Bayesian Design and Analysis of Small Multifactor Industrial Experiments

Deepa Kapoor

A Thesis submitted for the degree of Doctor of Philosophy

School of Mathematical Sciences

Queen Mary, University of London.

May 2011

Declaration

The work presented in this thesis is the result of original research carried out by myself under the supervision of Steven Gilmour, and has not been submitted for any other degree or award in any other university or educational establishment.

Signed:

Date:

For my mother, Nishi.

Acknowledgements

I would like to thank my supervisor Prof. Steven Gilmour, to whom I am immensely grateful for his academic guidance, invaluable support and suggestions over the course of my research and writing of this thesis. I also sincerely appreciate his prompt and critical review of my writing and his continual willingness to answer any questions or doubts that I have had throughout the PhD process.

This work would not have been possible without the CASE studentship provided by EPSRC and Pfizer Global R & D. I am most thankful to my industrial supervisor Phil Woodward for all his help and insightful discussions, and also to Debbie Kraus and Mark Whitlock for their time and assistance during my time spent at Pfizer. I also wish to thank my colleagues at the School of Mathematical Sciences for providing a pleasant working environment, and am very grateful to Maria Roopa Thomas for her friendship, providing welcome distractions and for helping make life in Queen Mary enjoyable and sociable.

I must express my gratitude to my friends and family for their enduring love, support and patience, and in particular my sisters Puja and Lakshmi who have continually believed in me and taught me to never give up. To my grandfather Satya Pal Chadha for teaching me that there is no substitute for honest, hard work and to always strive for the best. To Gaj for providing me with the necessary help, focus and determination

in order to see this PhD through to completion, for the welcome escape outside of the PhD, and without whom I could not have managed. Finally, I must acknowledge my mother Nishi, to whom I am indebted for her constant calming motherly care, belief in my ability and strength that has enabled me to overcome the struggles and difficulties that I have faced both personally and academically. Without them this thesis would not have been able to be completed.

Abstract

Unreplicated two level fractional factorial designs are a common type of experimental design used in the early stages of industrial experimentation. They allow considerable information about the effects of several factors on the response to be obtained with a relatively small number of runs.

The aim of this thesis is to improve the guidance available to experimenters in choosing a good design and analysing data. This is particularly important when there is commercial pressure to minimise the size of the experiment.

A design is usually chosen based on optimality, either in terms of a variance criterion or estimability criteria such as resolution. This is given the number of factors, number of levels of each factor and number of runs available. A decision theory approach is explored, which allows a more informed choice of design to be made. Prior distributions on the sizes of effects are taken into consideration, and then a design chosen from a candidate set of designs using a utility function relevant to the objectives of the experiment. Comparisons of the decision theoretic methods with simple rules of thumb are made to determine when the more complex approach is necessary.

Fully Bayesian methods are rarely used in multifactor experiments. However there is virtually always some prior knowledge about the sizes of effects and so using this in a Bayesian data analysis seems natural. Vague and more informative priors are

explored.

The analysis of this type of experiment can be impacted in a disastrous way in the presence of outliers. An analysis that is robust to outliers is sought by applying different model distributions of the data and prior assumptions on the parameters. Results obtained are compared with those from standard analyses to assess the benefits of the Bayesian analysis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivations for Thesis

The aim of this thesis is to improve the guidance available to experimenters in industry in choosing a good design and in analysing data. This will be done by applying Bayesian methods when there is prior knowledge about the sizes of effects that can be incorporated at the analysis stage. Decision theory will be used at the design stage and will also use prior knowledge in a formal way. This research is particularly important in circumstances when there is commercial pressure to minimise the size of the experiment.

In this thesis, the industrial experimentation discussed will be both in general terms, and in reference to the non-clinical statistical work carried out at Pfizer Global R & D which is the major motivation for this work. In this chapter, an overview of the history of industrial experimentation will be given as well as an introduction to factorial and fractional factorial experiments, decision theory and Bayesian methodology along with concepts, terminology and notation which shall be used throughout the thesis.

An outline of what will appear in each of the chapters is also given.

1.2 History of Industrial Experimentation

Experimentation forms a major part of scientific research and is a process that allows an experimenter to discover what happens to an output or *response* when the settings of the input variables are changed. Experimental design can be referred to as the general body of knowledge, encompassing techniques, concepts and terms, that an experimenter will use to carry out experiments more efficiently. This is from the initial design stage to the implementation stage of the experiment right through to the analysis of the data. Experimental design in the modern day can be traced back to the notable work carried out by R. A. Fisher at Rothamsted Experimental Station located in Hertfordshire, England where he began working in 1919. This work later culminated in his significant publications *Statistical Methods for Research Workers* (1925) and *The Design of Experiments* (1935) which became standard points of reference for both scientists and statisticians alike. Fisher's work at Rothamsted was concerned with problems in an agricultural setting and it is in that same background that the use of the more efficient factorial designs rather than the onerous change one factor at a time approach were first seen with practical applications. Significant theory and applications were introduced by Fisher (1935), Yates (1937) and Finney (1945), all contributions of which were also motivated by problems in agriculture and biology. The benefits of being able to vary levels of factors and consider all the possible treatment level combinations simultaneously were immediately seen and this work was then later extended to industrial experimentation.

The nature of agricultural experimentation means that the experiments tend to be large in scale, lengthy and have to deal with issues such as field and soil variation,

which resulted in the development of concepts and techniques such as blocking, randomisation, fractional factorial designs and the use of analysis of variance. In comparison, industrial experiments tend to take less time but each experimental run can be costly, placing the need on the experiment to be much smaller in size. After applying previous techniques developed from agriculture to experiments in the industrial setting, it was soon realised that further techniques and concepts had to be developed to overcome the unique features of industrial experiments. Soon after World War II G. E. P. Box and co-workers at Imperial Chemical Industries notably lead the way in introducing experimental design methods specific to the chemical and process industry and *The Design and Analysis of Industrial Experiments* (1960) was published as a result. New techniques presented dealt with process modelling and optimization, where previously focus had been placed upon treatment comparisons, the primary objective of agricultural experiments. Sequential tests were also discussed due to time and cost considerations mentioned previously, and lead to the introduction of new techniques for the planning of experiments such as central composite designs and optimal designs, the analysis of which rely greatly upon regression modelling and graphical methods. Plackett-Burman designs (1946) were introduced as another way of dealing with run size economic considerations. However, these saturated designs used for screening a large number of factors are useful when only main effects are of interest and interactions can be assumed to be negligible. Despite these developments made, the principles of experimental design were not at first as widely used in the Western world in comparison to Japan. Japanese industry has long recognised the importance of experimental design methods in the efficient development of products in industry and explains much of its success.

There has recently been a resurgence of interest in experimental design in industry in the United States and United Kingdom over the last 20 years or so and is largely thanks to the work from the early 1950s of the Japanese engineer, G. Taguchi, several publications of which have gone on to be translated into English. Taguchi's approach

of robust parameter design was to improve a system (a product or process) in order to make it less sensitive to 'hard to control' factors. Many of the designs recommended by Taguchi call for the run size to be minimized in order to consequently minimize the cost of the experiment. This is such as regular fractional factorial designs and the recommended reduced-run arrays which are similar to the saturated designs developed by Plackett and Burman.

More recently, a significant contribution to the field of experimental design has been provided by Wu and Hamada (2000). They realised that many new methodological developments had been made since the publication of *Statistics for Experimenters* by Box, Hunter and Hunter in 1978 and wanted to introduce these developments to a greater audience. Some of the methodologies covered include robust parameter design, the more widespread use of the minimum aberration criterion and designs with complex aliasing considering both a frequentist and Bayesian analysis. The analysis strategy for designs with complex aliasing presented was based on previous research conducted by Hamada and Wu (1992), and the Bayesian variable selection strategy also for designs with complex aliasing presented was based upon the research of Chipman, Hamada and Wu (1997). Their book has also been successful in presenting experimental methodologies that originated much earlier than the publication by Box, Hunter and Hunter (1978), which did not receive much attention in previous publications. Some of these methodologies that have been presented in greater detail in this book are Plackett-Burman designs, nonregular designs and the construction of orthogonal main-effect plans through collapsing factors.

1.3 Industrial Experiments at Pfizer Global R & D

Pfizer are a global research-based pharmaceutical company that primarily carry out research, development and manufacturing of drugs at a number of sites worldwide. The types of experiments used in non-clinical research include *screening experiments*, a filtering process carried out at the initial stages of an investigation where a large number of factors that are thought to potentially have some effect on the process in hand are investigated and the few found to be 'active' are then focussed on in further experimentation. Active factors are typically considered to be those factors that have large effects in relation to noise and so are an explanation for process variation that occurs. Of the many factors being investigated only a few are thought to be truly important which is described as the condition of *factor sparsity* (Box and Meyer, 1986a and 1986b).

Also carried out are experiments concerned with optimising a process, at the later stages of an investigation when the few true 'active' factors have been identified and the correct combination of levels of the factors are sought to optimise the process in hand. This will typically be to optimise the yield, and thus minimise the loss and also minimise cost, all aspects which are of importance to the company. An example of such an experiment is in a blending process where a drug is to be manufactured in a large scale blender which has three different variables, i.e. factors, that can be adjusted; pre-blend speed, final blend speed and final blend time. Each setting has a low and a high level and the amount of drug yield produced varies as a direct result of the combination of settings used by the blender, and so it is necessary to identify the optimum combination of factor levels in order to maximise the amount of the drug produced.

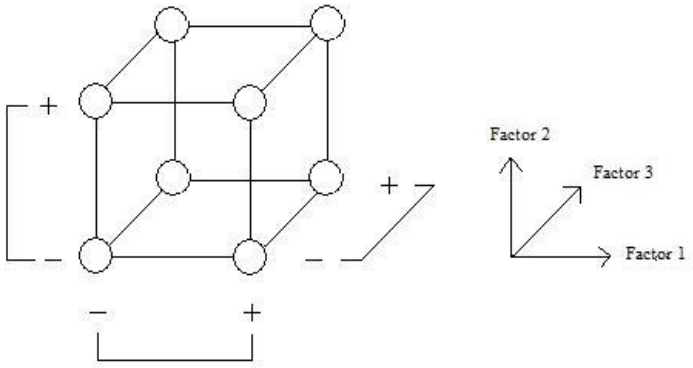
This thesis will focus on both the screening and optimization stages of non-clinical pharmaceutical research. The research will focus on providing guidelines to experimenters regarding strategies to choose designs with respect to screening when a 'good' design is where maximum information is exploited in terms of true active and inactive factors being correctly identified. Strategies will also look at how designs perform in the context of optimization, the purpose of which is to correctly identify the optimum combination of levels of factors, and this is taking into consideration prior knowledge the experimenter has regarding the process. Research is also carried out into the analysis of data from fractional factorial experiments using Bayesian methods.

1.4 Concepts and Terminology

1.4.1 Factorial Experiments

A thorough explanation of full 2^k factorial designs was given by Box and Hunter (1961a, 1961b), detailing that all combinations of two levels for each of k variables comprised a full 2^k factorial design. The variables in question may be continuous, thus the two levels being a low and high level of the variable (e.g. low or high temperature of filtration); alternatively the variable may be qualitative, the two versions corresponding to two types of the variable (e.g. new or old filter cloth). Two versions of notation expressing the runs and variables comprising the experimental design were also set out (Table 1.1). The first notation identified the variables by capital letters, the two versions of the variable denoted by the presence or absence of the corresponding lower case letter. The second notation, more commonly used throughout the literature, sees the variables denoted by numbers and the two versions denoted by a minus and a plus sign, or by minus and plus one. Although only the two sets of notation were set out in this literature and Table 1.1, another common notation expressing the runs and variables is a mixture of the two where variables are denoted by capital letters and the two versions denoted by a minus and a plus sign, or by minus and plus one. The more common usage of these latter two notations may be due to the ability to view the experimental design geometrically (Figure 1.1) and its natural extension to constructing fractional factorial designs which will be looked at later. The set of levels of all factors given for each run thus indicates the conditions under which that particular trial must be carried out and is called the *treatment combination*. When using the latter notation of minus and plus signs, the list of experimental runs is called the *design matrix*. However, the term design matrix is sometimes used to denote the matrix including all treatment combinations and with the first column as the column of all plus signs to denote the intercept. This matrix shall be denoted

Figure 1.1: Geometric Representation of the 2^3 Factorial Design



by **X** and the derivation of this is described later in this section. In the case of a full factorial design, where the model includes all main effects and interaction effects, **X** will be a $2^k \times 2^k$ matrix. Throughout this thesis I shall refer to the design matrix as the matrix **X** which will represent the matrix with all treatment combinations and has the first column of all plus signs for the intercept included, when considering a first-order model. The runs in Table 1.1 are listed in standard order, where for each of the k variables, the k^{th} column has 2^{k-1} alternating minus and plus signs.

