adjacent vector (one that differs from $\boldsymbol{\delta}^c$ in only one component) that gives the largest improvement over $o(\boldsymbol{\delta}^c)$. The best adjacent indicator becomes the current indicator, and the process is repeated until no further improvements can be achieved. The set $\boldsymbol{D}_{\boldsymbol{\delta}^0}$ of starting indicators $\boldsymbol{\delta}^0$ for the GS algorithm can have an important influence on the results obtained with the algorithm. We have found that a maximin design (Santner et al., 2003) on the model space $\{0,1\}^p$ is a better choice for $\boldsymbol{D}_{\boldsymbol{\delta}^0}$ than a randomly generated design because GS with $\boldsymbol{D}_{\boldsymbol{\delta}^0}$ tends to consistently give good models whereas GS started from random indicators do not. To generate a maximin design, we generate a large number (usually 10,000) of random designs, where the components of each design are sampled from independent Bernoulli distributions with success probability 0.5. For each design, we compute the minimum pairwise distance between design runs, and then pick the design with the maximum pairwise distance. We recommend taking $\boldsymbol{D}_{\boldsymbol{\delta}^0}$ to be a design of 100 runs because we have found that results obtained with this choice tend to be good (finds most or all good models) and consistent (very similar results for all replicates). However, our experience indicates that the design size may need to be increased for

experiments with $p > 30$ and $n < p$.

The local search (LS) algorithm searches around the good local optimal solutions ($\boldsymbol{\delta}^{opt}$ that satisfies $o(\boldsymbol{\delta}^{opt}) \geq MACV$) for good models ($\boldsymbol{\delta}$ that satisties $o(\boldsymbol{\delta}) \geq MACV$). We take $MACV = o(\boldsymbol{\delta}^{*}_{GS}) - \ln t$, where $\boldsymbol{\delta}^{*}_{GS}$ is the HPP model found by GS and t is a constant that we choose (note that $\boldsymbol{\delta}^{*}_{GS}$ is not necessarily the true HPP model over the entire model space). This implies that we are only interested in models that have posterior probabilities at least $1/t$ times that of the HPP model. Note that Madigan and Raftery (1994) have also proposed a search technique to find graphical models that meet the $MACV$ requirement. The LS algorithm builds up a set of models I whose log posterior odds is at least $MACV$, and then conducts further local searches around each of those models. The main difference between GS and LS is that LS is willing to explore sub-optimal models (i.e. those whose log posterior odds is at least $MACV$), while GS only follows paths through the model space that increase the posterior probability. We have found that when there is substantial model uncertainty and t is large, the set I can grow very large, causing the algorithm to be unable to terminate within a reasonable amount of time. To circumvent this problem, we terminate LS when I contains 10^4 or more models.

GLS has a few potential advantages over Gibbs sampling (Heaton and Scott, 2010). Firstly, it can be more efficient computationally. Since LS systematically explores neighboring models, it does not reevaluate the same models. While GS may revisit the same models when restarted from different points, it terminates each time it finds a local optimum. In contrast, Gibbs sampling would revisit high probability models frequently. Secondly, when there is a large number of good models that contain very different effects (i.e., when there is a large number of good locally optimal solutions), GLS can efficiently

discover these models. In contrast, an MCMC chain based on Gibbs sampling may get stuck around one of the locally optimal solutions. This creates the need for convergence assessment using multiple starting points and running Markov chains of long lengths.

4.5.3 Estimation of Posterior Model Probabilities

Note that the posterior probability of a model $\boldsymbol{\delta}$ is $p(\boldsymbol{\delta}|\boldsymbol{Y}) = \gamma exp[o(\boldsymbol{\delta})]$, where γ is a normalization constant. We use the following method to estimate posterior probability. Let $\boldsymbol{\delta}_1^*, \dots, \boldsymbol{\delta}_{U_t}^*$, where $p(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) > \cdots > p(\boldsymbol{\delta}_{U_t}^*|\boldsymbol{Y})$, be the models found by GLS with $MACV = o(\boldsymbol{\delta}_{GS}^*) - \ln t$. Then,

$$p(\boldsymbol{\delta}_j^*|\boldsymbol{Y}) \leq q_t(\boldsymbol{\delta}_j^*) = exp[o(\boldsymbol{\delta}_j^*)]/\{\textstyle\sum_{i=1}^{U_t} exp[o(\boldsymbol{\delta}_i^*)]\}. \quad (4.16)$$

If for each t in a sequence of increasing values t, GS finds the same HPP model (so that $o(\boldsymbol{\delta}_{GS}^*)$ is constant and $MACV$ decreases), and GLS finds all models $\boldsymbol{\delta}$ such that $o(\boldsymbol{\delta}) \geq MACV$, then $q_t(\boldsymbol{\delta}_j^*)$ is monotone decreasing and $\lim_{t\to\infty} q_t(\boldsymbol{\delta}_j^*) = p(\boldsymbol{\delta}_j^*|\boldsymbol{Y})$. This gives a method to estimate $p(\boldsymbol{\delta}_j^*|\boldsymbol{Y})$:

1. Run GLS for a sequence $\{t_1, \dots, t_k\}$ of increasing values of t. We recommend taking $t_l = Tl$ with $T = 50$ and $k = 4$.
2. Stop if either $\{q_{t_1}(\boldsymbol{\delta}_1^*), \dots, q_{t_k}(\boldsymbol{\delta}_1^*)\}$ is not monotone nonincreasing or $\boldsymbol{\delta}_{GS}^* \neq \boldsymbol{\delta}_1^*$ or the top M models do not remain the same (if less than M models are found when $t = t_l$, then check that no additional models found with $t = t_{l+1}$ meet the $MACV$ requirement with $t = t_l$). In this case, the size of $\boldsymbol{D}_{\boldsymbol{\delta}^0}$ needs to be increased because GLS does not give consistent results.
3. Estimate $p(\boldsymbol{\delta}_j^*|\boldsymbol{Y})$ with $\tilde{p}(\boldsymbol{\delta}_j^*|\boldsymbol{Y}) = q_{t_k}(\boldsymbol{\delta}_j^*)$. Alternatively, use the model $\Phi^{-1}[q_t(\boldsymbol{\delta}_1^*)] = \alpha_0 + \alpha_1/\sqrt{t}$, where $\Phi(\cdot)$ is the standard normal cdf. In this case,

estimate $p(\boldsymbol{\delta}_1^*|\boldsymbol{Y})$ with $\hat{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = \Phi(\hat{\alpha}_0)$, where $\hat{\alpha}_0$ is the least squares estimate of α_0, and $p(\boldsymbol{\delta}_j^*|\boldsymbol{Y}), j > 1$ with $\hat{p}(\boldsymbol{\delta}_j^*|\boldsymbol{Y}) = [\hat{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y})/\tilde{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y})]\tilde{p}(\boldsymbol{\delta}_j^*|\boldsymbol{Y})$. Note that $p(\varphi|\boldsymbol{Y}, \boldsymbol{\delta} \in \mathcal{M})$ and $p(\beta_i|\boldsymbol{Y}, \boldsymbol{\delta} \in \mathcal{M})$ can be computed with $\mathcal{M}$ equal to the top m models in $\{\boldsymbol{\delta}_1^*, \dots, \boldsymbol{\delta}_{U_{t_k}}^*\}$ such that $\sum_{i=1}^{m} \hat{p}(\boldsymbol{\delta}_i^*|\boldsymbol{Y})$ is large.

Remark 1: We recommend $T = 50$ and $k = 4$ because we have found that $q_{200}(\boldsymbol{\delta}_1^*)$ is often quite close to $p(\boldsymbol{\delta}_1^*|\boldsymbol{Y})$. The decrease in $q_t(\boldsymbol{\delta}_1^*)$ tends to be slow for t greater than 100. Moreover, as t increases, the set of models that satisfy the $MACV$ requirement increases, which can make GLS time-consuming.

Remark 2: Step 2 ensures that $\boldsymbol{D}_{\boldsymbol{\delta}^0}$ is large enough so that GS consistently finds the HPP model returned by GLS, and GLS consistently finds the top M models, where M should be large enough to include the top few models with high posterior probabilities. We take $M \geq 10$ for all examples in this chapter.

Remark 3: The model $\Phi^{-1}[q_t(\boldsymbol{\delta}_1^*)] = \alpha_0 + \alpha_1/\sqrt{t}$, where $\alpha_1 > 0$, is accurate for estimating $p(\boldsymbol{\delta}_1^*|\boldsymbol{Y})$ in simulations we performed. Moreover, it is also well-behaved. Since $\lim_{t\to\infty}(\alpha_0 + \alpha_1/\sqrt{t}) = \alpha_0$, $p(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = q_\infty(\boldsymbol{\delta}_1^*)$ is estimated with $\Phi(\alpha_0)$, which is in $(0,1)$.

4.5.4 Other Computation Issues

Computation of $p(\boldsymbol{\delta}|\boldsymbol{Y})$ given in (4.10) and the posterior mean and credible intervals for the components of $\boldsymbol{\beta}_{\boldsymbol{\delta}^*}$ via (4.12) and (4.13) require evaluation of low dimensional integrals. Because we use a discrete uniform prior for c, all integrals with respect to c in (4.10)-(4.14) can easily be replaced with finite sums. However, computation of (4.10)-(4.13) requires integration with respect to φ. We compute this uni-

dimensional integral using the Gaussian quadrature method (Miller and Rice, 1983). Since $p(\varphi)$ has a beta distribution and all moments of the beta distribution can be computed with a simple recursive formula, we can easily compute an L-point discrete approximation to $p(\varphi)$ such that the first $2L-1$ moments of the discrete approximation match that of the beta distribution. Details of the procedure for doing this is given in Section D.3 of Appendix D. Denote the support points of the discrete approximation by $\xi_1, \ldots, \xi_L$ and the corresponding probability masses by $w_1, \ldots, w_L$. Then, this discrete approximation is used in place of $p(\varphi)d\varphi$ to compute integrals. An L-point approximation will give exact values of integrals with respect to $p(\varphi)d\varphi$ if the integrand is a polynomial of degree $2L-1$ or less. Thus, the use of a small L is sufficient to approximate the integrals (4.10)-(4.13) accurately. In this chapter, we use $L = 5$.

The integrand of the unnormalized posterior probabilities $p(\boldsymbol{\delta}|\boldsymbol{Y})$ given by (4.10) tends to be extremely small in value. Thus, to improve computation accuracy, the logarithm of the integrand should be computed at the support points of $p(c)$ and the discrete approximation of $p(\varphi)$. Then, a constant should be subtracted from each of the quantities before they are exponentiated and summed. We take the constant as the value obtained in the first evaluation of the integrand in the computation of $p(\mathbf{0}|\boldsymbol{Y})$.

4.6 Examples

This section gives simulation studies of the performance of the proposed method and analyzes some real split plot RPD experiments using the method. The first example presents a simulation study of the Type I and Type II errors of the proposed method. We also study the effectiveness of GLS as a search algorithm and the accuracy of the method for estimating posterior model probabilities given in Section 4.5.3. The second example

is the well-known Ina tile experiment, which is a fractional factorial split plot experiment. In the third example, we analyze an RPD experiment for a heat-exchanger fan casing, where the signal and noise factors are subplot factors and the control factors are wholeplot factors. Experiments with signal and noise factors as subplot factors and control factors as wholeplot factors are useful and convenient in RPD. They are useful because the control-by-noise and control-by-signal interactions are estimated with lower statistical error. They are convenient because it is easier to test one product under multiple noise/signal conditions than to test different products under each noise/signal condition. The fourth example analyzes a rather complex split unit experiment reported by Taguchi (1987). In the second and third examples, we compare the results obtained using a $beta(2,2)$ prior for φ (see Section 4.4) and the results obtained with $\varphi = 0$ to assess the consequences of ignoring the split plot error structure. We call the HPP model obtained with the former and latter priors the split plot analysis (SPA) model and completely randomized analysis (CRA) model respectively.

4.6.1 Example 4.1: Simulation Experiments

4.6.1.1 Study of Type I and Type II Error Achieved with Proposed Approach

We perform simulation experiments to study the frequentist performance of the proposed method under the default choice of priors. In the first experiment, the split plot design is obtained by crossing a 2^3 design for wholeplot factors A, B and C, and a 2^1 design for subplot factor D. There are seven wholeplot effects and eight subplot effects (15 factorial effects in total). Data are generated from two mean models: a model with active effects B,C,AB,D,AD,CD,BCD (**Model 1**) and a model with active effects B,C,AB,D,AD,CD (**Model 2**). Model 1 contains three wholeplot effects (B,C,AB) and

four subplot effects (D,AD,CD,BCD) whereas Model 2 contains one less subplot effect. We fix $\sigma^2 = 1$, set all inactive effects to 0, and set all model coefficients of active effects to $\beta = \eta\sqrt{\sigma_{ave}^2}$, where σ_{ave}^2 is the average of the variances of the generalized least squares estimators of the nonzero coefficients. Both η and φ are changed according to a 3^2 factorial design shown in Figure 4.1. Each of the nine combinations in the 3^2 design is plotted as a circle in the figure. For each (η, φ), we perform 100 simulations. The exact value of β depends on both η and φ but it mainly depends on η. In particular, β is in the range of 0.71-0.75 when $\eta = 3$, 1.41-1.5 when $\eta = 6$, and 2.12-2.25 when $\eta = 9$. In each simulation, we generate a set of response values and apply the FS algorithm. We use FS because in our experience, both FS and GLS always give the same best models for two-level regular fractional factorial designs. We compare the results with Lenth's method (Wu and Hamada, 2009; Lenth, 1989) because it is a popular method for testing effects in unreplicated fractional factorial designs. Critical points that give individual error rates (IER) (see Wu and Hamada (2009) for a definition) of 0.1 and 0.05 for Lenth's method are employed and the same IER is applied to test wholeplot and subplot effect estimates.

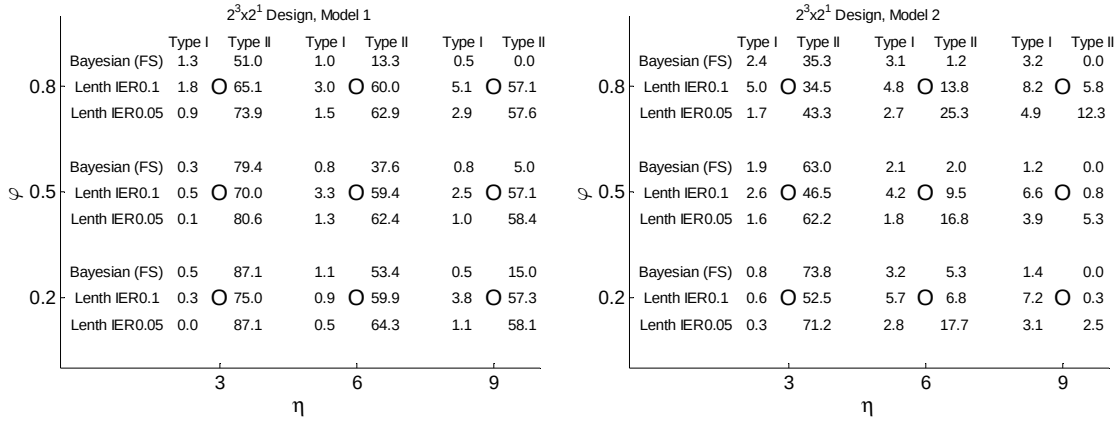

Figure 4.1: Type I and Type II Error Rates for Proposed Method (Bayesian (FS)) and Lenth's Method with IER=0.1 and IER=0.05 (Lenth IER0.1, Lenth IER0.05); $(\eta, \varphi) \in \{3,6,9\} \times \{0.2,0.5,0.8\}$; Model 1 (left) and Model 2 (right)

For each simulation run, the **Type I** and **Type II** errors of the HPP model are computed. The Type I error is the percentage of inactive effects identified as active (number of inactive effects in HPP model/number of inactive effects) and the Type II error is the percentage of active effects identified as inactive. The *averages* of the Type I and Type II errors for the 100 simulation runs for each combination of (η, φ) are plotted in Figure 4.1. Similar plots that give detailed breakdown of the Type I and Type II errors in terms of wholeplot and subplot effects are given in Section D.1 of Appendix D. In Figure 4.1, the proposed method is denoted by Bayesian (FS) and Lenth's method with IER=0.1 is denoted by "Lenth IER0.1" (similarly for IER=0.05). Figure 4.1 shows that the Type I error is relatively low for all methods and combinations of (η, φ). The Type II error changes more dramatically. For $\eta = 3$, the Type II errors for all methods are intolerably large (for Model 1, it ranges from 50 to 90 percent). Therefore, it is not meaningful to rank order the methods. However, as η increases to 6 or 9, the Type II error of the proposed method becomes acceptable except when $(\eta, \varphi) = (6,0.2)$ or $(6,0.5)$ for Model 1. The proposed method performs better as φ increases. It can also be seen that the error rates for Model 2 are in general lower than those for Model 1 since Model 2 is sparser.

Figure 4.1 indicates that when η equals 6 or 9, or when $\varphi = 0.8$, the proposed method outperforms Lenth's method. The improvement in Type II error can be very large (up to about 57 percent when $(\eta, \varphi) = (9,0.8)$) in the case of Model 1 and sizeable in the case of Model 2. For Model 2, the proposed method dominates Lenth's method with IER=0.1 when $\eta \geq 6$. In addition, it has Type I error comparable to that of Lenth's method with IER= 0.05 but smaller Type II errors when $\eta \geq 6$. For Model 1, the Type II

error for Lenth's method remains large even as η increases. This large Type II error is due to the failure of the method in detecting active subplot effects (Figure D.2, Appendix D shows that the Type II error for subplot effects is almost always 100 percent). The reason is that Lenth's method relies on the median of absolute values of the effects to screen out large effects, and for Model 1, there are four active subplots out of a total of eight subplots effects. Consequently, the median must be contaminated with active subplot effects. In comparison, the proposed method does not suffer from this problem. However, it does incur a higher wholeplot Type II error for $(\eta, \varphi) = (6,0.2), (6,0.5), (9,0.2)$. For Model 2, the gains in Type II error achieved with the proposed method mainly arise from the improvement in wholeplot Type II error for $\eta = 6,9$ (Figure D.1, Appendix D).

For the second experiment, the experimental design is a split plot design constructed from an $\mathrm{OA}(27, 3^9)$ (Table D.1, Appendix D) used by Taguchi (1987) (page 264). This design has $N = 9$ wholeplots and $r = 3$ subplots per wholeplot, and it is not a crossed array. Factors A-D are wholeplot factors and factors E-J are subplot factors (there is no factor I). The set of candidate effects consists of the linear and quadratic effects of each factor and all linear-by-linear interactions. This gives a total of 54 effects, 14 of which are wholeplot effects. Data are generated from two mean models: **Model 3**, which has active effects Al, El, Aq, Eq, AlEl, and **Model 4**, which has active effects Bl, Fl, Jl, Bq, Gq, AlGl, BlDl, DlEl, DlJl. Model 3 satisfies the strong heredity principle (Wu and Hamada, 2009) but Model 4 does not. Moreover, Model 4 is also larger. Thus, we would expect variable selection to be difficult with Model 4. Values of σ^2, β, and the experimental design for (η, φ) are as in the first experiment. For each (η, φ), we perform

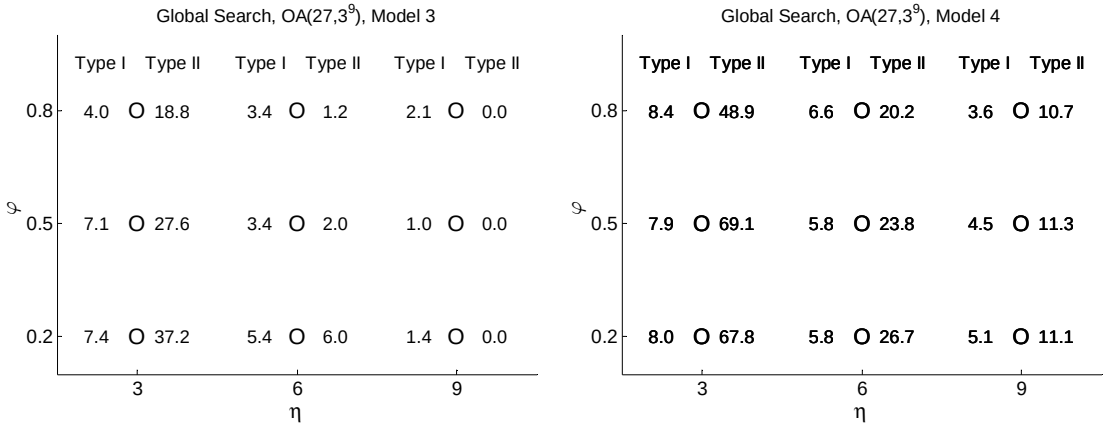

Figure 4.2: Type I and Type II Error Rates for Best Model Found with GS; $(\eta, \varphi) \in \{3,6,9\} \times \{0.2,0.5,0.8\}$; Model 3 (left) and Model 4 (right)

50 simulations. We perform a smaller number of simulations because they are very time consuming. In each simulation, we apply GS and FS. We found that GS almost always gives a HPP model with higher log posterior odds, and such a model tends to be better in terms of Type I and Type II errors (see next section). Note that we use GS instead of GLS because we focus only on evaluating the HPP model that is found and we find it computationally convenient. The HPP model $\boldsymbol{\delta}_1^*$ found with GLS is almost always the HPP model $\boldsymbol{\delta}_{GS}^*$ found with GS and the former is at least as good as the latter.

Figure 4.2 plots the Type I and Type II errors achieved with the HPP model returned by GS. Plots that breakdown the Type I and Type II errors in terms of wholeplot and subplot effects are given in Section D.1 of Appendix D. For Model 3, all errors, except possibly the Type II errors at $\eta = 3$, are small. For Model 4, the Type I errors are small but the Type II errors are very high for $\eta = 3$. However, the Type II errors are tolerable for $\eta = 6,9$.

In summary, the simulations above suggest that the proposed method can do well in terms of Type I error in many cases since it maintains a Type I error below 10 percent for all experiment runs. The Type II error would be low when the effects are sufficiently

large ($\eta \geq 6$) and the true model is sufficiently sparse. Even in the case of unreplicated fractional factorial split plot designs, it is advantageous to use the proposed method since it can outperform Lenth's method.

4.6.1.2 Comparison of GS and FS

In this section, we compare GS (with 100 starting points obtained from a maximin design) and FS in terms of the capabilities of the algorithms in finding good models. We have mentioned that both algorithms seem to always give the same best models for two-level regular fractional factorials. However, we have found that for designs with correlated effect estimates, FS can yield suboptimal models that are very different from the HPP model found by GS. This point is aptly demonstrated by an analysis of some of the results of the simulations reported in Section 4.6.1.1, specifically those for Model 4.

In each of the 50 simulations for a given (η, φ), GS and FS are applied to the same randomly drawn response vector. This allows a pairwise comparison of GS and FS. Let $\boldsymbol{\delta}_{GS}^*$ and $\boldsymbol{\delta}_{FS}^*$ be the HPP models found with GS and FS. Then, we have 50 pairs of $o(\boldsymbol{\delta}_{GS}^*) - o(\boldsymbol{\delta}_{FS}^*)$ for each experiment run. The mean $E[o(\boldsymbol{\delta}_{GS}^*) - o(\boldsymbol{\delta}_{FS}^*)]$ of $o(\boldsymbol{\delta}_{GS}^*) - o(\boldsymbol{\delta}_{FS}^*)$ is given in the fourth column of Table 4.1. This should be compared with the mean $E[o(\boldsymbol{\delta}_{GS}^*)]$ and standard deviation $std[o(\boldsymbol{\delta}_{GS}^*)]$ of $o(\boldsymbol{\delta}_{GS}^*)$ given in the second and third columns of Table 4.1. It can be seen that $E[o(\boldsymbol{\delta}_{GS}^*) - o(\boldsymbol{\delta}_{FS}^*)]$ is relatively large compared to $E[o(\boldsymbol{\delta}_{GS}^*)]$ and $std[o(\boldsymbol{\delta}_{GS}^*)]$. The number of times $o(\boldsymbol{\delta}_{GS}^*) - o(\boldsymbol{\delta}_{FS}^*) \geq 0$ and the number of times $o(\boldsymbol{\delta}_{GS}^*) - o(\boldsymbol{\delta}_{FS}^*) > 0$ are given in the fifth and sixth columns of Table 4.1.We see that except in one simulation, $o(\boldsymbol{\delta}_{GS}^*)$ is at least as good as $o(\boldsymbol{\delta}_{FS}^*)$ and the former is larger in value most of the time.

Table 4.1: Summary Statistics for $o(\boldsymbol{\delta}_{GS}^*)$ and $o(\boldsymbol{\delta}_{GS}^*) - o(\boldsymbol{\delta}_{FS}^*)$

Run	$o(\boldsymbol{\delta}_{GS}^*)$		$o(\boldsymbol{\delta}_{GS}^*) - o(\boldsymbol{\delta}_{FS}^*)$		
	Mean	Std. Dev.	Mean	No. ≥0	No. >0
1	3.89	2.80	1.05	50	30
2	4.02	3.11	0.88	49	31
3	6.61	2.85	2.36	50	44
4	8.84	3.14	4.08	50	48
5	10.20	3.27	5.18	50	49
6	13.26	2.90	7.92	50	50
7	13.73	2.66	7.87	50	50
8	15.21	2.90	9.03	50	50
9	19.50	3.37	10.77	50	50

Table 4.2: Mean Increases in Number of Type I and Type II Errors of $\boldsymbol{\delta}_{FS}^*$ over $\boldsymbol{\delta}_{GS}^*$

Run	η	φ	Mean Difference	
			Type I	Type II
1	3	0.2	-1.46	1.46
2	3	0.5	-0.96	1.22
3	3	0.8	-0.78	2.44
4	6	0.2	1.56	3.24
5	6	0.5	1.44	3.06
6	6	0.8	1.92	2.96
7	9	0.2	2.92	3.24
8	9	0.5	3.26	3
9	9	0.8	5.02	2.32

We now demonstrate that $\boldsymbol{\delta}_{GS}^*$ tends to have lower Type I and Type II errors than $\boldsymbol{\delta}_{FS}^*$. In each of the 50 simulations for a given (η, φ), the *number* of Type I and Type II errors made by $\boldsymbol{\delta}_{GS}^*$ and $\boldsymbol{\delta}_{FS}^*$ are computed and the increases incurred by $\boldsymbol{\delta}_{FS}^*$ over $\boldsymbol{\delta}_{GS}^*$, i.e., Type I$(\boldsymbol{\delta}_{FS}^*)$ − Type I$(\boldsymbol{\delta}_{GS}^*)$ and Type II$(\boldsymbol{\delta}_{FS}^*)$ − Type II$(\boldsymbol{\delta}_{GS}^*)$ are computed. Table 4.2 gives the mean increases for each (η, φ). Based on Wilcoxon's signed-rank test, all differences are significant at the 0.05 level. Except for two cases (Type I errors for Run 2 and Run 3), the p-values are less than or equal to 0.001. Note that the mean increases in Table 4.2 should be compared with the number of inactive effects, which is 45, and the

number of active effects, which is 9. We see from Table 4.2 that the mean increase in Type I error is actually negative for runs with $\eta = 3$, i.e., the Type I error of $\boldsymbol{\delta}_{FS}^{*}$ is lower than that of $\boldsymbol{\delta}_{GS}^{*}$ for $\eta = 3$. However, for all other cases $\boldsymbol{\delta}_{GS}^{*}$ outperform $\boldsymbol{\delta}_{FS}^{*}$ in both Type I and Type II errors. This indicates that the gains of GS over FS in finding models with better log posterior odds can be associated with important gains in frequentist properties. Thus, GS is worth its extra computation cost.

4.6.1.3 Study of Effectiveness of GLS and Estimation of Posterior Model Probabilities

We consider the 12-run Plackett-Burman split plot design given in Table D.2 in Appendix D. Two responses given by Y^1 and Y^2 in the table are analyzed. The set of effects consists of all main effects and two-factor interactions. Thus, there are a total of six main effects and 15 two-factor interactions. This gives about 2.1×10^6 models. The posterior probability and log posterior odds of the top 10 models are given in Table 4.3. We see that for Y^2, the posterior distribution of $\boldsymbol{\delta}$ is concentrated at a single model and models around it. The top 10 models account for 77% of the posterior mass. For Y^1, no single model stands out; the top 10 models account for only 37% of the posterior mass of $\boldsymbol{\delta}$ and those models look quite different. Table 4.3 also demonstrate the potential dangers of model selection based on $o(\boldsymbol{\delta})$: one cannot know how much of the posterior mass of $\boldsymbol{\delta}$ is concentrated on a group of top models simply by looking at their log posterior odds.

To determine how good GLS is at finding good models, we apply GLS with 12 values of t as given in Table 4.4. The table gives the actual number of models that have log posterior odds at least $Q(t) = o(\boldsymbol{\delta}^*) - \ln t$, where $\boldsymbol{\delta}^*$ is the true HPP model, and the number of models found by GLS. For all t, GS finds $\boldsymbol{\delta}^*$, i.e., $\boldsymbol{\delta}_{GS}^{*} = \boldsymbol{\delta}^*$, so that $MACV = Q(t)$. We see from Table 4.4 that for Y^1, GLS misses at most two good models, which

happens when $t = 140$ and $t = 150$. For Y^2, GLS does not miss any good model. Moreover, in all cases, GLS does not miss any of the top 10 models.

In Figure 4.3, we plot $q_t(\boldsymbol{\delta}_1^*)$ (upper bound) versus t and the true value $p(\boldsymbol{\delta}_1^*|\boldsymbol{Y})$. The figure shows that the decrease in $q_t(\boldsymbol{\delta}_1^*)$ is slow when $t \geq 100$. For Y^1, we obtain $\tilde{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = q_{200}(\boldsymbol{\delta}_1^*) = 0.133$; for Y^2, $\tilde{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = 0.579$. The estimates are sufficiently close to the true values (see Figure 4.3) for practical purposes. Fitting the model $\Phi^{-1}[q_t(\boldsymbol{\delta}_1^*)] = \alpha_0 + \alpha_1/\sqrt{t}$ to the data for $t = 50{,}100{,}150{,}200$ yields the estimates $\hat{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = 0.102$ for Y^1 and $\hat{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = 0.522$ for Y^2, which are close to the true values.

Table 4.3: Top 10 Models for Y^1 and Y^2

Y^1			Y^2		
Models, $\boldsymbol{\delta}$	$o(\boldsymbol{\delta})$	$p(\boldsymbol{\delta}\|\boldsymbol{Y})$	Models, $\boldsymbol{\delta}$	$o(\boldsymbol{\delta})$	$p(\boldsymbol{\delta}\|\boldsymbol{Y})$
A,B,C,AF,BC,CF	7.674	0.0889	A,AB,AC,AD	11.863	0.5145
C,AF,BC,CF	7.504	0.0749	A,AB,AC,AD,CD	10.138	0.0917
AD,AE,BC,BF,CD,CE	6.918	0.0417	A,AB,AC,AD,AF	9.084	0.0320
BC	6.773	0.0361	A,F,AB,AC,AD	9.041	0.0306
B,C,AF,BC,CF	6.682	0.0329	A,AB,AC,AD,DE	8.963	0.0283
C,AF,BC,CD,CF,DF	6.533	0.0284	A,AB,AC,AD,BD	8.678	0.0213
C,AF,BC	6.404	0.0250	A,B,AB,AC,AD	8.517	0.0181
A,C,AF,BC,CF	6.266	0.0217	A,E,AB,AC,AD	8.133	0.0123
B,C,AF,BC	5.597	0.0111	A,AB,AC,AD,EF	8.100	0.0120
C,AF,BC,CD,CF	4.996	0.0061	A,AB,AC,AD,BF	8.092	0.0118
Sum of Probabilities		0.367	Sum of Probabilities		0.773

Table 4.4: Number of Models Found with GLS and Actual Number of Models that Satisfy $o(\boldsymbol{\delta}) \geq o(\boldsymbol{\delta}^*) - \ln t$ for Y^1 and Y^2

t		20	40	50	60	80	100	120	140	150	160	180	200
Y^1	No. of Models	15	49	68	93	115	133	154	167	177	185	197	213
	Actual No. of Models	15	49	68	94	115	134	155	169	179	186	198	214
Y^2	No. of Models	5	7	14	19	19	19	20	21	24	25	25	26
	Actual No. of Models	5	7	14	19	19	19	20	21	24	25	25	26

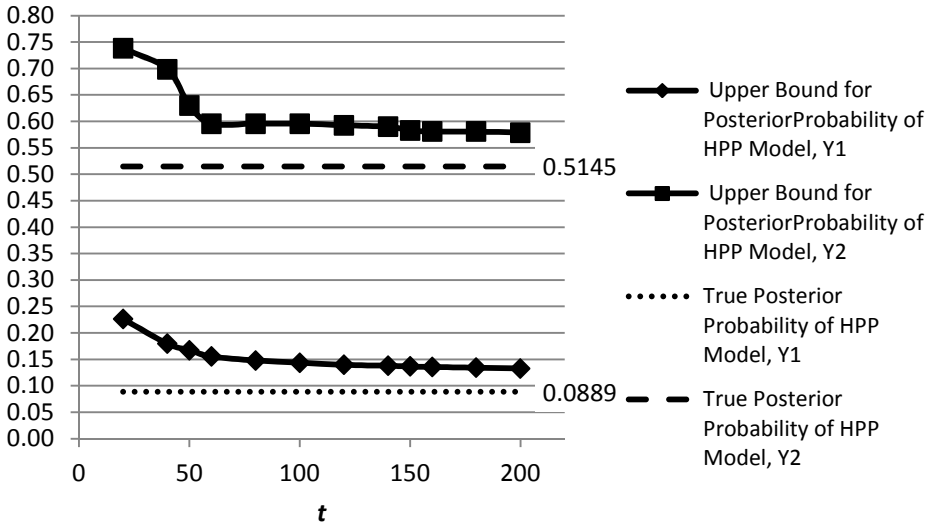

Figure 4.3: Plot of $q_t(\boldsymbol{\delta}_1^*)$ versus t for Y^1 and Y^2

Now, we consider the 12-run Plackett-Burman split plot design given in Table D.3 in Appendix D, which is the same as the design in Table D.2 except that there are five instead of six factors. Two responses given by Y^3 and Y^4 in the table are analyzed. The set of effects consists of all main effects and two factor interactions. This gives 32768 models.

Table 4.5: Top 10 Models for Y^3 and Y^4

Y^3			Y^4		
Models	$o(\boldsymbol{\delta})$	$p(\boldsymbol{\delta}\|\boldsymbol{Y})$	Models	$o(\boldsymbol{\delta})$	$p(\boldsymbol{\delta}\|\boldsymbol{Y})$
C	1.316	0.0684	A,B,C,AB,AC	3.716	0.3525
A,C	1.136	0.0571	A,B,C,AB,AC,CD	2.416	0.0961
A,B,C,D,E,AB,AC	0.980	0.0488	A,C,AB,AC	2.350	0.0899
A,C,E	0.881	0.0443	A,B,C,AB,AC,BE	1.513	0.0389
C,E	0.554	0.0319	C,AB,AC	1.175	0.0278
A,B,C,E	0.259	0.0238	A,B,C,E,AB,AC	0.692	0.0171
A,B,C,E,AB	0.104	0.0203	A,B,C,AB,AC,BD	0.421	0.0131
A,B,C	0.048	0.0192	A,B,C,AB,AC,AD	0.421	0.0131
ϕ	0.000	0.0183	A,C,AB,AC,CD	0.411	0.0129
A,C,AB	-0.003	0.0183	A,B,C,AB,AC,DE	0.399	0.0128
Sum of Probabilities		0.350	Sum of Probabilities		0.674

Table 4.6: Number of Models Found with GLS and Actual Number of Models that Satisfy $o(\boldsymbol{\delta}) \geq o(\boldsymbol{\delta}^*) - \ln t$ for Y^3 and Y^4

t		20	40	50	60	80	100	120	140	150	160	180	200
Y^3	No. of Models	43	84	114	127	172	183	201	221	230	247	282	309
	Actual No. of Models	43	84	114	127	172	183	202	221	231	247	283	309
Y^4	No. of Models	5	13	17	20	24	27	34	38	40	41	42	48
	Actual No. of Models	5	13	17	20	24	27	34	38	40	41	42	48

The posterior probability and log posterior odds of the top 10 models are given in Table 4.5. Table 4.6 gives the actual number of models that have log posterior odds at least $Q(t) = o(\boldsymbol{\delta}^*) - \ln t$ ($\boldsymbol{\delta}^*$ is the true HPP model), and the number of models found by GLS for 12 different values of t. For all t, GS finds $\boldsymbol{\delta}^*$, i.e., $\boldsymbol{\delta}^*_{GS} = \boldsymbol{\delta}^*$, so that $MACV = Q(t)$. We see from Table 4.6 that GLS almost always finds all good models. In Figure 4.4, we plot $q_t(\boldsymbol{\delta}_1^*)$ versus t and the true value $p(\boldsymbol{\delta}_1^*|\boldsymbol{Y})$. For Y^3, we obtain $\tilde{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = q_{200}(\boldsymbol{\delta}_1^*) = 0.0845$ whereas for Y^4, we obtain $\tilde{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = 0.419$. The estimates are

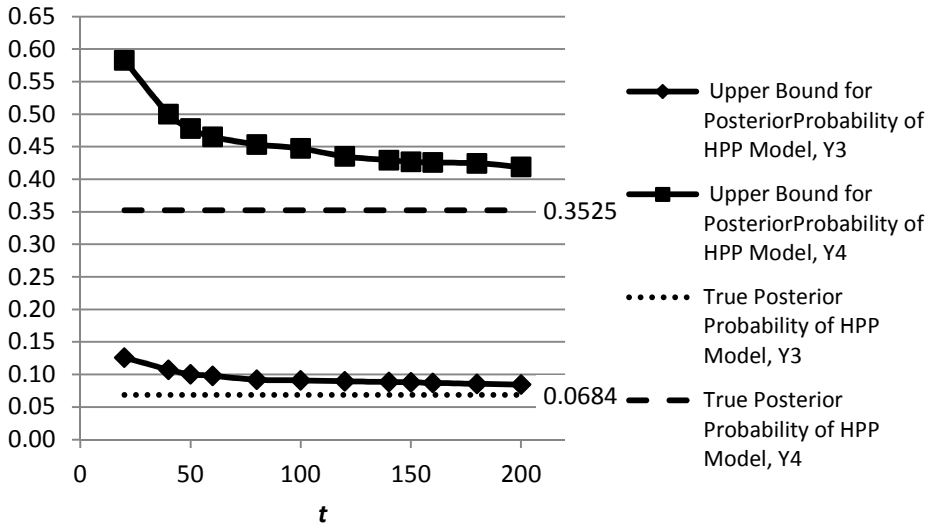

Figure 4.4: Plot of $q_t(\boldsymbol{\delta}_1^*)$ versus t for Y^3 and Y^4

sufficiently close the true values (see Figure 4.4) for most practical purposes. If we fit the model $\Phi^{-1}[q_t(\boldsymbol{\delta}_1^*)] = \alpha_0 + \alpha_1/\sqrt{t}$ using data for $t = 50,100,150,200$, then we obtain the estimates $\hat{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = 0.0716$ for Y^3 and $\hat{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = 0.362$ for Y^4, which are very close to the true values.

4.6.2 Example 4.2: Ina Tile Experiment

The Ina tile experiment (Taguchi, 1987) studies eight factors (labeled D, B, A, F, E, C, G, and H) in 16 runs. The noise factor H is the subplot factor. There are $N = 8$ wholeplots and $r = 2$ subplots per wholeplot. We take the set of candidate effects for model selection as the main effects of all wholeplot/ control factors, the main effect of the subplot/noise factor, and all control-by-noise interactions. These effects are estimable if all other interactions are assumed negligible. We work with the transformed response given by Bisgaard and Sutherland (2003). Table 4.7 presents the first five steps taken by FS with $\varphi \sim beta(2,2)$. The term added to the model at each step is given in the second column and the log posterior odds $o(\boldsymbol{\delta})$ is given in the third column. Results for $\varphi = 0$ are given in the fourth and fifth columns. For $\varphi \sim beta(2,2)$, the model {A, H} obtained at the second step (shaded cells) has the largest value of $o(\boldsymbol{\delta})$, i.e., the SPA model is {A, H}. It is seen that the CRA model is also {A, H}. The split plot analysis does not fully agree with Bisgaard and Sutherland's (2003) analysis (based on eyeballing half-normal plots), which finds that AH is active. However, AH is the third effect added by FS and $o(\boldsymbol{\delta}) = 1.72$ for the model {A, H, AH} is not too small compared to $o(\boldsymbol{\delta}) = 2.24$ for the model {A, H}. Moreover, if we increase $\rho = p(\delta_i = 1)$ from 0.25 to 0.4, then the SPA model would be {A, H, AH}, as is easily verified by adding $[\ln(0.4/0.6) - \ln(0.25/0.75)]\|\boldsymbol{\delta}\|_1 = 0.6931\|\boldsymbol{\delta}\|_1$ to each value in the third column of Table 4.7. In

Table 4.7: FS Results for Ina Tile Experiment

Step	Split Plot[1]		Completely Randomized[2]	
	Added Term	$o(\boldsymbol{\delta})$	Added Term	$o(\boldsymbol{\delta})$
1	H	2.0129	H	-0.14872
2	A	2.2427	A	1.3301
3	AH	1.7203	E	0.32268
4	E	0.61579	D	-0.5205
5	D	-0.39481	AH	-1.3311

[1]FS algorithm applied with $\varphi \sim beta(2,2)$. [2]FS algorithm applied with $\varphi = 0$.

contrast, if the split plot error structure is ignored, AH is only fifth in the list of effects to enter the model and the model with five effects has a smaller posterior probability than the null model.

For this example, FS chooses at each step either the largest wholeplot effect (ranked by the size of the factorial effect estimates) or the largest subplot effect not in the model to enter the model. This behavior has been observed to hold in our analyses of regular two-level designs. From a frequentist point of view, the behavior is sensible because for such designs, all wholeplot (respectively subplot) effect estimates have the same variance, and all effect estimates are independent. Note that for such designs, the generalized least squares estimates are equivalent to the ordinary least squares estimates.

To estimate posterior model probabilities, we run the probability estimation procedure given in Section 5.3 with $t = 50,100,150,200$ and $M = 30$. For the four values of t, there are 101, 199, 291, and 358 models that have log posterior odds at least $Q(t)$. The true HPP model $\boldsymbol{\delta}^*$ is {A,H}. GLS correctly finds all models in each case (thus, the choice of M is of no consequence and $\boldsymbol{\delta}_1^* = \boldsymbol{\delta}^*$). We find that $\tilde{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = 0.0913$, $\hat{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = 0.0686$, and $\sum_{j=1}^{358} \hat{p}(\boldsymbol{\delta}_j^*|\boldsymbol{Y}) = 0.751$. These estimates are close to the true values $p(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = 0.0726$ and $\sum_{j=1}^{358} p(\boldsymbol{\delta}_j^*|\boldsymbol{Y}) = 0.795$. Because $p(\boldsymbol{\delta}_1^*|\boldsymbol{Y})$ is small,

model uncertainty is high. Since the set $\mathcal{M}$ of top 358 models accounts for an estimated 75% of the posterior mass of $\boldsymbol{\delta}$, we can estimate the marginal posterior probability of each effect conditional on $\boldsymbol{\delta} \in \mathcal{M}$. This gives $\hat{p}(\delta_i = 1|\boldsymbol{Y}) = \sum_{j:\boldsymbol{\delta}_j^* \in \mathcal{M}_i} \hat{p}(\boldsymbol{\delta}_j^*|\boldsymbol{Y}) / \sum_{j=1}^{358} \hat{p}(\boldsymbol{\delta}_j^*|\boldsymbol{Y}) = \sum_{j:\boldsymbol{\delta}_j^* \in \mathcal{M}_i} \tilde{p}(\boldsymbol{\delta}_j^*|\boldsymbol{Y}) / \sum_{j=1}^{358} \tilde{p}(\boldsymbol{\delta}_j^*|\boldsymbol{Y})$, where $\mathcal{M}_i = \left\{\boldsymbol{\delta}_j^* = (\delta_{j1}^*, \dots, \delta_{jp}^*)^T \in \mathcal{M}: \delta_{ji}^* = 1\right\}$. A bar graph of $\hat{p}(\delta_i = 1|\boldsymbol{Y})$ and $p(\delta_i = 1|\boldsymbol{Y})$ is given in Figure 4.5, which shows that the probability estimates for the top ranking effects are close to the true values. Despite the high model uncertainty, we see from Figure 4.5 that we can be quite certain that effects H and A are active. Interestingly, the ranking of effects based on marginal posterior probabilities does not fully agree with the order they are entered in FS. For example, G is ranked 6 based on the marginal posterior probabilities while FH is the 6th effect to be entered by FS. However, the top five effects match the order they are entered by FS.

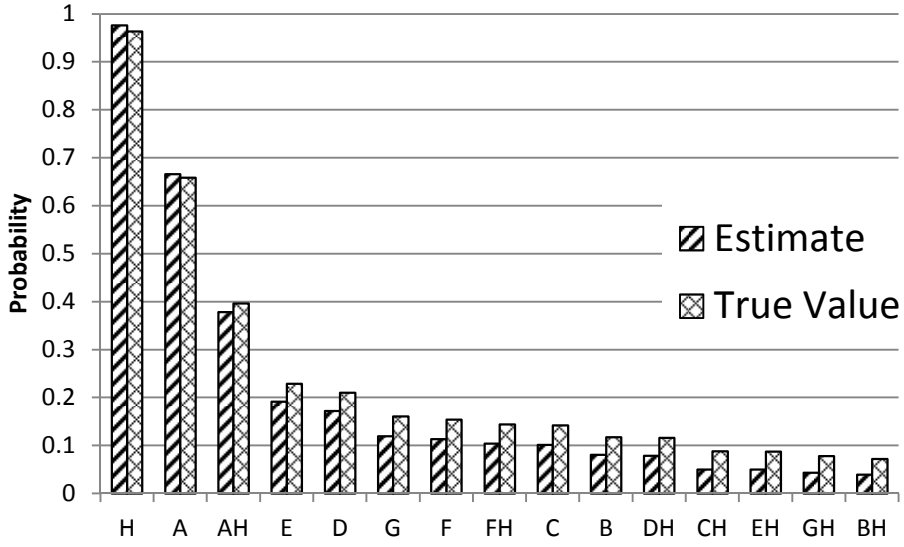

Figure 4.5: Bar Chart of Estimates and True Values of Marginal Posterior Probability of Each Effect

Before we end the section, we point out that FS is useful for regular two-level designs because it provides a ranking of factorial effects and a quantitative measure of relative importance (log posterior odds) that supplements a half-normal plot analysis. While GLS provides better search capability than FS and allows estimation of marginal posterior probabilities, it is computationally more intensive than FS. This computational advantage can be significant for large fractional factorial designs.

4.6.3 Example 4.3: Design of Heat-Exchanger Fan Casing of Clothes Dryer

In this example, we analyze an experiment for the robust design of a heat-exchanger fan casing for clothes dryers (Kim, 1999). The control array is a 3^{4-1} design and the signal-noise array is a 2×3 full factorial. The control factors A, B, C, D are parameters of the fan casing design. The signal factor M is the rotation speed of the fan and the noise factor N, which has two levels, is the amount of clothes in the drum. The response is the flow rate of the heat-exchanger fan. Data for this experiment are given in Table D.4 in Appendix D. Changing the control factors require manufacturing a new fan casing but changing the signal and noise factors are clearly very easy. Thus, nine fan casing designs are tested under six combinations of signal and noise factor conditions, i.e., the experiment is a split plot experiment with control factors as wholeplot factors and signal and noise factors as subplot factors.

The set of effects of interest are chosen as follows. The main effect for the two-level noise factor is its linear main effect. For all other factors (which each has three levels), linear and quadratic main effects are constructed from the orthogonal linear and quadratic contrasts given in Supplementary Section G of Wu and Hamada (2009). We also entertain all interactions between linear main effects. There are a total of $p = 26$

effects and $n = 54$ subplots. However, the full model cannot be estimated with frequentist methods because there are only $N = 9$ wholeplots and 14 wholeplot effects.

We apply GLS to six choices of priors: **Prior 1**: the default choice given in Section 4.4; **Prior 2**: $\varphi = 0$; **Prior 3**: $(a, b) = (1,3)$; **Prior 4**: $(a, b) = (3,1)$; **Prior 5**: $p(c) = unif\{0.5^2, 0.75^2, 1^2, 2^2\}$ (i.e., $p(c)$ is the uniform distribution on $\{0.5^2, 0.75^2, 1^2, 2^2\}$); **Prior 6**: $p(c) = unif\{2^2, 3^2, 4^2, 5^2\}$. For Priors 2- 6, all prior parameters that are not given are set to their default choices. The objectives are to assess the effects of ignoring the split plot error structure (Prior 2) and the sensitivity of results to prior specification (Priors 3- 6). Prior 3 gives $E(\varphi) = 0.25$, $E(\sigma_s^2/\sigma_w^2) = E[(1-\varphi)/\varphi] = \infty$, and $\text{median}\{\sigma_s^2/\sigma_w^2\} = 3.847$. Prior 4 gives $E(\varphi) = 0.75$, $E(\sigma_w^2/\sigma_s^2) = E[\varphi/(1-\varphi)] = \infty$, and $\text{median}\{\sigma_w^2/\sigma_s^2\} = 3.847$. For Prior 5, active effects have a prior that is a mixture of normal distributions with standard deviations 0.5 to 2 times σ (conditionally on σ). Finally, for Prior 6, active effects have a prior that is a mixture of normal distributions with standard deviations 2 to 5 times σ.

In each of the six applications of GLS, we set $MACV = o(\boldsymbol{\delta}_{GS}^*) - \ln 6$. The top two models obtained for each prior are given in Table 4.8 together with their log posterior odds. In the table, the linear and quadratic main effects of factor X are denoted by Xl and Xq respectively, whereas the linear by linear interaction of factors X and Y is denoted by XlYl. The number of models found is given in the bottom row of the table. For Prior 4, only one model with log posterior odds at least $MACV$ is found.

We see that the SPA model (HPP model $\boldsymbol{\delta}_1^*$ for Prior 1) contains 11 effects, whereas the CRA model (HPP model for Prior 2) contains 13 effects. The SPA model contains the DlMl term whereas the CRA model does not. Although the CRA model

Table 4.8: Model Selection Results for Fan Casing Experiment. A shaded cell represents an active term and terms not listed are excluded from all models.

Effects	Prior 1		Prior 2		Prior 3		Prior 4		Prior 5		Prior 6	
	Default		$\varphi = 0$		$a = 1,$ $b = 3$		$a = 3,$ $b = 1$		$p(c) = unif\{0.5^2, 0.75^2, 1^2, 2^2\}$		$p(c) = unif\{2^2, 3^2, 4^2, 5^2\}$	
	$\boldsymbol{\delta}_1^*$	$\boldsymbol{\delta}_2^*$	$\boldsymbol{\delta}_1^*$	$\boldsymbol{\delta}_2^*$	$\boldsymbol{\delta}_1^*$	$\boldsymbol{\delta}_2^*$	$\boldsymbol{\delta}_1^*$	$\boldsymbol{\delta}_2^*$	$\boldsymbol{\delta}_1^*$	$\boldsymbol{\delta}_2^*$	$\boldsymbol{\delta}_1^*$	$\boldsymbol{\delta}_2^*$
Al												
Bl												
Cl												
Dl												
Nl												
Ml												
Aq												
Bq												
Dq												
AlCl												
BlCl												
BlNl												
BlMl												
ClNl												
ClMl												
DlMl												
NlMl												
Log Posterior Odds	127.1	125.3	122.9	122.8	123.5	122.6	129.5	-	112.8	112.7	127.6	125.9
$P(\boldsymbol{\delta}_1^*\|\boldsymbol{Y})$ $/P(\boldsymbol{\delta}_2^*\|\boldsymbol{Y})$	5.78		1.11		2.46		-		1.04		5.78	
No. Models Found	2		8		3		1		6		2	

contains all 10 other effects in the SPA model, it also contains three additional wholeplot effects: Aq, Dq, AlCl. Thus, the significance of wholeplot effects appears to be exaggerated if correlations are ignored.

The results obtained with Priors 1,3,4,6 are similar. The HPP model $\boldsymbol{\delta}_1^*$ is the same for each prior and the second best model $\boldsymbol{\delta}_2^*$ for Priors 1, 3, and 6 are the same. For Prior 4, the fact that GLS cannot find any model different from $\boldsymbol{\delta}_{GS}^* = \boldsymbol{\delta}_1^*$ that has

posterior probability at least $P(\boldsymbol{\delta}_{GS}^{*}|\boldsymbol{Y})/6$ suggests that model uncertainty may be low for this choice of prior. For Priors 1 and 6, the ratio of the posterior probabilities of the HPP and second best models $P(\boldsymbol{\delta}_{1}^{*}|\boldsymbol{Y})/P(\boldsymbol{\delta}_{2}^{*}|\boldsymbol{Y}) = exp[o(\boldsymbol{\delta}_{1}^{*}) - o(\boldsymbol{\delta}_{2}^{*})]$ equals 5.78. This means that the HPP model is significantly better than the second best model. For Prior 3, the HPP model is about 2.5 times better than the second best model. Thus, for Priors 1,3,4,6, the HPP model is significantly better than the other models. For Prior 5, which include only the small values $0.5^2, 0.75^2, 1^2$ in the support of $p(c)$, there is higher model uncertainty since six models are found with GLS and $\boldsymbol{\delta}_{1}^{*}$ and $\boldsymbol{\delta}_{2}^{*}$ appear to be equally good. We conclude that the model selection results are insensitive to reasonable choices of a and b, and to whether we include the values of $0.5^2, 0.75^2, 1^2$ in the support of the discrete uniform distribution for $p(c)$. This suggests that we can be quite confident about the results obtained with Prior 1. In general, a sensitivity analysis can give more confidence to results obtained with the default priors if it indicates that the results are insensitive to prior parameter changes.

For Prior 1, we run the procedure given in Section 5.3 with $t = 50,100,150,200$ and $M = 10$. For the four values of t, GLS finds 16, 23, 29, and 35 models. Because the total number of models is $2^{26} \approx 67.1 \times 10^6$, it is difficult to evaluate all models. We find that $\tilde{p}(\boldsymbol{\delta}_{1}^{*}|\boldsymbol{Y}) = q_{200}(\boldsymbol{\delta}_{1}^{*}) = 0.504$, $\hat{p}(\boldsymbol{\delta}_{1}^{*}|\boldsymbol{Y}) = 0.458$, $\hat{p}(\boldsymbol{\delta}_{2}^{*}|\boldsymbol{Y}) = 0.079$, and $\hat{p}(\boldsymbol{\delta}_{3}^{*}|\boldsymbol{Y}) = 0.053$. Thus, the HPP model accounts for most of the posterior mass of the indicator. This justifies post model selection inference conditioned on the event that $\boldsymbol{\delta}_{1}^{*}$ is the true model or the event that either $\boldsymbol{\delta}_{1}^{*}$ or $\boldsymbol{\delta}_{2}^{*}$ is the true model.

Table 4.9 gives point estimates and equal tail 95% credible intervals for the effects in the SPA model conditioned on the event that the SPA model is the correct

model. The percentage increase column in Table 4.9 gives the width of the SPA interval (credible interval constructed with Prior 1) as a percentage of the width of the CRA interval (credible interval obtained with Prior 2). We see that the wholeplot effects Bl, Cl, and Dl have SPA intervals that are more than twice as wide as their CRA intervals. On the other hand, the lengths of the SPA intervals for the subplot effects are about 30 percent the lengths of their CRA intervals. These observations suggest that it can be misleading to use interval estimates that do not account for correlation. Interestingly the point estimates obtained with Prior 2 are very close to those obtained with Prior 1. This agrees with the result that least squares and generalized least squares estimates are equivalent for any design that is a Cartesian product of designs for wholeplot and subplot factors (Letsinger et al., 1996).

Table 4.9: 95% Credible Intervals for Effects in SPA Model, Fan Casing Experiment

Effect	Prior 1[1]			Prior 2[2]			Percentage Increase[3]
	LCL	Mean	UCL	LCL	Mean	UCL	
Bl	0.3601	0.3757	0.3913	0.3704	0.3771	0.3838	232.7%
Cl	0.1140	0.1296	0.1452	0.1234	0.1301	0.1368	232.6%
Dl	-0.0630	-0.0474	-0.0318	-0.0543	-0.0476	-0.0409	232.6%
Nl	-0.1475	-0.1455	-0.1436	-0.1521	-0.1454	-0.1387	28.8%
Ml	0.2022	0.2041	0.2060	0.1973	0.2040	0.2107	28.8%
BlNl	-0.0246	-0.0227	-0.0208	-0.0294	-0.0227	-0.0160	28.8%
BlMl	0.0320	0.0339	0.0358	0.0272	0.0339	0.0406	28.8%
ClNl	-0.0142	-0.0122	-0.0103	-0.0189	-0.0122	-0.0055	28.8%
ClMl	0.0150	0.0169	0.0189	0.0102	0.0169	0.0236	28.8%
DlMl	-0.0069	-0.0050	-0.0031	-0.0117	-0.0050	0.0017	28.8%
NlMl	0.0130	0.0150	0.0169	0.0083	0.0150	0.0217	28.8%

[1]Posterior mean and equal tail 95% credible intervals with $\varphi \sim beta(2,2)$. [2]Posterior mean and equal tail 95% credible intervals with $\varphi = 0$. [3]SPA credible interval width as a percentage of CRA credible interval width.

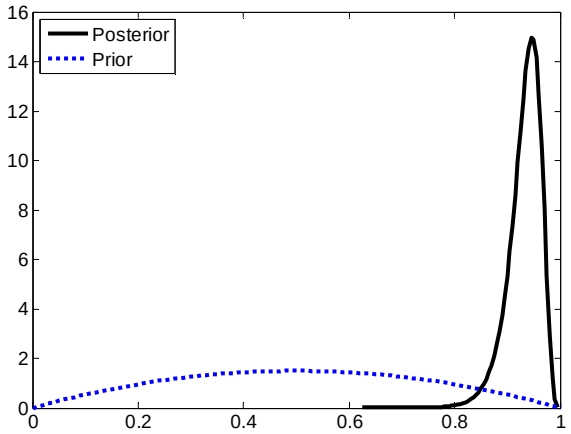

Figure 4.6: Posterior Distribution of Correlation Parameter ($\mathcal{M} = \{\boldsymbol{\delta}_1^*, \boldsymbol{\delta}_2^*\}$)

For Prior 1, the posterior density $p(\varphi|\boldsymbol{Y}, \boldsymbol{\delta} \in \mathcal{M})$, where $\mathcal{M} = \{\boldsymbol{\delta}_1^*, \boldsymbol{\delta}_2^*\}$, is plotted in Figure 4.6. The $beta(2,2)$ prior for φ is also plotted in the figure. The figure shows that subplot observations in the same wholeplot are highly correlated, i.e., the subplot variance is small compared to the wholeplot variance. Because model uncertainty is low, essentially the same plot of $p(\varphi|\boldsymbol{Y}, \boldsymbol{\delta} \in \mathcal{M})$ is obtained if we take $\mathcal{M} = \{\boldsymbol{\delta}_1^*\}$.

4.6.4. Example 4.4: Carbon Powder for Transmitter

Taguchi (1987) proposes methods to construct split plot designs, which he calls split unit designs. Split unit designs are designs constructed by reordering the rows of orthogonal arrays and assigning wholeplot and subplot factors to appropriate columns in such a way that each r consecutive runs for wholeplot factors have the same levels. For these designs, if main effects and two-factor interactions are entertained, the number of effects can be more than the number of runs. Therefore, variable selection is an important step in the analysis of split unit designs.

Taguchi (1987) (Chapter 9) gives an experiment to improve the properties of carbon powder that is placed in the transmitter of a telephone. There are a few interesting

features of the experiment. The design is not a crossed array, there are more effects than observations, and the levels of some of the factors are unequally spaced. The nine factors and their levels are given in Table 4.10. Four responses are observed in the experiment. We shall only mention the results for the bulk specific gravity response. The design matrix and responses are given in Table D.5 in Appendix D. The wholeplot factors are factors A-D, and the subplot factors are factors E-J. There are $N = 9$ wholeplots and $r = 3$ subplots per wholeplot. Factor J is a noise factor and all other factors are control factors. Factors C, D, H, and J are qualitative factors. The main effects for factors C, D, and H are included in the set of candidate effects. For factor J, we introduce two dummy variables J1 and J2 representing the difference in mean response between the top and middle and between the bottom and middle positions of the vessel. For the quantitative factors A, B, E, F, G, the linear main effect columns are obtained by coding the factor levels so that the low level is -1 and the high level is 1. Because the three levels for factors B, E, F, G are not even spaced, the middle levels of the factors are not 0. The quadratic main effect columns for factors A, B, E, F, G are obtained by squaring the linear main effects column. We also include all interactions between one of the noise effects J1, J2 and a linear effect of factors A-H in the set of candidate effects. This gives $2 \times 8 = 16$ control-by-noise interactions. Thus, we have $p = 31$ effects but only $n = 27$ runs.

To estimate posterior model probabilities, we run the probability estimation procedure given in Section 5.3 with $t = 50,100,150,200$ and $M = 20$. For these four values of t, GLS finds 152, 295, 448, and 611 models that meet the $MACV$ requirement. We find that $\tilde{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = q_{200}(\boldsymbol{\delta}_1^*) = 0.0683$, $\hat{p}(\boldsymbol{\delta}_1^*|\boldsymbol{Y}) = 0.0478$. The top 10 models

Table 4.10: Factors and Their Levels, Transmitter Carbon Powder Experiment

Factor	Levels		
	1	2	3
A: Heat Treatment Temperature ^{o}C	1000	1100	1200
B: Rate of Increase of Heat Treatment Temperature (^{o}C/hr)	25	50	100
C: Heat Treatment Atmosphere	N_2	H_2	
D: Flow Quantity of Gases (cm^3/min)	Low (400 for N_2 and 800 for H_2)	High (800 for N_2 and 1600 for H_2)	
E: Rate of Increase of Carbonization Temperature (^{o}C/hr)	25	50	100
F: Oxidation Time (hours)	2	12.6	80
G: Rate of Increase of Oxidation Temperature (^{o}C/hr)	25	50	100
H: Particle Size (mesh)	40-50	50-70	
J: Position Within Vessel	top	middle	bottom

found by GLS, their log posterior odds, and their posterior probability estimates are given in Table 4.11. A plot of the cumulative sum of $\hat{p}(\boldsymbol{\delta}_j^*|\boldsymbol{Y})$ versus j for the 611 models found with $t = 200$ is given in Figure 4.7. It can be seen that the top 150 models account for only approximately 50% of the posterior mass of $\boldsymbol{\delta}$. Thus, model uncertainty is high. The set $\mathcal{M}$ of top 611 models can be used for estimating marginal posterior probabilities of effects because it accounts for an estimated 70% of the posterior mass of $\boldsymbol{\delta}$. Estimates of the marginal posterior probability of each effect conditional on $\boldsymbol{\delta} \in \mathcal{M}$ are plotted in Figure 4.8. We see that despite the high model uncertainty, we can be quite sure that all effects in the top model {Cl, Fl, Hl, Fq, BlJ2, ClJ2, HlJ1} are active. The HPP model consists of all effects with marginal posterior probabilities greater than 0.5. In Figure 4.9, we plot $p(\varphi|\boldsymbol{Y}, \boldsymbol{\delta} \in \mathcal{M})$. The posterior mode of $p(\varphi|\boldsymbol{Y}, \boldsymbol{\delta} \in \mathcal{M})$ is approximately 0.3.

Table 4.11: Top 10 Models Together with Their Log Posterior Odds and Posterior Probability Estimates

Model	$o(\boldsymbol{\delta})$	$\tilde{p}(\boldsymbol{\delta}\|\boldsymbol{Y})$	$\hat{p}(\boldsymbol{\delta}\|\boldsymbol{Y})$
Cl,Fl,Hl,Fq,BlJ2,ClJ2,HlJ1	17.873	0.0683	0.0478
Cl,Fl,Hl,Fq,Gq,BlJ2,ClJ2,HlJ1	17.608	0.0524	0.0367
Cl,Fl,Hl,Eq,Fq,Gq,BlJ2,ClJ2,HlJ1	17.172	0.0339	0.0237
Cl,Fl,Hl,Fq,ClJ2,HlJ1	17.090	0.0312	0.0218
Cl,Fl,Hl,Fq,Gq,BlJ2,ClJ2,DlJ2,ElJ2,HlJ1	16.687	0.0209	0.0146
Cl,Fl,Hl,Fq,Gq,BlJ2,ClJ2,ElJ2,HlJ1	16.546	0.0181	0.0127
Cl,Fl,Hl,Eq,Fq,BlJ2,ClJ2,HlJ1	16.387	0.0155	0.0108
Cl,Fl,Hl,Fq,BlJ2,ClJ2,ElJ2,HlJ1	16.301	0.0142	0.0099
Cl,Fl,Hl,Fq,HlJ1	16.283	0.0139	0.0097
Cl,Fl,Hl,Fq,BlJ2,ClJ2,GlJ1,HlJ1	16.215	0.0130	0.0091

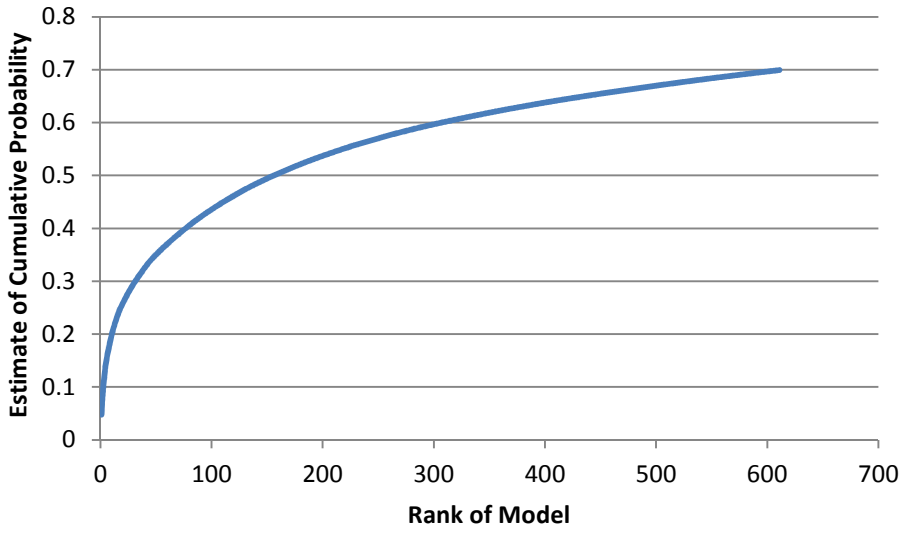

Figure 4.7: Plot of Cumulative Probability Estimate versus Rank of Model

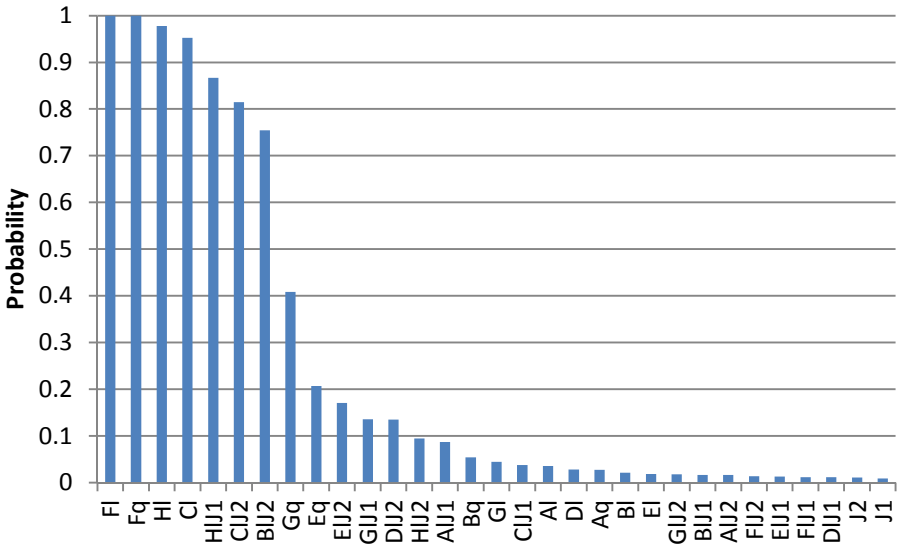

Figure 4.8: Bar Chart of Estimates of Marginal Posterior Probability of Each Effect

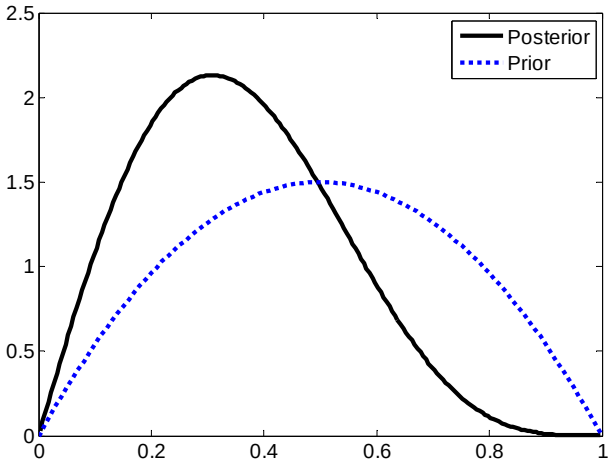

Figure 4.9: Posterior Distribution of Correlation Parameter for Bulk Specific Gravity Response

4.7 Conclusions

This chapter proposes a Bayesian method for model selection in split plot experiments. It is especially useful when the number of effects is comparable to the number of runs, which is quite common in RPD. RPD experiments run as split plot experiments with noise and signal factors as subplot/wholeplot factors and control factors as wholeplot/subplot factors automatically gives more power for detecting control-by-

noise and control-by-signal interactions than a completely randomized design. These interactions are important for achieving the objectives of RPD.

We employ a Bayesian hierarchical model with an indicator vector for model selection. It can be viewed as a generalization of the Bayesian models introduced by George and McCulloch (1993) and Box and Meyer (1993). Unlike Bayesian variable selection in iid error settings, the proposed model includes the correlation φ between subplot observations in the same wholeplot. A new algorithm called GLS is proposed to find good models and to estimate the posterior probability of the models.

Simulation results presented in the chapter show that the proposed method has low Type I error in most cases and low Type II error when the active effects are sufficiently large and the true model is sufficiently sparse. The proposed method can perform better than Lenth's method on unreplicated fractional factorial split plot experiments. It is also demonstrated that GLS performs well at finding models with log posterior odds above a threshold and that the proposed procedure for estimating posterior model probabilities is accurate. Three real examples are analyzed in the chapter: the Ina tile experiment, the heat-exchanger fan casing experiment and the carbon powder experiment. The problem of analyzing a highly fractionated split plot experiment has received scant attention in the statistical literature. The proposed method can help to popularize the use of fractionated split plot experiments.

4.8 References

Bingham, D., and Goh, J. (2012). "Bayesian Variable Selection for Split-Plot Designs with Complex Aliasing," Unpublished Manuscript.

Bingham, D., and Sitter, R. (2003). "Fractional Factorial Split-Plot Designs for Robust Parameter Experiments," *Technometrics*, 45(1), 80-89.

Bisgaard, S., and Sutherland, M. (2003). “Split Plot Experiments: Taguchi’s Ina Tile Experiment Reanalyzed,” *Quality Engineering*, 16(1), 157-164.

Box, G.E.P. and Jones, S. (2001). “Split-plot Designs for Robust Product and Process Experimentation,” *Quality Engineering*, 13(1), 127-134.

Box, G.E.P. and Meyer, R.D. (1993). “Finding the Active Factors in Fractionated Screening Experiments,” *Journal of Quality Technology*, 25(2), 94-105.

Chipman, H., Hamada, M., and Wu, C.F.J. (1997). “A Bayesian Variable-Selection Approach for Analyzing Designed Experiments with Complex Aliasing,” *Technometrics*, 39(4), 372-381.

Gelfand, A.E., Hills, S.E., Racine-Poon, A., Smith, A.F.M. (1990). “Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling,” *Journal of the American Statistical Association*, 85(412), 972-985.

George, E.I., and R.E. McCulloch. (1993). “Variable Selection Via Gibbs Sampling,” *Journal of the American Statistical Association*, 88(423), pp. 881-889.

George, E.I., and R.E. McCulloch. (1997). “Approaches for Bayesian Variable Selection,” *Statistica Sinica*, 7, 339-373.

Giesbrecht, F.G. and M.L. Gumpertz. (2004). *Planning, Construction, and Statistical Analysis of Comparative Experiments*. New York: Wiley.

Gilmour, S.G. and P. Goos. (2009). “Analysis of Data from Non-Orthogonal Multistratum Designs in Industrial Experiments,” *Applied Statistics*, 58(4), 467-484.

Goos, P. (2002). *The Optimal Design of Blocked and Split-Plot Experiments*. New York: Springer-Verlag.

Goos, P., Langhans, I., and Martina, V. (2006). “Practical Inference from Industrial Split-Plot Designs,” *Journal of Quality Technology*, 38(2), 162-179.

Heaton, M.J. and Scott, J.G. (2010). “Bayesian Computation and the Linear Model,” in *Frontiers of Statistical Decision Making and Bayesian Analysis in Honor of*

James O. Berger, ed. Chen, M.H., Dey, D.K., Müller, P., Sun, D., and Ye, K., New York: Springer.

Hill, B.M. (1965). "Inference about Variance Components in the One-Way Model," *Journal of the American Statistical Association*, 60(311), 806-825.

Jones, B. and Nachtsheim, C.J. (2009). "Split-Plot Designs: What, Why, and How," *Journal of Quality Technology*, 41(4), 340-361.

Kim, J. (1999). "Optimum Design of a Heat-Exchanger-Fan Casing of Clothes Dryer Using the Taguchi Method," *KSME International Journal*, 13(9), 962-972.

Lenth, R.V. (1989). "Quick and Easy Analysis of Unreplicated Factorials," *Technometrics*, 31, 469-473.