Table 1.1: Alternative Notations for the 2^3 Factorial Design

Run Number	Notation 1	Notation 2		
	Variables ABC	Variables **1 2 3**		
1	1	-	-	-
2	a	+	-	-
3	b	-	+	-
4	ab	+	+	-
5	c	-	-	+
6	ac	+	-	+
7	bc	-	+	+
8	abc	+	+	+

A method for calculating the estimates for the main effects and interaction effects was given by Yates (1937). Yates's algorithm, a systematic tabular method, involves the addition and subtraction of successive pairs in a column, beginning with the column of observations y listed in standard order. After writing down the treatment combinations as given in notation 1 and the responses in standard order, the first entry of the first column is derived by adding the sum of the first two responses. The second entry is the sum of the second pair of responses, i.e. observations 3 and 4. This is repeated until all pairs of the observation column are used and the top half of the new column created. The lower half is then derived by taking the differences of these pairs, the first entry in the lower half being the difference of response 2 and response 1, the next entry the difference of response 3 and response 4 and so on. The second new column is then derived from column 1 using the same addition and subtraction process and the overall process is illustrated in Table 1.2 for clarity. For a 2^k design, this process of creating new columns is repeated up until there are k columns, excluding the original column of observations y. For the final k^{th} column, the first entry is divided by N, the total number of observations, to obtain the grand mean and all other entries are divided by $\frac{N}{2}$ to obtain the size of the effect for the corresponding treatment or the *data contrast*. Although this algorithm provides a relatively straightforward approach to the calculation of effects, it is not without its drawbacks and can be prone to numerical errors.

Table 1.2: Calculation of Effects using Yates's Algorithm

Run	Treatment Combination	y	Column 1	Column 2	Column 3	Effects
1	(1)	2	12	32	56	$\bar{y} = 7.0$
2	a	10	20	24	12	$A = 3.0$
3	b	8	14	12	4	$B = 1.0$
4	ac	12	10	0	-8	AC = -2.0
5	c	6	8	8	-8	C = -2.0
6	ac	8	4	-4	-12	AC = -3.0
7	bc	6	2	-4	-12	BC = -3.0
8	abc	4	-2	-4	0	$ABC = 0$

An alternative to Yates's method was given by Box and Hunter and provided a longer, but more basic method. For a 2^k design, a $2^k \times 2^k$ matrix, $\mathbf{X}$, is derived from the design matrix (Table 1.3). Each of the k columns corresponds to the main and interaction effects, and each of the k rows corresponds to an experimental run. The ij interaction column of $\mathbf{X}$ is generated by multiplying the elements of the i and j columns of the design matrix, and the ijk interaction similarly calculated. The first column, called the *identity* and denoted **I** will consist of all plus signs and is used to calculate the estimated mean. The estimate of an interaction effect, say **ij**, is calculated by taking the product of the elements of the response column y and the corresponding elements of the ij column, and then taking the sum of the resulting elements. This final value is then divided by $\frac{N}{2}$ to give the estimated effect. It should be noted that these methods only work for orthogonal designs. To illustrate, the **2 3** interaction effect for the data given in Table 1.3 is

$$\mathbf{2 \quad 3} = \frac{1}{4}(2 + 10 - 8 - 12 - 6 - 8 + 6 + 4) = -3.0.$$

Table 1.3: Design Matrix

Design Matrix			X								y
1	**2**	**3**	**I**	**1**	**2**	**3**	**12**	**13**	**23**	**123**	**y**
-	-	-	+	-	-	-	+	+	+	-	2
+	-	-	+	+	-	-	-	-	+	+	10
-	+	-	+	-	+	-	-	+	-	+	8
+	+	-	+	+	+	-	+	-	-	-	12
-	-	+	+	-	-	+	+	-	-	+	6
+	-	+	+	+	-	+	-	+	-	-	8
-	+	+	+	-	+	+	-	-	+	-	6
+	+	+	+	+	+	+	+	+	+	+	4

1.4.2 Fractional Factorial Experiments

The number of runs required for a full factorial 2^k design increases progressively as the size of k increases, and the cost of carrying out a complete factorial design is usually beyond the resources of an experimenter. For example, a complete 2^7 factorial design will require 128 runs and to carry out such an experiment will be both time-consuming and extremely costly, and so very rarely would a full factorial design be used in practice. These are a couple of reasons that to use a *fractional factorial design* is of extreme benefit where only a subset or *fraction* of the full factorial design is used.

Consider the 2^7 factorial design where there are 127 degrees of freedom available and only 7 of the 127 degrees of freedom correspond to the estimation of main effects and 21 to the estimation of two-factor interactions. Owing to certain principles, namely the *effect sparsity principle*, resulting from the factor sparsity principle, and the *hierarchical ordering principle* (Wu and Hamada, 2000), an experimenter can assume tentatively that higher-order interactions, say greater than two-factor interactions, are negligible, thus allowing them to obtain information on main effects and low order interactions by only running a fraction of the full design. These principles respectively

state that out of the many effects calculated due to factors and their interactions, only a small number will be important, and that lower order effects are more likely to be important than higher order effects. The idea of effect sparsity is particularly important in the case of screening experiments, since it is likely that many of the factors being investigated will have little or no effect on the response, and those that have been identified as important will then be focussed on more carefully in further experimentation.

Consider a half-fraction of a 2^4 design (Table 1.4); this is therefore denoted as a 2^{4-1} design since it only consists of 8 runs.

Table 1.4: Half-fraction of a 2^4 Design

1	**2**	**3**	**4**
-	-	-	-
+	-	-	+
-	+	-	+
+	+	-	-
-	-	+	+
+	-	+	-
-	+	+	-
+	+	+	+

It is obvious that for factors **1**, **2** and **3** the levels of the factors have just been chosen according to standard ordering, but how have the levels of factor **4** been allocated? Notice that the column of minus and plus signs for **4** is the product of column **1**, column **2** and column **3**. Thus, **4** is assigned the column **123** and **4** is said to be *aliased* with the **123** interaction. This aliasing relation is $\mathbf{4} = \mathbf{123}$, called the *design generator*, or $\mathbf{I} = \mathbf{1234}$ which is called the *defining relation* of this 2^{4-1} design. Since the defining relation **1234** has word length 4, the design is said to have resolution IV, and can also be fully defined as a 2^{4-1}_{IV} design.

Resolution describes the degree to which the main effects are aliased with other main effects and interaction effects and in the case of the resolution IV design, main effects are not aliased with any other main effect or with any two-factor interaction. However, two-factor interactions are aliased with one another. Aliasing implies a confusion of effects where one is not able to distinguish the estimate of one particular effect from another. In the 2^{4-1} design considered, if the main effect of **4** was large, the experimenter cannot be sure if this is due to the influence of the main effect of **4** or the three-factor interaction **123**, as the design is actually estimating **4**+**123**. The designs of resolution III, IV and V are the most common and are as follows:

Resolution III Design Main effects are not aliased with other main effects, but main effects are aliased with two-factor interactions. Two-factor interactions may be aliased with each other.

Resolution IV Design Main effects are not aliased with other main effects or two-factor interactions, but two-factor interactions are aliased with each other.

Resolution V Design Main effects and two-factor interactions are not aliased with other main effects or two-factor interactions, but two-factor interactions are aliased with three-factor interactions.

A main effect or two-factor interaction is *clear* if it is not aliased with main effects or two-factor interactions and is *strongly clear* if it is not aliased with main effects, two-factor interactions or three-factor interactions (Wu and Chen, 1992). In general, a resolution p design has main effects aliased with $(p-1)$-factor interactions and so a design with high resolution is desirable to ensure that too many assumptions do not need to be made regarding which interactions are negligible.

A fractional factorial design will be denoted as a 2^{k-p} design, with k factors all at two levels and a 2^{-p} fraction of the full 2^k factorial design. As mentioned previously,

a design will be determined by its particular defining relation and specifically by p defining words. The group resulting from the p defining words is called the *defining contrast subgroup* and consists of the identity element $\mathbf{I}$ and $2^p - 1$ words.

Consider the 2^{5-2} design which has design generators $\mathbf{4} = \mathbf{12}$ and $\mathbf{5} = \mathbf{13}$. So factors **1, 2** and **3** consist of minus and plus signs according to standard ordering and factor **4** is assigned the column of two-factor interaction **12**, and **5** assigned the two-factor interaction **13**. Thus, the two defining relations for this design are $\mathbf{I} = \mathbf{124}$ and $\mathbf{I} = \mathbf{135}$. By multiplying the two defining words, **124** and **135** together (using the rule that $\mathbf{1}^2 = \mathbf{I}$, $\mathbf{2}^2 = \mathbf{I}$, etc.), the defining contrast subgroup is obtained. For this design the defining contrast subgroup is

$$\mathbf{I} = \mathbf{124} = \mathbf{135} = \mathbf{2345}.$$

This method can also be used to obtain aliasing patterns, simply by multiplying each term in the defining contrast subgroup by the main effect or interaction effect that one is interested in.

The *maximum resolution criterion* (Box and Hunter, 1961a and 1961b) can be used to discriminate between designs by choosing a design which has defining words with a longer length. This will thus imply aliasing of higher order effects. For a 2^{k-p} design, let A_i denote the number of words of length i in its defining contrast subgroup. Then the vector

$$W = (A_3, \ldots, A_k)$$

is called the *wordlength pattern* of the design (Wu and Hamada, 2000). The resolution of a 2^{k-p} design can be defined as the smallest r such that $A_r \geq 1$, i.e. the length of the shortest word in the defining contrast subgroup. Then the maximum resolution criterion will choose the 2^{k-p} design with maximum resolution.

The *minimum aberration criterion* (Fries and Hunter, 1980) can be used to further

discriminate and distinguish between two designs of the same resolution. This is done by discriminating against the design which has more aliasing as a result of the design's particular defining relation. The minimum aberration criterion is as follows:

> For any two 2^{k-p} designs d_1 and d_2, let r be the smallest integer such that $A_r(d_1) \neq A_r(d_2)$. Then d_1 is said to have less aberration than d_2 if $A_r(d_1) < A_r(d_2)$. If there is no design with less aberration than d_1, then d_1 has minimum aberration.

An 'optimal' fractional factorial design will usually be chosen based on some criteria such as maximum resolution and/or minimum aberration.

1.4.3 Analysis of Factorial Two-Level Experiments

Daniel (1959) gave a method which allowed the half-normal plot to be used to aid in the interpretation of two-level factorial experiments. This involved plotting the empirical cumulative distribution of the orthogonal contrasts computed from a 2^k experiment; the half-normal plot can then be used to estimate the standard deviation of error and make judgements regarding the true nature of the observed effects. However, this method did not come without its warnings - words of caution given with the half-normal plot concerned subjective biases and "that it is not offered as a general substitute for the analysis of variance." This graphical method will be referred to as the Daniel plot in this chapter. A more formal procedure was suggested by Daniel (1959), to be used alongside the somewhat subjective method of visually inspecting half-normal plots, later modified by Zahn (1975a, 1975b) and then further investigated and adjusted by Olguin and Fearn (1997).