Letsinger, J.D., Myers, R.H., and Lentner, M. (1996). "Response Surface Methods for Bi-Randomizatiaon Structures," *Journal of Quality Technology*, 28, 381-397.

Madigan, D. and Raftery, A.E. (1994). "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window," *Journal of the American Statistical Association*, 89(428), 1535-1546.

Miller, A.C., and Rice, T.R. (1983), "Discrete Approximations of Probability Distributions," *Management Science*, 29(3), 352-362.

Næs, T., Aastveit, A.H., and Sahni, N.S. (2007). "Analysis of Split-plot Designs: An Overview and Comparison of Methods," *Quality and Reliability Engineering International*, 23(5), 801-820.

O'Hara, R.B. and Sillanpää, M.J. (2009). "A Review of Bayesian Variable Selection Methods: What How and Which," *Bayesian Analysis*, 4(1), 85-118.

Santner, T.J., Williams, B.J., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. New York: Springer.

Sun, L., Hsu, J.S.J., Guttman, I., and Leonard, T. (1996). "Bayesian Methods for Variance Component Models," *Journal of the American Statistical Association*, 91(434), 743-752.

Taguchi, G. (1987). *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs*, Vols. 1&2, UNIPUB/Kraus International Publications, White Plains NY.

Tiao, G.C. and Tan, W.Y. (1965). "Bayesian Analysis of Random Effect Models in the Analysis of Variance. I. Posterior Distribution of Variance-Components," *Biometrika*, 52(1/2), 37-53.

Vounatsou, P. and A.F.M. Smith. (1997). "Simulation-based Bayesian Inferences for Two-Variance Components Linear Models," *Journal of Statistical Planning and Inference*, 59, 139-161.

Wolfinger, R.D. and R.E. Kass. (2000). "Nonconjugate Bayesian Analysis of Variance Component Models," *Biometrics*, 56, 768-774.

Wu, C.F.J. and Hamada, M.S. (2009). *Experiments: Planning, Analysis, and Optimization*. 2nd Edition. Wiley: New York.

CHAPTER 5

MINIMAX DESIGNS FOR FINITE DESIGN REGIONS

5.1 Introduction

"Space-filling" designs, which can loosely be interpreted as designs that are representative of the design region (e.g., uniformly distributed over the region), are widely used in computer experiments. Because these designs are not obtained by optimizing a model-based criterion, they are model-independent. Due to their representativeness and model-independence, space-filling designs are widely considered to be natural choices if the experimenter has no idea about an appropriate model before the experiment. They are useful for building Gaussian process (GP) and other nonparametric emulators (Chen et al., 2006; Santner et al, 2003) in cases where the computer code output is believed to be highly nonlinear and a parametric form for modeling the output cannot be specified.

Methods for generating space-filling designs are generally developed for one of three specific types of design regions: hypercubes, constrained continuous regions, and finite design regions. A constrained continuous region is defined by inequalities while a finite design region consists of a finite set of candidate points. Hypercube regions have received much attention (e.g., Morris and Mitchell (1995), Tang (1993), McKay et al. (1979)). A few methods have also been proposed to generate space-filling designs for constrained continuous regions (e.g., Stinstra et al. (2003), Draguljić (2012)) and finite regions (e.g., Marengo and Todeschini (1992), Kennard and Stone (1969), Royle and Nychka (1998)).

The notion of "space-filling" can be made more precise with distance-based criteria. Two important ones are the minimax and maximin criteria. Johnson et al. (1990) prove some optimality properties of minimax and maximin distance designs for constructing GP emulators. They prove the important result that for finite design regions, minimax designs are G-optimal in an asymptotic sense. While maximin designs have received considerable attention (e.g, Morris and Mitchell (1995)), minimax designs have received scant attention. John et al. (1995) discuss two-level minimax designs. Kennard and Stone (1969) give a sequential heuristic method for constructing designs based on the minimax criterion.

Despite the lack of attention on minimax designs, intuition suggests that they are good designs. They minimize the maximum distance of a candidate point to the nearest design point, which is desirable for two reasons. First, since emulators of nonlinear functions tend to incur higher prediction errors at points further from the design, minimax designs tend to minimize the worst-case prediction error. Second, because a minimax design is close to every candidate point, it is spatially representative of the candidate set; thus, it provides protection against erroneous modeling assumptions.

In this chapter, we address the problem of constructing minimax designs for finite design regions. We give a procedure for constructing *global optimal* minimax designs that is practical for a small number of candidate points (up to one or two hundred). It is based on solving a binary linear program (BLP), called set covering location problem (SCLP) in the operations research literature, given certain values of the covering radius. We prove that a minimax design can be found by solving SCLP at a discontinuity of the function that maps the covering radius to the optimal objective function value. The set of

minimax distances is the set of discontinuities of this function. Because the set of discontinuities is a subset of the distinct pairwise distances between candidate points, the discontinuities can be found by solving SCLP at the midpoints of the ordered distinct pairwise distances. Because the computational cost for solving SCLP increases drastically with the number of candidate points, the above procedure is infeasible when the candidate set is large. To overcome this problem, we give a heuristic procedure for constructing near-minimax designs that is affordable for large candidate sets. It relies on the well-known heuristic method proposed by Hochbaum (1982) for solving SCLP.

The proposed method has many potential applications in computer experiments because finite candidate sets arise naturally in many design problems. We briefly mention these applications and defer a detailed discussion to Section 5.2.2. First, in validation and calibration of a computer model with observation data, a preliminary experiment may consist of computer model runs at a subset of the input data points (observations on the input variables). Second, in sensitivity analysis or uncertainty quantification problems, the joint distribution of the inputs may be complex. In this case, either the input data or samples drawn from a nonparametric estimate of the distribution can be taken as the design region. Third, in the construction of nested space-filling designs (Qian, 2009), the current layer is the candidate set for the next layer. Thus, the proposed method can be used to construct all except the first layer of a nested space-filling design. Fourth, when a continuous design region is constrained, it can be easier to build a design from a finite candidate set used as a proxy for the region.

The problem of constructing designs for finite design regions is an important practical problem in physical experimentation (Anderson-Cook and Robinson, 2009;

Kennard and Stone, 1969) and in spatial statistics (Royle and Nychka, 1998). Marengo and Todeschini (1992) and Royle and Nychka (1998) employ exchange algorithms to generate designs based on the maximin criterion and a distance-based coverage criterion respectively. Anderson-Cook and Robinson (2009) employ an exchange algorithm to generate D-optimal designs that do not have replicates. In this chapter, we shall abbreviate Kennard and Stone's (1969) algorithm by KS and Marengo and Todeschini's (1992) algorithm by MT.

The remainder of the chapter is organized as follows: In Section 5.2, we define minimax designs for finite design regions, illustrate their value as space-filling designs, and give computer experiment applications. Section 5.3 introduces the set covering location problem and relates it to the construction of minimax designs. We describe the key results and present a procedure for finding minimax designs. Important results are stated as propositions and proved in Appendix E.1. A heuristic procedure for generating minimax designs from large candidate sets is given in Section 5.4. In Section 5.5, we give three examples to illustrate the proposed methods. Section 5.6 gives concluding remarks.

5.2 Minimax Design and Its Potential Applications

5.2.1 Minimax Design and Minimax Criterion

In this section, we define minimax designs for a finite candidate set of design points and we briefly illustrate the advantages of minimax designs over maximin designs.

We first give some definitions. Let $\chi = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_N\}$ be a finite candidate set of design points, where all $\boldsymbol{u}_i$ are *distinct*. A **design** $D = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ is a subset of χ. The distance between $\boldsymbol{u}_i$ and D is $d(\boldsymbol{u}_i, D) = \min\{\|\boldsymbol{u}_i - \boldsymbol{x}_1\|, \dots, \|\boldsymbol{u}_i - \boldsymbol{x}_n\|\}$. The distance between χ and D, which we call the **distance** of D is

$d(D) = \max\{d(\boldsymbol{u}_1, D), \ldots, d(\boldsymbol{u}_N, D)\}$. An n-point **minimax design** D_n^* satisfies $d(D_n^*) = \min\{d(D): D \subseteq \chi, |D| = n\}$ and $d_n^* = d(D_n^*)$ is called the **minimax distance** for n-point designs. On the other hand, an n-point **maximin design** maximizes the minimum pairwise distance between design points $\Delta(D) = \min\{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|: j \neq i\}$. Note that $\|\cdot\|$ can represent any distance metric. However, in the examples in this chapter, we shall take $\|\cdot\|$ as the *Euclidean* norm.

We shall now illustrate some differences between the minimax and maximin criteria, and also an application of the main results of this chapter. Consider the problem of selecting a design from the 9x9 grid plotted as black dots in each of the four diagrams in Figure 5.1. In Figure 5.1, we plot global optimal minimax designs of sizes 4, 9, and 20 constructed with the proposed method. We also plot maximin designs of sizes 4, 9, and 20, which are obtained using MT (note that MT is not guaranteed to find globally optimal maximin designs).

A comparison of each pair of minimax and maximin designs of the same size in Figure 5.1 indicates that the former tends to have points close to every candidate point whereas the latter tends to contain many boundary points. It is seen that pulling the boundary points of 4 and 9 point maximin designs some distance towards the center point gives minimax designs. This reduces the distance of some of the interior points to the design while not increasing the distance of the boundary points to the design by too much. One interesting feature of the 20-point minimax design is that we can reach at least one design point from every candidate point by taking at most a horizontal or a vertical step. In other words, the distance of the design is the minimum distance between distinct candidate points. In contrast, the 20-point maximin design, which has a larger minimum

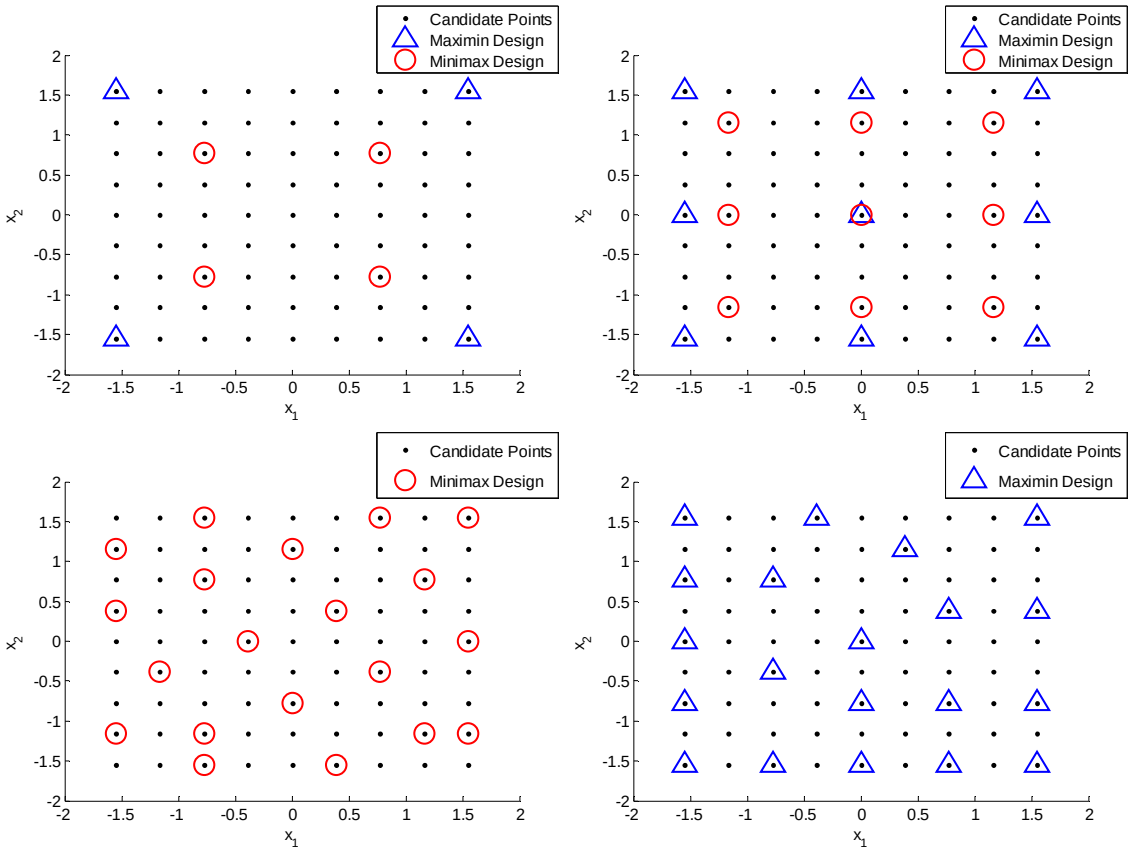

Figure 5.1: Top Left: Minimax and Maximin Designs of Size Four; Top Right: Minimax and Maximin Designs of Size Nine; Bottom Left: Minimax Design of Size 20; Bottom Right: Maximin Design of Size 20

pairwise distance between design points, does not share this desirable property. It leaves larger unfilled gaps in the design region.

The above examples illustrate the intuitive appeal of the minimax criterion. A minimax design ensures that there is a design point near every portion of the design region, i.e., the design points are spread over the candidate set. As such, a minimax design is spatially representative of the candidate set and is good for constructing emulators. When the true function is smooth but highly nonlinear, minimax designs can be expected to perform better than maximin designs in building nonparametric emulators. The reason is as follows: When changes in the function are smooth but without

systematic trends, predictions at candidate points closer to the design are likely to be more accurate than predictions at points further from the design. Since minimax designs minimize the maximum distance of a candidate point to the design, they tend to give small worst-case prediction errors. This may be viewed as an intuitive explanation of the theoretical results in Johnson et al. (1990). A similar explanation is given in Santner et al. (2003), page 149. In contrast, a maximin design does not guarantee that all candidate points would be close to a design point. Thus, it can incur some large prediction errors.

5.2.2 Potential Applications

The proposed method for constructing minimax designs has many potential applications. When inputs and outputs to computer codes are measurable physical quantities, data on these physical quantities are often available. In these cases, minimax designs chosen from the input data can be useful for a few reasons. First, the differences between the computer output and the response data at selected input points can be observed directly. This provides a reliable way to estimate the bias in the computer model, which is crucial for the validation and calibration of the computer code (Bayarri et al., 2007). Second, data for the inputs can be used to define a design region that is more realistic than hypercubes specified based on expert knowledge. In this case, the data may not be “nice” in the sense that its convex hull is not close to an ellipse or rectangle. When the data are not nice, it is more convenient and realistic to define the region of interest as all points close (in a vague sense) to the data rather than a rectangle that envelops the data. Thus, a design region consisting of all the data points automatically covers the region of interest well and a minimax design chosen from the input data can be good for constructing emulators for predicting in the region of interest.

In some cases, the experimenter may be interested in sensitivity analysis (Oakley and O'Hagan, 2004) or uncertainty propagation (Oakley and O'Hagan, 2002) with complex input distributions. This can involve modeling the input data with kernel or copula estimators, generating samples from the estimated distribution, and running the computer code at the sample points to obtain a random sample of the response. If the computer runs are expensive, it may be desirable to run the computer code at only a small subset of a large sample from the estimated distribution, construct a GP emulator based on the experiment data, and use this to predict the values of the response at all other sample points.

Another application of the proposed method for constructing minimax designs is in generating nested space-filling designs (Qian et al., 2009; Qian, 2009). A nested space-filling design with e layers consists of space-filling designs $D^1, D^2, \ldots, D^e$ such that $D^1 \supset D^2 \supset \cdots \supset D^e$. Such designs can be constructed by first generating a space-filling design D^1, and then applying the proposed method sequentially to construct minimax designs $D^2, \ldots, D^e$, where D^{j+1} is constructed by taking $\chi = D^j$.

Nested space-filling designs are useful for computer experiments with multiple levels of fidelity. A special case of multifidelity experiments is calibration and validation experiments, in which the highest level of fidelity is the physical experiment. In a multifidelity experiment, D^1 is for the experiment at the lowest level of fidelity, D^2 is for the experiment at the next higher level of fidelity and so on. Designs of smaller size are used for experiments with higher fidelity because experiment cost increases as the level of fidelity increases. Qian (2009) argues that nested designs are desirable for multifidelity experiments because they allow direct observation of the differences between responses

from different fidelity levels, and thus allow more accurate modeling of the differences. Note that nested space-filling designs can also be used for running computer experiments in a batch sequential manner.

Finally, the proposed method can be used to construct minimax designs from grids overlaid on a continuous region. However, the minimax designs obtained depend on the choice of the candidate set and may not be minimax with respect to the continuous region. A sparse grid may yield designs that are significantly different from designs that are minimax for the continuous region. On the other hand, the proposed method can be infeasible for dense grids.

In each of the above applications, an important advantage of using a minimax design over other space-filling designs is that they tend to minimize the maximum prediction variance of the GP emulator over the candidate set.

5.3 Construction of Minimax Designs via Solution of Set Covering Location Problem

In this section, we shall first introduce SCLP and its formulation as a BLP. Then, we discuss key results of this chapter without proof, and provide procedures for finding minimax designs based on the key results. Proofs of the key results are given in Appendix E.1.

SCLP is a classical problem in location analysis introduced by Hakimi (1965). A brief description of the problem is as follows: We are given a set of N_1 candidate locations for constructing facilities of a particular type and N_2 locations of demand points. Each facility has infinite capacity but can only serve demand nodes less than or equal to S unit distance from the facility, where S is called the *covering radius*. The problem is to determine the minimal number of locations at which to construct facilities

so that all demand nodes are served. Toregas et al. (1971) formulated SCLP as a BLP, which enables the problem to be solved efficiently. For a recent reference, see Snyder (2011).

In this chapter, we shall be concerned with only the *special case* of SCLP where $N_1 = N_2 = N$ and the set of candidate location for facilities is the same as the set of demand points. The BLP formulation of this problem has decision variables $I_1, \dots, I_N$. If $I_j = 1$, then a facility is constructed at location j and $I_j = 0$ otherwise. This SCLP is formulated as a BLP below.

Set Covering Location Problem (SCLP)

$$z(S) = \min \sum_{j=1}^{N} I_j$$

$$\text{s.t. } \sum_{j \in \Omega_i} I_j \geq 1, i = 1, \dots, N,$$

$$I_j \in \{0,1\}, j = 1, \dots, N.$$

$$\Omega_i = \{j \in \{1, \dots, N\} | h_{ij} \leq S\},$$

$$h_{ij} = distance\ between\ locations\ i\ and\ j,$$

$$S = covering\ radius.$$

Note that Ω_i is the set of indices of locations within a distance S from location i. The constraint $\sum_{j \in \Omega_i} I_j \geq 1$ requires that demand point i be a distance of no more than S from the closest facility. We shall write the constraints $\sum_{j \in \Omega_i} I_j \geq 1, i = 1, \dots, N$ compactly as $\boldsymbol{A}(S)\boldsymbol{I} \geq \boldsymbol{1}$, where $\boldsymbol{I} = (I_1, \dots, I_N)^T$, and $\boldsymbol{A}(S)$ is the $N \times N$ constraint matrix with elements $A_{ij} = 1$ if $j \in \Omega_i$ and $A_{ij} = 0$ otherwise.

We can use SCLP to find a design such that all candidate points are at a distance at most S from the closest design point. To do this, we interpret the candidate points as

the candidate facility locations (which are also demand points). A design $D \subseteq \chi$ is a set of facility locations, which corresponds to a (possibly infeasible) solution $\boldsymbol{I} \in \Lambda$ to SCLP, where Λ is the set of all nonzero binary N-vectors. Let $\{j_1, \dots, j_n\} = \{j \in \{1, \dots, N\}: I_j = 1\}$. Then, $\boldsymbol{I}$ corresponds to the design $D = \{\boldsymbol{u}_{j_1}, \dots, \boldsymbol{u}_{j_n}\}$. In this chapter, we shall sometimes write $D(\boldsymbol{I})$ or $\boldsymbol{I}(D)$ to make the correspondence clear. SCLP gives a design of minimum size such that every candidate point is at most a distance of S units away from a design point.

The **key results** of this chapter are as follows: We prove that a minimax design can be found by solving SCLP at a value α_i of the covering radius S (or slightly greater than α_i) such that any further decrease in S would cause a jump in the optimal objective function value, i.e., α_i is a discontinuity of the function $z(S)$, which maps the covering radius S to the optimal objective function value. Any design obtained by solving SCLP at $S = \alpha_i$ is a minimax design with minimax distance α_i. Moreover, the set of minimax distances is the set of discontinuities of the function $z(S)$. A key observation used to prove the results is that $d(D) \leq S$ if and only if $\boldsymbol{I}(D)$ is a feasible solution of SCLP with covering radius S.

The function $z(S)$ is a right continuous and nonincreasing function with range $\{1, \dots, N\}$. It has finite number of discontinuities $\omega = \{\alpha_1, \dots, \alpha_M: 0 = \alpha_1 < \cdots < \alpha_M\}$, which as mentioned above, is also the set of minimax distances (as a convention, we refer to $\alpha_1 = 0$ as a discontinuity). Define $\psi = \{h_{ij} = \|\boldsymbol{u}_i - \boldsymbol{u}_j\|: i, j = 1, \dots, N\}$ to be the set of pairwise distances between design points and let $\psi' = \{h^1, \dots, h^m\}$, where $0 = h^1 < \cdots < h^m$, be the set of distinct values in ψ. Because $\omega \subseteq \psi'$, i.e., the set of discontinuities is a subset of the distinct pairwise distances between points in χ, identification of the

discontinuities is easy. It is only necessary to solve SCLP at the midpoint between consecutive values of ψ' to determine the discontinuities. Minimax designs can be found by solving SCLP at a point of discontinuity $S = \alpha_i$ or for practical purposes (i.e., to guard against numerical errors), slightly greater than it but less than the next point in ψ'. Any solution of SCLP at $S = \alpha_i$ is an $n_i = z(\alpha_i)$-point minimax design and conversely any n_i-point minimax design is a solution of SCLP at $S = \alpha_i$. For any $n \in [n_i, n_{i-1} - 1]$, where $n_{i-1} = z(\alpha_{i-1})$, an n-point minimax design has minimax distance $d_n^* = \alpha_i$ and can be constructed by adding $n - n_i$ points to an n_i-point minimax design.

For illustration, consider the case where $\chi = \{(0,0), (1,0), (0,1), (0.5,0.5)\}$. Then, it is obvious that the 1-point minimax design is $D_1^* = \{(0.5,0.5)\}$, 2-point and 3-point minimax designs can be obtained by adding to D_1^* one and two other points from χ respectively, and the 4-point minimax design is $D_4^* = \chi$. The minimax distances are $d_1^* = 1/\sqrt{2} = d_2^* = d_3^*$, and $d_4^* = 0$. We have $\psi' = \{0, 1/\sqrt{2}, 1, \sqrt{2}\}$. If we solve SCLP at $S \in \{\tilde{h}^1, \tilde{h}^2, \tilde{h}^3, \tilde{h}^4\}$, where $\tilde{h}^1 \in [0, 1/\sqrt{2})$, $\tilde{h}^2 \in [1/\sqrt{2}, 1)$, $\tilde{h}^3 \in [1, \sqrt{2})$, and $\tilde{h}^4 \in [\sqrt{2}, \infty)$, i.e., one point in each interval formed by partitioning $[0, \infty)$ with consecutive points of ψ', then we would obtain the plot of $z(S)$ versus S shown in Figure 5.2. The figure shows that $z(S)$ has a discontinuity at $\alpha_2 = 1/\sqrt{2}$; note that by convention, $\alpha_1 = 0$ is also a discontinuity. The two discontinuities are the minimax distances. We also see that $z(\alpha_1) = 4$ and $z(\alpha_2) = 1$. This implies that minimax designs with one to three points have minimax distance α_2. The one-point minimax design can be found by solving SCLP with $S = \tilde{h}^2$.

We now present a computational procedure for efficiently creating a plot of $z(S)$ versus S. A minimax design can be obtained by solving SCLP at any identified

discontinuity of the plot.

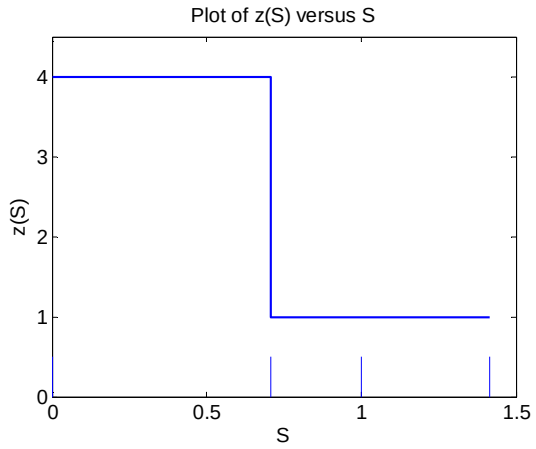

Figure 5.2: Plot of $z(S)$ versus S for $\chi = \{(0,0), (1,0), (0,1), (0.5,0.5)\}$. Long vertical lines extending upwards from the abscissa are values in ψ'

Procedure A: (for creating a plot of $z(S)$ versus S)

1. Specify an interval $[L, U]$. Compute ψ and ψ'. Compute $\Theta = \{H^1, \dots, H^m\}$, where $H^i = \left(h^i + h^{i+1}\right)/2, i = 1, \dots, m-1$, $H^m = 1.1h^m$. Find l and u such that $h^{l-1} < L \leq h^l$ and $h^u \leq U < h^{u+1}$.
2. Solve SCLP with $S = H^u$. Set $k = 2$.
3. Determine if $\boldsymbol{A}(H^{u-k+1})\boldsymbol{I}^*(H^{u-k+2}) \geq \mathbf{1}$, where $\boldsymbol{I}^*(H^{u-k+2})$ is the optimal solution for SCLP with $S = H^{u-k+2}$, by checking the equivalent condition $d\left(D\left(\boldsymbol{I}^*(H^{u-k+2})\right)\right) \leq H^{u-k+1}$. If the condition is satisfied, set $\boldsymbol{I}^*(H^{u-k+1}) = \boldsymbol{I}^*(H^{u-k+2})$ and go to Step 5; otherwise, go to Step 4.
4. Solve SCLP with $S = H^{u-k+1}$.
5. If $u - k + 1 = l$, **stop**. Otherwise, set $k = k + 1$. Go to Step 3.

Remark 3.1: If $z(H^{q-1}) > z(H^q)$, then h^q is a discontinuity of $z(S)$. Thus, minimax designs can be found by solving SCLP at $S = H^q$. If $z(H^{q-1}) > z(H^q) + 1$, minimax

designs of size $n \in [z(H^q)+1, z(H^{q-1})-1]$ can be obtained by adding $n - z(H^q)$ points to a minimax design of size $z(H^q)$. We suggest that the extra points be added at points $\boldsymbol{u}_i$ such that $d(\boldsymbol{u}_i, D)$ equals the minimax distance h^q.

Remark 3.2: Procedure A allows us to determine whether $h^{l+1}, \ldots, h^u$ are discontinuities of $z(S)$. We cannot tell from the results whether h^l is a discontinuity.

Remark 3.3: The solution of SCLP at $S = H^1$ is χ and a solution at $S = H^m$ is any one-point design.

Remark 3.4: If $\boldsymbol{A}(H^{q-1})\boldsymbol{I}^*(H^q) \geq \mathbf{1}$, $\boldsymbol{I}^*(H^q)$ is a feasible solution for SCLP with $S = H^{q-1}$. This implies that $z(H^{q-1}) \leq z(H^q)$. Since $H^{q-1} < H^q$, any feasible solution at $S = H^{q-1}$ must also be feasible at $S = H^q$. It follows that $z(H^{q-1}) \geq z(H^q)$ and $\boldsymbol{I}^*(H^q)$ is optimal at $S = H^{q-1}$. The step of checking whether an optimal solution remains optimal as S is reduced (Step 3) can yield significant computational savings over an alternative procedure that solves SCLP each time S is reduced. By Proposition E.2 in Appendix E.1, $\boldsymbol{A}(H^{q-1})\boldsymbol{I}^*(H^q) \geq \mathbf{1}$ if and only if $d\left(D\left(\boldsymbol{I}^*(H^q)\right)\right) \leq H^{q-1}$. The advantage of checking the latter condition is that $d\left(D(\boldsymbol{I}^*)\right)$ need only be computed once for an $\boldsymbol{I}^*$. This can be more efficient than the former condition, which requires computing $\boldsymbol{A}(H^q)$ for each q.

Remark 3.5: In practice, we would want to plot $z(S)$ over an interval $[L, U]$ such that the range $[z(H^u), z(H^l)-1]$ contains design sizes of interest to the experimenter. Suitable choices of $[L, U]$ can be obtained by say, trial and error. Alternatively, use Procedure B given below.

Remark 3.6: The set ψ' of unique values of ψ depends on the precision used in computing the elements of ψ. We compute the elements of ψ to $DP = 12$ decimal places

in Procedure A.

Input variables for an experiment are often quantities that vary over very different ranges. This is because in many experiments, these variables are different physical quantities measured in different units. Moreover, physical units can be written in different scales (e.g., kilometer, meter, centimeter). Because of these reasons, we recommend always *standardizing each variable*, i.e., subtract the mean from each value and divide the result by the standard deviation so that the sum of squares is N. Throughout the chapter, the candidate set χ consists of the standardized candidate points.

A global optimal solution to SCLP can be obtained by using a linear programming (LP) based branch-and-bound algorithm. One such algorithm is the bintprog function in Matlab. However, the time required by such algorithms to solve SCLP can increase dramatically with an increase in the size of the $N \times N$ constraint matrix $\boldsymbol{A}$. Note that we use $\boldsymbol{A}(S)$ and $\boldsymbol{A}$ interchangeably. A commonly used strategy to speed up the solution of SCLP is the removal of redundant constraints and decision variables (rows and columns of the constraint matrix) (Caprara and Toth, 2000). Define $C_k = \{i: A_{ik} = 1\}$ and $R_i = \{k: A_{ik} = 1\}$. Then, I_k is redundant if there exists $l \neq k$ such that $C_k \subseteq C_l$ since this implies that the set of points that is covered by point k is a subset of the points that is covered by point l. On the other hand, constraint i is redundant if there exists $j \neq i$ such that $R_i \supseteq R_j$ since this implies that if point j is covered, then so is point i. Any constraint matrix obtained after removal of some redundant rows and columns is called a reduced constraint matrix and the corresponding BLP is called a reduced version of SCLP. A reduced constraint matrix may be further reduced by removing redundant rows and columns. Note that the definitions of redundant rows and

columns extend in a straightforward manner to any reduced constraint matrix and a nonredundant row/column in any reduced constraint matrix may become redundant after removal of redundant columns/rows. We also call any reduced version of a reduced constraint matrix a reduced constraint matrix, and the corresponding BLP a reduced version of SCLP. It can be shown that SCLP and any reduced version of it have the same optimal objective function values, and an optimal solution to the latter can always be converted into an optimal solution to the former by setting to zero the decision variables that have been removed (see Appendix E.4). In the Matlab code for Procedure A that we provide, rows/columns of the constraint matrix are reduced via the efficient technique of making pairwise comparisons sequentially and removing a row/column as soon as it is found to be redundant. Our code makes pairwise comparisons of rows first. We call the BLP obtained by this reduction procedure *Reduced SCLP* and we denote its constraint matrix as $\widetilde{\boldsymbol{A}}$. Note, however, that any other reduced version of SCLP can be employed. Reduced SCLP can be much smaller than SCLP. We have found that in some instances, more than half of the rows and columns of the constraint matrix are removed by the reduction procedure.

Algorithms for solving BLP's often give only one optimal solution. However, the set of $z(h^k)$-point minimax designs can be obtained, at least in principle, using the following technique. Let $\boldsymbol{I}_1^* = (I_{11}^*, \ldots, I_{N1}^*)$ be the optimal solution obtained by solving SCLP at $S = H^k$. Then, we add the constraint $\sum_{j:I_{j1}^*=0} I_j + \sum_{j:I_{j1}^*=1}(1 - I_j) \geq 1$ to SCLP to "cut off" $\boldsymbol{I}_1^*$. This would force the program to yield an alternative solution. Suppose we have obtained solutions $\boldsymbol{I}_1^*, \ldots, \boldsymbol{I}_q^*$. Then, to obtain another, we add the constraint $\sum_{j:I_{jq}^*=0} I_j + \sum_{j:I_{jq}^*=1}(1 - I_j) \geq 1$ to the program, which together with the previously

added constraints, would cut off $\boldsymbol{I}_1^*, \ldots, \boldsymbol{I}_q^*$. This process is repeated until the program yields optimal objective function values greater than $z(h^k)$ or becomes infeasible. Of course, we can also terminate the search after a specified number of alternative solutions are found. We have found that it is often feasible to find n_a alternative minimax designs if $N + n_a < 250$ (so that there are less than 250 constraints). Alternative minimax designs can be found based on Reduced SCLP so that all design points are restricted to candidate points corresponding to the columns of $\widetilde{\boldsymbol{A}}$. However, this can yield far fewer alternative designs. One reason for the existence of alternative minimax designs is that it is often possible to substitute a few design points with some other candidate points close to them without changing the distance of the design. This is to be expected since the minimax criterion only measures the worse-case distance of the candidate points to the design. Another reason is that for rotationally symmetric candidate sets, any rotation of a minimax design would produce another minimax design.

We suggest choosing among alternative minimax designs using a secondary criterion. One obvious criterion that can be used is the maximin criterion. Another criterion, which we have found to produce visually appealing are designs, is the minimization of $V(D) = \text{var}\{\delta(\boldsymbol{x}_1), \ldots, \delta(\boldsymbol{x}_n)\}$, where $\delta(\boldsymbol{x}_i) = \min\{\|\boldsymbol{x}_i - \boldsymbol{x}_j\| : j \neq i, j = 1, \ldots, n\}$. This criterion is the variance of the minimum distances of each design point to another. Thus, a design that minimizes $V(D)$ would tend to look evenly spaced. For example, the four and nine point minimax designs given in Figure 5.1 achieve a minimum value of $V(D) = 0$. Note that rotations of any (minimax) design are equivalent in terms of any criterion based solely on (Euclidean) distances between design points, which includes the maximin and $V(D)$ criteria. A class of minimax designs that is of

theoretical interest (Johnson et al., 1990) is the highest index designs. However, we have frequently found that all alternative designs yield an index of 1.

Before we end this section, we present a bisection procedure for finding a tight interval $[S_L, S_U]$ that contains the minimax distance d_n^*. A slight modification of the interval can be used as input to Procedure A to determine the exact value of d_n^*. We add the constraint $\sum_{j=1}^{N} I_j = n$ to SCLP and call the resulting BLP n-SCLP. The bisection procedure is based on solving n-SCLP and is given below.

Procedure B: (Bisection procedure for finding a tight interval containing d_n^*)

1. If $n = N$, return $d_n^* = 0$ and terminate. Compute ψ' and Θ. Set $S_L = H^1$ and $S_U = H^m$.
2. Set $S_C = (S_L + S_U)/2$ and determine whether n-SCLP at S_C is feasible.
3. If n-SCLP is feasible, set $S_U = S_C$. If it is infeasible, set $S_L = S_C$.
4. If $\left|[S_L, S_U] \cap \psi'\right| \leq t$, stop. The interval $[S_L, S_U]$ contains d_n^*. Use $[L, U] = \begin{cases} [H^{l-1}, S_U], l \geq 2 \\ [0, S_U], l = 1 \end{cases}$, where $h^l \leq S_L < h^{l+1}$, as input to Procedure A. Otherwise, return to Step 2.

Remark 3.7: We can determine whether n-SCLP is feasible by working only with Reduced SCLP. Let N_c denote the number of columns of the constraint matrix $\widetilde{\boldsymbol{A}}$ of Reduced SCLP and let Reduced SCLP with the added constraint that all its decision variables sum to n be called Reduced n-SCLP. Then, if $N_c \leq n$, n-SCLP is feasible. If $N_c > n$, then n-SCLP is feasible if and only if Reduced n-SCLP is feasible.

Remark 3.8: If $[L, U] = \begin{cases} [H^{l-1}, S_U], l \geq 2 \\ [0, S_U], l = 1 \end{cases}$, where $h^l \leq S_L < h^{l+1}$, is given as input to Procedure A, then Procedure A will solve SCLP at $H^l, \dots, H^u$, where $u \leq l + t$. Thus, the

parameter t controls the number of times SCLP needs to be solved by Procedure A. We suggest choosing $t \geq 10$.

5.4 Algorithms for Large Candidate Sets

Solving SCLP as an integer program can be very time consuming when N is large. Because SCLP is an NP-hard problem (Current et al., 2002), there is no known fast (polynomial time) algorithm for solving it. In our experience with Matlab's bintprog function, only problems of size up to about 200 can be solved quickly (in a few minutes) on a MacBook Pro laptop for all values of S. If the candidate set is a grid, it can also be difficult to find global optimal minimax designs for N between around 150 to 200. As N grows, the number of elements of ψ' can increase rapidly and its values become more finely spaced. Thus, SCLP may need to be solved many times to construct a plot of $z(S)$. In this section, we propose modifications to Procedure A that overcome these problems.

To overcome the limitation of solving SCLP as an integer program, we employ a heuristic procedure. Many heuristic algorithms have been proposed to solve large scale SCLP's. Classical heuristics include the ones proposed by Chvatal (1979) and Hochbaum (1982). These are simple heuristics with guaranteed worst-case performance. We have implemented Chvatal's (1979) and Hochbaum's (1982) heuristics together with Grossman and Wool's (1997) redundancy elimination procedure. The redundancy elimination procedure removes redundant design points from a heuristic solution. Based on our experience, the combination of Hochbaum's and Grossman and Wool's heuristics tend to give better solutions than the combination of Chvatal's and Grossman and Wool's heuristics. Hence, we shall focus on the former heuristic procedure, which we call HGW.

We first introduce some notation. The number of rows of the reduced constraint

matrix $\widetilde{\boldsymbol{A}}$ is denoted by N_r (recall that the number of columns of $\widetilde{\boldsymbol{A}}$ is denoted by N_c) and the (i,j) element of $\widetilde{\boldsymbol{A}}$ is denoted by $\tilde{A}_{ij}$. We assume without loss of generality that the columns of $\widetilde{\boldsymbol{A}}$ correspond to decision variables $I_1, \dots, I_{N_c}$. The HGW heuristic is given below.

HGW Heuristic

1. Solve the LP relaxation of Reduced SCLP. The LP is obtained by replacing the constraints $I_j \in \{0,1\}, j = 1, \dots, N_c$ with the constraints $0 \leq I_j \leq 1, j = 1, \dots, N_c$. Let $\boldsymbol{I}^{LR} = \left(I_1^{LR}, \dots, I_{N_c}^{LR}\right)$ denote the optimal solution to the LP relaxation.
2. Let $v = \max\left\{\sum_{j=1}^{N_c} \tilde{A}_{ij} : i = 1, \dots, N_r\right\}$ (v is the maximum row sum of $\widetilde{\boldsymbol{A}}$) and find $F = \left\{j : I_j^{LR} \geq 1/v\right\}$. Set $I_j^H = \begin{cases} 1, j \in F \\ 0, j \notin F \end{cases}$. Then, it can be shown that $\boldsymbol{I}^H = \left(I_1^H, \dots, I_{N_c}^H\right)$, which is a heuristic solution proposed by Hochbaum (1982), is feasible for Reduced SCLP. Set $\boldsymbol{I}^{HGW} = \boldsymbol{I}^H$.
3. Calculate the redundancy $r_i = \sum_{j=1}^{N_c} \tilde{A}_{ij} I_j^{HGW}$ for $i = 1, \dots, N_r$, i.e., for each constraint associated with the rows of $\widetilde{\boldsymbol{A}}$. Set $\Xi = \left\{j : I_j^{HGW} = 1\right\}$. For each $j \in \Xi$, calculate the minimal redundancy $mr_j = \min\left\{r_i : \tilde{A}_{ij} = 1\right\}$. Find j^* such that $mr_{j^*} = \max\left\{mr_j : j \in \Xi\right\}$ (break ties with the minimum index rule).
4. If $mr_{j^*} \geq 2$, set $I_{j^*}^{HGW} = 0$ and return to Step 3. Otherwise, **stop**.

In Step 1, the LP relaxation of Reduced SCLP rather than that of SCLP is solved. We recommend doing this because we can expect better results by applying Hochbaum's heuristic to Reduced SCLP than SCLP. However, one may choose not to reduce SCLP to save computation time. The method used to solve the LP relaxation can have an effect on

the solution returned by HGW because different methods can give different optimal solutions. This chapter uses the simplex algorithm implemented in Matlab.

We modify Procedure A so that it can be applied to find near-minimax designs of size n from large candidate sets. The procedure, which includes the use of HGW as a key ingredient, is given below.

Procedure C (for constructing near-minimax designs)

1. Specify a tentative design size n. Set $U = f_u d_n^{KS}$ and $L = f_l d_n^{KS}$, where d_n^{KS} is the distance of the n-point design obtained with KS. Compute each element of ψ to DP decimal places. Determine ψ' and compute Θ. Find l and u such that $h^{l-1} < L \leq h^l$ and $h^u \leq U < h^{u+1}$.
2. Solve SCLP with $S = H^u$ using HGW. Set $k = 2$.
3. Check if $d\left(D\left(\hat{\boldsymbol{I}}(H^{u-k+2})\right)\right) \leq H^{u-k+1}$, where $\hat{\boldsymbol{I}}(H^{u-k+2})$ is the heuristic solution for SCLP with $S = H^{u-k+2}$. If the condition is satisfied, set $\hat{\boldsymbol{I}}(H^{u-k+1}) = \hat{\boldsymbol{I}}(H^{u-k+2})$ and go to Step 5; otherwise, go to Step 4.
4. Solve SCLP with $S = H^{u-k+1}$ using HGW.
5. If $u - k + 1 = l$, **stop**. Otherwise, set $k = k + 1$. Go to step 3.
6. Denote by $\hat{z}(S)$ the step function whose value at $S \in [h^q, h^{q+1})$ is the sum of the components of $\hat{\boldsymbol{I}}(H^q)$. Plot the monotone decreasing step function $\zeta(S)$ with set of discontinuities $\omega_\zeta = \{h^t : \hat{z}(h^q) > \hat{z}(h^t), l \leq q < t \leq u\}$, $\zeta(h^q) = \hat{z}(h^q)$ for all $h^q \in \omega_\zeta$, and $\zeta(h^l) = \hat{z}(h^l)$. Let $n' \in \{\zeta(h^q) : h^q \in \omega_\zeta\}$ and let $t(n') \in \{l+1, \ldots, u\}$ be such that $\zeta(h^{t(n')}) = n'$. Then, the heuristic solution $\hat{\boldsymbol{I}}\left(H^{t(n')}\right)$ is an n'-point *near-minimax* design and $h^{t(n')}$ is the distance of the design (accurate

up to DP decimal places). By examining a plot of $\zeta(S)$, the user can specify the value of n' to obtain near-minimax designs of several different sizes close to n.

Remark 4.1: The choice of f_u and f_l should be obtained by trial and error. We suggest trying $f_u \in [0.9,1]$ and $f_l \in [0.5,0.7]$. We also suggest that f_l be chosen so that $n < n_{max} = \max\{\zeta(h^q): h^q \in \omega_\zeta\}$ and only n'-point near-minimax designs with $n' < n_{max}$ be used. The discussion below justifies the suggestions. Of course, f_u should be chosen so that $n_{min} = \min\{\zeta(h^q): h^q \in \omega_\zeta\} \leq n$.

In the remainder of this section, we discuss the details of Procedure C. If we execute Procedure C by specifying $[L, U] = [0, H^m]$, we would get an estimate of $z(S)$ which we denote by $\hat{z}_{[0,H^m]}(S)$. It is possible that for some H^q, $\hat{z}_{[0,H^m]}(H^q) \neq z^{HGW}(H^q)$, where $z^{HGW}(H^q)$ is the objective function value obtained by solving SCLP with HGW at H^q, due to the checking in Step 3 of Procedure C. Unlike $z(S)$, $\hat{z}_{[0,H^m]}(S)$ may not be a monotonic function of S. Thus, a design obtained at a discontinuity of $\hat{z}_{[0,H^m]}(S)$ may not be a good design. However, a good design of size n can be found at the discontinuity $\hat{\alpha}(n)$ of $\hat{z}_{[0,H^m]}(S)$ such that $\hat{z}_{[0,H^m]}(\hat{\alpha}(n) - \varepsilon) > \hat{z}_{[0,H^m]}(\hat{\alpha}(n)) = n$ for all $\varepsilon > 0$ (note that $\hat{\alpha}(n)$ may not exist for certain n). We call $\hat{\alpha}(n)$ a *minimal discontinuity* of $\hat{z}_{[0,H^m]}(S)$. The construction of $\hat{z}_{[0,H^m]}(S)$ can be too costly because m can be large. To overcome this problem, we apply Procedure C by specifying $[L, U]$ to be a narrower interval. This leads to an estimate $\hat{z}_{[L,U]}(S)$ of $z(S)$. In the rest of the chapter, we shall denote $\hat{z}_{[L,U]}(S)$ by $\hat{z}(S)$ since the interval $[L, U]$ is often clear from the discussion. For finding n-point near-minimax designs, we recommend computing $\hat{z}(S)$ with $[L, U] = [f_l d_n^{KS}, f_u d_n^{KS}]$, where d_n^{KS} is the distance of the n-point KS design,

$f_u \in [0.9,1]$ and $f_l \in [0.5,0.7]$.

The discontinuity of $\hat{z}(S)$ that is of interest is the minimal discontinuity $h^{t(n)}$, i.e., $h^{t(n)}$ satisfies $\hat{z}(h^q) > \hat{z}(h^{t(n)}) = n\,, l \le q < t(n) \le u$. The heuristic solution $\hat{\boldsymbol{I}}\left(H^{t(n)}\right)$ gives an n-point near-minimax design with distance $h^{t(n)}$. It is not necessarily a solution that is found by solving SCLP at $H^{t(n)}$ with HGW; due to Step 3 of Procedure C, it could be a solution obtained with HGW at $S = H^{t(n)+k}$ but is found feasible for all $S = H^{t(n)}, \dots, H^{t(n)+k-1}$. Note that $h^{t(n)}$ may not exist. In this case, we can construct an n'-point near-minimax design, where n' is an integer close to n such that $h^{t(n')}$ exists. A graph of $\hat{z}(S)$ versus S may not show clearly the distance of near-minimax designs because $\hat{z}(S)$ may not be a monotonic function of S. However, the problem can be rectified by constructing the monotone nonincreasing step function $\zeta(S)$ that envelops $\hat{z}(S)$ from below in the tightest possible way, i.e., $\zeta(S)$ has set of discontinuities $\omega_\zeta = \{h^t: \hat{z}(h^q) > \hat{z}(h^t), l \le q < t \le u\}$, and function values $\zeta(h^q) = \hat{z}(h^q)$ for all $h^q \in \omega_\zeta$, and $\zeta(h^l) = \hat{z}(h^l)$. It is clear that if $L' < L$, the minimal discontinuity of $\hat{z}_{[L',U]}(S)$ that gives an n-point near-minimax design is always less than or equal to the corresponding minimal discontinuity of $\hat{z}_{[L,U]}(S)$ (assuming both exist). Thus, the value of L or f_l should be small enough so that a good n-point near-minimax design is not missed. If n' is close to $n_{max} = \max\{\zeta(h^q): h^q \in \omega_\zeta\}$ so that $h^{t(n')}$ is close to L, then it is possible that a better near-minimax design of size n' or less can be found by further decreasing L because it may happen that for some $S < L$, $\hat{z}(S) \le n'$. Thus, near-minimax designs of size n' near n_{max} should be avoided.

As N gets larger, the size of ψ' increases while the differences between values in

ψ' get smaller. When N is large, some of the differences can be so small that it is of no practical interest to distinguish those values. Moreover, because there are many values in ψ', SCLP may need to be solved many times even if $[L, U]$ is narrow. In this chapter, we compute the elements of ψ to $DP = 12$ decimal places when $N \leq 200$ and to $DP = 2$ decimal places when $N > 200$. Larger values of DP give more accurate results but we have found that good near-minimax designs can be obtained with Procedure C even if $DP = 2$.

It is found that when $N > 2000$, the Matlab implementation of Procedure C incurs high memory usage and long computation time on the author's laptop, which is a MacBook Pro with 2.4GHz Intel® Core™ Duo Processor and 4GB of RAM. The former problem (high memory usage) can cause the computer to freeze when the code is executed. It is due to the precomputing and storing of ψ and $\boldsymbol{A}(S)$. Note that $\boldsymbol{A}(S)$ (or $\widetilde{\boldsymbol{A}}(S)$) is a required input to Matlab's linear program solver and the precomputed ψ is used to quickly generate $\boldsymbol{A}(S)$ for various values of S.

5.5 Examples and Comparisons

This section gives three examples to demonstrate the application of the proposed methods to three different design problems. The first example is a real problem in a building energy simulation project involving the author. A small dataset on the input variables is available and it indicates that a highly irregular design region is appropriate. Since it is difficult to define a continuous design region that envelops the data, it is convenient to construct minimax designs from the data. The second example employs the proposed method to construct a nested space-filling design. The third example examines construction of near-minimax designs from a larger dataset ($N = 458$) using the

proposed heuristic method.

We emphasize that minimax designs are intended to be initial designs for computer experiments (which are almost always sequential in nature). Choice of sample size for an initial design has been discussed by Box (1993), Box et al. (1978), and Loeppky et al. (2009). Box (1993) and Box et al. (1978) recommend 25% of the experimental budget (which we call Box's rule-of-thumb). Loeppky et al. (2009) recommend a design of size $10p$ ($10p$ rule-of-thumb), where p is the dimension/number of variables. After an initial experiment has been run, an analysis of the data will indicate whether the sample size is adequate. If the experimental objective is prediction, prediction intervals (Tan and Wu, 2012; Santner et al, 2003) can be used to determine whether follow-up runs are needed.

In the examples, we compare the proposed method with KS, MT, and KS-MF, where KS-MF is an algorithm that improves KS designs with the modified Fedorov algorithm (MF). These algorithms are discussed in Appendix E.2. Note that MF is one of the best algorithms for finding optimal designs on finite candidate set (Cook and Nachtsheim, 1980; SAS Institute, 2010).

5.5.1 Example 5.1: Urban Heat Island Effect

In this problem, it is desired to construct a nonlinear emulator for a computer code used to predict urban heat island intensity, which is important for uncertainty quantification of building energy consumption (Sun et al., 2012). Interest centers on studying the effect of four variables, i.e., canyon height, canyon ratio, vegetation area fraction, and built-up area fraction, on urban heat island intensity. These variables are geometric parameters that characterize the layout of buildings within a city (see

Appendix E.3). It is not easy for the engineers to specify a design region for the four variables because reasonable values must correctly reflect the correlations between the variables.

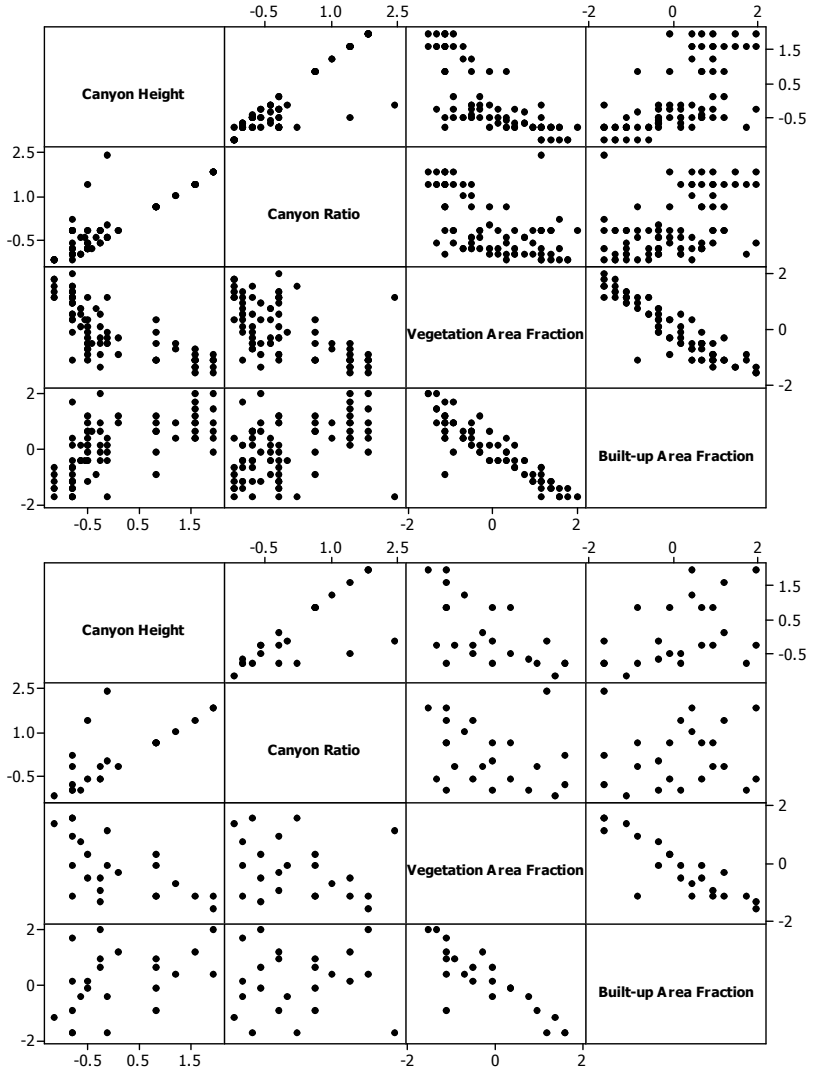

Figure 5.3: Top: 80 Data Points for the Four Urban Layout Geometry Parameters; Bottom: 23-Point Minimax Design that Minimizes $V(D)$

Data on the geometric parameters for $N = 80$ real urban areas are available. The engineers express interest in predicting urban heat island intensity at the data points because this would allow direct validation of the computer model with measured data

from real urban areas. Since the four variables vary over quite different ranges, we first center and standardize each variable. The standardized data are plotted in the top of Figure 5.3. It is immediately clear from the figure that the region of interest cannot be fitted into a hypercube or hypersphere. Thus, the design region is taken as the set of data points. Moreover, because running the computer code at all 80 data points is costly, a minimax design constructed using the data as candidate set is particularly useful.

There are a total of 2383 distinct pairwise distances between data points with values from 0 to 6.5932. We apply Procedure A with $[L, U] = [0.4,1.6]$. There are 547 values in $\psi' \cap [0.4,1.6]$, which we denote by $\{h^{u-546}, \dots, h^u\}$. Procedure A only requires about 9 seconds to find an optimal solution to SCLP for all $S \in \{H^{u-546}, \dots, H^u\}$ on the author's laptop. A plot of $z(S)$ versus S over $[0.4,1.6]$ is given in Figure 5.4. Discontinuities are plotted as dots in Figure 5.4. It can be seen that there are less than 40 discontinuities since $z(S)$ ranges from 6 to 43 within the interval $[0.4,1.6]$. It is not possible to determine whether h^{u-546} is a discontinuity from the information we have since this would require solving SCLP at $S = h^{u-547}$. Each h^j that is a discontinuity of $z(S)$ is the minimax distance for a $z(h^j)$-point minimax design. A $z(h^j)$-point minimax design can be obtained by solving SCLP at $S = H^j$. Figure 5.4 shows that minimax designs of sizes 6 to 42 have minimax distances in the interval $[0.4,1.6]$.

For comparison, we compute designs of size 6-42 using KS, MT, and KS-MF. MT uses 50 random starting designs to generate designs of each size. We have verified that different sets of 50 random starting designs tend to yield MT designs with similar distances (across replicates). KS requires a fraction of a second to compute all designs and their distances, KS-MF requires two seconds, and MT requires about 30 seconds. In

Figure 5.4, we plot the design size versus distance. It is clear from the figure that none of the KS and MT designs are (global optimal) minimax designs. We also see that MT designs can be competitive with KS designs but they have slightly larger distance values in most cases.

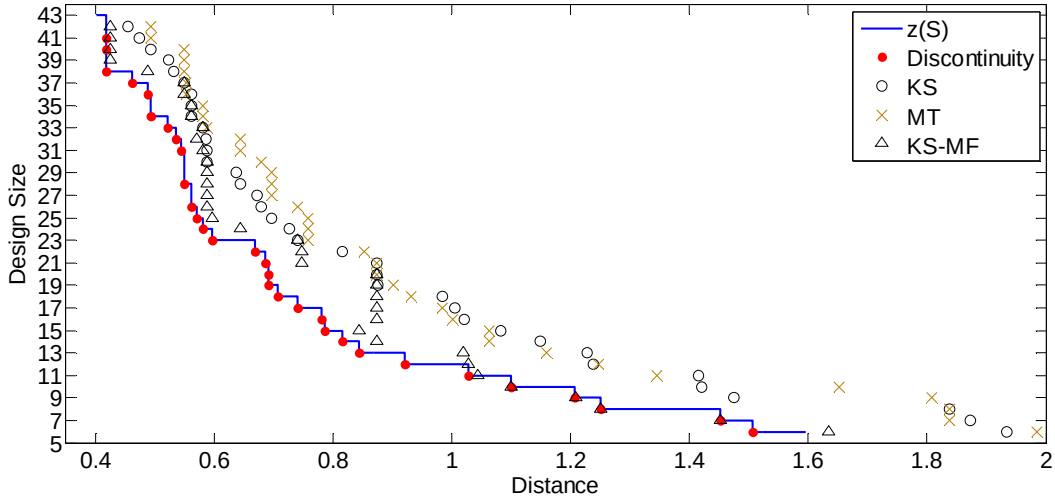

Figure 5.4: Plot of $z(S)$ versus S, and Design Size versus Distance for KS, MT, and KS-MF Designs, Urban Heat Island Problem

KS-MF performs quite well (i.e., MF is effective at improving KS designs) in this example. Three of the KS-MF designs are minimax designs (those of size 7, 8, and 10; the one of size 9 is not) while several others have close to minimax distances. However, some of the KS-MF designs perform unsatisfactorily compared to minimax designs (e.g., those of sizes 16 to 24). Observe that a KS-MF design of size 23 achieves a distance of $d_{17}^{*} = 0.74021$, which is the distance of a minimax design of size 17. Thus, KS-MF designs are not guaranteed to perform well even for small candidate sets (they do not perform well for larger N as demonstrated in Examples 5.2 and 5.3). Furthermore, there is no obvious way to check the goodness of the KS-MF designs other than applying the proposed method, and the computational advantage of KS-MF is not of practical

significance for small candidate sets. Consequently, it seems preferable to use the proposed method for small candidate sets.

The HGW heuristic gives excellent performance in this example. It yields a global optimal solution for each $S \in \{H^{u-546}, \ldots, H^u\}$, i.e., $z^{HGW}(S)$ is identical to $z(S)$ on $[0.4,1.6]$. As a result, the estimate $\hat{z}(S)$ obtained with Procedure C is identical to $z(S)$ on $[0.4,1.6]$, and designs produced by the procedure are minimax designs. In contrast, the combination of Chvatal's (1979) and Grossman and Wool's (1997) heuristics yields suboptimal solutions for many $S \in \{H^{u-546}, \ldots, H^u\}$.

For the urban heat island experiment, Box's rule-of-thumb suggests an initial design of size about 20 (which is 25% of the total budget of 80). Results of the computations used to construct Figure 5.4 show that $z(S)$ jumps from 23 to 24 as S is decreased beyond 0.59573. In particular, we find that $z(S) = 24$ for all $S \in [0.58046,0.59573)$ and $z(S) = 23$ for all $S = [0.59573, 0.66781)$. Thus, 0.59573 is the minimax distance of 23-point minimax designs, i.e., $d_{23}^* = 0.59573$. Note that 0.59573 and 0.59641 are two consecutive values in ψ'. Thus, to obtain a 23-point minimax design, we solve SCLP at (0.59573+0.59641)/2=0.59607. The search for alternative minimax designs is terminated when 150 designs are found. This takes about 83 seconds. Among the 150 designs, the value of the maximin criterion, i.e., $\Delta(D)$, varies from 0.544331 to 0.653584, and five designs achieve the maximum value. The value of the $V(D)$ criterion varies from 0.14464 to 0.17989 and only one design D^{**} gives this value. The design D^{**}, which has a value of $\Delta(D)$ of 0.636245, is plotted in Figure 5.3. A comparison of the top and bottom plots of Figure 5.3 indicates that D^{**} is spatially representative of the candidate set.

5.5.2 Example 5.2: Nested Space-Filling Design

This example illustrates the use of the proposed method for constructing nested space-filling designs. We suppose it is desired to construct a two-layer nested design for a problem with dimension $p = 10$. The $10p$ rule-of-thumb (Loeppky et al., 2009) suggests the use of designs of size 100. Since the first layer D^1 is used for the experiment on the low fidelity computer model, which is cheap to run, it seems reasonable to choose D^1 to be of size 2x100=200. On the other hand, since the second layer D^2 is used for the experiment on the high fidelity computer model, which is expensive to run, it is economical to choose D^2 to be of size 100/2=50. The first layer D^1 is constructed as follows. We generate 10,000 random Latin hypercube designs in $[0,1]^{10}$ of size 200, select the design that is best with respect to the maximin criterion, and set D^1 to be the standardized version of that design.

To construct D^2, we use $\chi = D^1$ to build 50-run minimax designs. Applying Procedure B with $n = t = 50$ gives the interval $[L, U] = [2.713445536765, 2.744064182816]$ (note that we give the values to $DP = 12$ decimal places). The computation takes 6 seconds. By applying Procedure A to compute $z(S)$ over $[L, U]$, we find that $d_{50}^* = 2.717545205187$, and the algorithm returns $H^q = 2.7177296575255$ as the value of S to use to solve for alternative minimax designs of size 50. The computation takes 87 seconds. A total of 20 alternative minimax designs are found in 70 seconds. All except two designs achieve the largest $\Delta(D)$ value of about 1.902. The minimum $V(D)$ value of 0.107 is achieved by only one design D^{**}. The design D^{**} also achieves the largest $\Delta(D)$. Thus, it seems desirable to take $D^2 = D^{**}$. Two-dimensional projections of D^2 are plotted in Figure 5.5. We do not

plot D^1 because its projections densely cover all of $\left[(0-0.5)/\left(1/\sqrt{12}\right),(1-0.5)\left(1/\sqrt{12}\right)\right]^2 = \left[-\sqrt{3},\sqrt{3}\right]^2$.

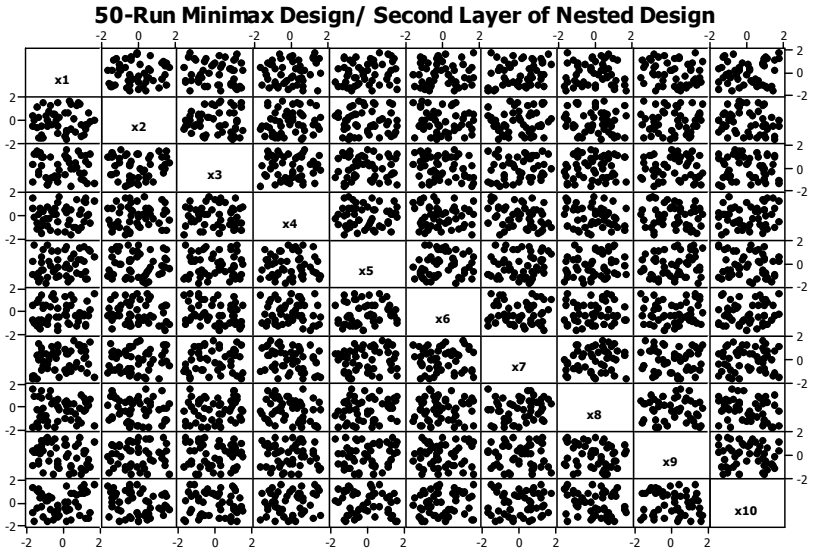

Figure 5.5: Matrix Plot of 50-Run Minimax Design D^2

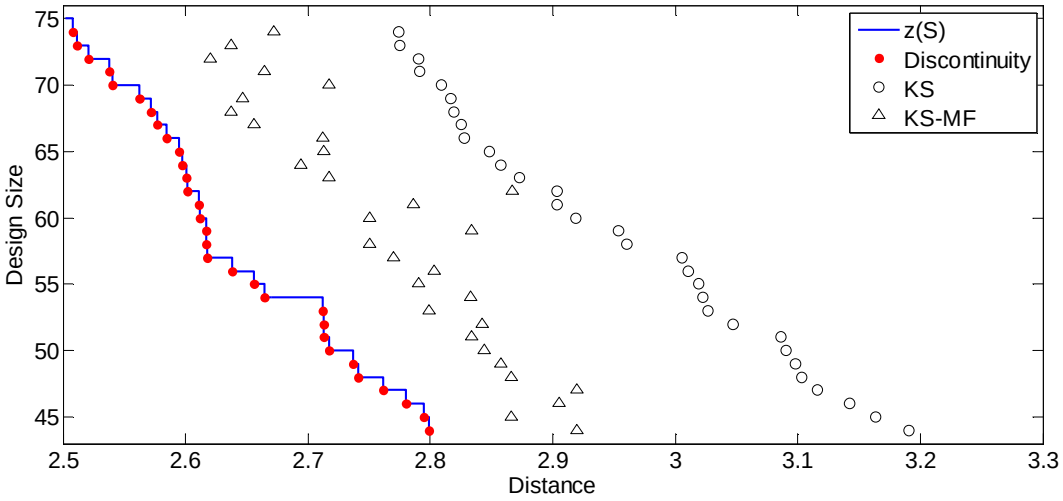

Figure 5.6: Plot of $z(S)$ versus S, and Design Size versus Distance for KS and KS-MF Designs

We construct a plot of $z(S)$ versus S over [2.5,2.8], as given in Figure 5.6. This takes about 330 seconds. It can be seen that minimax distances $d_{44}^*, \ldots, d_{74}^*$ are in the

interval $[2.5,2.8]$. We also plot the distances of designs of sizes $44, \ldots, 74$ obtained with KS and KS-MF in Figure 5.6. KS-MF takes about 61 seconds to find all the designs and distances. Although KS-MF is an improvement over KS, none of the KS-MF designs are global optimal minimax designs. An undesirable feature of KS-MF is that the KS-MF design distance is not a monotone function of design size. Moreover, KS-MF has significantly poorer performance compared to minimax designs. For example, the 65-point KS-MF design achieves $d_{52}^{*} = 2.71274$. The increase in sample size of 13 is substantial.

5.5.3 Example 5.3: Forest Fires

In this example, we apply the proposed heuristic procedure to construct near-minimax designs for a large candidate set. We consider a hypothetical problem of building an emulator of a computer model of forest fires, such as the model employed by Miller and Urban (1999). We use the forest fire dataset collected by Cortex and Morais (2007), which can be downloaded from the UCI machine learning repository at http://archive.ics.uci.edu/ml/datasets/Forest+Fires to define the candidate set. This helps create a realistic design region, and also allow direct validation of the computer predictions. The coordinate and time variables are removed from the data, giving nine remaining variables. One of the variables is the total burned area, which is the response. The eight remaining variables are the quantities employed in the computation of the Canadian forest fire weather index (Cortex and Morais, 2007). Appendix E.3 describes the variables. We assume that these eight variables are inputs to our fictitious forest fire computer model.

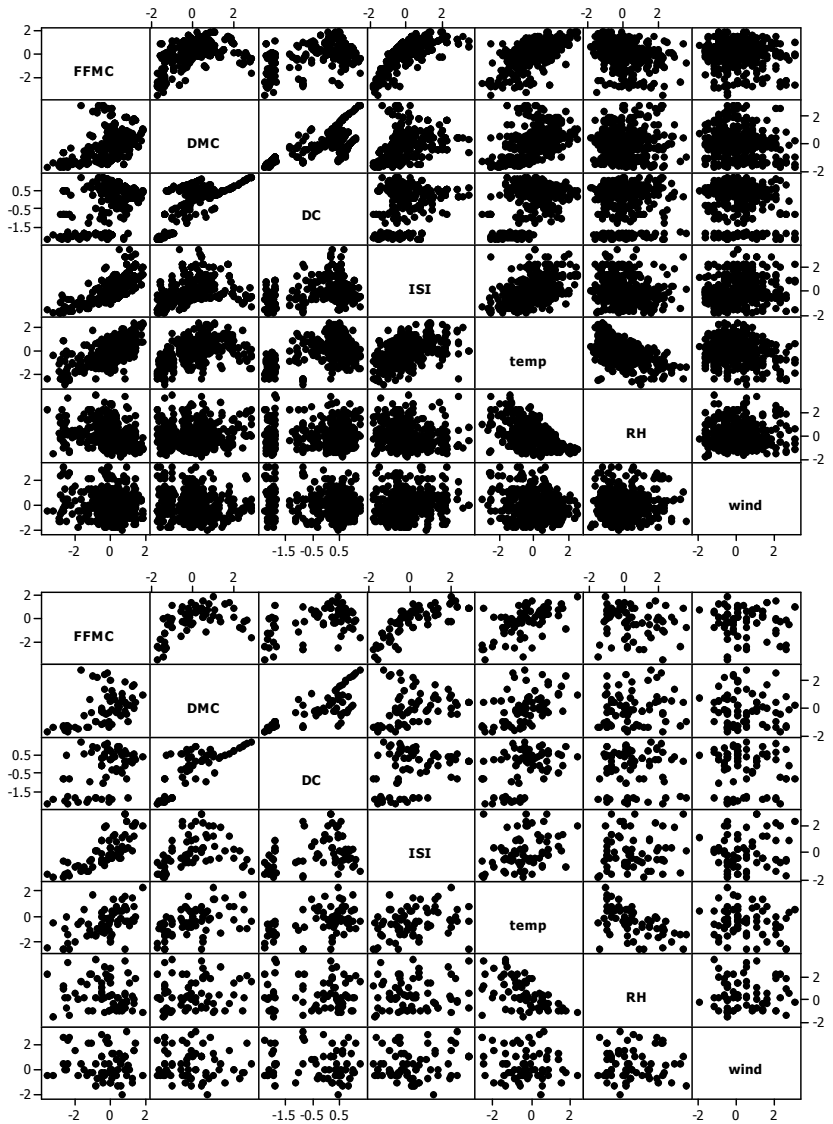

Figure 5.7: Matrix Plot of Candidate Set (Top) and 69-Run Near-Minimax Design (Bottom), Forest Fire Data

There are a total of 517 data points and several outliers can be seen when two-dimensional projections of the data are plotted. While it is of interest to run computer models at locations of outliers in a validation exercise, it may not be desirable to include these outliers in the candidate set for several reasons. First, the outliers can significantly influence the standardization of the variables. Second, if these outliers are included in the candidate set, some of them would be included in a minimax design because they are far

from the other candidate points. However, the computer code output may behave very differently at the locations of the outliers, which may have an adverse effect on the emulator predictions at non-outlier locations.

In the 517-point data, the variable rain is positive in eight instances and equals zero in all other instances. Thus, these eight outlying data points are removed and the variable rain is fixed at zero in the computer experiment. We now have a 509-point dataset with seven variables. For this dataset, we apply the boxplot to screen out extreme outliers in each variable. Any data point with a variable that falls outside the outer fences of its boxplot is discarded. Finally, repeated points are removed from the data. A total $N = 458$ data points remain after removal of all outlying and repeated points. Two-dimensional projections of the standardized data are plotted in the top of Figure 5.7. We see that the data concentrate on irregularly shaped regions. For example, in the FFMC versus DMC plot, the points concentrate on a "curved" region and in the DMC versus DC plot, the points concentrate on a highly irregular region. Note also that in the DC versus ISI, temp, RH, and wind plots, the points concentrate on two disjoint regions.

It is difficult to construct global optimal minimax designs due to the size of the candidate set. Since there are $p = 7$ variables, a 70-run minimax design would be a suitable choice for an initial design according to the $10p$ rule-of-thumb. Thus, we apply Procedure C with $U = d_{70}^{KS}$ and $L = 0.7d_{70}^{KS}$. This yields a plot of $\hat{z}(S)$ and $\zeta(S)$ over $[L, U]$ shown in Figure 5.8.

In Figure 5.8, discontinuities of $\zeta(S)$ are plotted as dots and $\hat{z}(S)$ is plotted with a dotted line. Note that $\hat{z}(S)$ is actually a continuous curve. The portions where it overlaps with $\zeta(S)$ are superimposed with the solid line that represents $\zeta(S)$. As illustrated in

Figure 5.8, $\hat{z}(S)$ is not monotonic. The monotonic lower envelope $\zeta(S)$ is an estimate of $z(S)$ and its discontinuities give the distances of near-minimax designs. We see from the figure that there is a discontinuity (indicated by an arrow) in $\zeta(S)$ at $h^q = 1.51$ and $\zeta(h^q) = 69$. The next largest near-minimax design corresponds to the discontinuity $h^{q-1} = 1.49$ and it has size $\zeta(h^{q-1}) = 73$. Figure 5.7 plots the near-minimax design of size 69, which has a distance of $h^q = 1.51$ up to $DP = 2$ decimal places (a more precise calculation gives a value of 1.511322). It can be seen that the design is representative of the candidate set.