Literature by Box, Hunter and Hunter (1978) and Box and Draper (1987) presented the use of the full normal plot, a normal plot of the signed contrasts, as a preferential

alternative to the half-normal plot. The normal plot of effects is perhaps one of the most common methods for detecting active effects. This is due to ANOVA methods not being easily able to use as a result of the lack of degrees of freedom available for estimation of the variance of experimental error, which is particularly true in the case of saturated designs. However, as highlighted with the half-normal plots, there still remain problems associated with interpreting normal plots due to the subjectivity involved.

Box and Meyer (1986a) provided a more formal analysis of unreplicated fractional factorials that could be used in association with graphical analyses such as normal probability plotting, to supplement such plots. This extended from the work of Daniel that made use of the factor sparsity principle which Box and Meyer used much more explicitly in their paper. A simple theoretical model takes account of the Pareto principle, "the law of the vital few", that the majority of the process variation is explained by a small proportion of the variables, and is thus associated with factor sparsity, which states that of the total number of factors being investigated, only a small proportion of these factors will have effects that are large. A posterior probability that each contrast is active is computed and a posterior probability plot produced, now more popularly known as a Bayes plot. Their method can be summarized as follows. Suppose an effect β_i $(i = 1, \ldots, v)$ is active with probability α, where active effects are iid $N(0, \sigma_\beta^2)$ and inert effects β_i (also known as passive or inactive) are 0. Let $\mathbf{T} = (T_1, \ldots, T_v)$ be the vector of v estimated effects and where necessary are standardized so that given β, they all have the same unknown variance σ^2. Thus, for an inert effect $T_i = e_i$ and for an active effect $T_i = \beta_i + e_i$ where e_i are iid $N(0, \sigma^2)$ error terms. Let $k^2 = (\sigma^2 + \sigma_\beta^2)/\sigma^2$, and then $T_1, \ldots, T_v$ are iid from the scale-contaminated normal distribution denoted by $(1 - \alpha)N(0, \sigma^2) + \alpha N(0, k^2\sigma^2)$. That is, the effect T_i is normally distributed with zero mean and variance σ^2 with probability $1 - \alpha$ and has much larger variance $k^2\sigma^2$ with probability α.

Let $a_{(r)}$ be the event that a set of r effects out of the total v effects are active and $\mathbf{T}_{(r)}$ be the corresponding estimated effects for those r active effects. Then assuming *a priori* that $\log \sigma$ is locally independent and uniform, the posterior probability that $\mathbf{T}_{(r)}$ comprises the active effects is

$$p(a_{(r)}|\mathbf{T}, \alpha, k) \propto \left[\frac{\alpha k^{-1}}{1-\alpha}\right]^r \left[1 - \varphi f_{(r)}\right]^{-v/2} \tag{1.1}$$

where $\varphi = 1 - 1/k^2$ and $f_{(r)} = \mathbf{T}_{(r)}^{\mathrm{T}}\mathbf{T}_{(r)}/\mathbf{T}^{\mathrm{T}}\mathbf{T}$, the fraction of the sum of squares associated with $\mathbf{T}_{(r)}$.

Then, the marginal probability p_i that an effect i is active given $\mathbf{T}$, α and k is

$$p_i = \sum_{(r):\, i\, active} p(a_{(r)}|\mathbf{T}, \alpha, k), \tag{1.2}$$

i.e. to compute p_i for $i = 1, \ldots, v$ the probabilities in (1.1) must be computed for all 2^v possible events $a_{(r)}$.

Box and Meyer also presented a much quicker way of computing the posterior probabilities $\{p_i\}$, rather than having to sum over all 2^v combinations. Given the assumption that $T_1, \ldots, T_v$ are iid from $(1-\alpha)N(0, \sigma^2) + \alpha N(0, k^2\sigma^2)$, then the posterior probability that a single estimated effect T_i comes from the wider distribution $N(0, k^2\sigma^2)$, given σ, is

$$P_{i|\sigma} = \frac{\alpha \frac{1}{k} \exp\left\{\frac{-T_i^2}{2k^2\sigma^2}\right\}}{\alpha \frac{1}{k} \exp\left\{\frac{-T_i^2}{2k^2\sigma^2}\right\} + (1-\alpha)\exp\left\{\frac{-T_i^2}{2\sigma^2}\right\}}. \tag{1.3}$$

This formula is obtained by direct application of Bayes' Theorem. To compute the unconditional posterior probability p_i that effect T_i is active, the parameter σ must be integrated out of (1.3) over its posterior distribution $p(\sigma|\mathbf{T})$. That is

$$p_i = \int_0^\infty p_{i|\sigma} p(\sigma|\mathbf{T})\, d\sigma. \tag{1.4}$$

The posterior distribution of σ is

$$p(\sigma|\mathbf{T}) \propto \sigma^{-n} \prod_{j=1}^{v} \left[(1-\alpha)\exp\left\{\frac{-T_j^2}{2\sigma^2}\right\} + \alpha\frac{1}{k}\exp\left\{\frac{-T_j^2}{2k^2\sigma^2}\right\}\right]. \tag{1.5}$$

The product in (1.5) can then be expanded into the sum of 2^v terms and substituted into the integral (1.4), the integration being done analytically to obtain the expression (1.2) for p_i. However, the integral can be computed much more efficiently by numerical integration rather than expanding the integrand in (1.4) to compute the probabilities $\{p_i\}$.

Lenth (1989) presented a competing technique to the Box-Meyer method for assessing the sizes of contrasts in unreplicated fractional factorial designs. This method uses robust estimation of the standard deviation of estimated factorial effects to define a measure called the *pseudo standard error* (PSE). Let $\kappa_1, \kappa_2, \ldots, \kappa_m$ denote the contrasts of interest and $c_1, c_2, \ldots, c_m$ denote the corresponding estimates. Then

$$s_0 = 1.5 \times \underset{j}{\text{median}}\, |c_j|$$

and the pseudo standard error of the contrasts is defined as

$$PSE = 1.5 \times \underset{|c_j|<2.5s_0}{\text{median}}\, |c_j|.$$

That is, a trimmed median attempts to remove contrasts associated with active or non-zero effects. The PSE is then used to calculate a *margin of error* (ME) and a *simultaneous margin of error* (SME) as follows

$$ME = t_{0.975;d} \times PSE$$

$$SME = t_{\gamma;d} \times PSE$$

where

$$\gamma = (1 + 0.95^{1/m})/2.$$

Let m denote the number of contrasts being estimated and d denote the df which are usually taken to be $d = m/3$. As stated ME is a margin of error, and this is in fact for the estimates corresponding to the m contrasts and is with approximately 95%

confidence. It may be the case that due to several inferences being made simultaneously, and with a large number of contrasts that one or two estimates associated with inactive contrasts may exceed the ME. This would in fact lead to an incorrect conclusion, and so to take into account this possibility the SME is defined. The subject of the poor approximations of the critical values for ME and SME was discussed by Olguin and Fearn (1997). It was noted that these critical values were computed on the basis that the ratios of contrasts to PSE are distributed approximately as t with $m/3$ degrees of freedom. However, these approximations were found to be inaccurate, at least in the tails of the distribution and evidence of this deficiency in Lenth's method is demonstrated to some extent in later results in Chapter 3. Despite this, the competing technique of Lenth's method has proved to be somewhat effective, is relatively simple computationally and also has the added advantage of the results being given in terms of the original units of measurement.

The graphical methods that have been outlined, namely the methods of Daniel (1959), Box and Meyer (1986a) and Lenth (1989) (Figures 1.2 - 1.4) are illustrated with tensile strength data taken from Taguchi and Wu (1980). The experiment taken from Taguchi and Wu (1980) is a 2^{9-5}_{III} design. The factors investigated were Rods (A), Period (B), Material (C), Thickness (D), Angle (E), Opening (F), Current (G), Method (H) and Preheating (J), and the response measured was tensile strength. The column allocation of factors and responses are given in Table 1.5. The values used to generate the Daniel, Bayes and Lenth Plot are displayed in Table 1.6, i.e. estimated effects, their associated posterior probabilities, margin of error (ME) and simultaneous margin of error (SME).

For Daniel's plot, the half-normal score is shown on the horizontal scale and the absolute value of the estimated effects on the vertical scale. Each point plotted is labelled by its corresponding column number. The Bayes plot has the horizontal scale denoting each of the columns. The bar drawn at each column is the probability

that the column is associated with an active contrast, either as a result of a single effect or a linear combination of effects depending upon the alias string. Lenth's plot, similarly to Bayes plot, has numbers $1, \ldots, 15$ on the horizontal scale denoting each of the columns of the design. Reference lines are drawn on the plot at both $\pm$ME and $\pm$SME and are used to decide whether a factor is active or inactive. If a bar at any of the columns, $1, \ldots, 15$, extends beyond the $\pm$SME, then it can be concluded that contrast is clearly active. If a bar does not extend beyond $\pm$ME, then that particular contrast is not active. If a bar falls in between $\pm$ME and $\pm$SME, then there is uncertainty surrounding the decision whether the contrast is active or not and the experimenter must decide whether to run further experiments or use scientific knowledge to draw the most appropriate conclusion.

In all three plots it can clearly be seen that contrasts, or columns, 14 and 15 are clearly active, and all other factors are inactive. The active contrasts correspond to B and $-C$.

Table 1.5: Column Allocation and Results for Tensile Strength Experiment

	Column Allocation																Tensile Strength
Run	**I**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**y**
1	+	-	-	+	-	+	+	-	-	+	+	-	+	-	-	+	43.7
2	+	+	-	-	-	-	+	+	-	-	+	+	+	+	-	-	40.2
3	+	-	+	-	-	+	-	+	-	+	-	+	+	-	+	-	42.4
4	+	+	+	+	-	-	-	-	-	-	-	-	+	+	+	+	44.7
5	+	-	-	+	+	-	-	+	-	+	+	-	-	+	+	-	42.4
6	+	+	-	-	+	+	-	-	-	-	+	+	-	-	+	+	45.9
7	+	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+	42.2
8	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	40.6
9	+	-	-	+	-	+	+	-	+	-	-	+	-	+	+	-	42.4
10	+	+	-	-	-	-	+	+	+	+	-	-	-	-	+	+	45.5
11	+	-	+	-	-	+	-	+	+	-	+	-	-	+	-	+	43.6
12	+	+	+	+	-	-	-	-	+	+	+	+	-	-	-	-	40.6
13	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	44.0
14	+	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-	40.2
15	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	42.5
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	46.5
		D	H		G	$-F$			A	$-E$				J	B	$-C$	

Table 1.6: Estimated Effects, Their Associated Posterior Probabilities, ME and SME.

Column	Effect	p
1	0.13	0.03
2	-0.15	0.03
3	0.30	0.05
4	0.15	0.03
5	0.40	0.08
6	-0.30	0.02
7	0.37	0.07
8	0.40	0.08
9	-0.05	0.02
10	0.42	0.09
11	0.13	0.03
12	0.13	0.03
13	-0.37	0.07
14	2.15	1.00
15	3.10	1.00
ME = 0.58	SME = 1.17	

Figure 1.2: Daniel's Plot of Tensile Strength Data

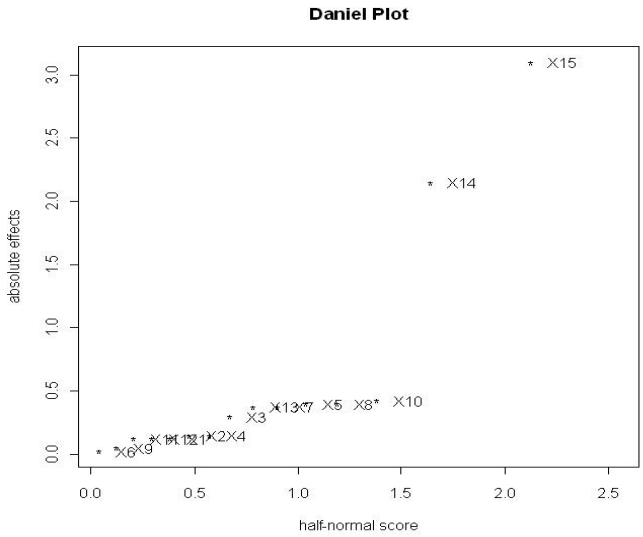


Figure 1.3: Bayes Plot of Tensile Strength Data

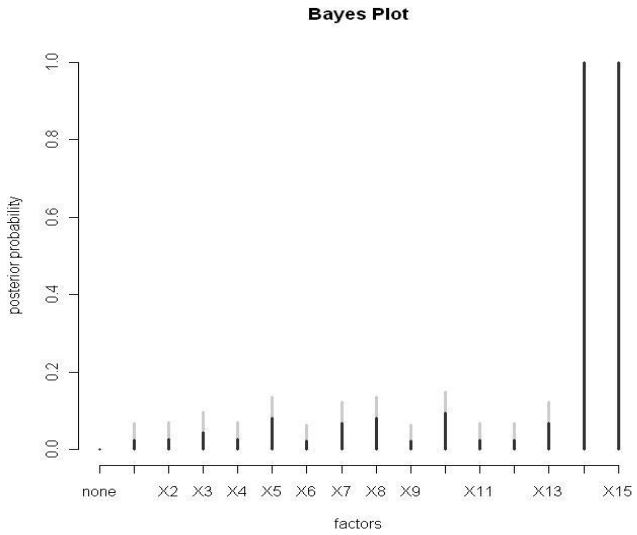
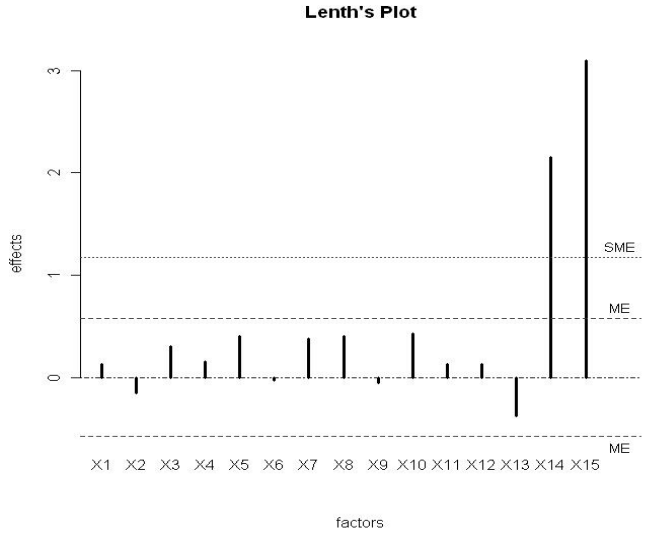


Figure 1.4: Lenth's Plot of Tensile Strength Data



1.4.4 Decision Theory

Decision theory is a formal theory of decision making under uncertainty. A decision problem typically consists of possible *actions* $\{a_i\}_{a_i \in A}\, i = 1, \ldots, m$, possible *states* $\{s_j\}_{s_j \in S}\, j = 1, \ldots, n$ and possible *consequences* $\{c_{i,j}\}_{c_{i,j} \in C}$ where

$$a_i + s_j \longrightarrow c_{i,j}.$$

That is, taking some course of action a_i when the state of nature is s_j, leads to the consequence $c_{i,j}$. The question of which is the best action to take can be answered, according to decision theory, by measuring the 'goodness' of the consequences with a *utility function* $U(c)$. The best action, a^*, will be that which maximizes the expected utility. That is,

$$a^* = \arg\max_a \{E\left[U(c)|a\right]\}.$$

These types of decision problems can be represented by a decision table, an example of which is displayed in Table 1.7.

Table 1.7: General Form of a Decision Table

	States			
Actions	s_1	s_2	$\cdots$	s_n
a_1	$c_{1,1}$	$c_{1,2}$	$\cdots$	$c_{1,n}$
a_2	$c_{2,1}$	$c_{2,2}$	$\cdots$	$c_{2,n}$
.	.	.	$\cdots$	.
a_m	$c_{m,1}$	$c_{m,2}$	$\cdots$	$c_{m,n}$

1.4.5 Bayesian Methodology

Bayesian theory is an interpretation of probability which is determined by the degree of belief a person has in a hypothesis and also suggests that using Bayes' Theorem, we

can update this degree of belief in light of new information. People who at one time disagree about the likeliness or truth of a hypothesis may then agree after sharing experiences of the situation. They are not coerced into agreeing with one another but the experiment in hand forces them to agree and subjective theory provides some explanation as to how this agreement comes about. Axioms are given to personal degrees of belief and any person can compare and also quantify their beliefs. We then have some consequence of the axioms where we let h denote a hypothesis, d denote some initial data about h and e denote the result of an experiment after d is known. We can then try to establish the confidence we will have in h given d alone (initial or prior confidence); the confidence we would have in e if we already know h and d but not e (likelihood of e in light of h); and the confidence in h after we learn of e (posterior confidence). An application of Bayes' Theorem is that it can be proved that the posterior belief is proportional to the prior belief and the likelihood. Variations in belief represent the influence of experience and a consequence of Bayes' Theorem describes how we should learn from the experimental data e. As we carry out experiments of the same kind our degree of confidence will become increasingly more stable. Also, when we have two people who have greatly different initial confidences, so long as they agree on what is possible and on the outcomes of the individual experiments, Bayes' Theorem proves that not only do their opinions become more stable but also their differing views will coincide increasingly over time. These ideas underpin the very core of Bayesian inference and can be more formally represented as

$$p(\theta|y) \propto p(\theta)p(y|\theta),$$

where θ is an unknown parameter of interest and y is some observed data. Thus the *posterior density* $p(\theta|y)$ is proportional to the *prior distribution* $p(\theta)$ and *sampling distribution* $p(y|\theta)$.

The use of Bayesian methods in the clinical work of the pharmaceutical industry

has only really caught on over the last 25 years, the reluctance largely due to heavy regulations particularly when testing drugs and carrying out clinical trials on patients. The industry has now warmed somewhat to these ideas with major world regulatory authorities stating that Bayesian procedures can be considered when reasons for their use are clear and also given that conclusions are seen to be robust. Despite the more free use of Bayesian methods, they are still met with some scepticism regarding the subjective measure of belief in comparison to the more traditional nature of frequentist statistics based on repeated sampling. However, this does not apply so much in the case of non-clinical work where regulatory practices have not been as stringent, and statisticians need to justify their choice of prior perhaps only to other colleagues working in their area or on the project, rather than to a regulatory body. This does not mean that the use of Bayesian methods is completely unproblematic in industrial experiments; there still remains the problem of the choice of prior which best reflects the experimenters' knowledge about the process in hand and the size of the parameter estimates.

1.5 Outline of Chapters

Multifactor designs, including fractional factorial designs, are one of the most widely used statistical contributions to industrial experimentation. They are used in many manufacturing and processing industries, including the pharmaceutical industry, where they are used in both pre-clinical research and in process improvement in manufacturing. They are increasingly recognised by scientists and engineers as allowing considerable information about the effects of several factors on the response to be obtained with a relatively small number of runs, which is extremely important given that a single run can be time-consuming and costly.

An important decision for an experimenter is in choosing the design for their experiment, which will greatly influence the results they obtain and consequently what they are able to analyse as a result. However, the experimenters' prior knowledge is not usually exploited and used in a formal manner at the design stage. In chapter 2, literature on optimal designs is introduced and decision-theoretic ideas explored that take into account the prior knowledge that an experimenter may have. The key utility functions that will be used to discriminate between designs are introduced. As mentioned, prior knowledge will be taken into consideration and designs compared on the basis of the usual least squares model and analysis, and also looking at comparing designs when the alternative Lenth's method of analysis is used. This research will involve the use of utility functions relevant to the objectives of the experiment, say maximising the amount of yield, and the design which has the maximum utility is chosen as the 'best'.

The methods set out in chapter 2 are then applied in chapter 3 with a series of examples to illustrate the concepts of the utility functions. The chapter begins with a small example comparing two designs studying three factors at two levels with the purpose of seeking the optimal treatment combination. A larger case is then presented

where a 12 and 16 run design, studying 5 factors at two levels, are compared with the objective of finding a design which is most efficient in terms of identifying active and inactive effects. Finally, the matter of optimization and obtaining the optimal treatment combination is again considered with a larger example. Two designs are compared where 5 factors are investigated at two-levels in a 16 run experiment, and the same factors investigated at three-levels in an 18 run experiment with the purpose of identifying which design is most effective.

An introduction to the concepts and most notable literature of Bayesian analysis are given in chapter 4 and these ideas are then discussed in relation to the analysis of data from fractional factorial experiments. There is virtually always some prior knowledge about the sizes of effects and using this in a fully Bayesian data analysis seems natural. The use of fairly vague priors are explored, as well as more informative priors. A typical assumption made in fractional factorial experiments *a priori* is that only a small number of all the effects being investigated will be truly important, which could be considered to be an extreme form of prior knowledge and consequently leads to the justification of the use of Bayesian analysis.

Another problem faced by experimenters is the possibility of the dataset containing outliers. The analysis of fractional factorial experiments can be impacted severely in the presence of an outlier, which will be considered to be an observation which is suspected of not being generated from the mechanism which produced the majority of observations. Outliers are an unavoidable circumstance; recording errors and temporary changes in experimental conditions are just a few reasons for their occurrence. The possibility that they may occur is always present and so an analysis that is more robust to outliers is sought. In chapter 5 an analysis that is more robust to outliers is sought and a comparison of methods of analysis on some historical datasets is made. This involves looking at both non-saturated and saturated designs, the latter of which introduces complications such as the problem of a high level of aliasing being

introduced. A simulation study will be conducted to test the ideas developed.

The main conclusions of the thesis are summarised in chapter 6 and further research and additional problems that extend from this thesis are also given.

Chapter 2

A Decision Theory Approach to the Choice of Factorial Design

2.1 Introduction

The statistical tools available for designing experiments usually choose a design which is optimal. This is either in terms of a variance criterion such as D-efficiency, or in terms of estimability criteria such as resolution and aberration, given the number of factors, the number of levels of each factor and the number of runs available. However, in practice these are not given but are to be decided by the experimenter and this is usually done in an informal manner by comparing what can be obtained from various designs of different sizes and then choosing a design based on what seems sensible. Also, the usual optimal design criteria and utility functions do not provide the correct criteria for experimenters to answer certain questions, such as which design will maximise yield or correctly identify active factors.

A decision-theoretic approach is explored in this chapter, which will allow a more

informed choice of design to be made. Prior distributions on the sizes of effects will be taken into consideration, and a design (including the number of runs, factors and levels) can then be chosen from a candidate set of designs using a utility function relevant to the objectives of the experiment. For example, a screening experiment to screen factors for further experimentation, would require the need to correctly identify the true active and inactive factors and so a utility function combining the proportion of active and inactive factors declared correctly would be used. This differs from that of an (usually one-off) experiment where the factors involved in the process have been identified and the optimal treatment combination may be sought in order to optimize the process. An example of optimizing the process may be to maximise yield or other similar things such as to maximise purity, to minimise waste, etc. A gain function may be considered in order to compare the yield obtained from the estimated optimum factor level combination in comparison to standard operating conditions, and a loss function would be considered to be the yield obtained at the true optimum factor level combination compared to the yield obtained at the estimated optimum factor level combination.

2.2 Standard Optimality Criteria

The theory of the optimum design of experiments was formally discussed by Atkinson, Donev and Tobias (2007), and the most commonly used optimality criteria described. An optimality criterion provides a measure of how good a design is, and *information-based* criteria and *distance-based* criteria are the two general types of criteria available. The distance-based criteria correspond to the idea of filling the candidate space as well as possible. The candidate points comprise a point cloud in p-dimensional Euclidean space where p is the number of terms in the model. Information-based criteria are related to the information matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ for the design and proposed model. The criteria of optimality set out by Atkinson et al. focussed more on the information-based criteria.