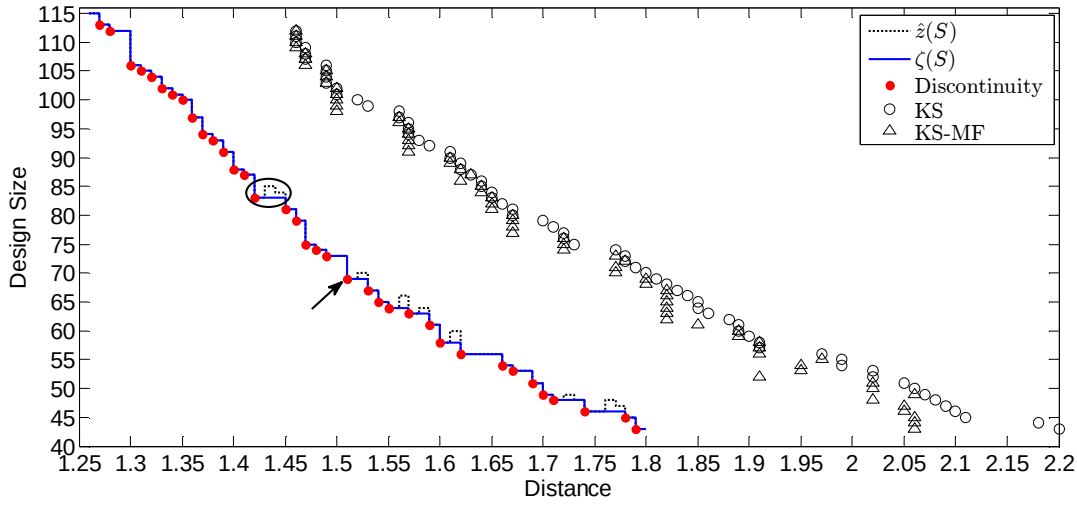

Figure 5.8: Plot of $\hat{z}(S)$ and $\zeta(S)$ versus S, and Design Size versus Distance for KS and KS-MF Designs

The drawback of using near-minimax designs of size n_{max} (defined in Remark 4.1) can be understood by looking at the circled portion of Figure 5.8. If $[L, U] = [1.43,1.8]$, then a near-minimax design of size $n_{max} = 84$ with distance 1.44 is obtained. However, we see from the figure that we can obtain an 83-point near-minimax design with distance 1.42.

Figure 5.8 gives the distances of near-minimax designs of sizes 43-113, which are

all greater than 1.25. For comparison, we plot the design size versus distance for KS and KS-MF designs of sizes 43-112 (excluding $n_{max} = 113$). KS-MF takes about 500 seconds to run. In contrast, Procedure C takes only about 180 seconds. It is clear from the figure that for the large candidate set in this example, MF is unable to significantly improve KS designs. In addition, it can be seen that KS-MF designs perform poorly compared to near-minimax designs. To achieve the same distance as a near-minimax design, the design size for KS-MF would need to be significantly larger. For example, a 68-point KS-MF design has distance 1.8 whereas a 43-point near-minimax design has distance of 1.79. As another example, to achieve the distance of 1.51 of the 69-point near-minimax design in Figure 5.7, the KS-MF design size would have to be 98; this would give a design with distance 1.50.

5.6 Conclusions

This chapter proposes a method for constructing minimax designs from finite candidate sets. The method is based on solving a BLP, i.e., SCLP, at specially chosen values of the covering radius. We prove that the set of minimax distances is the set of discontinuities of $z(S)$. We show that solving SCLP at or slightly to the right of discontinuity points gives minimax designs. To overcome computational limitations in solving large SCLPs, we give a heuristic procedure that is useful for constructing near-minimax designs for large candidate sets. Three examples are given to illustrate the usefulness of the proposed method. The first example is a real problem with a small candidate set. Minimax distances are determined via a plot of $z(S)$ versus S, and alternative minimax designs of a specified size are found. The second example illustrates the use of the proposed method for constructing nested space-filling designs. In the third

example, we employ the proposed heuristic procedure to a problem with a large candidate set. In the first and third examples, the candidate points define highly irregular regions, and the minimax designs are spatially representative of the candidate set.

The algorithms developed in this chapter are very different from existing algorithms for finding optimal designs on finite candidate set. They are based on solving finite sequences of SCLP's with gradually decreasing covering radius. Since SCLP is a very well-studied problem in the operations research literature, there exists a wide variety of powerful exact and heuristic tools for solving the problem. Some of these tools are employed in the chapter to solve the minimax design problem while others can be tested in future research. Even though existing algorithms for finding optimal designs on finite candidate can be used to find minimax designs, there is no guarantee that these algorithms can produce good designs. In fact, they do not perform well on moderate and large candidate sets. This may be due to the fact that the number of alternative n-point designs increases as the number of candidate points increases. Examples 5.1-5.3 demonstrate that as N increases, the performance of KS-MF designs relative to minimax or near-minimax designs deteriorates. For Example 5.1 ($N = 80$), many of the KS-MF designs have distances close to minimax distances but some of the designs perform poorly. For Example 5.2 ($N = 200$), none of the KS-MF designs perform well relative to minimax designs although KS-MF designs are significantly better than KS designs. Finally, in Example 5.3 ($N = 458$), Procedure C is shown to produce near-minimax designs that are substantially better than KS-MF designs in a much shorter time. Moreover, the KS-MF designs yield negligible improvements over the KS designs.

A few areas require further research. First, extension of the proposed method to

construct designs for continuous regions is needed. Second, clever computational techniques and programming tricks can be used to improve the code for Procedure C so that it can work for larger candidate sets. The need to store the entire constraint matrix may be eliminated with delayed generation of the rows and columns of the constraint matrix (Bersimas and Tsitsiklis, 1997). In addition, state-of-the-art LP solvers can reduce the time required by Procedure C. For example, CPLEX is capable of solving LP's with more than 10^5 constraints and variables in reasonable time (see Bixby (2002)). Third, extension of the heuristic procedure to generate more than one near-minimax designs of each size is needed. Fourth, we can substitute HGW with more modern and powerful heuristics for solving SCLP, such as the one proposed by Caprara et al. (1999).

5.7 References

Anderson-Cook, C.M. and Robinson, T.J. (2009). "A Designed Screening Study with Prespecified Combinations of Factor Settings," *Quality Engineering*, 21(4), 392-404.

Bayarri, M.J., Berger, J.O., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J., Lin, C., and Tu, J. (2007). "A Framework for Validation of Computer Models," *Technometrics*, 49(2), 138-154.

Bertsimas, D. and Tsitsiklis, J.N. (1997). *Introduction to Linear Optimization*, Belmont, MA: Athena Scientific.

Bixby, R.E. (2002). "Solving Real-World Linear Programs: A Decade and More of Progress," *Operations Research*, 50(1), 3-15.

Box, G.E.P. (1993). "Sequential Experimentation and Sequential Assembly of Designs," *Quality Engineering*, 5(2), 321-330.

Box, G.E.P., Hunter, W.G., Hunter, J.S. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, New York: Wiley.

Caprara, A, Fischetti, M., and Toth, P. (1999). "A Heuristic Method for the Set Covering Problem," *Operations Research*, 47(5), 730-743.

Caprara, A. and Toth, P. (2000). "Algorithms for the Set Covering Problem," *Annals of Operations Research*, 98, 353-371.

Chen, V.C.P., Tsui, K., Barton, R.R., and Meckesheimer, M. (2006). "A Review on Design, Modeling and Applications of Computer Experiments," *IIE Transactions*, 38(4), 273-291.

Chvatal, V. (1979). "A Greedy Heuristic for the Set-Covering Problem," *Mathematics of Operations Research*, 4(3), 233-235.

Cook, D.R. and Nachtsheim, C.J. (1980). "A Comparison of Algorithms for Constructing Exact D-Optimal Designs," *Technometrics*, 22(3), 315-324.

Cortex, P. and Morais, A. (2007). "A Data Mining Approach to Predict Forest Fires using Meteorological Data." In Neves, J. , Santos, M. F. and Machado, J. Eds., *New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, December, Guimarães, Portugal*, 512-523.

Current, J., Daskin, M., and Schilling, D. (2002). "Discrete Network Location Models," in *Facility Location: Applications and Theory*, eds. Drezner, Z. and Hamacher, H.W., Berlin: Springer-Verlag, pp. 81-118.

Draguljić, D., Santner, J., and Dean, A.M. (2012). "Noncollapsing Space-Filling Designs for Bounded Nonrectangular Regions," *Technometrics*, 54(2), 169-178.

Erell, E., Pearlmutter, D., and Williamson, T. (2010). *Urban Microclimate: Designing the Spaces Between Buildings*, London: Earthscan.

Grossman, T. and Wool, A. (1997). "Computational Experience with Approximation Algorithms for the Set Covering Problem," *European Journal of Operational Research*, 101, 81-92.

Hakimi, S.L. (1965). "Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph," *Operations Research*, 12(3), 450-459.

Hochbaum, D.S. (1982). “Approximation Algorithms for the Set Covering and Vertex Cover Problems,” *SIAM Journal of Computing*, 11(3), 555-556.

John, P.W.M., Johnson, M.E., Moore, L.M., and Ylvisaker, D. (1995). “Minimax Distance Designs in Two-Level Factorial Experiments,” *Journal of Statistical Planning and Inference*, 44, 249-263.

Johnson, M.E., Moore, L.M., and Ylvisaker, D. (1990). “Minimax and Maximin Distance Designs,” *Journal of Statistical Planning and Inference*, 26, 131-148.

Kennard, R.W. and Stone, L.A. (1969). “Computer Aided Design of Experiments,” *Technometrics*, 11(1), 137-148.

Loeppky, J.L., Sacks, J., and Welch, W.J. (2009). “Choosing the Sample Size of a Computer Experiment: A Practical Guide,” *Technometrics*, 51(4), 366-376.

Marengo, E. and Todeschini, R. (1992). “A New Algorithm for Optimal, Distance-based Experimental Design,” *Chemometrics and Intelligent Laboratory Systems*, 16, 37-44.

McKay, M.D., Beckman, R.J., and Conover, W.J. (1979). “A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code,” *Technometrics*, 21(2), 239-245.

Miller, C. and Urban, D.L. (1999). “A Model of Surface Fire, Climate and Forest Pattern in the Sierra Nevada, California,” Ecological Modelling, 114, 113-135.

Morris, M.D. and Mitchell, T.J. (1995). “Exploratory Designs for Computational Experiments,” *Journal of Statistical Planning and Inference*, 43, 381-402.

Oakley, J.E. and O’Hagan. (2002). “Bayesian Inference for the Uncertainty Distribution of Computer Model Outputs,” *Biometrika*, 89(4), 769-784.

Oakley, J.E. and O’Hagan. (2004). “Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach,” *Journal of the Royal Statistical Society (Series B)*, 66(3), 751-769.

Qian, P.Z.G. (2009). “Nested Latin Hypercube Designs,” *Biometrika*, 96(4), 957-970.

Qian, P.Z.G., Tang, B., and Wu, C.F.J. (2009). "Nested Space-Filling Designs for Computer Experiments with Two Levels of Accuracy," *Statistica Sinica*, 19, 287-300.

Royle, J.A. and Nychka, D. (1998). "An Algorithm for the Construction of Spatial Coverage Designs with Implementation in SPLUS," *Computers and Geosciences*, 24(5), 479-488.

Santner, T.J., Williams, B.J., and Notz, W.I. (2003). *The Design and Analysis of Computer Experiments*, New York: Springer-Verlag.

SAS Institute (2010). *SAS/QC(R) 9.2 User's Guide*, Cary, NC: SAS Institute.

Snyder, L.V. (2011). "Covering Problems," in *Foundations of Location Analysis*. Edited by Eiselt, H.A. and Marianov, V. New York: Springer-Verlag.

Stinstra, E., Hertog, D., Stehouwer, P., and Vestjens, A. (2003). "Constrained Maximin Designs for Computer Experiments," *Technometrics*, 45(4), 340-346.

Sun,Y., Heo, Y., Xie, H., Tan, M., Wu, C.F.J., and Augenbroe, G. (2011). "Uncertainty Quantification of Microclimate Variables in Building Energy Simulation," *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association, Sydney*, 2423-2430.

Tang, B. (1993). "Orthogonal Array-Based Latin Hypercubes," *Journal of the American Statistical Association*, 88(424), 1392-1397.

Tan, M.H.Y. and Wu, C.F.J. (2012). "Robust Design Optimization with Quadratic Loss Derived From Gaussian Process Models," *Technometrics*, 54(1), 51-63.

Toregas, C., Swain, R., ReVelle, C., and Bergman, L. (1971). "The Location of Emergency Service Facilities," *Operations Research*, 19(6), 1363-1373.

APPENDIX A

SUPPLEMENTARY MATERIAL FOR CHAPTER 1

A.1 Simulation Procedure for Assessing Effectiveness of Direct Selective Assembly Procedure I

1. Create the $kN \times N^k$ constraint matrix $\boldsymbol{A}$ for Program A, where $\boldsymbol{A}$ can be obtained in the following way:
   i. Generate a N^k full factorial design $\boldsymbol{F} = (F_{ij})$, where the N levels are labelled as $1,2, \dots, N$.
   ii. For the $1st$ row of $\boldsymbol{A}$, set the element in the jth column equal 1 if $F_{j1} = 1$ and zero otherwise. For the $2nd$ row of $\boldsymbol{A}$, set the element in the jth column equal 1 if $F_{j1} = 2$ and zero otherwise. Continue in this fashion until the Nth row of $\boldsymbol{A}$. Then, generate the $(N+1)th$ to $(2N)th$ row of $\boldsymbol{A}$ based on the $2nd$ column of $\boldsymbol{F}$. Continue until all kN rows of $\boldsymbol{A}$ have been generated.
2. For $i = 1, \dots, k$, draw a random sample of size N, $\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{iN}$ from distribution $F_{\boldsymbol{X}_i}$.
3. Compute all $Q_{i_1 i_2 \cdots i_k} = \min\{L(\boldsymbol{Y}_{i_1 i_2 \cdots i_k}, \boldsymbol{T}), \sum_{j=1}^{k} c_j^S\}$, where $\boldsymbol{Y}_{i_1 i_2 \cdots i_k} = \boldsymbol{f}(\boldsymbol{X}_{1i_1}, \dots, \boldsymbol{X}_{ki_k})$. Write the costs in a vector $\boldsymbol{Q}$, where the jth element is $Q_{i_1 i_2 \cdots i_k}$ if the jth row of $\boldsymbol{F}$ is $(i_1, i_2, \dots, i_k)$.
4. Solve Program A, which can be written compactly as $\min_{\boldsymbol{D}}\{\boldsymbol{Q}'\boldsymbol{D}: \boldsymbol{A}\boldsymbol{D} = \boldsymbol{1}, \boldsymbol{D} \in \mathbb{Z}^{N^k}, \boldsymbol{D} \geq \boldsymbol{0}\}$, where $\boldsymbol{D}$ is the vector of decision variables, and $\boldsymbol{1}$ is a kN vector of 1's.
5. Repeat steps 2-4 M times. Estimate the expected quality cost of a batch by $\widehat{QC}_{SA} =$

$\frac{1}{M}\sum_{i=1}^{M} QC_i^*$, where QC_i^* denotes the optimal objective function value obtained in the ith replicate.

A.2 Illustrative Figures of Assemblies Discussed in Examples 1.1-1.4

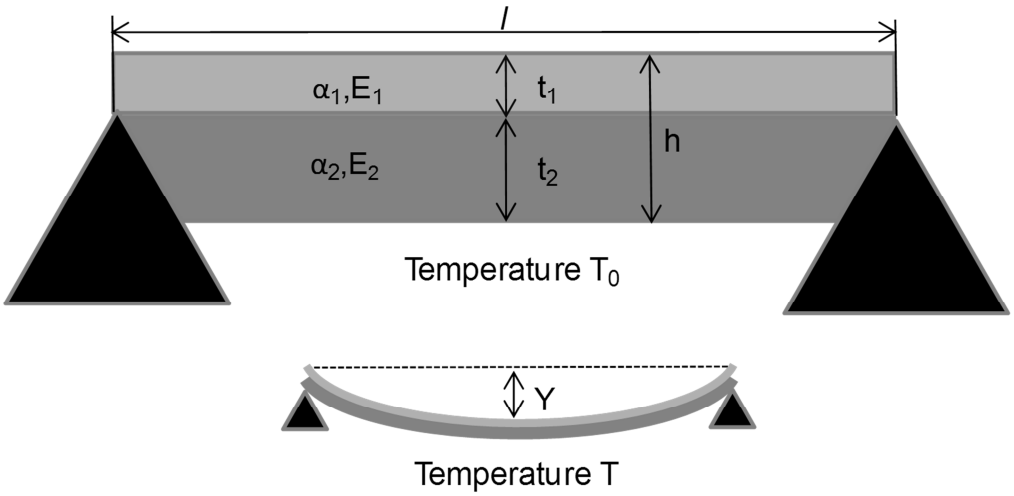

Figure A.1: Bimetal Thermostat

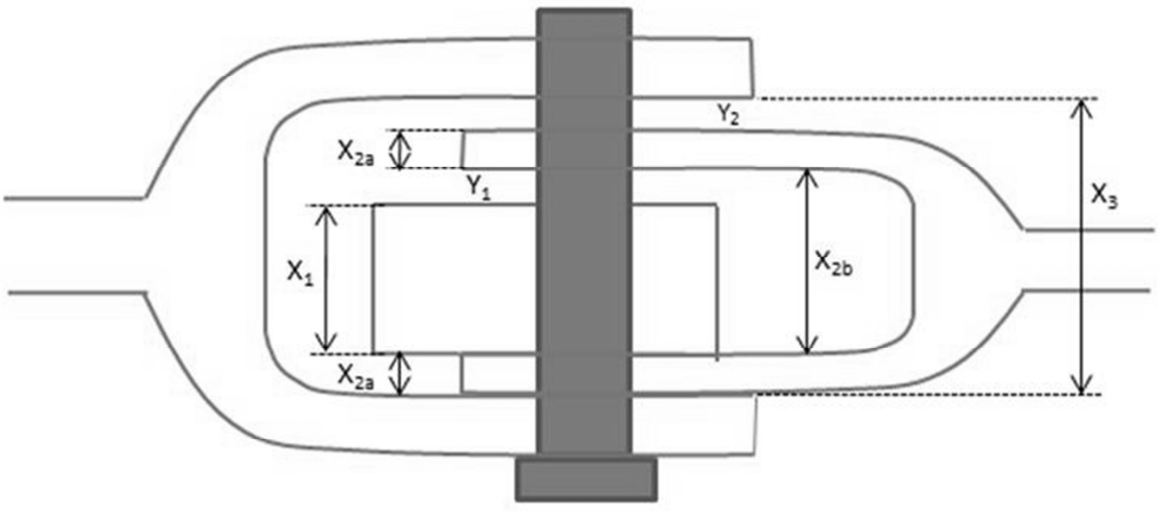

Figure A.2: Knucle Joint Assembly

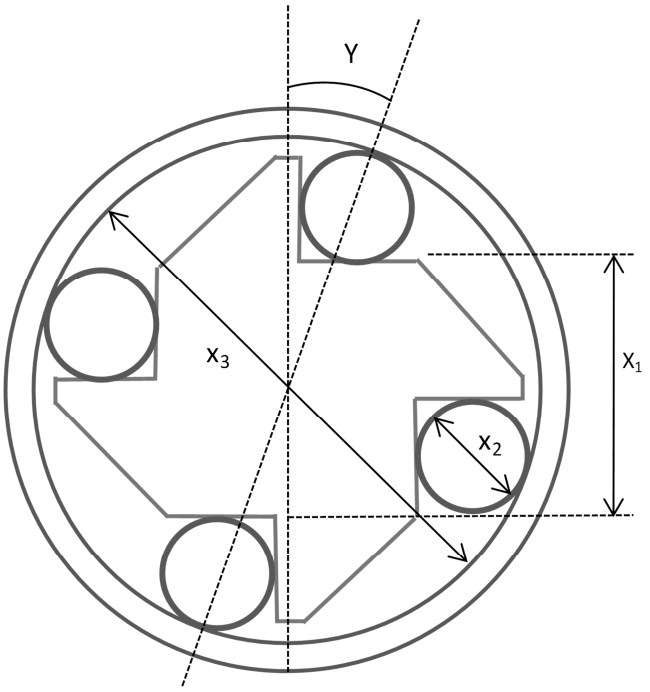

Figure A.3: Fortini's Clutch

A.3 Simulation Procedure for Assessing Effectiveness of Fixed Bin Selective Assembly

Procedure II

Given a combination of $\boldsymbol{n} \in \omega_{\boldsymbol{n}}$ and bin-formation rule, do the following.

1. Generate the $(\sum_{i=1}^{k} n_i) \times \prod_{i=1}^{k} n_i$ constraint matrix $\boldsymbol{A}$. This is done as follows:
   i. Generate an $n_1 \times \cdots \times n_k$ full factorial design $\boldsymbol{F} = (F_{ij})$, where the n_j levels of the jth column are labelled as 1,2, … , n_j.
   ii. For the $1st$ row of $\boldsymbol{A}$, set the element in the jth column equal 1 if $F_{j1} = 1$ and zero otherwise. For the $2nd$ row of $\boldsymbol{A}$, set the element in the jth column equal 1 if $F_{j1} = 2$ and zero otherwise. Continue in this fashion until the $n_1 th$ row of $\boldsymbol{A}$. Then, generate the $(n_1 + 1)th$ to $(n_1 + n_2)th$ row of $\boldsymbol{A}$ based on the $2nd$ column of $\boldsymbol{A}$. Continue until all $\sum_{i=1}^{k} n_i$ rows of $\boldsymbol{A}$ have been generated.

2. Determine the b_{ij}'s based on the bin-formation rule.
3. Compute all $Q_{i_1 i_2 \cdots i_k} = \min\{E[L(\boldsymbol{Y}_{i_1 i_2 \cdots i_k}, \boldsymbol{T})], \sum_{j=1}^{k} c_j^S\}$ and write the costs in a vector $\boldsymbol{Q}$, where the jth element is $Q_{i_1 i_2 \cdots i_k}$ if the jth row of $\boldsymbol{F}$ is $(i_1, i_2, \ldots, i_k)$.
4. Generate $\boldsymbol{\eta} = (N_{11}, \ldots, N_{1n_1}, N_{21}, \ldots, N_{2n_2}, \ldots, N_{k1}, \ldots, N_{kn_k})$. Note that $(N_{i1}, \ldots, N_{in_i})$ has a multinomial distribution with probabilities $p_{ij} = F_{X_i}(b_{ij}) - F_{X_i}(b_{i,j-1})$.
5. Solve Program B, which can be written compactly as $\min_{\boldsymbol{D}} \{\boldsymbol{Q}'\boldsymbol{D}: \boldsymbol{A}\boldsymbol{D} = \boldsymbol{\eta}, \boldsymbol{D} \in \mathbb{Z}^{\prod_{i=1}^{k} n_i}, \boldsymbol{D} \geq \boldsymbol{0}\}$.
6. Repeat Steps 4-5 M times. Denote the optimal objective function value obtained in the ith replicate as QC_i^*. We can estimate the expected total cost of a batch by $\widehat{TC} = \frac{1}{M}\sum_{i=1}^{M} QC_i^* + C(\boldsymbol{n}, N)$, and the expected quality cost by $\widehat{QC}_{SA} = \frac{1}{M}\sum_{i=1}^{M} QC_i^*$.

A.4 Simulation Procedure for Estimating Expected Total Cost of Alternative Bin Designs

Procedure III

1. Select the confidence level $1 - \alpha$, practically significant difference δ, number of replicates $M = r$ for Procedure II and first stage sample size $m_1 \geq 2$. Set $t = t_{(1-\alpha/2)^{1/(\gamma-1)}, m_1 - 1}$ and $h = h\left(1 - \frac{\alpha}{2}, m_1, \gamma\right)$, where h is Rinott's (see Rinott, 1978) constant and γ is the number of alternatives to be compared, i.e. $|\omega_{alt}|$.
2. For each $\in \omega_{alt}$, do the following.
   i. For $l = 1, \ldots, m_1$, run Procedure II with $M = r$. Denote the estimate of the total cost for alternative i obtained in the lth run by $\widehat{TC}_{il}$.
   ii. Compute the sample mean $\overline{TC}_i^{(1)} = \sum_{l=1}^{m_1} \widehat{TC}_{il} / m_1$ and sample variance $s_i^2 =$

$\sum_{l=1}^{m_1}\left(\widehat{TC}_{il} - \overline{TC}_i^{(1)}\right)^2 /(m_1 - 1)$.

3. Compute $W_{ij} = t\sqrt{\left(s_i^2 + s_j^2\right)/m_1}$ for all $i \neq j$.

4. Set $I = \{i: i \in \omega_{alt}\ and\ \overline{TC}_i^{(1)} \leq \overline{TC}_j^{(1)} + \left(W_{ij} - \delta\right)^+ \forall i \neq j\}$.

5. If I contains a single index, then stop and return that combination of $\boldsymbol{n}$ and bin-formation rule as best.

6. Otherwise, for all $\in I$, compute $m_2 = \max\left\{m_1, \left(\frac{hs_i}{\delta}\right)^2\right\}$.

7. For all $i \in I$, run Procedure II $(m_2 - m_1)$ additional times with $M = r$ and compute the sample mean $\overline{TC}_i^{(2)} = \sum_{l=1}^{m_2} \widehat{TC}_{il} /m_2$.

8. Select as best the alternative $i \in I$ with smallest $\overline{TC}_i^{(2)}$.

APPENDIX B

SUPPLEMENTARY MATERIAL FOR CHAPTER 2

B.1 Cumulant Generating Function, Cumulants, and Moments of Quadratic Forms

Lemma B.1: Let $\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim N\left(\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\right)$, where $\boldsymbol{V}_{11}$ is positive definite, and $\boldsymbol{H}$ be a symmetric matrix. Write $\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_{11} & \boldsymbol{H}_{12} \\ \boldsymbol{H}_{21} & \boldsymbol{H}_{22} \end{pmatrix}$, where $\boldsymbol{H}_{11}$ is a square matrix with the same number of rows as $\boldsymbol{Y}_1$ and $\boldsymbol{H}_{22}$ is a square matrix with the same number of rows as $\boldsymbol{Y}_2$. Denote the minimum and maximum eigenvalues of $\boldsymbol{V}_{11}^{1/2}\boldsymbol{H}_{11}\boldsymbol{V}_{11}^{1/2}$ by λ'_{min} and λ'_{max} respectively. Then, the quadratic form $Q = \boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y}$ has the cumulant-generating function given by

$$K(t) = t(\boldsymbol{\mu}^T\boldsymbol{H}\boldsymbol{\mu}) - \frac{1}{2}\ln[\det(\boldsymbol{I} - 2t\boldsymbol{V}_{11}\boldsymbol{H}_{11})] + 2t^2(\boldsymbol{\mu}_2^T\boldsymbol{H}_{21} + \boldsymbol{\mu}_1^T\boldsymbol{H}_{11})(\boldsymbol{V}_{11}^{-1} - 2t\boldsymbol{H}_{11})^{-1}(\boldsymbol{H}_{12}\boldsymbol{\mu}_2 + \boldsymbol{H}_{11}\boldsymbol{\mu}_1) \quad \text{(B.0)}$$

for $t \in (-|2\min(\lambda'_{min}, 0)|^{-1}, [2\max(\lambda'_{max}, 0)]^{-1})$.

Proof: This result follows easily from Equation (2.6) of Feuerverger and Wong (2000).

Theorem B.1: Let $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{V})$, where $\boldsymbol{V}$ is positive semidefinite, and let $\boldsymbol{V}^{1/2}\boldsymbol{H}\boldsymbol{V}^{1/2} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T$ be the spectral decomposition of $\boldsymbol{V}^{1/2}\boldsymbol{H}\boldsymbol{V}^{1/2}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix with diagonal elements $\lambda_1 \leq \cdots \leq \lambda_l$. Then, the cumulant-generating function of $Q = \boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y}$ is given by

$$K(t) = t\boldsymbol{\mu}^T\boldsymbol{H}\boldsymbol{\mu} - \frac{1}{2}\ln[\det(\boldsymbol{I} - 2t\boldsymbol{V}\boldsymbol{H})] + 2t^2\boldsymbol{\mu}^T\boldsymbol{H}\boldsymbol{V}^{1/2}(\boldsymbol{I} - 2t\boldsymbol{V}^{1/2}\boldsymbol{H}\boldsymbol{V}^{1/2})^{-1}\boldsymbol{V}^{1/2}\boldsymbol{H}\boldsymbol{\mu}$$

$$= t\boldsymbol{\mu}^T\boldsymbol{H}\boldsymbol{\mu} - \frac{1}{2}\sum_{i=1}^{l}\ln(1-2t\lambda_i) + \sum_{i=1}^{L}\frac{2t^2u_i^2}{1-2t\lambda_i}, \quad \text{(B.1)}$$

where u_i is the ith element of the vector $\boldsymbol{\Gamma}^T\boldsymbol{V}^{1/2}\boldsymbol{H}\boldsymbol{\mu}$, and the expression holds for $t \in \left((-|2\min(\lambda_1, 0)|^{-1}, [2\max(\lambda_l, 0)]^{-1})\right)$. Moreover, we have

$$K'(t) = \boldsymbol{\mu}^T\boldsymbol{H}\boldsymbol{\mu} + \sum_{i=1}^{L}\frac{\lambda_i}{1-2t\lambda_i} + 4\sum_{i=1}^{L}\frac{t(1-t\lambda_i)u_i^2}{(1-2t\lambda_i)^2}, \quad \text{(B.2)}$$

$$K''(t) = 2\sum_{i=1}^{L}\frac{\lambda_i^2}{(1-2t\lambda_i)^2} + 4\sum_{i=1}^{L}\frac{u_i^2}{(1-2t\lambda_i)^3}, \quad \text{(B.3)}$$

and

$$K'''(t) = 8\sum_{i=1}^{L}\frac{\lambda_i^3}{(1-2t\lambda_i)^3} + 24\sum_{i=1}^{L}\frac{u_i^2\lambda_i}{(1-2t\lambda_i)^4}. \quad \text{(B.4)}$$

Proof: Let $\boldsymbol{B}$ be an orthogonal matrix such that $= \boldsymbol{Z} = \begin{pmatrix}\boldsymbol{Z}_1\\ \boldsymbol{Z}_2\end{pmatrix} \sim N(\boldsymbol{M}, \boldsymbol{W}) =$

$N\left(\begin{pmatrix}\boldsymbol{M}_1\\ \boldsymbol{M}_2\end{pmatrix}, \begin{pmatrix}\boldsymbol{W}_{11} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\right)$, where $\boldsymbol{W}_{11}$ is diagonal matrix with strictly positive diagonal elements, $\boldsymbol{M} = \boldsymbol{B}\boldsymbol{\mu}$, and $\boldsymbol{W} = \boldsymbol{B}\boldsymbol{V}\boldsymbol{B}^T$. Hence, we have $Q = \boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y} = \boldsymbol{Y}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{H}\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{Y} = \boldsymbol{Z}^T\boldsymbol{E}\boldsymbol{Z}$, where $\boldsymbol{E} = \boldsymbol{B}\boldsymbol{H}\boldsymbol{B}^T$.

Let us write

$$\boldsymbol{B} = \begin{pmatrix}\boldsymbol{B}_1\\ \boldsymbol{B}_2\end{pmatrix}, \boldsymbol{M} = \begin{pmatrix}\boldsymbol{M}_1\\ \boldsymbol{M}_2\end{pmatrix} = \boldsymbol{B}\boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{B}_1\boldsymbol{\mu}\\ \boldsymbol{B}_2\boldsymbol{\mu}\end{pmatrix},$$

$$\boldsymbol{W} = \begin{pmatrix}\boldsymbol{W}_{11} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix} = \boldsymbol{B}\boldsymbol{V}\boldsymbol{B}^T = \begin{pmatrix}\boldsymbol{B}_1\boldsymbol{V}\boldsymbol{B}_1^T & \boldsymbol{B}_1\boldsymbol{V}\boldsymbol{B}_2^T\\ \boldsymbol{B}_2\boldsymbol{V}\boldsymbol{B}_1^T & \boldsymbol{B}_2\boldsymbol{V}\boldsymbol{B}_2^T\end{pmatrix}, \text{ and}$$

$$\boldsymbol{E} = \boldsymbol{B}\boldsymbol{H}\boldsymbol{B}^T = \begin{pmatrix}\boldsymbol{B}_1\boldsymbol{H}\boldsymbol{B}_1^T & \boldsymbol{B}_1\boldsymbol{H}\boldsymbol{B}_2^T\\ \boldsymbol{B}_2\boldsymbol{H}\boldsymbol{B}_1^T & \boldsymbol{B}_2\boldsymbol{H}\boldsymbol{B}_2^T\end{pmatrix} = \begin{pmatrix}\boldsymbol{E}_{11} & \boldsymbol{E}_{12}\\ \boldsymbol{E}_{21} & \boldsymbol{E}_{22}\end{pmatrix}.$$

All identity matrices are denoted by $\boldsymbol{I}$, and the dimension of each $\boldsymbol{I}$ shall be clear from the context. By (B.0), we have

$K(t) = \log[M(t)]$

$$= t(\boldsymbol{M}^T\boldsymbol{E}\boldsymbol{M}) - \frac{1}{2}\ln[\det(\boldsymbol{I} - 2t\boldsymbol{W}_{11}\boldsymbol{E}_{11})]$$

$$+2t^2(\boldsymbol{M}_2^T\boldsymbol{E}_{21} + \boldsymbol{M}_1^T\boldsymbol{E}_{11})(\boldsymbol{W}_{11}^{-1} - 2t\boldsymbol{E}_{11})^{-1}(\boldsymbol{E}_{12}\boldsymbol{M}_2 + \boldsymbol{E}_{11}\boldsymbol{M}_1)$$

$$= t(\boldsymbol{\mu}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{H}\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{\mu}) - \frac{1}{2}\ln[\det(\boldsymbol{I} - 2t\boldsymbol{W}\boldsymbol{E})]$$

$$+2t^2(\boldsymbol{\mu}^T\boldsymbol{B}_2^T\boldsymbol{B}_2\boldsymbol{H}\boldsymbol{B}_1^T + \boldsymbol{\mu}^T\boldsymbol{B}_1^T\boldsymbol{B}_1\boldsymbol{H}\boldsymbol{B}_1^T)(\boldsymbol{W}_{11}^{-1} - 2t\boldsymbol{E}_{11})^{-1}(\boldsymbol{B}_1\boldsymbol{H}\boldsymbol{B}_2^T\boldsymbol{B}_2\boldsymbol{\mu} + \boldsymbol{B}_1\boldsymbol{H}\boldsymbol{B}_1^T\boldsymbol{B}_1\boldsymbol{\mu})$$

$$= t(\boldsymbol{\mu}^T\boldsymbol{H}\boldsymbol{\mu}) - \tfrac{1}{2}\ln[\det(\boldsymbol{I} - 2t\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{H}\boldsymbol{B}^T)] + 2t^2\boldsymbol{\mu}^T\boldsymbol{H}\boldsymbol{B}_1^T(\boldsymbol{W}_{11}^{-1} - 2t\boldsymbol{E}_{11})^{-1}\boldsymbol{B}_1\boldsymbol{H}\boldsymbol{\mu}$$

$$= t(\boldsymbol{\mu}^T\boldsymbol{H}\boldsymbol{\mu}) - \tfrac{1}{2}\ln[\det(\boldsymbol{I} - 2t\boldsymbol{V}\boldsymbol{H})] + 2t^2\boldsymbol{\mu}^T\boldsymbol{H}\boldsymbol{V}^{1/2}(\boldsymbol{I} - 2t\boldsymbol{V}^{1/2}\boldsymbol{H}\boldsymbol{V}^{1/2})^{-1}\boldsymbol{V}^{1/2}\boldsymbol{H}\boldsymbol{\mu},$$

where the last equality follows from the fact that

$$\boldsymbol{V}^{1/2}(\boldsymbol{I} - 2t\boldsymbol{V}^{1/2}\boldsymbol{H}\boldsymbol{V}^{1/2})^{-1}\boldsymbol{V}^{1/2}$$

$$= \boldsymbol{B}_1^T\boldsymbol{W}_{11}^{1/2}\boldsymbol{B}_1(\boldsymbol{I} - 2t\boldsymbol{B}_1^T\boldsymbol{W}_{11}^{1/2}\boldsymbol{B}_1\boldsymbol{H}\boldsymbol{B}_1^T\boldsymbol{W}_{11}^{1/2}\boldsymbol{B}_1)^{-1}\boldsymbol{B}_1^T\boldsymbol{W}_{11}^{1/2}\boldsymbol{B}_1$$

$$= \boldsymbol{B}_1^T\boldsymbol{W}_{11}^{1/2}\boldsymbol{B}_1(\boldsymbol{B}^T\boldsymbol{B} - 2t\boldsymbol{B}_1^T\boldsymbol{W}_{11}^{1/2}\boldsymbol{E}_{11}\boldsymbol{W}_{11}^{1/2}\boldsymbol{B}_1)^{-1}\boldsymbol{B}_1^T\boldsymbol{W}_{11}^{1/2}\boldsymbol{B}_1$$

$$= \boldsymbol{B}_1^T\boldsymbol{W}_{11}^{1/2}\boldsymbol{B}_1\boldsymbol{B}^T\left[\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}\end{pmatrix} - 2t\begin{pmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{pmatrix}\boldsymbol{W}_{11}^{1/2}\boldsymbol{E}_{11}\boldsymbol{W}_{11}^{1/2}(\boldsymbol{I} \quad \boldsymbol{0})\right]^{-1}\boldsymbol{B}\boldsymbol{B}_1^T\boldsymbol{W}_{11}^{1/2}\boldsymbol{B}_1$$

$$= \boldsymbol{B}_1^T\boldsymbol{W}_{11}^{1/2}(\boldsymbol{I} \quad \boldsymbol{0})\begin{pmatrix}\boldsymbol{I} - 2t\boldsymbol{W}_{11}^{1/2}\boldsymbol{E}_{11}\boldsymbol{W}_{11}^{1/2} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{pmatrix}\boldsymbol{W}_{11}^{1/2}\boldsymbol{B}_1$$

$$= \boldsymbol{B}_1^T\boldsymbol{W}_{11}^{1/2}(\boldsymbol{I} - 2t\boldsymbol{W}_{11}^{1/2}\boldsymbol{E}_{11}\boldsymbol{W}_{11}^{1/2})^{-1}\boldsymbol{W}_{11}^{1/2}\boldsymbol{B}_1 = \boldsymbol{B}_1^T(\boldsymbol{W}_{11}^{-1} - 2t\boldsymbol{E}_{11})^{-1}\boldsymbol{B}_1.$$

Note that (B.1) holds for $t \in (-|2\min(\lambda'_{min}, 0)|^{-1}, [2\max(\lambda'_{max}, 0)]^{-1})$, where λ'_{min} and λ'_{max} are the minimum and maximum eigenvalues of $\boldsymbol{W}_{11}^{1/2}\boldsymbol{E}_{11}\boldsymbol{W}_{11}^{1/2}$. Since $\boldsymbol{B}\boldsymbol{V}^{1/2}\boldsymbol{H}\boldsymbol{V}^{1/2}\boldsymbol{B}^T = \begin{pmatrix}\boldsymbol{W}_{11}^{1/2}\boldsymbol{E}_{11}\boldsymbol{W}_{11}^{1/2} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}$, we have $\lambda'_{min} = \lambda_1$ if $\lambda'_{min} < 0$ and $\lambda'_{max} = \lambda_l$ if $\lambda'_{max} > 0$.