To illustrate these information-based criteria let us first consider the general linear model, which will be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$$

where $\mathbf{y} = (y_1, ..., y_N)^{\mathrm{T}}$ is the $N \times 1$ vector of responses, $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta p)^{\mathrm{T}}$ is the $(p+1) \times 1$ vector of regression coefficients, $\epsilon = (\epsilon_1, ..., \epsilon_N)^{\mathrm{T}}$ is the $N \times 1$ vector of errors, and $\mathbf{X}$ is the $N \times (p+1)$ design matrix and N is the total number of runs. p is the number of explanatory variables for the multiple linear regression model but not in the case of a second order polynomial model.

Then for this model, the sum of squares to be minimized is

$$S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

and the least squares estimator of $\boldsymbol{\beta}$ satisfies the $p+1$ least squares equations

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^{\mathrm{T}}\mathbf{y}.$$

The matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is then the *information matrix* for $\boldsymbol{\beta}$. The covariance matrix of the least squares estimator for a non-singular design is

$$var(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1},$$

and from this

$$\left|(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\right| = 1/\left|\mathbf{X}^{\mathrm{T}}\mathbf{X}\right|$$

is called the generalized variance of $\hat{\boldsymbol{\beta}}$. Thus, the information matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, a measure of the information from the experiment, is of importance relating to optimal design theory. It is proportional to the inverse of the variance-covariance matrix for the least-squares estimators of the model parameters.

The most widely accepted criteria for optimal experimental designs are now set out:

A-optimality Minimizes $\mathrm{tr}[(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}]$, that is the average of the variances of the parameter estimates is minimized.

D-optimality Maximizes $\left|\mathbf{X}^{\mathrm{T}}\mathbf{X}\right|$, which is the same as minimizing the generalized variance of the parameter estimates.

G-optimality This is based on the variance of prediction across the design region, which is proportional to $\mathbf{x}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}$. This minimizes the maximum variance of the predicted values, i.e. minimizes $\mathbf{x}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}$ over all points $\mathbf{x}$ in the design region.

Generally, a good design which 'minimizes' the variance $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$, or 'maximizes' the information $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is sought.

The standard optimality criteria do not always provide appropriate criteria with which to discriminate between designs. As the optimality criteria are based on some functions of the information matrix, the optimality of a given design is therefore model

dependent. Since an experimenter must specify a model for the design and the number of design points prior to the optimal design being generated, this means that the optimal design generated will actually only be optimal for that model.

It is also worth noting that an approach adopted to compare designs of different sizes is to scale by N, the total number of observations for the design (Anderson-Cook, 2005). Scaling by N in this way allows one to compare designs of different sizes on an information per observation basis. However, this does impose a very specific and often unrealistic utility function and as with all optimality criteria should be used with both caution and judgement.

Design efficiency, as specificied by Atkinson, Donev and Tobias (2007), is a measure of how good a design is in comparison to the optimal design. For D-optimality let the optimum design be ξ_D^*, and then the D-efficiency of any other design ξ is

$$\text{Eff}^D(\xi) = \left\{ \frac{\left|\mathbf{X}^{\mathrm{T}}\mathbf{X}(\xi)\right|}{\left|\mathbf{X}^{\mathrm{T}}\mathbf{X}(\xi_D^*)\right|} \right\}^{1/(p+1)}.$$

Since optimality, and therefore also design efficiency is model dependent, it is impossible to compare designs estimating different models. Upon calculating the D-efficiency for a particular candidate design, it is impossible to compare this measure to the D-efficiency for a design estimating a different model. This is also true for A-efficiency, G-efficiency, etc. and in the case of the Bayesian versions which shall be outlined later in this section. This leads to the need for some measure which can appropriately discriminate between designs, even in the case where they are estimating different models.

The optimality criteria discussed thus far have been concerned with the optimality in the case of standard least squares. Now we will consider experimental design from a Bayesian viewpoint. Chaloner and Verdinelli (1995) provided a comprehensive review of the literature on Bayesian experimental design. They also noted that experimental

design is a situation where it is meaningful within Bayesian theory to average over the sample space, although as the sample has not yet been observed, averaging over what is unknown applies. Following on from Raiffa and Schlaifer (1961), Lindley (1972) presented a decision-theoretic approach to experimental design. Lindley's argument suggested that the most effective way to design experiments would be to specify a utility function to reflect the objective of the experiment. Then, one should regard the choice of design rather as a decision problem and select a design that maximizes the expected utility. This argument is given more formally as follows. A design ξ is to be chosen from the design space Ξ, and some data y will be observed from a sample space Y. Given y, the selection of ξ from Ξ and a terminal decision d from the set of decisions, D, the action space, will be made. The unknown parameters are θ and the parameter space Θ. A general utility function can then be defined in the following form of $U(d, \theta, \xi, y)$.

For any design ξ, the expected utility of the best decision is given by

$$U(\xi) = \int_Y \max_{d \epsilon D} \int_\Theta U(d, \theta, \xi, y) p(\theta|y, \xi) p(y|\xi) \, d\theta \, dy.$$

The Bayesian experimental design is then given by the design ξ^* maximizing $U(\xi)$:

$$U(\xi^*) = \max_{\xi \epsilon \Xi} \int_Y \max_{d \epsilon D} \int_\Theta U(d, \theta, \xi, y) p(\theta|y, \xi) p(y|\xi) \, d\theta \, dy.$$

Previously, we were considering optimality criteria with regards to least squares, that could be used to discriminate between designs. An alternative way of discriminating between designs in a Bayesian fashion is to consider a utility function which considers the expected gain in Shannon information given by an experiment (Shannon, 1948) following from Lindley (1956). It was proposed that a design should be chosen that maximises the Kullback-Leibler distance, that is the expected gain in Shannon information which is equivalent to maximising a distance between the posterior and prior

distribution,

$$\int \log \frac{p(\theta|y,\xi)}{p(\theta)} p(y,\theta|\xi)\, d\theta\, dy.$$

Since the prior distribution does not depend upon the design ξ, the design actually maximises the expected gain in Shannon information of the posterior distribution and the following is the utility function for this;

$$U(\xi) = \int \log\{p(\theta|y,\xi)\} p(y,\theta|\xi)\, d\theta\, dy.$$

In the case when utility is based on Shannon information, then the Bayesian optimal design ξ^* does in fact satisfy

$$\xi^* = \arg\max_{\xi \epsilon \Xi} |\mathbf{X}^{\mathrm{T}}\mathbf{X}(\xi) + R|$$

for a linear model, and this is where R is a known $k \times k$ matrix, when θ is normally distributed with mean θ_0 and variance-covariance matrix $\sigma^2 R^{-1}$. Thus, when utility is based on Shannon information then this utility reduces to maximizing $|\mathbf{X}^{\mathrm{T}}\mathbf{X}(\xi) + R|$ and is known as Bayesian D-optimality. This differs from non-Bayesian D-optimality which maximizes $|\mathbf{X}^{\mathrm{T}}\mathbf{X}(\xi)|$. However, when n the number of observations is large, i.e. $n \to \infty$, or when there is little prior information about the unknown parameters θ, then the Bayesian D-optimality does in fact reduce to the classical D-optimality.

Bayesian A-optimality maximizes $-\mathrm{tr}[A(\mathbf{X}^{\mathrm{T}}\mathbf{X}(\xi) + R)^{-1}]$, where A is a symmetric non-negative definite matrix. This is a generalization of a non-Bayesian optimality criterion that minimizes $\mathrm{tr}[A(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}]$. When $A = I$, the identity matrix, A-optimality is recovered and the Bayesian A-optimality criterion is then a generalization of the non-Bayesian A-optimality criterion. Similarly to its non-Bayesian counterpart, it reduces the variances of the parameter estimates and this criterion also corresponds to the quadratic loss utility function (Chaloner and Verdinelli, 1995).

The criterion provided by utility functions such as Shannon information, may not always be appropriate such as in cases when comparing designs of different sizes. Due

to the nature of this utility maximising the expected gain of a particular design, when comparing two designs of different sizes, the design which is larger will usually have a greater expected gain and thus be chosen as optimal. This may not always be the case though, if the larger design is a poor choice of design. Thus, it is of interest to know how much more information is gained from an experiment when carrying out an extra n runs and a utility to quantify this gain in information is necessary.

As well as the Shannon utility function, or expected Kullback-Leibler distance, not being a natural criterion to use in some cases as the larger design will typically result in being the optimal design, it also does not relate to the questions experimenters will want to answer. The Kullback-Leibler distance does not seem an appropriate measure when having to choose between various designs that are estimating different models. It also does not translate naturally to answering questions experimenters may have concerning cost; it does not seem natural to ask "how does this Kullback-Leibler distance relate to the cost of carrying out n extra runs?".

Also mentioned in the previous chapter were the criteria used to choose an optimal fractional factorial design in terms of resolution and aberration. Specifically, an optimal design would have maximum resolution and/or minimum aberration. However, it can be shown that there are situations where these criteria are not always sensible to judge a design on their own. A downfall of the maximum resolution criterion is that two designs may have the same resolution, but due to different wordlength patterns, one of the designs may have more aliasing than the other. Likewise, the minimum aberration criterion is not without its faults. Consider two designs, d_1 and d_2, which have the same resolution and $A_r(d_1) < A_r(d_2)$. It is possible that although d_1 has minimum aberration, d_2 can have more clear interaction effects which would deem it superior to d_1, although not in terms of minimum aberration. The estimation of more clear effects would be considered in this situation to be a more desirable property than the length of defining words.

Both non-Bayesian and Bayesian criteria, typically used to discriminate between and ultimately choose an optimal design relative to the objective of the experiment, have been presented. Some of the drawbacks of these criteria have also been illustrated. It is therefore important to develop a utility function that incorporates the principal goals of the experiment. A utility is sought that can identify a design with minimal aliasing of effects. It is also important to quantify the worth of carrying out the extra n runs and whether this extra cost and time spent is of worth in comparison to carrying out an experiment with a smaller design. This is when designs of different sizes are to be considered. A utility is also sought to aid with determining the optimal design when there are several candidate designs of the same size, estimating different models. These utilities must be relevant to the experimenter's questions.

2.3 Utility Functions for Optimization

2.3.1 Loss and Gain Functions

As discussed previously, a utility function appropriate for the experimenter's objective is necessary. In the case of industrial experimentation in the pharmaceutical industry the principal aim is typically to first identify the key few factors having an effect on the process in hand, and then after further focussing on these active factors, to identify the optimal treatment combination. Optimal in this context will normally mean the treatment combination resulting in maximum yield, or alternatively, minimising loss of yield during manufacturing processes.

Let us consider the objective of identifying the optimal treatment combination when a fractional factorial experiment has been carried out. A full factorial experiment allows all effects to be estimated clearly, but due to time and cost constraints this is almost always an impossibility to carry out. Thus, the fractional factorial experiment will be employed in the knowledge that not all effects can be estimated and of those that can be estimated, some may not be estimated clearly. Thus, the experimenter wants to be sure the treatment combination that has been estimated to be optimal, gives a response as close to the true optimal response as possible. That is, we want to minimise the difference between the response at the true optimal treatment combination, and the response at the estimated optimal treatment combination. This is minimising the loss.

Alternatively, the experimenter may be more concerned with wanting to improve on standard operating conditions that are already in place. So, we would want to maximise the difference between the response at the estimated optimal treatment combination, and the expected response at the standard operating treatment combi-

nation currently in use. That is, maximising the gain in yield due to the knowledge obtained from the experiment.

We can define for any treatment combination, $\mathbf{x}$, a loss function that gives the loss between the yield due to the true optimal treatment combination $\mu_{\mathbf{x}_{opt}}$ and the yield due to the particular treatment combination $\mathbf{x}$ that we are interested in. A gain utility function can similarly be defined as the function that gives the gain in yield between using the treatment combination $\mathbf{x}$ we are interested in and the yield due to the standard operating treatment combination currently in use, $\mu_{\mathbf{x}_0}$. For any treatment combination $\mathbf{x}$, these functions can then be specified more formally as:

$$U_{Loss}(\mathbf{x}) = \mu_{\mathbf{x}_{opt}} - \mu_{\mathbf{x}}$$

$$U_{Gain}(\mathbf{x}) = \mu_{\mathbf{x}} - \mu_{\mathbf{x}_0}.$$

We are interested in the loss and gain for the estimated optimal treatment combination, $\hat{\mathbf{x}}_{opt}$. The concepts outlined for any treatment combination $\mathbf{x}$ can be also be written more succinctly for the estimated optimal treatment combination with the following utility functions:

$$U_{Loss}(\hat{\mathbf{x}}_{opt}) = \mu_{\mathbf{x}_{opt}} - \mu_{\hat{\mathbf{x}}_{opt}}$$

$$U_{Gain}(\hat{\mathbf{x}}_{opt}) = \mu_{\hat{\mathbf{x}}_{opt}} - \mu_{\mathbf{x}_0}.$$

This is where $\mu_{\mathbf{x}_{opt}}$ is the expected response of the optimal treatment combination, $\mu_{\hat{\mathbf{x}}_{opt}}$ the expected response for estimated optimal treatment combination, and $\mu_{\mathbf{x}_0}$ the expected response at standard operating conditions. In the situation where the factors being considered are all at two levels, the objective would be to find which of the 2^k possible treatments results in the maximum response. This equates to locating the estimated optimal treatment combination, $\mu_{\hat{\mathbf{x}}_{opt}}$, that maximizes, or nearly maximizes,

the response. When considering loss, this is the distance between the responses at the two points $\mu_{\mathbf{x}_{opt}}$ and $\mu_{\hat{\mathbf{x}}_{opt}}$, i.e. looking at $\mu_{\mathbf{x}_{opt}} - \mu_{\hat{\mathbf{x}}_{opt}}$ and when considering gain this is $\mu_{\hat{\mathbf{x}}_{opt}} - \mu_{\mathbf{x}_0}$. However, when all factors are continuous and a first-order model is inadequate, the quantities $U_{Loss}(\hat{\mathbf{x}}_{opt})$ and $U_{Gain}(\hat{\mathbf{x}}_{opt})$ depend on the curvature of the surface and it is therefore necessary to consider *response surface methodology* to locate the optimum point on the response surface.

Both the key quantities, $U_{Loss}(\hat{\mathbf{x}}_{opt})$ and $U_{Gain}(\hat{\mathbf{x}}_{opt})$ depend upon the unknown parameters $\boldsymbol{\beta}$. By using subjective distributions for $\boldsymbol{\beta}$, since it will most likely be the case that the experimenter will have some prior knowledge about the unknown parameters, a Bayesian viewpoint can be incorporated into the decision theoretic approach. In using Bayes' Theorem and placing prior distributions on $\boldsymbol{\beta}$ a posterior distribution can be outlined for the loss and gain utilities, namely $U_{Loss}(\hat{\mathbf{x}}_{opt})|\mathbf{y}$ and $U_{Gain}(\hat{\mathbf{x}}_{opt})|\mathbf{y}$. It is not trivial to obtain these posterior distributions analytically, thus it can be approximated by simulating from the distribution of $\boldsymbol{\beta}|\mathbf{y}$ and then obtaining $U_{Loss}(\mathbf{x})$ and $U_{Gain}(\mathbf{x})$ for each realization of $\boldsymbol{\beta}$.

The utilities $U_{Loss}(\hat{\mathbf{x}}_{opt})$ and $U_{Gain}(\hat{\mathbf{x}}_{opt})$ will give the loss or gain in terms of units of yield (or some other similar measure that the experimenter is considering), however for ease of understanding it may be more natural to express the gain in terms of 'relative gain', which will be denoted as

$$Relative\,Gain = \frac{\mu_{\hat{\mathbf{x}}_{opt}} - \mu_{\mathbf{x}_0}}{\mu_{\mathbf{x}_{opt}} - \mu_{\mathbf{x}_0}}$$

where the gain $\mu_{\hat{\mathbf{x}}_{opt}} - \mu_{\mathbf{x}_0}$ can be either a negative or positive value. The relative gain utility function will express the gain, i.e. the estimated improvement on standard operating conditions, in relation to the maximum gain that would have been made if the true optimal treatment combination had been used compared to standard operating conditions. It would be expected that an experimenter hopes the gain will be a positive quantity, since one would ideally want the response obtained at the estimated

optimal treatment combination to improve upon the response obtained at standard operating conditions. However, it is possible that the estimated optimal treatment combination is in fact worse than standard operating conditions thus resulting in a negative quantity.

It should be noted that when typically referring to a utility function, a 'utility' is called as such if maximizing the function. However, in the case of the loss function mentioned earlier we in fact want to minimize this. Therefore if referring to this loss in terms of a utility, we should in fact state this as $-U_{Loss}(\mathbf{x})$, i.e. we want to maximize $-U_{Loss}(\mathbf{x})$. However, where results are given in the next chapter, they are for $U_{Loss}(\mathbf{x})$, giving a measure of the difference between the estimated value and the desired, optimum, value. It should also be noted that although the gain, loss and relative gain are on different scales, if a design is found to be optimum in terms of maximizing the gain or relative gain, then this design would in fact be optimum in terms of minimizing the loss as all quantities are equivalent.

It is in fact $E(Utility)$ that we shall be using, since we do not know μ, the expectation being taken over the prior. $E(Utility)$ shall be used in the case of utility functions for both optimization and screening.

2.3.2 Relevant Literature

The utility functions discussed in the previous section extend from the work of Gilmour and Mead (1995). They obtained the posterior distribution of the difference between the expected response at the true optimum and the expected response at the predicted optimum, using a Bayesian analysis. This posterior distribution was then used to make a decision when to stop experimentation in the sequential design of response surface and fractional factorial experiments. Their method can be

summarized as follows. Assume that the response of interest, y, is modelled by

$$y = Y(\mathbf{x}) + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$ and independent. The vector $\mathbf{x}$ contains the levels of several qualitative or quantitative factors of interest. Also assume that $Y(\mathbf{x})$ is a linear model

$$Y(\mathbf{x}) = \mathbf{z}^{\mathrm{T}}\boldsymbol{\beta},$$

where each element in $\mathbf{z}$ is a function of the elements of $\mathbf{x}$. The design matrix $\mathbf{X}$ contains the vector $\mathbf{z}_i^{\mathrm{T}}$ for the i^{th} run as its i^{th} row. $\boldsymbol{\beta}$ is the vector of parameters of the model. Then, the value $\mathbf{x}_{max}$ of $\mathbf{x}$ is the value for which the expected response $Y(\mathbf{x})$ is a maximum and the series of experiments should be stopped when the experimenter is confident that the predicted optimal combination maximizes, or nearly maximizes, the response. Thus, let $\hat{\mathbf{x}}_{max}$ be an estimate of $\mathbf{x}_{max}$, where $\hat{\mathbf{x}}_{max}$ can be obtained by the method of least squares of $\boldsymbol{\beta}$. Then the quantity for assessing $\hat{\mathbf{x}}_{max}$ is

$$L(\hat{\mathbf{x}}_{max}) = Y(\mathbf{x}_{max}) - Y(\hat{\mathbf{x}}_{max}).$$

That is, $L(\hat{\mathbf{x}}_{max})$ is the difference between the expected response at the true optimum and the expected response at the predicted optimum and is identical to what has been referred to as the loss function in §2.3.1.

Gilmour and Mead then further extended this work (Gilmour and Mead, 2003) where a Bayesian A-optimality criterion was proposed for choosing designs in the case of sequential experimentation where each stage consists of a small number of runs and the objective is to optimize a response.

In addition, Müller et al. (2006) presented a formal Bayesian decision-theoretic approach to clinical trial design. In this paper they proposed a Bayesian decision-theoretic approach to a phase II dose-response finding study where they used a utility function to formalize learning about the unknown dose-response curve for the

adaptive dose allocation. The problem of computing an optimal dose for the next patients was considered. In this case let N denote the number of currently accrued patients and K denote the maximum number of patients who are recruited into the trial on one day. Then the optimal doses to be assigned to the next K patients, $i = N + 1, \ldots, N + K$ were to be computed. It was assumed that Z_j, $j = 1, \ldots, J$ is the range of allowable doses and $\tilde{z}_k$ is the dose to be determined for the future patient $N + K$. Then it was shown that $U_k(z_j)$ is the expected utility of decision $\tilde{z}_k = Z_j$ for a future patient, i.e. $U_k(z_j)$ expresses how much deciding on dose Z_j is worth to us and further, the solution to the optimal dose problem was formalized as

$$\tilde{z}_k = \arg\max_{Z_j} U_k(Z_j).$$

Thus, the dose which in expectation maximizes the utility defined by posterior variance on the key parameters was recommended.

The approach considered in this chapter, and later applied with several case studies in the following chapter, incorporates the decision-theoretic methodology and the concept of the posterior distribution of the difference between the optimal response and response from the predicted optimum. These methods have primarily been applied in sequential design, particularly in factorial experiments in industrial research, and decision theory application in clinical trials as displayed from the relevant literature indicated. However, these methods have scarcely been applied in the context of screening and fractional factorial designs used in industrial experimentation. This approach will now be applied in the case of screening experiments at the initial stages of experimentation, and optimization experiments at the latter stage of the process.

2.3.3 Considerations Regarding Formulating Priors

So, the utility function proposed (§2.3.1) will give the loss or gain of a design, which can be seen as a measure of the efficiency of the design, which is how well the experimental design answers the questions we are interested in. Where standard optimality criteria are dependent upon the information matrix, such as in the commonly used D-optimality criterion, the proposed utilities will require the specification of design matrices for the candidate designs, and also prior knowledge on the sizes of effects to be set out. This can either be in the form of prior distributions, which would be more commonly used, or given in the form of an actual effect size such as $\beta_1 = 10$, if an experimenter had strong prior knowledge.

After presenting some issues that are to be taken into consideration when formulating priors in this section, the algorithm for the utility function and summary of how the utility function operates computationally will be outlined in the next section.

Formally quantifying the prior knowledge an experimenter has about sizes of effects can prove to be difficult in practice. Most experimenters won't state this knowledge in terms of prior distributions, but will instead have some idea about the range within which they think the size of effect will lie. It is most often reasonable to assume the size of effect can be represented by a normal distribution unless there is strong reasoning from the experimenter that their belief in the size of effect is not symmetrically distributed. Otherwise, the prior mean can be taken as the value lying in the middle of this specified range and the standard deviation taken as $\frac{1}{4}$ of the range. This is given that the range within which the size of the effect is expected to lie can be compared to the central 95% region of the prior distribution.

The form of the prior can also vary according to other aspects such as whether to treat effects separately or assume some dependency between effects as is often the

case. Usually, an experimenter will treat each effect independently, however this is not always the case. According to the principle of factor sparsity, only a small number of the many factors initially considered will contribute significantly to variation in the response. Thus, if each effect is treated separately it would be possible for all effect sizes to result in a large effect size, and therefore for all factors to be considered as active which would not obey the 'law of the vital few'. The lack of information obtained would not provide the experimenter with any further insight and would result in the experiment not being of any benefit, particularly in the case of screening experiments where the principal aim is to screen the many factors for the important few.

Another point to consider is the dependency between the size of an interaction effect and the size of one of its associated parent effects. In view of the effect heredity principle, it would seem sensible to incorporate some prior that takes into account the implication that whether an interaction effect is active or not does depend upon whether another effect is active.

Having formalized the prior distributions for the sizes of effects, the next step is to then determine the candidate designs that the experimenter wishes to consider. This may be to decide between designs which all have the same number of runs but each individual candidate design being able to estimate different effects, or to compare the benefits of candidate designs of different run sizes.