Corollary B.1: Let $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{V})$, where $\boldsymbol{V}$ is positive semidefinite. Then, the cumulants for the quadratic form $Q = \boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y}$ are given by

$$K_r = 2^{r-1}(r-1)!\{trace[(\boldsymbol{HV})^r] + r\boldsymbol{\mu}^T(\boldsymbol{HV})^{r-1}\boldsymbol{H\mu}\}, r \in \mathbb{N}. \quad \text{(B.5)}$$

Proof: This result follows by expanding $(\boldsymbol{I} - 2t\boldsymbol{V}^{1/2}\boldsymbol{HV}^{1/2})^{-1}$ and $-\frac{1}{2}\ln[\det(\boldsymbol{I} - 2t\boldsymbol{VH})]$ in the expression for $K(t)$ as power series in t (see Equations (5.10) and (5.11) in Khuri (2009)).

Theorem B.2: Let $\boldsymbol{Y}$ be a multivariate t random variable with v degrees of freedom, location parameter $\boldsymbol{\mu}$, and scale matrix $\boldsymbol{V}$, i.e., $\boldsymbol{Y} =^d \boldsymbol{Z}/\sqrt{\chi_v^2/v} + \boldsymbol{\mu}$, where $\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{V})$, χ_v^2 is a chi-squared random variable with v degrees of freedom and $\boldsymbol{Z}$ and χ_v^2 are independent. Then, $Q = \boldsymbol{Y}^T\boldsymbol{HY}$ has first three central moments given by

$M_1 = E(Q) = c_1a_1 + b_1, v > 2,$

$M_2 = E[(Q - M_1)^2] = c_2(a_1^2 + a_2) + c_1(2a_1b_1 + b_2) + b_1^2 - M_1^2, v > 4,$

$M_3 = E[(Q - M_1)^3] = c_3(a_1^3 + 3a_1a_2 + a_3) + c_2(3a_1^2b_1 + 3a_2b_1 + 3a_1b_2 + b_3) + c_1(3a_1b_1^2 + 3b_1b_2) + b_1^3 - 3M_1M_2 - M_1^3, v > 6,$

where $a_r = 2^{r-1}(r-1)!\,trace[(\boldsymbol{HV})^r]$, $b_r = 2^{r-1}r!\,\boldsymbol{\mu}^T(\boldsymbol{HV})^{r-1}\boldsymbol{H\mu}$, and $c_r = [(1-2/v)(1-4/v)\cdots(1-2r/v)]^{-1}$.

Proof: We have $Q = \boldsymbol{Y}^T\boldsymbol{HY} =^d \left(\boldsymbol{Z}/\sqrt{\chi_v^2/v} + \boldsymbol{\mu}\right)^T\boldsymbol{H}\left(\boldsymbol{Z}/\sqrt{\chi_v^2/v} + \boldsymbol{\mu}\right)$. Thus, $\boldsymbol{Z}/\sqrt{\chi_v^2/v} + \boldsymbol{\mu}\big|\chi_v^2 =^d N(\boldsymbol{\mu}, (v/\chi_v^2)\boldsymbol{V})$. It follows that $Q|\chi_v^2 = \left(\boldsymbol{Z}/\sqrt{\chi_v^2/v} + \boldsymbol{\mu}\right)^T\boldsymbol{H}\left(\boldsymbol{Z}/\sqrt{\chi_v^2/v} + \boldsymbol{\mu}\right)\Big|\,\chi_v^2$ has cumulants given by

$K_r = 2^{r-1}(r-1)!\{(v/\chi_v^2)^r trace[(\boldsymbol{HV})^r] + (v/\chi_v^2)^{r-1}r\boldsymbol{\mu}^T(\boldsymbol{HV})^{r-1}\boldsymbol{H\mu}\}$

$= a_r(v/\chi_v^2)^r + b_r(v/\chi_v^2)^{r-1} = a_rJ^r + b_rJ^{r-1},$

where $J = v/\chi_v^2$. Now, we can obtain $E(Q^r|\chi_v^2)$ using the relationship between moments and cumulants. Since $E(Q^r|\chi_v^2)$ can be expressed as a polynomial of degree r in J, we

can obtain the raw moments $E(Q^r)$ using the fact that

$$E(J^r) = E\left[\left(\frac{v}{\chi_v^2}\right)^r\right] = [(1-2/v)(1-4/v)\cdots(1-2r/v)]^{-1} = c_r.$$

The central moments can then be obtained from the raw moments.

B.2 Data for Examples

Table B.1: Data for Example 2.1

BHF d_1	Friction d_2	Material d_3	Part Thickness d_4	Y_1	Y_2
13.6492	0.1405	1.0819	11.8017	16.1242	12.3581
13.7553	0.0904	0.9814	11.9484	16.5589	12.1747
13.7211	0.1534	0.8811	11.9619	12.5694	9.4674
2.3992	0.1512	0.9933	11.9606	18.442	11.0999
24.9033	0.097	1.0923	11.9473	19.8823	13.4837
2.5053	0.1005	1.0913	12.0388	17.8932	11.0234
25.0027	0.1466	1.0063	12.0067	17.0396	11.4995
2.4711	0.1521	0.991	12.0383	17.5935	10.8937
2.5291	0.1016	1.0999	12.0638	15.0679	11.1937
13.7791	0.0404	0.8937	11.9781	14.1995	11.092
24.997	0.0536	1.016	11.9351	17.7311	12.976
2.5072	0.0958	0.8941	12.0725	9.8603	10.0025
2.4983	0.0422	0.9933	11.8973	14.9086	9.9908
13.7572	0.0995	1.0031	12.0039	16.452	12.23
13.7483	0.0959	1.0025	12.0272	16.6234	12.2783
13.7372	0.0555	0.9081	11.9032	15.399	11.3933
2.4872	0.097	0.8889	12.0405	14.7347	9.3532
13.792	0.1035	1.0077	11.9982	16.3857	12.243
24.9737	0.0527	0.9922	11.9569	17.2571	12.6146
13.7809	0.1515	0.9056	12.0172	12.6997	9.7055
24.9978	0.0977	0.8982	12.0322	15.9928	11.2144
13.7721	0.0484	1.1079	11.9815	18.508	13.6613
13.7187	0.0964	0.9919	11.8797	16.5546	12.2819
24.9718	0.1483	1.0078	11.9887	17.0759	11.555
13.7858	0.1407	1.1077	12.0855	17.0869	13.951
25.0469	0.1037	0.9091	11.8921	16.4418	11.4263
13.7685	0.0428	1.0955	12.0642	18.7483	14.9717
13.7511	0.096	0.9936	12.0129	16.4303	12.1832
2.542	0.048	1.0014	11.9466	15.2362	10.1161
24.988	0.1004	1.0918	12.0695	18.7641	14.3375

Table B.2: Data for Example 2.3

A	B	C	n	m	FC	Y
502.1	710.8	0.0067	0.125	0.7	0.75	328.75
553.1	600.8	0.0268	0.234	1	0.31	346.64
604.1	490.8	0.0067	0.343	1.3	0.27	461.21
655.1	600.8	0.0134	0.234	1	0.325	349.87
502.1	490.8	0.0067	0.343	0.7	0.625	538.07
553.1	600.8	0.0134	0.016	1	0.82	297.08
604.1	710.8	0.0201	0.343	1.3	0.035	358.65
553.1	600.8	0.0134	0.234	1.6	0.15	380.1
553.1	820.8	0.0134	0.234	1	0.25	344.98
502.1	490.8	0.0201	0.343	1.3	0.2	434.53
553.1	600.8	0.0134	0.452	1	0.18	456.77
502.1	710.8	0.0067	0.343	1.3	0.108	433.48
604.1	490.8	0.0201	0.125	1.3	0.375	341.5
553.1	380.8	0.0134	0.234	1	0.63	442.39
502.1	710.8	0.0201	0.125	1.3	0.25	315.75
451.1	600.8	0.0134	0.234	1	0.44	436.15
604.1	710.8	0.0201	0.125	0.7	0.72	306.88
604.1	710.8	0.0067	0.125	1.3	0.325	325.09
604.1	710.8	0.0067	0.343	0.7	0.425	435.75
604.1	490.8	0.0067	0.125	0.7	0.79	342.89
502.1	710.8	0.0201	0.343	0.7	0.37	419.73
553.1	600.8	0.0134	0.234	0.4	0.85	394.89
553.1	600.8	0	0.234	1	0.47	459.84
502.1	490.8	0.0067	0.125	1.3	0.52	426.87
604.1	490.8	0.0201	0.343	0.7	0.56	415.21
502.1	490.8	0.0201	0.125	0.7	0.81	357.67
553.1	600.8	0.0134	0.234	1	0.375	385.27

B.3 Supplementary Figure for Section 2.7.1

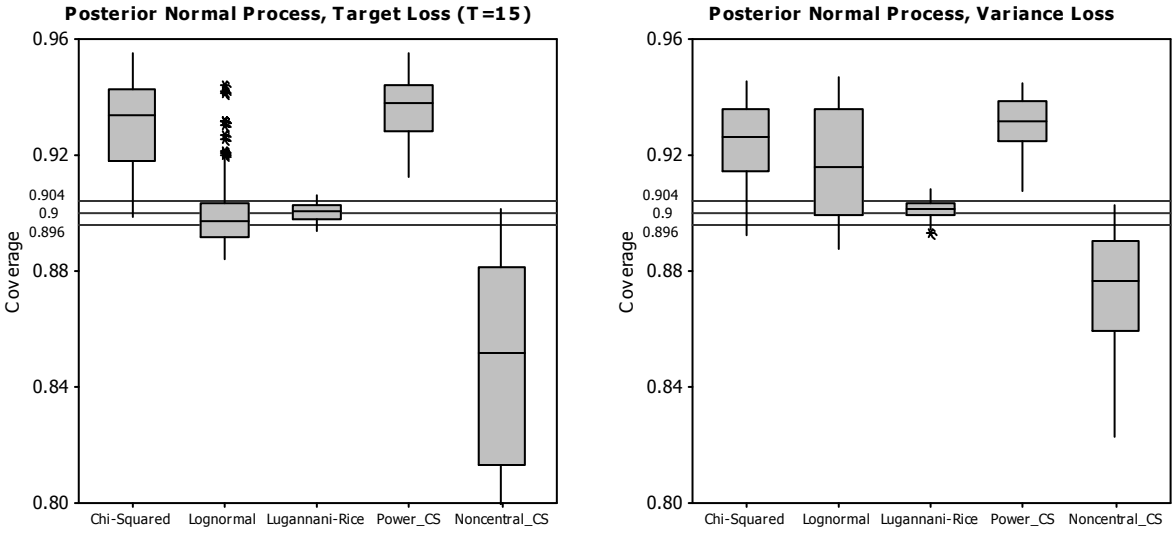

Figure B.1. Empirical Coverage of 90% Credible Intervals for Response Y_1, Posterior Normal Process. The empirical coverage is evaluated at 125 points on the control factor space $[0,1]^3$, $\boldsymbol{x}_c^0$ and $\boldsymbol{x}_c^1$.

B.4 Parameter Settings of Pattern Search Algorithm

We only changed the maximum number of iterations, tolerance of function, and tolerance of variable to make sure that the algorithm converges, as indicated by the exitflag output. All other settings remained as default values. For finding MLE's, we set

```
psoptimset('display','off','MaxIter',90000,'TolFun',10^-6,'TolX',10^-6)
```

For finding optimal settings in Example 2.1, all algorithm parameter values are default values. For Example 2.2, we use a 10-point grid to identify a close-to-optimal control factor setting, and then we refine the setting using a crude patternsearch with parameter values

```
psoptimset('display','off','MaxIter',90000,'TolFun',10^-3,'TolX',10^-3)
```

Careful use of other optimization packages should produce the same results.

APPENDIX C

SUPPLEMENTARY MATERIAL FOR CHAPTER 3

C.1 Covariance Matrices and Data for Examples

C.1.1 Covariance Matrix for Example 3.1

0.6952	0.2133	0.9125	0.3183	-0.023	0.0074	0.2711	0.5471	0.0857	-0.1293	-0.4068	-0.3014
0.2133	0.2033	0.4661	-0.0033	0.0425	0.128	0.0834	0.3103	0.0458	-0.16	-0.2934	-0.2345
0.9125	0.4661	1.7496	0.3168	0.0584	0.2637	0.3879	1.0961	0.0909	-0.4433	-0.877	-0.7561
0.3183	-0.0033	0.3168	1.045	-0.3363	-0.2946	0.3119	0.2712	0.2634	0.4661	0.077	0.3024
-0.023	0.0425	0.0584	-0.3363	0.2524	0.1839	-0.0717	0.0683	-0.111	-0.2382	-0.1413	-0.2186
0.0074	0.128	0.2637	-0.2946	0.1839	0.4899	0.0156	0.4161	-0.0313	-0.3132	-0.3254	-0.3152
0.2711	0.0834	0.3879	0.3119	-0.0717	0.0156	0.2783	0.3853	0.0363	-0.012	-0.104	-0.0991
0.5471	0.3103	1.0961	0.2712	0.0683	0.4161	0.3853	1.1345	0.1296	-0.3019	-0.6347	-0.4795
0.0857	0.0458	0.0909	0.2634	-0.111	-0.0313	0.0363	0.1296	0.3016	0.1946	-0.1296	0.1436
-0.1293	-0.16	-0.4433	0.4661	-0.2382	-0.3132	-0.012	-0.3019	0.1946	0.5211	0.3287	0.4919
-0.4068	-0.2934	-0.877	0.077	-0.1413	-0.3254	-0.104	-0.6347	-0.1296	0.3287	0.6865	0.4841
-0.3014	-0.2345	-0.7561	0.3024	-0.2186	-0.3152	-0.0991	-0.4795	0.1436	0.4919	0.4841	0.669

C.1.2 Covariance Matrices for the Example 3.2

Covariance Matrix 1

0.1964	0.1323	0.0615	-0.1305	0.1137	-0.0881
0.1323	1.8235	0.2952	-1.2099	0.3865	-0.1227
0.0615	0.2952	0.184	-0.2689	0.0944	-0.0752
-0.1305	-1.2099	-0.2689	1.1715	-0.4383	0.3145
0.1137	0.3865	0.0944	-0.4383	0.386	-0.1177
-0.0881	-0.1227	-0.0752	0.3145	-0.1177	0.2915

Covariance Matrix 2

0.511	0.0374	-0.176	-0.4448	0.624	0.835
0.0374	0.2593	0.0128	-0.3342	0.111	0.0288
-0.176	0.0128	0.1999	0.0381	-0.1952	-0.2217
-0.4448	-0.3342	0.0381	1.8928	-0.9916	-1.4821
0.624	0.111	-0.1952	-0.9916	1.1354	1.5288
0.835	0.0288	-0.2217	-1.4821	1.5288	2.6812

Table C.1: Data for Example 3.4

Var 1 (LYS)[1]	Var 2 (ARG)	Var 3 (ASP)	Var 4 (SER)	Var 5 (GLU)	Var 6 (PRO)	Var 7 (GLY)	Var 8 (ALA)	Var 9 (VAL)	Var 10 (PHA)	Var 11 (GABA)
0.48	5.81	2.12	4.68	0.78	12.41	0.31	0.96	0.18	0.2	4.73
0.47	5.25	2.75	4.42	0.88	14.72	0.3	1.04	0.19	0.22	3.96
0.42	4.98	2.79	3.85	0.75	12.13	0.32	0.99	0.15	0.2	3.94
0.35	4.79	2.79	3.39	0.81	12.77	0.25	0.75	0.16	0.15	3.69
0.43	4.92	2.88	3.53	0.78	13.11	0.25	0.91	0.16	0.15	4.23
0.4	5.61	2.26	3.39	0.69	12.69	0.2	1.06	0.16	0.18	3.76
0.35	4.54	2.96	3.89	0.88	14.01	0.24	0.86	0.16	0.12	3.92
0.34	3.82	2.86	3.63	0.86	15.73	0.22	1.34	0.14	0.12	2.88
0.27	3.42	2.27	4.81	0.9	8.99	0.23	1.43	0.1	0.1	2.68
0.39	3.6	2.99	5.03	0.92	13.71	0.28	1.99	0.13	0.1	2.88
0.37	3.39	2.78	5.96	0.84	12.92	0.24	1.76	0.12	0.14	3.01
0.26	2.72	3.82	6.03	1.17	7.18	0.15	1.3	0.11	0.07	3.4
0.24	3.13	3.35	5.76	0.96	6.75	0.21	1.14	0.11	0.08	2.43
0.2	2.15	3.28	5.8	1.04	5.34	0.22	1.06	0.12	0.08	2.41
0.26	2.89	3.67	6.34	1.22	5.87	0.18	1.1	0.14	0.12	2.4
0.52	5.53	2.97	3.37	0.78	10.74	0.24	0.96	0.1	0.16	3.4
0.42	5.07	3.06	4.32	0.91	15.37	0.47	1.32	0.16	0.2	3.63
0.45	5.46	3.06	4.68	0.84	16.52	0.39	1.35	0.14	0.18	3.89
0.47	5.79	2.91	4.44	0.8	16.21	0.35	1.2	0.2	0.18	4.52
0.44	2.52	2.4	4.09	0.72	12.81	0.28	0.86	0.18	0.23	4.43
0.48	5.14	2.66	4.04	0.94	16.77	0.33	0.97	0.22	0.23	4.9
0.49	4.77	2.42	5.92	1	15.62	0.34	1.93	0.5	0.15	4.05
0.37	4.35	3.04	5.07	0.87	15.81	0.31	2.08	0.19	0.1	4.17
0.36	4.01	2.37	3.93	0.76	11.28	0.22	0.75	0.12	0.12	3.27
0.46	4.26	2.51	7.29	1.07	18.57	0.37	2.67	0.19	0.1	2.95
0.34	3.46	2.2	3.8	0.93	11.73	0.26	1.4	0.18	0.1	3.06
0.34	4.13	2.72	6.01	0.95	13.96	0.34	2.3	0.1	0.08	3.06
0.31	3.7	2.77	5.29	0.85	10.8	0.22	1.68	0.1	0.01	2.61
0.3	3.18	2.54	5.04	0.95	11.25	0.21	1.84	0.1	0.01	2.48
0.3	3.57	2.45	5.7	1.06	12.28	0.26	1.53	0.1	0.1	2.46
0.3	3.31	2.53	5.21	0.88	9.1	0.23	1.37	0.08	0.01	2.55
0.3	3.13	2.82	5.85	1	10.31	0.21	1.55	0.1	0.08	2.69
0.33	3.1	3.01	7.15	1.04	12.71	0.23	1.79	0.09	0.1	3.52
0.32	3.84	3.79	6.08	1.01	10.13	0.18	1.3	0.09	0.01	3.67
0.3	3.75	2.83	6.24	0.71	6.2	0.16	1.2	0.05	0.08	3.01
0.26	3.34	3.46	7.01	1.02	6.68	0.2	1.52	0.1	0.08	2.18

[1]LYS=Lysine, ARG=Arginine, ASP=Aspartic acid, SER=Serine, GLU=Glutamine acid, PRO=Proline, GLY=Glycine, ALA=Alanine, VAL=Valine, PHA=Phenyl alanine, GABA=Gamma-amino butric acid

C.2 Additional Figures for Example 3.4

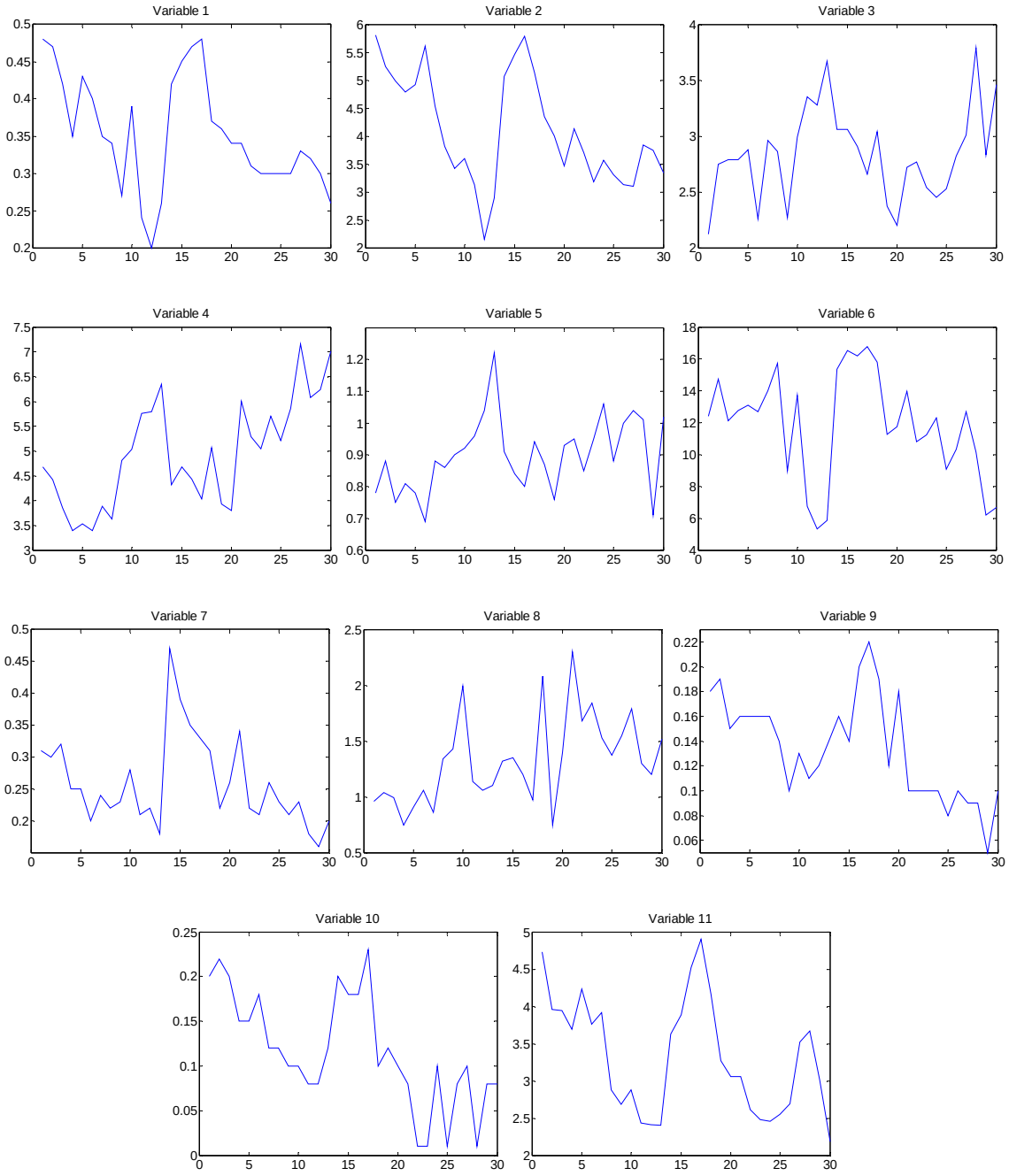

Figure C.1: Plot of Phase I and Phase II Observations for Example 3.4

C.3 Matlab Code for Implementing the Empirical Bayes Approach

Note: This code was used to obtain the results in Example 3.4.

```
function EmpiricalBayes
format short g
global PhaseI OC
%Phase I global variable should contain the phase I data
%OC global variable should contain the out-of-control sample
[N p]=size(PhaseI);
[n ~]=size(OC);

abststat=sqrt(n)*abs((mean(OC)-mean(PhaseI))./(std(PhaseI)));
ind=find(abststat>2);
lambda=abststat(ind);
h=mean(lambda)
a0=std(lambda)*sqrt(N/n);

Xbb=mean(PhaseI)';
Sigma=cov(PhaseI);
S=Sigma*(N-1);

Sf=(n-1)*cov(OC);
Xbf=mean(OC)';

C=h*sqrt(diag(Sigma))/sqrt(n);
a=max(max(a0,(h/2)*sqrt(N/n)-1),1)
a=a*ones(p,1);

sigma2hat=diag(Sigma)/(N);

Nruns=20000;
tau=10000;
q=Nruns-tau;

t=(N-p-2)/2;
v=n+2*t+p+1;

Del=zeros(p,1);
Prob=[repmat([0.25 0.5 0.25],p,1)];
mu=Xbb;
T=Sf+S;

for i=1:Nruns

Scale=inv(T+n*(Xbf-mu)*(Xbf-mu)');
Sigmainv=wishartrnd(Scale,v,p);
psi=sigma2hat.*(a.^(2*abs(Del)));
psiinv=diag(1./psi);
Theta=Xbb+Del.*C;
g=(psiinv+n*Sigmainv)\(psiinv*Theta+n*Sigmainv*Xbf);
V=inv(psiinv+n*Sigmainv);

mu=(mvnrnd(g',V))';
Smu(:,i)=mu;

for k=1:p
```

```
for j=1:3
    Del2(k)=j-2;
    psi2=sigma2hat(k)*(a(k)^(2*abs(Del2(k))));
    psiinv2=1/psi2;
    Theta2(k)=Xbb(k)+Del2(k)*C(k);
    probdel=Prob(k,j);
    const(j)=psiinv2^0.5*exp(-0.5*psiinv2*(mu(k)-Theta2(k))^2)*probdel;
end
Del(k)=randsample([-1 0 1],1,true,const);
end

SDel(:,i)=Del;
end

SDel2=SDel(:,(tau+1):Nruns);
for i=1:p
    ['Marginal posterior distribution of delta',num2str(i)]
    tabulate(SDel2(i,:))
end

count=zeros(3^p,1);

index=zeros(q,1);
for i=1:q
    for j=p:-1:2
        index(i,1)=index(i,1)+(SDel2(j,i)+1)*3^(j-1);
    end
    index(i,1)=index(i,1)+(SDel2(1,i)+2);
    count(index(i,1),1)=count(index(i,1),1)+1;
end

[val index2]=max(count);
posteriormode=indexconvert(index2,p)
display('Posterior probability of mode')
val/(Nruns-tau)

function wishartrnd=wishartrnd(Sigma,n,p)
if(n>=p)
L=chol(Sigma)';
A=zeros(p,p);
for i=1:p
A(i,i)=sqrt(chi2rnd(n-i+1));
end
for i=2:p
    for j=1:i-1
        A(i,j)=normrnd(0,1);
    end
end
temp=L*A;
wishartrnd=temp*temp';
%wishartrnd=L*A*A'*L';
elseif(n<=p-1)
for i=1:n+1
X(:,i)=(mvnrnd(zeros(1,p),Sigma))';
end
wishartrnd=(X-repmat(mean(X,2),1,n+1))*(X-repmat(mean(X,2),1,n+1))';
```

```
end

function indexconvert=indexconvert(no,p)
u=1;
remainder=no;
for i=p:-1:2
    remainder2=rem(remainder,3^(i-1));
    if(remainder2==0&u==1)
        index(i)=(remainder-remainder2)/(3^(i-1))-2;
        u=0;
    elseif(remainder==0&u==0)
        index(i)=1;
    else
        index(i)=(remainder-remainder2)/(3^(i-1))-1;
    end
    remainder=remainder2;
end
if(remainder==0)
    index(1)=1;
else
    index(1)=remainder-2;
end
indexconvert=index;
```

APPENDIX D

SUPPLEMENTARY MATERIAL FOR CHAPTER 4

D.1 Wholeplot and Subplot Type I and Type II Error Rates for Simulation in Section 4.6.1.1

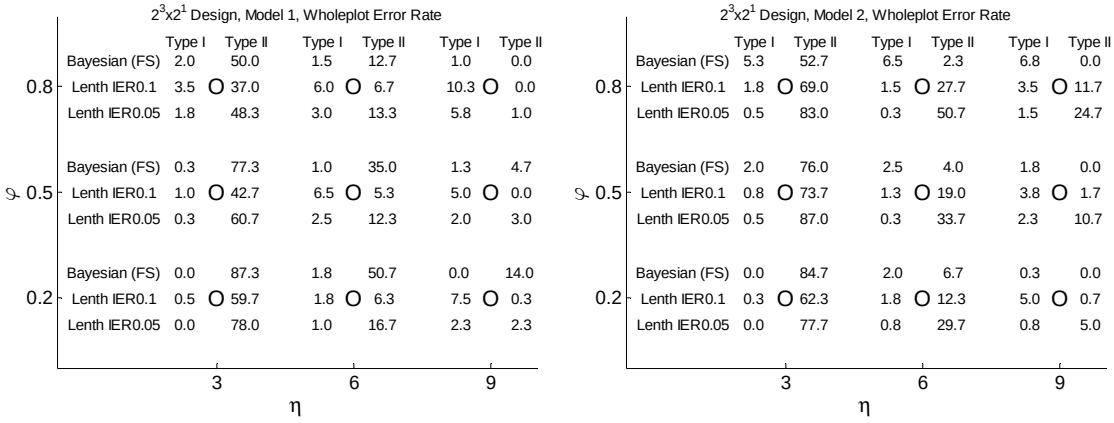

Figure D.1: Type I and Type II Wholeplot Error Rates for Proposed Method (Bayesian (FS)) and Lenth's Method with IER=0.1 and IER=0.05 (Lenth IER0.1, Lenth IER0.05); $(\eta, \varphi) \in \{3,6,9\} \times \{0.2,0.5,0.8\}$; Model 1 (left) and Model 2 (right)

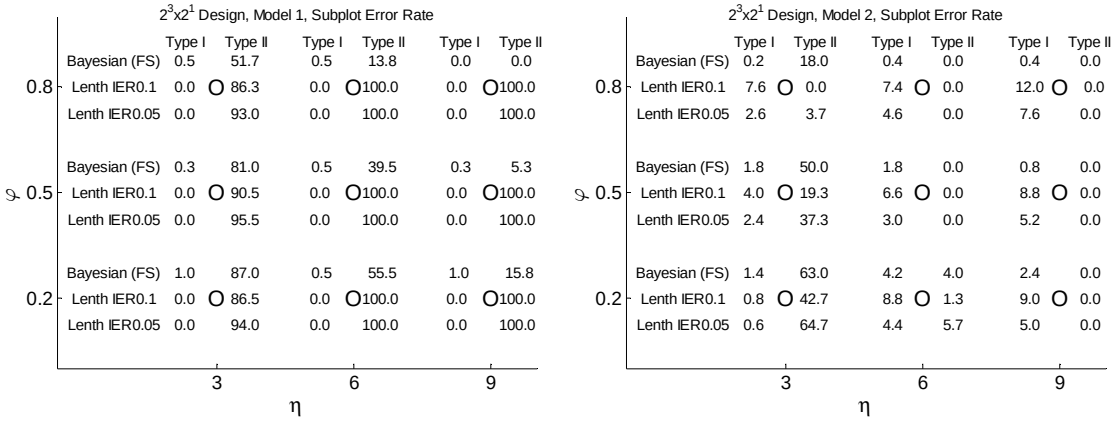

Figure D.2: Type I and Type II Subplot Error Rates for Proposed Method (Bayesian (FS)) and Lenth's Method with IER=0.1 and IER=0.05 (Lenth IER0.1, Lenth IER0.05); $(\eta, \varphi) \in \{3,6,9\} \times \{0.2,0.5,0.8\}$; Model 1 (left) and Model 2 (right)

Figure D.3: Type I and Type II Wholeplot Error Rates for Best Model Found with GS; $(\eta, \varphi) \in \{3,6,9\} \times \{0.2,0.5,0.8\}$; Model 3 (left) and Model 4 (right)

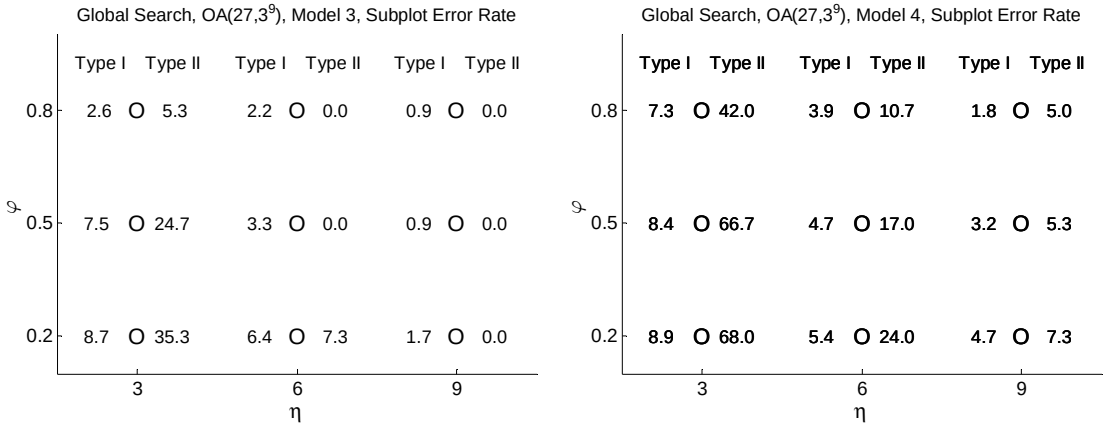

Figure D.4: Type I and Type II Subplot Error Rates for Best Model Found with GS; $(\eta, \varphi) \in \{3,6,9\} \times \{0.2,0.5,0.8\}$; Model 3 (left) and Model 4 (right)

D.2 Theoretical Results

Lemma D.1: For $\boldsymbol{R}_{\varphi}$ given by

$$\boldsymbol{R}_{\varphi} = \begin{pmatrix} \varphi\mathbf{1}\mathbf{1}^T + (1-\varphi)\boldsymbol{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \varphi\mathbf{1}\mathbf{1}^T + (1-\varphi)\boldsymbol{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \varphi\mathbf{1}\mathbf{1}^T + (1-\varphi)\boldsymbol{I} \end{pmatrix} =$$

$$\boldsymbol{I} \otimes \begin{pmatrix} 1 & \varphi & \cdots & \varphi \\ \varphi & 1 & \cdots & \varphi \\ \vdots & \vdots & \ddots & \vdots \\ \varphi & \varphi & \cdots & 1 \end{pmatrix},$$

$$\boldsymbol{R}_{\varphi}^{-1} = \boldsymbol{I}\otimes\begin{pmatrix} 1 & \varphi & \cdots & \varphi \\ \varphi & 1 & \cdots & \varphi \\ \vdots & \vdots & \ddots & \vdots \\ \varphi & \varphi & \cdots & 1 \end{pmatrix}^{-1} = \begin{pmatrix} A\boldsymbol{I} + B\boldsymbol{1}\boldsymbol{1}^T & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & A\boldsymbol{I} + B\boldsymbol{1}\boldsymbol{1}^T & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & A\boldsymbol{I} + B\boldsymbol{1}\boldsymbol{1}^T \end{pmatrix},$$

where $A = (1-\varphi)^{-1}$ and $B = -\frac{\varphi}{(1-\varphi)(1+(r-1)\varphi)}$. Moreover, we have $\boldsymbol{1}^T\boldsymbol{R}_{\varphi}^{-1}\boldsymbol{1} = n/[1+(r-1)\varphi]$.

Lemma D.2: Set $\overline{\boldsymbol{Y}} = \bar{Y}\boldsymbol{1}$, where $\bar{Y} = \sum_{i=1}^{N}\sum_{j=1}^{r} Y_{ij}\,/n$. Then, $(\overline{\boldsymbol{Y}} - \beta_0\boldsymbol{1})^T\boldsymbol{R}_{\varphi}^{-1}(\boldsymbol{Y} - \overline{\boldsymbol{Y}} - \boldsymbol{X}\boldsymbol{\beta}) = 0$ and $(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \beta_0\boldsymbol{1})^T\boldsymbol{R}_{\varphi}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \beta_0\boldsymbol{1}) = (\boldsymbol{Y} - \overline{\boldsymbol{Y}} - \boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{R}_{\varphi}^{-1}(\boldsymbol{Y} - \overline{\boldsymbol{Y}} - \boldsymbol{X}\boldsymbol{\beta}) + (\overline{\boldsymbol{Y}} - \beta_0\boldsymbol{1})^T\boldsymbol{R}_{\varphi}^{-1}(\overline{\boldsymbol{Y}} - \beta_0\boldsymbol{1})$.

Lemma D.3: Assume $\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{S}^{-1}$ has an inverse (this would be true if $\boldsymbol{S}$ is diagonal with positive elements on the diagonal so that $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} + \boldsymbol{S}^{-1}$ is positive definite). Define

$$RSS_{\boldsymbol{\delta},c,\varphi} = (\boldsymbol{Y} - \overline{\boldsymbol{Y}})^T\boldsymbol{R}_{\varphi}^{-1/2}\left\{\boldsymbol{I} - \boldsymbol{R}_{\varphi}^{-1/2}\boldsymbol{X}\left[\left(\boldsymbol{X}^T\boldsymbol{R}_{\varphi}^{-1}\boldsymbol{X}\right) + \boldsymbol{S}_{\boldsymbol{\delta},c}^{-1}\right]^{-1}\boldsymbol{X}^T\boldsymbol{R}_{\varphi}^{-1/2}\right\}\boldsymbol{R}_{\varphi}^{-1/2}(\boldsymbol{Y} - \overline{\boldsymbol{Y}})$$

and $\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi} = \left(\boldsymbol{X}^T\boldsymbol{R}_{\varphi}^{-1}\boldsymbol{X} + \boldsymbol{S}_{\boldsymbol{\delta},c}^{-1}\right)^{-1}$. Then,

$$\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{S}^{-1})^{-1}\boldsymbol{X}^T = (\boldsymbol{I} + \boldsymbol{X}\boldsymbol{S}\boldsymbol{X}^T)^{-1}, \tag{D1}$$

$$RSS_{\boldsymbol{\delta},c,\varphi} = (\boldsymbol{Y} - \overline{\boldsymbol{Y}})^T\left(\boldsymbol{R}_{\varphi} + \boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T\right)^{-1}(\boldsymbol{Y} - \overline{\boldsymbol{Y}}), \tag{D2}$$

$$\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi} = \boldsymbol{S}_{\boldsymbol{\delta},c} - \boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T\left(\boldsymbol{R}_{\varphi} + \boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T\right)^{-1}\boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}. \tag{D3}$$

Proof: Equations (D1) and (D3) follow from the Woodbury Matrix Identity. Equation (D2) follows from (D1).

Lemma D.4: $\left|\left(\boldsymbol{X}^T\boldsymbol{R}_{\varphi}^{-1}\boldsymbol{X}\right) + \boldsymbol{S}_{\boldsymbol{\delta},c}^{-1}\right|^{-1/2}\left|\boldsymbol{S}_{\boldsymbol{\delta},c}\right|^{-1/2}\left|\boldsymbol{R}_{\varphi}\right|^{-1/2} = \left|\boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T + \boldsymbol{R}_{\varphi}\right|^{-1/2}$.

Proposition D.1: The joint posterior distribution of $(\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}, c, \varphi)$ is

$$p(\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}, c, \varphi|\boldsymbol{Y}) \propto L(\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}, \varphi|\boldsymbol{Y})p(\beta_0)p(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}, c)p(\sigma^2)p(\boldsymbol{\delta})p(\varphi)p(c)$$

$$\propto (\sigma^2)^{-(n+p+v+2)/2}\left|\boldsymbol{R}_{\varphi}\right|^{-1/2} exp\left(-\frac{v\lambda + RSS_{\boldsymbol{\delta},c,\varphi}}{2\sigma^2}\right) exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta} - \boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}\right)^T\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi}^{-1}\left(\boldsymbol{\beta} -\right.\right.$$

$\boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}\big)\big\}\, exp\left(-\frac{(\beta_0-\bar{Y})^2}{2\sigma^2\left(\mathbf{1}^T\boldsymbol{R}_{\varphi}^{-1}\mathbf{1}\right)^{-1}}\right)\left|\boldsymbol{S}_{\boldsymbol{\delta},c}\right|^{-1/2} p(\boldsymbol{\delta})p(\varphi)p(c),$ (D4)

where $L(\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}, \varphi|\boldsymbol{Y})$ is the likelihood,

$\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi} = \boldsymbol{S}_{\boldsymbol{\delta},c} - \boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T\left(\boldsymbol{R}_{\varphi} + \boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T\right)^{-1}\boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}$, $\boldsymbol{m}_{\boldsymbol{\delta},c,\varphi} = \boldsymbol{G}_{\boldsymbol{\delta},c,\varphi}\boldsymbol{X}^T\boldsymbol{R}_{\varphi}^{-1}(\boldsymbol{Y} - \overline{\boldsymbol{Y}})$, $\overline{\boldsymbol{Y}} = \bar{Y}\mathbf{1}$,

$\bar{Y} = \sum_{i=1}^{N}\sum_{j=1}^{r} Y_{ij}\,/n$, $RSS_{\boldsymbol{\delta},c,\varphi} = (\boldsymbol{Y} - \overline{\boldsymbol{Y}})^T\left(\boldsymbol{R}_{\varphi} + \boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T\right)^{-1}(\boldsymbol{Y} - \overline{\boldsymbol{Y}})$, and $\mathbf{1}^T\boldsymbol{R}_{\varphi}^{-1}\mathbf{1} = n/[1 + (r-1)\varphi]$.

Proof: The result follows by writing down

$L(\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\delta}, \varphi|\boldsymbol{Y})p(\beta_0)p(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}, c)p(\sigma^2)p(\boldsymbol{\delta})p(\varphi)p(c)$ and applying Lemmas D.1-D.3.

Proposition D.2: The posterior distribution of $\boldsymbol{\delta}$ is given by

$p(\boldsymbol{\delta}|\boldsymbol{Y}) = \int_0^1\int_0^{\infty} p(\boldsymbol{\delta}, c, \varphi|\boldsymbol{Y})\, dcd\varphi \propto \int_0^1\int_0^{\infty}\left[\upsilon\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\right]^{-(n-1+\upsilon)/2}\left|\boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T + \boldsymbol{R}_{\varphi}\right|^{-1/2}\left(\mathbf{1}^T\boldsymbol{R}_{\varphi}^{-1}\mathbf{1}\right)^{-1/2} p(\boldsymbol{\delta})p(\varphi)p(c)dc\, d\varphi.$ (D5)

Proof: This follows by first integrating out $\beta_0, \boldsymbol{\beta}$, and σ^2 from the joint posterior (D4) given in Proposition D.1. The result is the integrand of the integral with respect to c and φ given above.

Proposition D.3: The posterior distribution of $\boldsymbol{\beta}$ conditional on $\boldsymbol{\delta}$ is given by

$p(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\delta}) = K\int_0^1\int_0^{\infty} t\big(\boldsymbol{\beta}; \boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}, (n+\upsilon-1)^{-1}\big(\upsilon\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\big)\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi}, n+\upsilon-1\big)$

$\big(\upsilon\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\big)^{-(n+\upsilon-1)/2}\left|\boldsymbol{R}_{\varphi} + \boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T\right|^{-1/2}\left(\mathbf{1}^T\boldsymbol{R}_{\varphi}^{-1}\mathbf{1}\right)^{-1/2} p(\varphi)p(c)dcd\varphi,$ (D6)

where

$t\left(\boldsymbol{\beta}; \boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}, \frac{(\upsilon\lambda+RSS_{\boldsymbol{\delta},c,\varphi})\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi}}{n+\upsilon-1}, n+\upsilon-1\right) = \frac{\Gamma\left(\frac{n+\upsilon-1+p}{2}\right)}{\Gamma\left(\frac{n+\upsilon-1}{2}\right)(n+\upsilon-1)^{p/2}\pi^{p/2}}\left|\frac{(\upsilon\lambda+RSS_{\boldsymbol{\delta},c,\varphi})\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi}}{n+\upsilon-1}\right|^{-1/2}$

$\left[1 + \frac{1}{n+\upsilon-1}\left(\boldsymbol{\beta} - \boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}\right)^T\left\{\frac{(\upsilon\lambda+RSS_{\boldsymbol{\delta},c,\varphi})\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi}}{n+\upsilon-1}\right\}^{-1}\left(\boldsymbol{\beta} - \boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}\right)\right]^{-(n+\upsilon-1+p)/2}$ (D7)

is the multivariate student t probability density function with mean $\boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}$, scale matrix $\left(v\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\right)\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi}/(n+v-1)$ and $n+v-1$ degrees of freedom, and

$K =$

$$\left[\int_0^1\int_0^\infty \left(v\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\right)^{-(n+v-1)/2}\left|\boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T + \boldsymbol{R}_\varphi\right|^{-1/2}\left(\mathbf{1}^T\boldsymbol{R}_\varphi^{-1}\mathbf{1}\right)^{-1/2}p(\varphi)p(c)\,dcd\varphi\right]^{-1}.$$

Remark: The distribution $\boldsymbol{\beta}|\boldsymbol{Y},\boldsymbol{\delta}$ is a mixture of multivariate t distributions with weight function

$$\omega(c,\varphi) = K\left(v\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\right)^{-(n+v-1)/2}\left|\boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T + \boldsymbol{R}_\varphi\right|^{-1/2}\left(\mathbf{1}^T\boldsymbol{R}_\varphi^{-1}\mathbf{1}\right)^{-1/2}p(\varphi)p(c).$$

Proof: This result follows by dropping all terms involving $\boldsymbol{\delta}$ only in (D4), and integrating out β_0, and σ^2. The result is the integrand in (D6).

Corollary D.1: The posterior mean of $\boldsymbol{\beta}$ conditional on $\boldsymbol{\delta}$ is given by

$$E(\boldsymbol{\beta}|\boldsymbol{Y},\boldsymbol{\delta}) = \int_0^1\int_0^\infty \boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}\,\omega(c,\varphi)dcd\varphi. \tag{D8}$$

Proof: This follows from (D6) and the fact that $\boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}$ is the mean of the density given in (D7).

Proposition D.4: Let m_i denote the ith component of $\boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}$, θ_i denote the ith diagonal element of $(n+v-1)^{-1}\left(v\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\right)\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi}$. Then,

$$p(\beta_i|\boldsymbol{Y},\boldsymbol{\delta}) = \int_0^\infty\int_0^1 \omega(c,\varphi)t(\beta_i; m_i, \theta_i, n+v-1)\,d\varphi dc, \tag{D9}$$

where $t(\beta_i; m_i, \theta_i, n+v-1)$ is the univariate t-distribution with mean m_i, scale θ_i, and $n+v-1$ degrees of freedom. Thus,

$$P(\beta_i \le x|\boldsymbol{Y},\boldsymbol{\delta}) = \int_0^\infty\int_0^1 \omega(c,\varphi)P\left(t_{n+v-1} \le (x-m_i)/\sqrt{\theta_i}\right)d\varphi dc, \tag{D10}$$

where t_{n+v-1} is a t random variable with mean 0, scale 1, and $n+v-1$ degrees of freedom.

Proof: This result follows from Proposition D.3 and the fact that if

$\boldsymbol{\beta} \sim t\left(\boldsymbol{\beta}; \boldsymbol{m}_{\boldsymbol{\delta},c,\varphi}, \frac{(\upsilon\lambda + RSS_{\boldsymbol{\delta},c,\varphi})\boldsymbol{G}_{\boldsymbol{\delta},c,\varphi}}{n+v-1}, n+v-1\right)$, then $(\beta_i - m_i)/\sqrt{\theta_i} \sim t_{n+v-1}$.

Proposition D.5: The posterior distribution of correlation parameter φ given $\boldsymbol{\delta} \in \mathcal{M}$ is

$$p(\varphi|\boldsymbol{Y}, \boldsymbol{\delta} \in \mathcal{M}) \propto \sum_{\boldsymbol{\delta}\in\mathcal{M}} \int_0^\infty \left[\upsilon\lambda + RSS_{\boldsymbol{\delta},c,\varphi}\right]^{-(n-1+v)/2}$$

$$\left|\boldsymbol{X}\boldsymbol{S}_{\boldsymbol{\delta},c}\boldsymbol{X}^T + \boldsymbol{R}_\varphi\right|^{-1/2}\left(\mathbf{1}^T\boldsymbol{R}_\varphi^{-1}\mathbf{1}\right)^{-1/2} p(c)dcp(\boldsymbol{\delta})p(\varphi). \quad \text{(D11)}$$

Proof: This result follows by integrating out $\beta_0, \boldsymbol{\beta}$, and σ^2 from (D4). The result is the integrand in (D11).

D.3 Gaussian Quadrature Method for Discretizing a Density

Miller and Rice (1983) propose the Gaussian quadrature method for constructing an L-point discrete approximation of a density. Let m_j be the jth moment of the distribution. Then, the procedure for constructing the discrete approximation is given in the steps below.

1. Solve the system of equations

$$\begin{pmatrix} m_0 & \cdots & m_{L-1} \\ \vdots & \ddots & \vdots \\ m_{L-1} & \cdots & m_{2L-2} \end{pmatrix}\begin{pmatrix} C_0 \\ \vdots \\ C_{L-1} \end{pmatrix} = \begin{pmatrix} -m_L \\ \vdots \\ -m_{2L-1} \end{pmatrix} \quad \text{(D12)}$$

   for $C_0, \ldots, C_{L-1}$.

2. Find the roots $(\xi_1, \ldots, \xi_L)$ of the polynomial $\sum_{i=0}^{L-1} C_i x^i + x^L$. These are the support points of the discrete approximation.

3. Solve the system of equations

$$\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ \xi_1^{L-1} & \cdots & \xi_L^{L-1} \end{pmatrix}\begin{pmatrix} w_1 \\ \vdots \\ w_L \end{pmatrix} = \begin{pmatrix} m_0 \\ \vdots \\ m_{L-1} \end{pmatrix} \quad \text{(D13)}$$

   for $w_1, \ldots, w_L$. The values $w_1, \ldots, w_L$ are the probability masses for the support points $\xi_1, \ldots, \xi_L$ respectively.

The discrete approximation given above would be simple to construct if the first $2L-1$ moments of the density are easy to compute. For the beta distribution with parameters a and b, the kth moment can be easily obtained from the recursion

$$m_k = \frac{k+a-1}{a+k+b-1} m_{k-1}, \tag{D14}$$

where $m_0 = 1$.

D.4 Data Used in Examples

Table D.1: Split Plot Design Derived from an $OA(27, 3^9)$ (for Example 4.1)

Wholeplot	Factor	A	B	C	D	E	F	G	H	J
	Subplot	Coded Factor Levels								
1	1	1	1	1	1	1	1	1	1	1
	2	1	1	1	1	2	2	2	2	2
	3	1	1	1	1	3	3	3	3	3
2	4	1	2	2	2	1	2	3	3	3
	5	1	2	2	2	2	3	1	1	1
	6	1	2	2	2	3	1	2	2	2
3	7	1	3	3	3	1	3	2	2	2
	8	1	3	3	3	2	1	3	3	3
	9	1	3	3	3	3	2	1	1	1
4	10	2	1	2	3	1	1	1	2	3
	11	2	1	2	3	2	2	2	3	1
	12	2	1	2	3	3	3	3	1	2
5	13	2	2	3	1	1	2	3	1	2
	14	2	2	3	1	2	3	1	2	3
	15	2	2	3	1	3	1	2	3	1
6	16	2	3	1	2	1	3	2	3	1
	17	2	3	1	2	2	1	3	1	2
	18	2	3	1	2	3	2	1	2	3
7	19	3	1	3	2	1	1	1	3	2
	20	3	1	3	2	2	2	2	1	3
	21	3	1	3	2	3	3	3	2	1
8	22	3	2	1	3	1	2	3	2	1
	23	3	2	1	3	2	3	1	3	2
	24	3	2	1	3	3	1	2	1	3
9	25	3	3	2	1	1	3	2	1	3
	26	3	3	2	1	2	1	3	2	1
	27	3	3	2	1	3	2	1	3	2

Table D.2: Split Plot Design and Data for the First Part of Section 4.6.1.3

Wholeplot	A	B	C	D	E	F	Y^1	Y^2
1	-1	-1	1	-1	1	1	26.0454	11.0454
	-1	-1	-1	1	-1	1	10.1488	19.1488
	-1	-1	-1	-1	-1	-1	16.5896	13.5896
2	-1	1	1	-1	1	1	13.454	14.454
	-1	1	-1	1	1	-1	24.1056	23.1056
	-1	1	1	1	-1	-1	16.2053	21.2053
3	1	-1	1	1	-1	1	31.6942	22.6942
	1	-1	-1	-1	1	-1	13.9312	26.9312
	1	-1	1	1	1	-1	25.7543	22.7543
4	1	1	-1	1	1	1	23.9856	16.9856
	1	1	-1	-1	-1	1	24.3905	23.3905
	1	1	1	-1	-1	-1	14.2034	25.2034

Table D.3: Split Plot Design and Data for the Second Part of Section 4.6.1.3

Wholeplot	A	B	C	D	E	Y^3	Y^4
1	-1	-1	1	-1	1	20.0454	26.0454
	-1	-1	-1	1	-1	8.1488	26.1488
	-1	-1	-1	-1	-1	10.5896	26.5896
2	-1	1	1	-1	1	19.454	29.454
	-1	1	-1	1	1	12.1056	30.1056
	-1	1	1	1	-1	14.2053	30.2053
3	1	-1	1	1	-1	33.6942	37.6942
	1	-1	-1	-1	1	29.9312	33.9312
	1	-1	1	1	1	37.7543	37.7543
4	1	1	-1	1	1	15.9856	25.9856
	1	1	-1	-1	-1	14.3905	26.3905
	1	1	1	-1	-1	24.2034	30.2034

Table D.4: Heat-Exchanger-Fan Casing Experiment

			M	1	1	2	2	3	3
			N	1	2	1	2	1	2
A	B	C	D						
1	1	1	1	1.23	0.99	1.4	1.19	1.56	1.38
1	2	2	2	1.8	1.48	2.05	1.76	2.28	2.03
1	3	3	3	2.31	1.89	2.61	2.23	2.9	2.57
2	1	2	3	1.29	1.02	1.47	1.24	1.65	1.44
2	2	3	1	2.02	1.66	2.29	1.97	2.55	2.26
2	3	1	2	2.09	1.73	2.35	2.04	2.61	2.33
3	1	3	2	1.49	1.19	1.7	1.44	1.9	1.67
3	2	1	3	1.61	1.31	1.79	1.54	1.99	1.77
3	3	2	1	2.26	1.87	2.55	2.2	2.83	2.52

Table D.5: Split Unit Design and Data

Wholeplot	Factor	A	B	C	D	E	F	G	H	J	Bulk Specific
	Subplot	Coded Factor Levels									Gravity
1	1	1	1	1	1	1	1	1	1	1	0.728
	2	1	1	1	1	2	2	2	2	2	0.634
	3	1	1	1	1	3	3	3	1	3	0.59
2	4	1	2	2	2	1	2	3	1	3	0.56
	5	1	2	2	2	2	3	1	1	1	0.538
	6	1	2	2	2	3	1	2	2	2	0.741
3	7	1	3	1	1	1	3	2	2	2	0.668
	8	1	3	1	1	2	1	3	1	3	0.742
	9	1	3	1	1	3	2	1	1	1	0.581
4	10	2	1	2	1	1	1	1	2	3	0.754
	11	2	1	2	1	2	2	2	1	1	0.502
	12	2	1	2	1	3	3	3	1	2	0.559
5	13	2	2	1	1	1	2	3	1	2	0.638
	14	2	2	1	1	2	3	1	2	3	0.657
	15	2	2	1	1	3	1	2	1	1	0.724
6	16	2	3	1	2	1	3	2	1	1	0.568
	17	2	3	1	2	2	1	3	1	2	0.754
	18	2	3	1	2	3	2	1	2	3	0.661
7	19	3	1	1	2	1	1	1	1	2	0.736
	20	3	1	1	2	2	2	2	1	3	0.515
	21	3	1	1	2	3	3	3	2	1	0.698
8	22	3	2	1	1	1	2	3	2	1	0.693
	23	3	2	1	1	2	3	1	1	2	0.607
	24	3	2	1	1	3	1	2	1	3	0.733
9	25	3	3	2	1	1	3	2	1	3	0.597
	26	3	3	2	1	2	1	3	2	1	0.768
	27	3	3	2	1	3	2	1	1	2	0.547

APPENDIX E

SUPPLEMENTARY MATERIAL FOR CHAPTER 5

E.1 Theoretical Results

In this section, we give important theoretical results that justify the proposed method for constructing minimax designs. Note that the results hold for any distance metric $\|\cdot\|$ used to measure the distance between candidate points.

We shall give seven propositions, which we summarize in this paragraph. Proposition E.1 needs no explanation. Proposition E.2 states the important fact that the set of feasible solutions of SCLP with $S = T$ is the set of all designs with distance less than or equal to T. In Proposition E.3, we prove that $z(S)$ is a right continuous, nonincreasing function with discontinuities that are a subset of the design distances. Proposition E.4 gives bounds for the minimax distance d_n^*. It says that if $S_1 < S_2$, and $z(S_1) > z(S_2)$, then for any $n \in [z(S_2), z(S_1) - 1]$, we have $d_n^* \in (S_1, S_2]$. It is easy to see that Proposition E.4 justifies Procedure B: n-SCLP is infeasible at $S = S_1$ if and only if $z(S_1) > n$, and it is feasible at $S = S_2$ if and only if $z(S_2) \leq n$. Proposition E.5 states that the discontinuities of $z(S)$ are a subset of ψ' and that the set of feasible solutions of SCLP remains the same for all $S \in [h^q, h^{q+1})$. Propositions E.3 and E.5 imply that to construct a plot of $z(S)$ versus S, we merely need to find the value of $z(S)$ at $H^1, \ldots, H^m$ (which justify Procedure A). Propositions E.6 and E.7 establish the relationship between minimax designs and SCLP, indicating how minimax designs can be obtained. First, if h^k is a discontinuity of $z(S)$, then $\boldsymbol{I}^*$ is an optimal solution for SCLP at $S \in [h^k, h^{k+1})$ if and only if $D(\boldsymbol{I}^*)$ is a $z(h^k)$-point minimax design. Second, for $n \in [z(h^k), z(h^{k-1}) - 1]$,

$d_n^* = h^k$ and an n-point minimax design can be obtained by adding $n - z(h^k)$ points to $D(\boldsymbol{I}^*)$. Third, the set of minimax distances is the set of discontinuities of $z(S)$.

Proposition E.1: For any $n = 1, \dots, N$, an n-point minimax design always exists.

Proof: This follows from the fact that there are only a finite number of designs.

Proposition E.2: D is a design with $d(D) \leq T$ if and only if $\boldsymbol{I}(D)$ is a feasible solution of SCLP with $S = T$.

Proof: Suppose D is a design with $d(D) \leq T$. For $i = 1, \dots, N$, $\min\{\|\boldsymbol{u}_i - \boldsymbol{x}_1\|, \dots, \|\boldsymbol{u}_i - \boldsymbol{x}_n\|\} = d(\boldsymbol{u}_i, D) \leq \max\{d(\boldsymbol{u}_i, D), \dots, d(\boldsymbol{u}_n, D)\} = d(D) \leq T$. Thus, there exists $\boldsymbol{x}_{k(i)} = \boldsymbol{u}_{j(i)} \in D$ such that $\|\boldsymbol{u}_i - \boldsymbol{u}_{j(i)}\| \leq T$. This implies that $\sum_{j \in \Omega_i} I_j \geq I_{j(i)} = 1$. Conversely, if $\boldsymbol{I}(D)$ is a solution of SCLP with $S = T$, then for $i = 1, \dots, N$, $\sum_{j \in \Omega_i} I_j \geq 1$. Thus, there exists $j(i) \in \Omega_i$ such that $I_{j(i)} = 1$, i.e., there exists $\boldsymbol{x}_{k(i)} = \boldsymbol{u}_{j(i)} \in D$ such that $\|\boldsymbol{u}_i - \boldsymbol{u}_{j(i)}\| \leq T$. It follows that $d(\boldsymbol{u}_i, D) = \min\{\|\boldsymbol{u}_i - \boldsymbol{x}_1\|, \dots, \|\boldsymbol{u}_i - \boldsymbol{x}_n\|\} \leq \|\boldsymbol{u}_i - \boldsymbol{x}_{k(i)}\| = \|\boldsymbol{u}_i - \boldsymbol{u}_{j(i)}\| \leq T$. Thus, $d(D) = \max\{d(\boldsymbol{u}_i, D), \dots, d(\boldsymbol{u}_n, D)\} \leq T$.

Corollary E.1: The set of feasible solutions of SCLP is given by $\{\boldsymbol{I} \in \Lambda: d(D(\boldsymbol{I})) \leq S\}$, where $\Lambda = \{\boldsymbol{I}_1, \dots, \boldsymbol{I}_r\}$ is the set of all nonzero binary N-vectors.

Proposition E.3: The function $z(S)$ is right continuous and nonincreasing with range $\{1, \dots, N\}$. It has a finite set of discontinuities ω which satisfies $\omega \subseteq \{d^1, \dots, d^R\}$, where $d^1, \dots, d^R$ are the distinct values of $d(D(\boldsymbol{I}_1)), \dots, d(D(\boldsymbol{I}_r))$.

Proof: The fact that $z(S)$ has range $\{1, \dots, N\}$ is obvious. By Corollary E.1, $z(S) = \min\{\|\boldsymbol{I}\|_1: \boldsymbol{I} \in \Lambda, d(D(\boldsymbol{I})) \leq S\}$. From this, it clearly follows that $z(S)$ is a nonincreasing function since $\{\boldsymbol{I} \in \Lambda: d(D(\boldsymbol{I})) \leq S_1\} \subseteq \{\boldsymbol{I} \in \Lambda: d(D(\boldsymbol{I})) \leq S_2\}$ if $S_1 < S_2$. It is also clear that the discontinuities of $z(S)$ is a subset of $\{d^1, \dots, d^R\}$. Suppose that $\boldsymbol{I}_j$ is an optimal

solution for SCLP with $S = S_j > d\left(D(\boldsymbol{I}_j)\right)$. Then, it must be an optimal solution for all $S \in \left[d\left(D(\boldsymbol{I}_j)\right), S_j\right]$. So, $z(S)$ is constant for $S \in \left[d\left(D(\boldsymbol{I}_j)\right), S_j\right]$. This implies that $z(S)$ is right continuous.

Lemma E.1: Let D be a design and let $D' = D \cup \{\boldsymbol{x}'\}$. Then, $d(D') \leq d(D)$.

Proof: This follows from the fact that $d(\boldsymbol{u}_i, D) = \min\{\|\boldsymbol{u}_i - \boldsymbol{x}_1\|, \dots, \|\boldsymbol{u}_i - \boldsymbol{x}_n\|\} \geq \min\{\|\boldsymbol{u}_i - \boldsymbol{x}_1\|, \dots, \|\boldsymbol{u}_i - \boldsymbol{x}_n\|, \|\boldsymbol{u}_i - \boldsymbol{x}'\|\} = d(\boldsymbol{u}_i, D')$.

Proposition E.4: If $S_1 < S_2$, and $z(S_1) > z(S_2)$, then the minimax distance d_n^* of a minimax design D_n^* with $n \in [z(S_2), z(S_1) - 1]$ points satisfies $d_n^* \in (S_1, S_2]$.

Proof: For $n' = z(S_2)$, a minimax design $D_{n'}^*$ with n' points must have minimax distance $d_{n'}^* \leq S_2$ since there exists a design with n' points that has distance at most S_2 (by Proposition E.2). For $n \in [z(S_2), z(S_1) - 1]$, let D_n be any n-point design with $D_{n'}^* \subseteq D_n$. Then, by Lemma E.1, we have $d(D_n) \leq d_{n'}^* \leq S_2$. Since $d_n^* \leq d(D_n)$, it follows that $d_n^* \leq S_2$.

It is clear that we must also have $d_n^* > S_1$ for any $n \in [z(S_2), z(S_1) - 1]$. Otherwise, by Proposition E.2, D_n^* is a feasible solution to SCLP with $S = S_1$, and $n < z(S_1)$, which is a contradiction.

Proposition E.5: Let the set of distinct values of $\psi = \{h_{ij} = \|\boldsymbol{u}_i - \boldsymbol{u}_j\| : i, j = 1, \dots, N\}$ be written as $\psi' = \{h^1, \dots, h^m\}$, where $0 = h^1 < \cdots < h^m$. Then, $\omega \subseteq \psi'$, where ω is the set of discontinuities of $z(S)$. Moreover, for any $q = 1, \dots, m$, the set of feasible solutions of SCLP is the same for all $S \in [h^q, h^{q+1})$, where $h^{m+1} = \infty$.

Proof: By Proposition E.3, $\omega \subseteq \{d^1, \dots, d^R\}$. Since $D(\boldsymbol{I}_j) = \{\boldsymbol{u}_k : I_{jk} = 1\}$, where I_{jk} is the kth component of $\boldsymbol{I}_j$, we have $d\left(\boldsymbol{u}_i, D(\boldsymbol{I}_j)\right) = \min\{\|\boldsymbol{u}_i - \boldsymbol{u}_k\| : \boldsymbol{u}_k \in D(\boldsymbol{I}_j)\} \in \psi'$

for all $i = 1, \dots, N$. Thus, $d\left(D(\boldsymbol{I}_j)\right) = \max\left\{d\left(\boldsymbol{u}_i, D(\boldsymbol{I}_j)\right) : i = 1, \dots, N\right\} \in \psi'$. It follows that $\omega \subseteq \{d^1, \dots, d^R\} \subseteq \psi'$ and the set of feasible solutions must remain the same for all $S \in [h^q, h^{q+1})$.

Remark E.1: Proposition E.5 says that the set of discontinuities of $z(S)$ can be identified by solving SCLP at midpoints between consecutive values in ψ'.

Remark E.2: It follows from Proposition E.5 that for any $q \in \{1, \dots, m\}$, a solution that is optimal for some $S \in [h^q, h^{q+1})$ must be optimal for all S in that interval.

Proposition E.6: Let $\boldsymbol{I}_1^*$ be an optimal solution of SCLP for $S_1 \in [h^{k-1}, h^k)$, $\boldsymbol{I}_2^*$ be an optimal solution for $S_2 \in [h^k, h^{k+1})$, and $n_1 = \|\boldsymbol{I}_1^*\|_1 = z(h^{k-1}) > n_2 = \|\boldsymbol{I}_2^*\|_1 = z(h^k)$. Then, for any $n \in [n_2, n_1 - 1]$, the minimax distance d_n^* of a minimax design with n points is $d_n^* = h^k$, $\boldsymbol{I}_2^*$ is an n_2-point minimax design, and an n-point minimax design can be obtained by adding $n - n_2$ points to $D(\boldsymbol{I}_2^*)$.

Proof: Since $\boldsymbol{I}_1^*$ is the optimal solution for all $S_1 = [h^{k-1}, h^k)$ and $\|\boldsymbol{I}_1^*\|_1 > \|\boldsymbol{I}_2^*\|_1$, $\boldsymbol{I}_2^*$ must be infeasible for SCLP for any $S < h^k$. By Proposition E.2, $D(\boldsymbol{I}_2^*)$ must have distance h^k, i.e., $d\left(D(\boldsymbol{I}_2^*)\right) = h^k$. Thus, a minimax design with n_2 points must have minimax distance $d_{n_2}^* \leq h^k$. However, we cannot have $d_{n_2}^* < h^k$ since the smallest designs with distance at most $d \in [h^{k-1}, h^k)$ have n_1 points. This implies that $d_{n_2}^* = h^k$ and $\boldsymbol{I}_2^*$ is a minimax design.

Let $n \in [n_2, n_1 - 1]$. Then, by Proposition E.4, $d_n^* \in (h^k - \varepsilon, h^k]$ for all $0 < \varepsilon \leq h^k - h^{k-1}$. Thus, $d_n^* = h^k$. By Lemma E.1, a minimax design of size n can be obtained by adding $n - n_2$ points to $D(\boldsymbol{I}_2^*)$.

Remark E.3: Proposition E.6 says that a minimax design can be obtained by solving

SCLP a small distance to the right of a discontinuity but before the next distance value in ψ'.

Remark E.4: If $n_1 > n_2 + 1$, then we must have more than one i such that $d(\boldsymbol{u}_i, D(\boldsymbol{I}_2^*)) = \max\{d(\boldsymbol{u}_i, D(\boldsymbol{I}_2^*)), \ldots, d(\boldsymbol{u}_N, D(\boldsymbol{I}_2^*))\} = d(D(\boldsymbol{I}_2^*))$.

Proposition E.7: If $h^k \in \psi'$ is a discontinuity of $z(S)$, then the set of optimal solution of SCLP for any $S \in [h^k, h^{k+1})$ is the set of $z(h^k)$-point minimax design. Moreover, the set of minimax distances is the set of discontinuities of $z(S)$.

Proof: By Proposition E.6, any optimal solution of SCLP at $S \in [h^k, h^{k+1})$ is a $z(h^k)$-point minimax design with distance h^k. Conversely, let D be a $z(h^k)$-point minimax design. Then, $d(D) = h^k$. Thus, $\boldsymbol{I}(D)$ is a feasible solution of SCLP for $S \in [h^k, h^{k+1})$. Since $\|\boldsymbol{I}(D)\|_1 = z(h^k)$, $\boldsymbol{I}(D)$ is an optimal solution.

The set of minimax distances is the set of distances for 1-point to N-point minimax designs. Note that the value of $z(S)$ changes from 1 to N as S is decreased from h^m to 0. This fact and Propositions E.5 and E.6 imply that the set of minimax distances is the set of discontinuities of $z(S)$.

E.2 Alternative Algorithms for Finding Space-Filling Designs on Finite Candidate Sets

KS starts with a two-point design that consists of points furthest apart. It then sequentially adds points with the largest distance to the design. Thus, KS can be viewed as a heuristic method for generating near-minimax designs. Designs obtained from KS can be improved by a modified Fedorov algorithm (hereafter abbreviated as MF) (Cook and Nachtsheim, 1980; SAS Institute, 2010). In each iteration of MF, design point $1, \ldots, n$ is exchanged (in that order) with a nondesign point that gives the largest increase in

distance. The algorithm stops when no improvement is made in an iteration. We call the combined method KS-MF. MT is a Fedorov algorithm (FA) (Cook and Nachtsheim, 1980; SAS Institute, 2010) that starts with a random design. At each step, it exchanges a design point $\boldsymbol{x}_i$ with a nondesign point $\boldsymbol{u}_k$ such that $\rho(\boldsymbol{u}_k; \boldsymbol{x}_i) = \max\{\rho(\boldsymbol{u}_l; \boldsymbol{x}_j) : \boldsymbol{x}_j \in D, \boldsymbol{u}_l \in \chi \backslash D\} > 1$, where $\rho(\boldsymbol{u}_l; \boldsymbol{x}_j) = \min\{\|\boldsymbol{u}_l - \boldsymbol{x}_q\| : q \neq j, q = 1, \ldots, n\} / \min\{\|\boldsymbol{x}_j - \boldsymbol{x}_q\| : q \neq j, q = 1, \ldots, n\}$. Because of this, we may view MT as a method for generating maximin designs.

SAS can be used to construct S and U optimal space-filling designs from finite candidate sets (SAS Institute, 2010). SAS constructs these designs by generating an initial design and then improving it. Options for generating an initial design include sequential and random search while two options for improving the initial design are FA and MF (which SAS Institute (2010) says usually give better designs than the other methods). Thus, SAS uses the same algorithms as KS-MF and MT. Moreover, since the S and U optimality criteria are modifications of the maximin and minimax criteria respectively, the S and U optimal designs produced by SAS are likely to be similar to the designs obtained with MT and KS-MF. As with Procedure C, high memory usage and a large amount of computation is incurred when MF or FA is used to optimize the minimax criterion (e.g., KS-MF) for large N. The former problem is due to storage of ψ. The latter problem arises from the large number of possible exchanges of a design point with a nondesign point, and the computation of the minimax criterion. If ψ cannot be stored in memory, its elements must be recomputed as needed, causing further increases in amount of computation. SAS User Guide (SAS Institute, 2010) states that: “the U-optimality criterion can be ***very*** difficult to optimize, especially if the matrix of all pairwise

distances between candidate points does not fit in memory." However, as shown in Example 5.3, Procedure C can outperform KS-MF significantly both in terms of design distance (the minimax criterion) and time needed. Moreover, Examples 5.1-5.3 suggest that the improvements in KS-MF designs over KS designs decrease and the performance of KS-MF designs relative to minimax or near-minimax designs deteriorates as N increases.

E.3 Description of Variables for Examples 5.1 and 5.3

Example 5.1: The canyon height is the average height of buildings along streets, the canyon ratio is the canyon height divided by the average street width, the vegetation area fraction is the percentage of total horizontal area covered by vegetation, and built-up area fraction is the percentage of total horizontal area covered by the roofs of buildings (see Erell et al. (2010)).

Example 5.3: see Table E.1.