Given the prior information available about the unknown parameters of interest θ, the data to be observed y, the designs to be considered ξ and the terminal decision, d, we are wanting to calculate $U(\xi)$. However, the integration must be done numerically and it is not trivial to evaluate these integrals analytically in the case when the variance is unknown and consequently a prior must be placed on this parameter. Thus, in such cases the evaluation of integrals shall be done through simulations. It should be

noted that the example for which the algorithm for the optimization utility function is given in the next section is with σ^2 known, and so actually in this case the integration would be relatively straightforward.

2.3.4 Algorithm for Optimization Utility Function

For sake of clarity, let us consider two candidate designs, A and B, each with three factors, *a, b* and *c*. We assume the general linear model

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

where the responses are distributed as follows:

$$\mathbf{y}_{abc} \sim N(\mu_{abc}, \sigma^2);\ \sigma^2\, known.$$

Also,

$$\mu_{abc} = \beta_0 + \beta_1 x_a + \beta_2 x_b + \beta_3 x_c + \beta_{12} x_a x_b + \beta_{13} x_a x_c + \beta_{23} x_b x_c + \beta_{123} x_a x_b x_c$$

where $x_a, x_b, x_c = -1$ or $+1$ corresponding to either the low or high level of the factor.

A step-by-step guide of the operations of the utility function is now given:

1. Specify prior $\boldsymbol{\beta}$, where

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_{12} \\ \beta_{13} \\ \beta_{23} \\ \beta_{123} \end{bmatrix}$$

and each of the parameter effects have distributions specified as:

$$\begin{aligned} \beta_0 &\sim N(\mu_0, \sigma_0^2) \\ \beta_1 &\sim N(\mu_1, \sigma_1^2) \\ &\vdots \\ \beta_{123} &\sim N(\mu_{123}, \sigma_{123}^2) \end{aligned}$$

Then simulate from this prior r times to obtain

$$\boldsymbol{\beta}_{(r)} = \begin{bmatrix} \beta_{0(1)} & \cdots & \beta_{0(r)} \\ \beta_{1(1)} & \cdots & \beta_{1(r)} \\ \beta_{2(1)} & \cdots & \beta_{2(r)} \\ \beta_{3(1)} & \cdots & \beta_{3(r)} \\ \beta_{12(1)} & \cdots & \beta_{12(r)} \\ \beta_{13(1)} & \cdots & \beta_{13(r)} \\ \beta_{23(1)} & \cdots & \beta_{23(r)} \\ \beta_{123(1)} & \cdots & \beta_{123(r)} \end{bmatrix}.$$

2. Create the $n \times (p+1)$ candidate design matrix X_i, where n = run size for fractional factorial design and i = A or B.

3. Simulate a set of errors $\epsilon_1, \ldots, \epsilon_n$ for each prior sample and for each candidate design where $\epsilon_i \sim N(0, \sigma^2)$ to obtain

$$\boldsymbol{\epsilon}_{(r)} = \begin{bmatrix} \epsilon_{1(1)} & \cdots & \epsilon_{1(r)} \\ \vdots & \ddots & \vdots \\ \epsilon_{n(1)} & \cdots & \epsilon_{n(r)} \end{bmatrix}.$$

4. Create the $N \times (p+1)$ design matrix X_{full}, where N = run size for full factorial design.

5. For each simulation $j = 1, \ldots, r$ and for each candidate design:

   (a) Use column j from the prior $\boldsymbol{\beta}_{(r)}$.

   (b) Thus, calculate $y_i = X_i\, \boldsymbol{\beta}_{(j)} + \boldsymbol{\epsilon}_{(j)}$ to obtain simulated responses $y_{A(r)}$ and $y_{B(r)}$.

   (c) Use simulated data $y_{A(r)}$ and $y_{B(r)}$, to fit the appropriate model to estimate $\hat{\beta}_{A(r)}$ and $\hat{\beta}_{B(r)}$.

   (d) Using the fitted model, predict $\hat{\mu}$ for all combinations, i.e. $\hat{\mu}_{---}, \ldots, \hat{\mu}_{+++}$ and find estimated optimum $\hat{\mathbf{x}}_{opt}$, i.e. the combination which gives maximum $\hat{\mu}$.

   (e) Obtain true expected responses assuming $\boldsymbol{\beta} = \boldsymbol{\beta}_{(r)}$ and $\boldsymbol{\mu} = X_{full}\boldsymbol{\beta}$. Calculate $\mu_{---}, \ldots, \mu_{+++}$.

   (f) Find μ with true expected optimum $\mathbf{x}_{opt}$.

   (g) Find $\mu_{\hat{\mathbf{x}}_{opt}}$.

   (h) Find $\mu_{\mathbf{x}_0}$, where $\mathbf{x}_0$ = standard operating conditions.

   (i) Hence calculate;

      i. $\mu_{\mathbf{x}_{opt}}$ - $\mu_{\hat{\mathbf{x}}_{opt}}$

      ii. $\mu_{\hat{\mathbf{x}}_{opt}}$ - $\mu_{\mathbf{x}_0}$

   (j) Calculate mean over r samples to obtain $E(\widehat{Utility})$.

2.4 Utility Functions for Screening

2.4.1 Identifying Active and Inactive Effects Correctly

The algorithm mentioned in the previous section would be of benefit to an experimenter in the optimization stage of an investigation. At this stage the experimenter will have narrowed down the factors they are working with and will want to pinpoint the optimum settings of each of these factors to best meet their objective. In the non-clinical setting of the pharmaceutical industry this could be to typically maximise the yield or purity of a drug, or to minimize waste.

However, during the initial stages of investigation when screening experiments are carried out, it is of primary interest to correctly determine the active factors having an effect and driving the process in hand. We can therefore conclude that in screening experiments it would be more serious to miss out on identifying potential active effects, than to incorrectly identify inactive effects as active. This differs to the situation of an optimization experiment, where it would be more serious to have identified inactive effects as active. This would therefore lead to more factors being investigated than are necessary in order to find the optimum treatment combination, and more settings being changed in the process, running the risk of more time and cost being spent than is needed. Of course, in both scenarios of screening and optimization, it would be preferred to not commit either of these errors, however it should be taken into consideration which of these errors would be more serious depending on whether the experiment is for the purposes of screening or optimization.

The more serious of errors in the case of optimization, i.e. declaring that an effect is active when actually it is not, can be thought of as a *Type I error*. In the case of screening, the more serious error is to fail to declare an effect active when in fact it

is, i.e. a *Type II error*. We will consider the null hypothesis to be that the absolute value of an effect is not large enough to be considered active. Then we can define the Type I and Type II errors as follows:

> Type I error = reject the null hypothesis when the null hypothesis is true.
>
> Type II error = fail to reject the null hypothesis when the null hypothesis is false.

Now, given prior distributions that have been specified on effect sizes and some candidate designs we are considering, it would be beneficial to calculate the efficiency of a design in terms of how well the active and inactive effects are declared correctly, and estimate the size of the Type I and II errors. A utility function will be specified in terms of a linear combination of the Type I and Type II errors with some weight, γ ($0 \leq \gamma \leq 1$), on these errors dependent upon the experimenters' belief of the seriousness of each of the errors relative to the experiment. From a Bayesian perspective, we would just want to minimize the posterior probabilities of these events, i.e. the Type I and Type II errors, occurring.

The experimenter may also be interested in taking into account the cost required per run or design. This was considered by DeGroot (1970) and Lindley (1972) where the sampling cost, c, per observation was set out and then incorporated into a cost function. These ideas can be encompassed into the utility function for screening where the cost, c, per run can be considered for the number of runs, n, in the design.

Thus, given that

$$U_1 = P(Type\ I\ error)$$
$$U_2 = P(Type\ II\ error)$$

a quantity used to assess how well a design performs at the screening stage can be

defined as

$$U = \gamma U_1 + (1 - \gamma)U_2.$$

As stated when discussing the loss and gain functions earlier, a utility function is one that typically maximizes something, therefore we actually want to take $-U$. This is due to $\gamma U_1 + (1-\gamma)U_2$ being a quantity that we want to minimize since we clearly wish to minimize any Type I and Type II error probabilities that may occur. However, as similarly stated in the previous section for the optimization results, the results presented in the following chapter are for $U = \gamma U_1 + (1 - \gamma)U_2$, i.e. the quantity that we wish to minimize rather than the utility $-U = -\{\gamma U_1 + (1 - \gamma)U_2\}$ which we would want to maximize.

If cost is to be taken into consideration then this utility, $-U$, is then extended to be $-cnU$.

2.4.2 Decision Table for Screening Utility

Another way to express a screening utility, a quantity with which to discriminate designs on how efficient they are at identifying active and inactive effects, is with a decision table as set out in Table 2.1. This table illustrates the states, actions and consequences in the case of an experiment where the three variables **1**, **2** and **3** are being investigated.

Table 2.1: Decision Table for Screening Utility

	States			
Actions	0	**1**	$\cdots$	**1 + 2 + $\cdots$ +123**
0	1	$-u(c_{0,1})$	$\cdots$	$-u(c_{0,1+2+\cdots+123})$
1	$-u(c_{1,0})$	1	$\cdots$	$-u(c_{1,1+2+\cdots+123})$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
1 + 2 + $\cdots$ + 123	$-u(c_{1+2+\cdots+123,0})$	$-u(c_{1+2+\cdots+123,1})$	$\cdots$	1

The states represent all the possibilities of which effects are truly active, where 0 denotes no effects are active, **1** denotes that only main effect **1** is active and so on. The actions given denote which effects the experimenter has decided to declare as being active. Each value shown in the body of the table gives the utility $u(c_{i,j})$ of the consequence rather than the consequence itself, which was previously set out in §1.4.4. The utility is for each action taken, given the true state for each particular row and column.

Let us denote c_{best} as the best possible consequence within the decision table and c_{worst} as the worst possible consequence. Then the origin and unit of a utility function are arbitrary (French, 1986) and we can set

$$-u(c_{best}) = 1$$

and

$$-u(c_{worst}) = 0.$$

It can be seen that when the experimenter correctly declares all of which effects are active, such as with $c_{0,0}$, $c_{1,1}$, and so on, the outcome of this decision will give a quantity of 1 as the Type I and II errors are 0. This is an intuitive result since clearly

if the 'truth' is that no effects are active and the experimenter makes the correct declaration that no effects are active then this is an efficient decision. The other possible consequences are where some effects have been correctly declared as active or inactive, but the action does not totally coincide with the true state. This is such as in the case of consequence $c_{1,1+2+\cdots+123}$ as given in the table. This is where the experimenter has taken the action to declare effect **1** as active but the true state is that all main effects and interactions involving variables **1**, **2** and **3** are active. Then the quantity of U associated with this consequence will have a value of 0 for U_1 as no inactive effects have been declared as active. However, the value of U_2 will be large as all effects except for **1** have failed to be declared as active. Similarly, for all consequences $c_{i,0}$, the associated quantity U will have a large value of U_1, although varying in size according to how many inactive effects have been incorrectly declared as active, and the value of U_2 will be 0 as there are no inactive effects declared as active incorrectly.

Consider another example where the true state is that main effect **1** is active, but the action is taken to declare all other effects, i.e. **2**, **3** ..., **123** as active, which leads to the consequence $c_{2+3+\ldots+123,1}$. Then the quantity U that we wish to minimize, will be worse than the situations where no effects are declared as active, or if **1** was declared active along with the inactive effects. That is, $c_{0,1}$ and $c_{1+2+3+\ldots+123,1}$ respectively. This is because for the consequence $c_{2+3+\ldots+123,1}$, the Type I error will be large due to **2**, **3** ..., **123** being declared as active incorrectly placing a greater weight on U_1 than if only some of the inactives had been declared as active. Also, there will be some Type II error as main effect **1** is failed to be declared as active, resulting in U_2 taking some non-zero value. Comparing this to the quantities for $c_{0,1}$ and $c_{1+2+3+\ldots+123,1}$, we see that the quantity U for $c_{0,1}$ would be smaller and therefore better than for $c_{2+3+\ldots+123,1}$. This is, since although the value U_2 will be the same, the value of U_1 will be 0. Similarly, for $c_{1+2+3+\ldots+123,1}$, the value of U_1 will be the same as for $c_{2+3+\ldots+123,1}$, however this time U_2 will be 0. Thus, $c_{2+3+\ldots+123,1}$ would be seen as a worse outcome

than the other two possibilities outlined.