Table E.1: Description of Variables in Forest Fire Dataset

Variable	Description
FFMC	Fine Fuel Moisture Code (represents the moisture content of surface litter)
DMC	Duff Moisture Code (represents the moisture content of shallow organic layers)
DC	Drought Code (represents the moisture content of deep organic layers)
ISI	Initial Spread Index (score that correlates with fire velocity spread)
temp	Outside temperature in ℃
RH	Outside relative humidity in %
wind	Outside wind speed in km/h
rain	Outside rain in mm/m^2
area	Total burned area in ha

E.4 Reduced Versions of SCLP

This section gives a detailed discussion of row and column reductions of SCLP. We work with the general SCLP, defined by

General SCLP

$$\min \sum_{j=1}^{W} I_j$$

$$\text{s.t. } \boldsymbol{BI} \geq \mathbf{1},$$

$$I_j \in \{0,1\}, j = 1, \ldots, W,$$

where $\boldsymbol{B}$ is an arbitrary matrix of 0's and 1's, and $\mathbf{1}$ is a vector of 1's.

Section E.4.1 defines redundant rows and columns of the constraint matrix $\boldsymbol{B}$. Section E.4.2 gives the column redundancy elimination procedure that we use. The row redundancy elimination procedure is similar. In Chapter 5, we obtain Reduced SCLP by applying the row redundancy elimination procedure followed by the column redundancy elimination procedure. In Section E.4.3, we prove that SCLP and any reduced version of it have the same optimal objective function value, and an optimal solution to the former can be obtained from an optimal solution to the latter by setting to zero decision variables that have been removed. Note that any reduced version of SCLP is a general SCLP. Section E.4.4 reports the gains we achieve by solving Reduced SCLP instead of SCLP for Example 5.1 and Example 5.2.

In the following sections, the notation $X \subset Y$, where X and Y are sets, means that X is a *strict* subset of Y. The notation $X \subseteq Y$ means that either $X \subset Y$ or $X = Y$.

E.4.1 Redundant Rows and Columns

Let $\boldsymbol{B}$ be the constraint matrix of a general SCLP (i.e., an arbitrary matrix of 0's

and 1's), and B_{ij} denote the (i, j) element of the matrix. Define $C_k = \{i: B_{ik} = 1\}$ and $R_i = \{k: B_{ik} = 1\}$. Then, column k of $\boldsymbol{B}$ is redundant if there exists $l \neq k$ such that $C_k \subseteq C_l$. On the other hand, row i of $\boldsymbol{B}$ is redundant if there exists $j \neq i$ such that $R_i \supseteq R_j$.

E.4.2 Column Redundancy Elimination Procedure

1. Set $CI = \{1, \dots, W\}$, $i = 1$, and $n = W$.
2. If $i < n$, set $k = i + 1$. Otherwise, **stop**.
3. Label the elements in CI as $j_1 < j_2 < \cdots < j_n$. Check the following conditions:
   i) If $C_{j_i} \subset C_{j_k}$, then set $CI = CI \backslash \{j_i\}$, $n = n - 1$ and go to Step 2.
   ii) If $C_{j_k} \subseteq C_{j_i}$, then set $CI = CI \backslash \{j_k\}$, $n = n - 1$ and go to Step 4.
   iii) If neither i) nor ii) holds, set $k = k + 1$ and go to Step 4.
4. If $k > n$, set $i = i + 1$ and go to Step 2. Otherwise, return to Step 3.

Let $CI = \{j_1, \dots, j_{W_c}\}$ be the remaining set of columns. Then, CI cannot be further reduced because j_i have been compared to $j_{i+1}, \dots, j_{W_c}$ for all $i = 1, \dots, W_c - 1$.

E.4.3 Relationship between SCLP and Reduced Versions of SCLP

Definition: A row reduced version of a general SCLP is obtained from the general SCLP by removing some redundant rows of the constraint matrix. A column reduced version of a general SCLP is obtained from the general SCLP by removing some redundant columns of the constraint matrix and corresponding decision variables.

Proposition E.8: Let Program R be any general SCLP that has a feasible solution. Then, both Program R and a row reduced version of it have the same set of feasible solutions.

Proof: Let $\boldsymbol{B}$ denote the constraint matrix of Program R and let $\widetilde{\boldsymbol{B}}$ denote the constraint matrix of the row reduced version of it. Let $\boldsymbol{I}$ denote a binary W-vector. In reducing $\boldsymbol{B}$ to $\widetilde{\boldsymbol{B}}$, only redundant rows are removed. Thus, $\boldsymbol{BI} \geq \mathbf{1}$ if and only if $\widetilde{\boldsymbol{B}}\boldsymbol{I} \geq \mathbf{1}$. In other words, $\boldsymbol{I}$ is a feasible solution of Program R if and only if it is a feasible solution of the row reduced version of Program R.

Proposition E.9: Let Program R be any general SCLP that has a feasible solution. Then, both Program R and a column reduced version of it have the same optimal objective function values. If $\tilde{\boldsymbol{I}}_A^*$ is the W- vector obtained from an optimal solution $\tilde{\boldsymbol{I}}^*$ of the column reduced SCLP by setting to zero decision variables that have been removed, then $\tilde{\boldsymbol{I}}_A^*$ is an optimal solution to Program R.

Proof: Let $\boldsymbol{B}$ denote the constraint matrix of Program R and let $\widetilde{\boldsymbol{B}}$ denote the constraint matrix of the column reduced version of it. Suppose that $\boldsymbol{I}^* = (1, \dots, 1, 0, \dots, 0)$ is an optimal solution of Program R, where the first n components equal 1 and the rest of the components equal 0. In reducing $\boldsymbol{B}$ to $\widetilde{\boldsymbol{B}}$, only redundant columns are removed. Let the indices of nonredundant columns of $\boldsymbol{B}$ be denoted by $j_1, \dots, j_{W_c}$. Construct a new binary W_c-vector $\boldsymbol{I}_R^*$ from $\boldsymbol{I}^*$ as follows: First, set the ith component of $\boldsymbol{I}_R^*$ equal to the $j_i th$ component of $\boldsymbol{I}^*$ for all i. For each column $l \in \{1, \dots, n\}$ that is removed from $\boldsymbol{B}$, choose a $j_{\tau(l)} \neq l$ such that the indices of 1's in column l is a subset of the indices of 1's in column $j_{\tau(l)}$. Change the $\tau(l)th$ component of $\boldsymbol{I}_R^*$ to 1.

By construction, $\|\boldsymbol{I}_R^*\|_1 \leq \|\boldsymbol{I}^*\|_1$**.** Moreover, since $\boldsymbol{BI}^* \geq \mathbf{1}$, we must have $\widetilde{\boldsymbol{B}}\boldsymbol{I}_R^* \geq \mathbf{1}$. Thus, the column reduced version of Program R is feasible and $\left\|\tilde{\boldsymbol{I}}^*\right\|_1 \leq \|\boldsymbol{I}_R^*\|_1 \leq \|\boldsymbol{I}^*\|_1$, where $\tilde{\boldsymbol{I}}^*$ is an optimal solution of the reduced program.

Now, it is clear that if $\tilde{\boldsymbol{B}}\tilde{\boldsymbol{I}} \geq \mathbf{1}$, where $\tilde{\boldsymbol{I}}$ is a binary W_c-vector, the W-vector $\tilde{\boldsymbol{I}}_A$ obtained from $\tilde{\boldsymbol{I}}$ by adding 0's must satisfy $\boldsymbol{B}\tilde{\boldsymbol{I}}_A \geq \mathbf{1}$. Thus, $\tilde{\boldsymbol{I}}_A^*$ (which is defined in the statement of the proposition) is a feasible solution to Program R and $\left\|\tilde{\boldsymbol{I}}_A^*\right\|_1 \geq \|\boldsymbol{I}^*\|_1$. On the other, by construction of $\tilde{\boldsymbol{I}}_A^*$ and the inequality $\left\|\tilde{\boldsymbol{I}}^*\right\|_1 \leq \|\boldsymbol{I}^*\|_1$ established in the previous paragraph, we have $\left\|\tilde{\boldsymbol{I}}_A^*\right\|_1 = \left\|\tilde{\boldsymbol{I}}^*\right\|_1 \leq \|\boldsymbol{I}^*\|_1$. We conclude that $\left\|\tilde{\boldsymbol{I}}_A^*\right\|_1 = \left\|\tilde{\boldsymbol{I}}^*\right\|_1 = \|\boldsymbol{I}^*\|_1$.

Thus, both Program R and the column reduced version of it have the same optimal objective function values and any optimal solution $\tilde{\boldsymbol{I}}^*$ to the column reduced version gives an optimal solution $\tilde{\boldsymbol{I}}_A^*$ to Program R.

Proposition E.10: Both SCLP and a reduced version of it (as defined in Chapter 5) have the same optimal objective function values. If $\tilde{\boldsymbol{I}}_A^*$ is the N-vector obtained from an optimal solution $\tilde{\boldsymbol{I}}^*$ of the reduced version of SCLP by setting to zero decision variables that have been removed, then $\tilde{\boldsymbol{I}}_A^*$ is an optimal solution to SCLP.

Proof: This follows from Proposition E.8 and Proposition E.9.

E.4.4 Sizes of Reduced SCLP and SCLP for Examples in Chapter 5

In Example 5.1, SCLP needs to be solved with a branch-and-bound algorithm a total of 85 times for $[L, U] = [0.4,1.6]$. Up to 80% of the 80 columns and 72.5% of the 80 rows of the constraint matrix are removed in the 85 instances of SCLP that are solved. The average number of rows of Reduced SCLP is 39.7 and the average number of columns is 30.9.

In Example 5.2, SCLP needs to be solved with a branch-and-bound algorithm a

total of 67 times for $[L, U] = [2.5,2.8]$. Up to 42% of the 200 columns and 27.5% of the 200 rows of the constraint matrix are removed in the 67 instances of SCLP that are solved. The average number of rows of Reduced SCLP is 154.3 and the average number of columns is 140.1.

REFERENCES

Anderson-Cook, C.M. and Robinson, T.J. (2009). "A Designed Screening Study with Prespecified Combinations of Factor Settings," *Quality Engineering*, 21(4), 392-404.

Apley, D.W., Liu, J., and Chen, W. (2006). "Understanding the Effects of Model Uncertainty in Robust Design with Computer Experiments," *Journal of Mechanical Design*, 128, 945-958.

Baldessari, B. (1967). "The Distribution of a Quadratic Form of Normal Random Variables," *The Annals of Mathematical Statistics*, 38(6), 1700-1704.

Bandelt, H.J., Maas, A., and Spieksma, F.C.R. (2004). "Local Search Heuristics for Multi-Index Assignment Problems with Decomposable Costs," *Journal of the Operational Research Society*, 55, 694-704.

Barnard, J., R. McCulloch, and X. Meng. (2000). "Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage," *Statistica Sinica*, 10, 1281-1311.

Bawaneh, M.A. (2007). "Determination of Material Constitutive Models Using Orthogonal Machining Tests," Ph.D. Dissertation, Wichita State University, Dept. of Industrial and Manufacturing Engineering.

Bayarri, M.J., Berger, J.O., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J., Lin, C., and Tu, J. (2007). "A Framework for Validation of Computer Models," *Technometrics*, 49(2), 138-154.

Beltran, L.A. (2006). "Nonparametric Multivariate Statistical Process Control Using Principal Component Analysis and Simplicial Depth," PhD Dissertation, University of Central Florida, Dept. of Industrial and Management Systems.

Bersimis, S., S. Psarakis, and J. Panaretos. (2007). "Multivariate Statistical Process Control Charts: An Overview," *Quality and Reliability Engineering International*, 23, 517-543.

Bertsimas, D. and Tsitsiklis, J.N. (1997). *Introduction to Linear Optimization*, Belmont, MA: Athena Scientific.

Bingham, D., and Goh, J. (2012). "Bayesian Variable Selection for Split-Plot Designs with Complex Aliasing," Unpublished Manuscript.

Bingham, D., and Sitter, R. (2003). "Fractional Factorial Split-Plot Designs for Robust Parameter Experiments," *Technometrics*, 45(1), 80-89.

Bisgaard, S., and Sutherland, M. (2003). "Split Plot Experiments: Taguchi's Ina Tile Experiment Reanalyzed," *Quality Engineering*, 16(1), 157-164.

Bixby, R.E. (2002). "Solving Real-World Linear Programs: A Decade and More of Progress," *Operations Research*, 50(1), 3-15.

Box, G.E.P. (1954). "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification," *The Annals of Mathematical Statistics*, 25(2), 290-302.

Box, G.E.P. (1993). "Sequential Experimentation and Sequential Assembly of Designs," *Quality Engineering*, 5(2), 321-330.

Box, G.E.P. and Jones, S. (2001). "Split-plot Designs for Robust Product and Process Experimentation," *Quality Engineering*, 13(1), 127-134.

Box, G.E.P. and Meyer, R.D. (1993). "Finding the Active Factors in Fractionated Screening Experiments," *Journal of Quality Technology*, 25(2), 94-105.

Box, G.E.P., and G.C. Tiao. (1973). *Bayesian Inference in Statistical Analysis*, Reading, MA: Addison-Wesley.

Box, G.E.P., Hunter, W.G., Hunter, J.S. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, New York: Wiley.

Butler, R.W. (2007). *Saddlepoint Approximations with Applications*, Cambridge, U.K.: Cambridge University Press.

Capizzi, G. and Masarotto, G. (2011). "A Least Angle Regression Control Chart for Multidimensional Data," *Technometrics*, 53(2), 285-296.

Caprara, A, Fischetti, M., and Toth, P. (1999). "A Heuristic Method for the Set Covering Problem," *Operations Research*, 47(5), 730-743.

Caprara, A. and Toth, P. (2000). "Algorithms for the Set Covering Problem," *Annals of Operations Research*, 98, 353-371.

Chan, K.C. and Linn, R.J. (1998). "A Grouping Method for Selective Assembly of Parts of Dissimilar Distributions," *Quality Engineering*, 11(2), 221-234.

Chang, P.B., Williams, B.J., Bhalla, K.S.B., Belknap, T.W., Santner, T.J., Notz, W.I., and Bartel, D.L. (2001). "Design and Analysis of Robust Total Joint Replacements: Finite Element Model Experiments with Environmental Variables," *Journal of Biomechanical Engineering*, 123, 239-246.

Chang, P.B., Williams, B.J., Notz, W.I., Santner, T.J., and Bartel, D.L. (1999). "Robust Optimization of Total Joint Replacements Incorporating Environmental Variables," *Journal of Biomechanical Engineering*, 121, 304-310.

Chen, P., and Koc, M. (2007). "Simulation of Springback Variation in Forming of Advanced High Strength Steels," *Journal of Materials Processing Technology*, 190, 189-198.

Chen, V.C.P., Tsui, K., Barton, R.R., and Meckesheimer, M. (2006). "A Review on Design, Modeling and Applications of Computer Experiments," *IIE Transactions*, 38(4), 273-291.

Chipman, H. (1998). "Handling Uncertainty in Analysis of Robust Design Experiments," *Journal of Quality Technology*, 30(1), 11-17.

Chipman, H., Hamada, M., and Wu, C.F.J. (1997). "A Bayesian Variable-Selection Approach for Analyzing Designed Experiments with Complex Aliasing," *Technometrics*, 39(4), 372-381.

Chvatal, V. (1979). "A Greedy Heuristic for the Set-Covering Problem," *Mathematics of Operations Research*, 4(3), 233-235.

Conti, S., Gosling, J.P., Oakley, J.E., and O'Hagan, A. (2009). "Gaussian Process Emulation of Dynamic Computer Codes," *Biometrika*, 96(3), 663-676.

Cook, D.R. and Nachtsheim, C.J. (1980). "A Comparison of Algorithms for Constructing Exact D-Optimal Designs," *Technometrics*, 22(3), 315-324.

Cortex, P. and Morais, A. (2007). "A Data Mining Approach to Predict Forest Fires using Meteorological Data." In Neves, J. , Santos, M. F. and Machado, J. Eds., *New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, December, Guimarães, Portugal*, 512-523.

Coullard, C.R., Gamble, A.B., and Jones, P.C. (1998). "Matching Problems in Selective Assembly Operations," *Annals of Operations Research*, 76, 95-107.

Current, J., Daskin, M., and Schilling, D. (2002). "Discrete Network Location Models," in *Facility Location: Applications and Theory*, eds. Drezner, Z. and Hamacher, H.W., Berlin: Springer-Verlag, pp. 81-118.

Daniels, H.E. (1987). "Tail Probability Approximations," *International Statistical Review*, 55(1), 37-48.

Dantzig, G.B. and Thapa, M.N. (2003). *Linear Programming 2: Theory and Extensions*, New York: Springer-Verlag.

Das, N., and V. Prakash. (2008). "Interpreting the Out-of-Control Signal in Multivariate Control Chart – a Comparative Study," *International Journal of Advanced Manufacturing Technology*, 37, 966-979.

Doganaksoy, N., F.W. Faltin, W.T. Tucker. (1991). "Identification of Out-of-Control Quality Characteristics in a Multivariate Manufacturing Environment," *Communications in Statistics – Theory and Methods*, 20(9), 2775-2790.

Draguljić, D., Santner, J., and Dean, A.M. (2012). "Noncollapsing Space-Filling Designs for Bounded Nonrectangular Regions," *Technometrics*, 54(2), 169-178.

Eischen, J.W. (1989). "Geometric Nonlinearity in a Bimaterial Strip," *Proceedings of the International Congress of Applied Mechanics*, Beijing, P.R.C., 1-7.

Erell, E., Pearlmutter, D., and Williamson, T. (2010). *Urban Microclimate: Designing the Spaces Between Buildings*, London: Earthscan.

Fang, X.D. and Zhang, Y. (1995). "A New Algorithm for Minimizing the Surplus Parts in Selective Assembly," *Computers and Industrial Engineering*, 28(2), 341-350.

Feng, C. and Kusiak, A. (1997). "Robust Tolerance Design with the Integer Programming Approach," *Journal of Manufacturing Science and Engineering*, 119, 603-610.

Feuerverger, A. and A.C.M. Wong. (2000). "Computation of Value-at-Risk for Nonlinear Portfolios," *Journal of Risk*, 3(1), 37-55.

Forouraghi, B. (2002). "Worst-Case Tolerance Design and Quality Assurance via Genetic Algorithms," *Journal of Optimization Theory and Applications*, 113(2), 251-268.

Fuchs, C. and Kenett, R. (1998). *Multivariate Quality Control*, New York: Marcel Dekker.

Furnival, G.M. and Wilson, R.W. (1974). "Regressions by Leaps and Bounds," *Technometrics*, 16(1), 499-511.

Gelfand, A.E. and A.F.M. Smith. (1990). "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85(410), 398-409.

Gelfand, A.E., Hills, S.E., Racine-Poon, A., Smith, A.F.M. (1990). "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling," *Journal of the American Statistical Association*, 85(412), 972-985.

George, E.I., and R.E. McCulloch. (1993). "Variable Selection Via Gibbs Sampling," *Journal of the American Statistical Association*, 88(423), 881-889.

George, E.I., and R.E. McCulloch. (1997). "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7, 339-373.

Giesbrecht, F.G. and M.L. Gumpertz. (2004). *Planning, Construction, and Statistical Analysis of Comparative Experiments*, New York: Wiley.

Gilbert, K.C. and Hofstra, R.B. (1988). "Multidimensional Assignment Problems," *Decision Sciences*, 19(2), 306-321.

Gilmour, S.G. and P. Goos. (2009). "Analysis of Data from Non-Orthogonal Multistratum Designs in Industrial Experiments," *Applied Statistics*, 58(4), 467-484.

Ginsburg, H., and Ben-Gal, I. (2006). "Designing Experiments for Robust-Optimization Problems: the V_S-Optimality Criterion," *IIE Transactions*, 38, 445-461.

Goos, P. (2002). *The Optimal Design of Blocked and Split-Plot Experiments*. New York: Springer-Verlag.

Goos, P., Langhans, I., and Martina, V. (2006). "Practical Inference from Industrial Split-Plot Designs," *Journal of Quality Technology*, 38(2), 162-179.

Grossman, T. and Wool, A. (1997). "Computational Experience with Approximation Algorithms for the Set Covering Problem," *European Journal of Operational Research*, 101, 81-92.

Hakimi, S.L. (1965). "Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph," *Operations Research*, 12(3), 450-459.

Haley, K.B. (1963). "The Multi-Index Problem," *Operations Research*, 11(3), 368-379.

Handcock, M.S., and Stein, M.L. (1993). "A Bayesian Analysis of Kriging," *Technometrics*, 35(4), 403-410.

Hawkins, D.M. (1991). "Multivariate Quality Control Based on Regression-Adjusted Variables," *Technometrics*, 33(1), 61-75.

Heaton, M.J. and Scott, J.G. (2010). "Bayesian Computation and the Linear Model," in *Frontiers of Statistical Decision Making and Bayesian Analysis in Honor of James O. Berger*, ed. Chen, M.H., Dey, D.K., Müller, P., Sun, D., and Ye, K., New York: Springer.

Hill, B.M. (1965). "Inference about Variance Components in the One-Way Model," *Journal of the American Statistical Association*, 60(311), 806-825.

Hochbaum, D.S. (1982). “Approximation Algorithms for the Set Covering and Vertex Cover Problems,” *SIAM Journal of Computing*, 11(3), 555-556.

Imhof, J.P. (1961). “Computing the Distribution of Quadratic Forms in Normal Variables,” *Biometrika*, 48(3/4), 419-426.

Iwata,S., Matsui, T., and McCormick, S.T. (1998). “A Fast Bipartite Network Flow Algorithm for Selective Assembly,” *Operations Research Letters*, 22, 137-143.

Jeang, A. and Chang C.L. (2002). “Combined Robust Parameter and Tolerance Design Using Orthogonal Arrays,” *International Journal of Advanced Manufacturing Technology*, 19, 442-447.

Jellison, J.L. (1975). “Effect of Surface Contamination on the Thermocompression Bondability of Gold,” *IEEE Transactions on Parts, Hybrids, and Packaging*, 11(3), 206-211.

John, P.W.M., Johnson, M.E., Moore, L.M., and Ylvisaker, D. (1995). “Minimax Distance Designs in Two-Level Factorial Experiments,” *Journal of Statistical Planning and Inference*, 44, 249-263.

Johnson, M.E., Moore, L.M., and Ylvisaker, D. (1990). “Minimax and Maximin Distance Designs,” *Journal of Statistical Planning and Inference*, 26, 131-148.

Jones, B. and Nachtsheim, C.J. (2009). “Split-Plot Designs: What, Why, and How,” *Journal of Quality Technology*, 41(4), 340-361.

Kannan, S.M., Asha, A., and Jayabalan, V. (2005). “A New Method in Selective Assembly to Minimize Clearance Variation for a Radial Assembly Using Genetic Algorithm,” *Quality Engineering*, 17(4), 595-607.

Kannan, S.M. and Jayabalan, V. (2002). “A New Grouping Method for Minimizing the Surplus Parts in Selective Assembly,” *Quality Engineering*, 14(1), 67-75.

Kannan, S.M., Jeevanantham, A.K., and Jayabalan, V. (2008). “Modeling and Analysis of Selective Assembly Using Taguchi’s Loss Function,” *International Journal of Production Research*, 46(15), 4309-4330.

Kapur, K.C., and Cho, B. (1996). “Economic Design of the Specification Region for Multiple Quality Characteristics,” *IIE Transactions*, 28(3), 237-248.

Kennard, R.W. and Stone, L.A. (1969). “Computer Aided Design of Experiments,” *Technometrics*, 11(1), 137-148.

Khuri, A.I. (2009). *Linear Model Methodology*, Boca Raton: Chapman & Hall.

Kim, J. (1999). “Optimum Design of a Heat-Exchanger-Fan Casing of Clothes Dryer Using the Taguchi Method,” *KSME International Journal*, 13(9), 962-972.

Koksoy, O., and Doganaksoy, N. (2003). “Joint Optimization of Mean and Standard Deviation Using Response Surface Methods,” *Journal of Quality Technology*, 35(3), 239-252.

Kwon, H., Kim, K., and Chandra, M.J. (1999). “An Economic Selective Assembly Procedure for Two Mating Components with Equal Variance,” *Naval Research Logistics*, 46, 809-821.

Lee, S.H. and Chen, W. (2007). “A Comparative Study of Uncertainty Propagation Methods for Black-Box Type Functions,” *Proceedings in IDETC/CIE*, Las Vegas, Nevada, 1-10.

Lee, S.H. and Kwak, B.M. (2005). “Response Surface Augmented Moment Method for Efficient Reliability Analysis,” *Structural Safety*, 28, 261-272.

Lehman, J.S., Santner, T.J., and Notz, W.I. (2004). “Designing Computer Experiments to Determine Robust Control Variables,” *Statistica Sinica*, 14, 571-590.

Lenth, R.V. (1989). “Quick and Easy Analysis of Unreplicated Factorials,” *Technometrics*, 31, 469-473.

Letsinger, J.D., Myers, R.H., and Lentner, M. (1996). “Response Surface Methods for Bi-Randomizatiaon Structures,” *Journal of Quality Technology*, 28, 381-397.

Lewis, R.M., and Torczon, V. (1999). “Pattern Search Algorithms for Bound Constrained Minimization,” *SIAM Journal of Optimization*, 9(4), 1082-1099.

Li, J., J. Jin, and J. Shi. (2008). "Causation-Based T^2 Decomposition for Multivariate Process Monitoring and Diagnosis," *Journal of Quality Technology*, 40(1), 46-58.

Liu, H., Tang, Y., and Zhang, H.H. (2009). "A New Chi-Square Approximation to the Distribution of Non-Negative Definite Quadratic Forms in Non-Central Normal Variables," *Computational Statistics and Data Analysis*, 53, 853-856.

Loeppky, J.L., Sacks, J., and Welch, W.J. (2009). "Choosing the Sample Size of a Computer Experiment: A Practical Guide," *Technometrics*, 51(4), 366-376.

Lugannani, R., and Rice, S. (1980). "Saddle Point Approximation for the Distribution of the Sum of Independent Random Variables," *Advances in Applied Probability*, 12, 475-490.

Madigan, D. and Raftery, A.E. (1994). "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window," *Journal of the American Statistical Association*, 89(428), 1535-1546.

Mansoor, E.M. (1961). "Selective Assembly: Its Analysis and Applications," *International Journal of Production Research*, 1(1), 13-24.

Marengo, E. and Todeschini, R. (1992). "A New Algorithm for Optimal, Distance-based Experimental Design," *Chemometrics and Intelligent Laboratory Systems*, 16, 37-44.

Mason, R.L., N.D. Tracy, and J.C. Young. (1995). "Decomposition of T^2 for Multivariate Control Chart Interpretation," *Journal of Quality Technology*, 27(1), 99-108.

Mason, R.L., N.D. Tracy, and J.C. Young. (1997). "A Practical Approach for Interpreting Multivariate T^2 Control Chart Signals," *Journal of Quality Technology*, 29(1), 396-406.

Matsura, S. and Shinozaki, N. (2007). "Optimal Binning Strategies under Squared Error Loss in Selective Assembly with Measurement Error," *Communications in Statistics – Theory and Methods*, 36(16), 2863-2876.

McKay, M.D., Beckman, R.J., and Conover, W.J. (1979). "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, 21(2), 239-245.

Mease, D., Sudjianto, A., and Nair, V.N. (2004). "Selective Assembly in Manufacturing: Statistical Issues and Optimal Binning Strategies," *Technometrics*, 46(2), 165-175.

Miller, A.C., and Rice, T.R. (1983), "Discrete Approximations of Probability Distributions," *Management Science*, 29(3), 352-362.

Miller, C. and Urban, D.L. (1999). "A Model of Surface Fire, Climate and Forest Pattern in the Sierra Nevada, California," Ecological Modelling, 114, 113-135.

Miro-Quesada, G., and Del Castillo, E. (2004). "A Bayesian Approach for Multiple Response Surface Optimization in the Presence of Noise Variables," *Journal of Applied Statistics*, 31(3), 251-270.

Montgomery, D.C. (2009). *Introduction to Statistical Quality Control* (6^{th} ed.), New York: Wiley.

Morris, M.D. and Mitchell, T.J. (1995). "Exploratory Designs for Computational Experiments," *Journal of Statistical Planning and Inference*, 43, 381-402.

Murphy, B.J. (1987). "Selecting Out of Control Variables with the T^2 Multivariate Quality Control Procedure," *The Statistician*, 36(2), 571-581.

Myers, R.H., Kim, Y., and Griffiths, K.L. (1997). "Response Surface Methods and the Use of Noise Variables," *Journal of Quality Technology*, 29(4), 429-440.

Myers, R.H., and Montgomery, D.C. (2002). *Response Surface Methodology* (2^{nd} ed.), New York: Wiley.

Næs, T., Aastveit, A.H., and Sahni, N.S. (2007). "Analysis of Split-plot Designs: An Overview and Comparison of Methods," *Quality and Reliability Engineering International*, 23(5), 801-820.

Nelson, B.L., Swann, J., Goldsman, D., and Song, W. (2001). "Simple Procedures for Selecting the Best Simulated System When the Number of Alternatives is Large," *Operations Research*, 49(6), 950-963.

O'Hagan, A. (1994). *Kendalls Advanced Theory of Statistics: Volume 2B, Bayesian Analysis*, London, U.K.: Edwards Arnold.

O'Hara, R.B. and Sillanpää, M.J. (2009). "A Review of Bayesian Variable Selection Methods: What How and Which," *Bayesian Analysis*, 4(1), 85-118.

Oakley, J.E. and O'Hagan. (2002). "Bayesian Inference for the Uncertainty Distribution of Computer Model Outputs," *Biometrika*, 89(4), 769-784.

Oakley, J.E. and O'Hagan. (2004). "Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach," *Journal of the Royal Statistical Society (Series B)*, 66(3), 751-769.

Patnaik, P.B. (1949). "The Non-Central χ^2 and F Distributions and Their Applications," *Biometrika*, 36, 202-232.

Peterson, J.J., and Kuhn, A.M. (2005). "Ridge Analysis with Noise Variables," *Technometrics*, 47(3), 274-283.

Pierskalla, W.P. (1968). "The Multidimensional Assignment Problem," *Operations Research*, 16(2), 422-431.

Pugh, G.A. (1992). "Selective Assembly with Components of Dissimilar Variance," *Computers and Industrial Engineering*, 23(4), 487-491.

Qian, P.Z.G. (2009). "Nested Latin Hypercube Designs," *Biometrika*, 96(4), 957-970.

Qian, P.Z.G., Tang, B., and Wu, C.F.J. (2009). "Nested Space-Filling Designs for Computer Experiments with Two Levels of Accuracy," *Statistica Sinica*, 19, 287-300.

Queyranne, M. and Spieksma, F.C.R. (1997). "Approximation Algorithms for Multi-Index Transportation Problems with Decomposable Costs," *Discrete Applied Mathematics*, 76, 239-253.

Rinott, Y. (1978). "On Two-Stage Selection Procedures and Related Probability-Inequalities," *Communications in Statistics – Theory and Methods*, **7**(8), 799-811.

Royle, J.A. and Nychka, D. (1998). "An Algorithm for the Construction of Spatial Coverage Designs with Implementation in SPLUS," *Computers and Geosciences*, 24(5), 479-488.

Runger, G.C., F.B. Alt, and D.C. Montgomery. (1996). "Contributors to a Multivariate Statistical Process Control Chart Signal," *Communications in Statistics – Theory and Methods*, 25(6), 2203-2213.

Sacks, J., Welch, W.J., Mitchell, T.J., and Wynn, H.P. (1989). "Design and Analysis of Computer Experiments," *Statistical Science*, 4(4), 409-435.

Santner, T.J., Williams, B.J., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. New York: Springer.

SAS Institute (2010). *SAS/QC(R) 9.2 User's Guide*, Cary, NC: SAS Institute.

Schmeiser, B. (1982). "Batch Size Effects in the Analysis of Simulation Output," *Operations Research*, 30(3), 556-568.

Schrijver, A. (1986). *Theory of Linear and Integer Programming*, New York: Wiley.

Singh, P.K., Jain, S.C., and Jain, P.K. (2005). "Advanced Optimal Tolerance Design of Mechanical Assemblies with Interrelated Dimension Chains and Process Precision Limits," *Computers in Industry*, 56(2), 179-194.

Snyder, L.V. (2011). "Covering Problems," in *Foundations of Location Analysis*. Edited by Eiselt, H.A. and Marianov, V. New York: Springer-Verlag.

Solomon, H., and Stephens, M.A. (1977). "Distribution of a Sum of Weighted Chi-Square Variables," *Journal of the American Statistical Association*, 72(360), 881-885.

Spedding, T.A. and Sun, G.Q. (1999). "Application of Discrete Event Simulation to the Activity Based Costing of Manufacturing Systems," *International Journal of Production Economics*, 58, 289-301.

Stinstra, E., Hertog, D., Stehouwer, P., and Vestjens, A. (2003). "Constrained Maximin Designs for Computer Experiments," *Technometrics*, 45(4), 340-346.

Stoer, J., and Bulirsch, R. (1993). *Introduction to Numerical Analysis* (2nd ed.), New York: Springer-Verlag.

Sun, L., Hsu, J.S.J., Guttman, I., and Leonard, T. (1996). "Bayesian Methods for Variance Component Models," *Journal of the American Statistical Association*, 91(434), 743-752.

Sun,Y., Heo, Y., Xie, H., Tan, M., Wu, C.F.J., and Augenbroe, G. (2011). "Uncertainty Quantification of Microclimate Variables in Building Energy Simulation," *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association, Sydney*, 2423-2430.

Taguchi, G. (1987). *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs*, Vols. 1&2, White Plains NY: UNIPUB/Kraus International Publications.

Tan, M.H.Y. and Wu, C.F.J. (2012). "Robust Design Optimization with Quadratic Loss Derived From Gaussian Process Models," *Technometrics*, 54(1), 51-63.

Tang, B. (1993). "Orthogonal Array-Based Latin Hypercubes," *Journal of the American Statistical Association*, 88(424), 1392-1397.

Thesen, A. and Jantayavichit, A. (1999). "Design and Evaluation of a Selective Assembly Station for High Precision Scroll Compressor Shells," *Proceedings of the 1999 Winter Simulation Conference*, Phoenix, Arizona, 694–700.

Tiao, G.C. and Tan, W.Y. (1965). "Bayesian Analysis of Random Effect Models in the Analysis of Variance. I. Posterior Distribution of Variance-Components," *Biometrika*, 52(1/2), 37-53.

Tierney, J.W. and Eischen, J.W. (1997). "Residual Stress Analysis of Bimaterial Strips Under Multiple Thermal Loading," *Journal of Electronic Packaging*, 119, 281-287.

Timm, N.H. (2002). *Applied Multivariate Analysis*, New York: Springer-Verlag.

Timoshenko, S. (1925). "Analysis of Bi-Metal Thermostats," *Journal of the Optical Society of America*, 11(3), 223-233.

Toregas, C., Swain, R., ReVelle, C., and Bergman, L. (1971). "The Location of Emergency Service Facilities," *Operations Research*, 19(6), 1363-1373.

Vanderbei, R.J. (2001). *Linear Programming: Foundations and Extensions* (2nd ed.), New York: Springer-Verlag.

Vounatsou, P. and A.F.M. Smith. (1997). "Simulation-based Bayesian Inferences for Two-Variance Components Linear Models," *Journal of Statistical Planning and Inference*, 59, 139-161.

Wang, K. and Jiang, W. (2009). "High-Dimensional Process Monitoring and Fault Isolation via Variable Selection," *Journal of Quality Technology*, 41(2), 247-258.

Williams, B.J., Santner, T.J., and Notz, W.I. (2000). "Sequential Design of Computer Experiments to Minimize Integrated Response Functions," *Statistica Sinica*, 10, 1133-1152.

Wolfinger, R.D. and R.E. Kass. (2000). "Nonconjugate Bayesian Analysis of Variance Component Models," *Biometrics*, 56, 768-774.

Wu, C., Chen, Z., and Tang, G. (1998). "Component Tolerance Design for Minimum Quality Loss and Manufacturing Cost," *Computers in Industry*, 35, 223-232.

Wu, C.F.J. and Hamada, M.S. (2009). *Experiments: Planning, Analysis, and Optimization* (2nd ed.), New York: Wiley.

Yeh, A.B., D.K.J. Lin, and McGrath, R.N. (2006). "Multivariate Control Charts for Monitoring Covariance Matrix: A Review," *Quality Technology and Quantitative Management*, 3(1), 415-436.

Zellner, A., and Chetty, V.K. (1965). "Prediction and Decision Problems in Regression Models from the Bayesian Point of View," *Journal of the American Statistical Association*, 60(310), 608-616.

Zhang, Y. and Fang, X.D. (1999). “Predict and Assure the Matchable Degree in Selective Assembly via PCI-based Tolerance,” *Journal of Manufacturing Science and Engineering*, 121(3), 494-500.

Zhang, Y., Krueger, D., Durst, R., Lee, R., Wang, D., Seeram, N., and Heber, D. (2009). “International Multidimensional Authenticity Specification (IMAS) Algorithm for Detection of Commercial Pomegranate Juice Adulteration,” *Journal of Agricultural and Food Chemistry*, 57, 2550-2557.

Zou, C., Jiang, W., and Tsung, F. (2011). “A LASSO-Based Diagnostic Framework for Multivariate Statistical Process Control,” *Technometrics*, 53(2), 297-309.

Zou, C. and Qiu, P. (2009). “Multivariate Statistical Process Control Using LASSO,” *Journal of the American Statistical Association*, 104(488), 1586-1596.