A Bayesian or subjective viewpoint can also be incorporated into this decision theory approach where the experimenter can represent their belief in each of the j possible states occurring through subjective probability distributions. The experimenter will have to make this decision based on their knowledge of the process in hand or through the use of historical data. Priors $P(s_j)$ are then placed on each of the j states and represent the belief that state j will occur.

Expected utilities for the action a_i can then be given by

$$E[u(a_i)] = \sum_{j=1}^{n} u(c_{i,j})P(s_j).$$

However, we are concerned with obtaining an expected utility with respect to each of the candidate designs that are being considered. Thus, the utility we would actually be interested in obtaining for each of the k candidate designs being considered would be

$$E[U(Design\ k)] = \sum_{i=1}^{m}\sum_{j=1}^{n} U(Design\ k|c_{i,j})u(c_{i,j})P(s_j).$$

The utility $U(Design\ k|c_{i,j})$ can be considered to be a new set of actions, where the experimenter has taken the action to use design k given that they have declared some set of effects as active (a_i) when there is some true state s_j.

2.4.3 Relevant Literature

Screening experiments investigate a large number of factors in order to identify those few key active effects that are driving the process and having an impact on the response of interest. Designs typically used for the purposes of screening are two-level fractional factorials, which were discussed in detail by Box and Hunter (1961a)

and Finney (1945). Plackett and Burman also notably introduced Plackett-Burman designs (1946). Their paper outlined the construction of economical designs with the run size a multiple of 4, but not a power of two. Plackett-Burman designs are very efficient screening designs when investigating only main effects, however in this situation the assumption that all interactions are negligible has to be made. Thus, main effects will be *confounded* (i.e. confused) with the two-factor interactions. Meyer et al. (1996) considered this problem, where they developed a method for designing a follow-up experiment to resolve ambiguity arising as a result of confounding amongst effects. This followed from previous work developing a Bayesian method based upon the idea of model discrimination in order to uncover active factors, which is a key area of exploration in this thesis.

The criteria proposed by Allen and Bernshteyn (2003) are for purposes relating to model-identification objectives, similarly to that of Meyer et al. (1996). In their paper, simulation optimization studies were used to evaluate the abilities of existing analysis methods to achieve model identification. The results motivated a new class of supersaturated designs, where the probability was maximized that stepwise regression would identify the important main effects. Marley and Woods (2010) also used simulation studies to evaluate designs and compare model selection methods for supersaturated experiments. Similarities can be drawn between the motivations of their paper and this thesis (Chapters 2 and 3), in that several designs of different sizes are studied, with the number and sizes of active effects being altered in order to assess the performance of design and model selection methods.

Another notable paper focusing on a relatively sophisticated approach for model identification using methods based on stochastic search variable selection (SSVS) is that of George and McCulloch (1993). Their increasingly standard assumptions were outlined in the paper, where the key features of a hierarchical model for variable selection were described. That is, each component of $\boldsymbol{\beta}$ (where $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta p)^{\mathrm{T}}$) is modeled

as coming from a mixture of two normal distributions with different variances, and this approach of using a *normal mixture* prior was also adopted by Wu and Hamada (2000). These assumptions are discussed in further detail in the following chapter and are prominently used in several of the simulation studies that are featured.

The approach taken in this chapter, incorporates the model-identification objectives similarly to that of Allen and Bernshteyn (2003) and Meyer et al. (1996), and is applied in the context of screening experiments. Simulation studies will be implemented in the following chapter with several case studies, which includes investigating a Plackett-Burman design and an unreplicated fractional factorial design. This is where the number and sizes of active effects are varied, and the design selection criteria outlined in the following section is then assessed.

2.4.4 Algorithm for Screening Utility Function

Let us now consider the algorithm that will be used to calculate the efficiencies of declaring active and inactive effects correctly based on the Type I and II errors as outlined in §2.4.1. For the sake of clarity, the two candidate designs being considered when illustrating the algorithm will be 5 factors, each at two levels, being studied in a 12 run design and a 16 run design. Thus, candidate design A, the 12 run design has the following design matrix, where the columns assigned to the 5 factors are from a Plackett-Burman design. Depending on whether only main effects are being estimated, then the following design matrix

$$\boldsymbol{X_A} = \begin{bmatrix} + & + & - & + & - & - \\ + & + & + & - & + & - \\ + & - & + & + & - & + \\ + & + & - & + & + & - \\ + & + & + & - & + & + \\ + & + & + & + & - & + \\ + & - & + & + & + & - \\ + & - & - & + & + & + \\ + & - & - & - & + & + \\ + & + & - & - & - & + \\ + & - & + & - & - & - \\ + & - & - & - & - & - \end{bmatrix}$$

will be used. In the case where both main effects and two-factor interactions are being estimated, then the following design matrix

$$\boldsymbol{X_A} = \begin{bmatrix} + & + & - & + & - & - & - & + & - & - & - & + & + & - & - & + \\ + & + & + & - & + & - & + & - & + & - & - & + & - & - & + & - \\ + & - & + & + & - & + & - & - & + & - & + & - & + & - & + & - \\ + & + & - & + & + & - & - & + & + & - & - & - & + & + & - & - \\ + & + & + & - & + & + & + & - & + & + & - & + & + & - & - & + \\ + & + & + & + & - & + & + & + & - & + & + & - & + & - & + & - \\ + & - & + & + & + & - & - & - & - & + & + & + & - & + & - & - \\ + & - & - & + & + & + & + & - & - & - & - & - & - & + & + & + \\ + & - & - & - & + & + & + & + & - & - & + & - & - & - & - & + \\ + & + & - & - & - & + & - & - & - & + & + & + & - & + & - & - \\ + & - & + & - & - & - & - & + & + & + & - & - & - & + & + & + \\ + & - & - & - & - & - & + & + & + & + & + & + & + & + & + & + \end{bmatrix}$$

is used. Similarly for candidate design B, the 16 run design, has one of the following design matrices depending on which effects are being estimated. The following design matrix is used when only main effects are estimated,

$$\boldsymbol{X_B} = \begin{bmatrix} + & - & - & - & - & + \\ + & + & - & - & - & - \\ + & - & + & - & - & - \\ + & + & + & - & - & + \\ + & - & - & + & - & - \\ + & + & - & + & - & + \\ + & - & + & + & - & + \\ + & + & + & + & - & - \\ + & - & - & - & + & - \\ + & + & - & - & + & + \\ + & - & + & - & + & + \\ + & + & + & - & + & - \\ + & - & - & + & + & + \\ + & + & - & + & + & - \\ + & - & + & + & + & - \\ + & + & + & + & + & + \end{bmatrix},$$

or the design matrix below is used when both main effects and two-factor interactions are estimated,

$$\boldsymbol{X_B} = \begin{bmatrix} + & - & - & - & - & + & + & + & + & - & + & + & - & + & - & - \\ + & + & - & - & - & - & - & - & - & - & + & + & + & + & + & + \\ + & - & + & - & - & - & - & + & + & + & - & - & - & + & + & + \\ + & + & + & - & - & + & + & - & - & + & - & - & + & + & - & - \\ + & - & - & + & - & - & + & - & + & + & - & + & + & - & - & + \\ + & + & - & + & - & + & - & + & - & + & - & + & - & - & + & - \\ + & - & + & + & - & + & - & - & + & - & + & - & + & - & + & - \\ + & + & + & + & - & - & + & + & - & - & + & - & - & - & - & + \\ + & - & - & - & + & - & + & + & - & + & + & - & + & - & + & - \\ + & + & - & - & + & + & - & - & + & + & + & - & - & - & - & + \\ + & - & + & - & + & + & - & + & - & - & - & + & + & - & - & + \\ + & + & + & - & + & - & + & - & + & - & - & + & - & - & + & - \\ + & - & - & + & + & + & + & - & - & - & - & - & - & + & + & + \\ + & + & - & + & + & - & - & + & + & - & - & - & + & + & - & - \\ + & - & + & + & + & - & - & - & - & + & + & + & - & + & - & - \\ + & + & + & + & + & + & + & + & + & + & + & + & + & + & + & + \end{bmatrix}.$$

This 2^{5-1} design is a regular fraction and has the design generator $\mathbf{5 = 1234}$.

Since it is reasonable to assume that higher order interactions than two-factor interactions are negligible due to the hierarchical ordering principle, in this illustration effects of order up to two-factor interactions will be considered.

A step-by-step outline of the algorithm is now given:

1. Specify prior $\boldsymbol{\beta}$ and then simulate from this prior r times to obtain $\boldsymbol{\beta}_{(r)}$.

2. For each effect i and each simulation $j = 1, \ldots,$ r, declare an effect $\beta_{i(j)}$ active if for some constant ω, $|\beta_{i(j)}| > \omega$.

3. Create a vector $\delta^*_{(j)}$ of size p, for each simulation $j = 1, \ldots, r$, whose entries are 1 if an effect is active and 0 otherwise.

4. From observing vector $\delta^*_{(j)}$, set q = number of active effects, s = number of inactive effects and p = total number of effects.

5. Create the $n_d \times (p + 1)$ candidate design matrix X_i, where n = run size for fractional factorial design, p = number of effects and d = design A or design B.

6. Simulate a set of r errors $\epsilon_1, \ldots, \epsilon_n$ for each prior sample and for each candidate design where $\epsilon_i \sim N(0, \sigma^2)$ to obtain

$$\boldsymbol{\epsilon}_{(r)} = \begin{bmatrix} \epsilon_{1(1)} & \cdots & \epsilon_{1(r)} \\ \vdots & \ddots & \vdots \\ \epsilon_{n(1)} & \cdots & \epsilon_{n(r)} \end{bmatrix}.$$

7. For each candidate design and each simulation $j = 1, \ldots, r$:

   (a) Use column j from the prior $\boldsymbol{\beta}_{(r)}$ and error $\boldsymbol{\epsilon}_{(r)}$, denoted $\boldsymbol{\beta}_{(j)}$ and $\boldsymbol{\epsilon}_{(j)}$.

   (b) Calculate $y_i = X_i\, \boldsymbol{\beta}_{(j)} + \boldsymbol{\epsilon}_{(j)}$ to obtain simulated responses $y_{A(r)}$ and $y_{B(r)}$.

(c) Use simulated data $y_{A(r)}$ and $y_{B(r)}$, to fit the appropriate model to estimate $\hat{\beta}_{A(r)}$ and $\hat{\beta}_{B(r)}$.

(d) By analysing $\hat{\beta}_{A(r)}$ and $\hat{\beta}_{B(r)}$ using some appropriate method, obtain for each simulated dataset, a vector $\delta_{(j)}$ of size p whose entries are 1 if an effect is active and 0 otherwise.

(e) By comparing each entry $k = 1, \ldots, p$ of $\delta^*_{(j)}$ and $\delta_{(j)}$ obtain:

i.
$$True\ Positive = \begin{cases} 1 & \text{if } \delta^*_{(j)}[k] = 1 \text{ and } \delta_{(j)}[k] = 1 \\ 0 & \text{otherwise.} \end{cases}$$

ii.
$$True\ Negative = \begin{cases} 1 & \text{if } \delta^*_{(j)}[k] = 0 \text{ and } \delta_{(j)}[k] = 0 \\ 0 & \text{otherwise.} \end{cases}$$

iii.
$$False\ Positive = \begin{cases} 1 & \text{if } \delta^*_{(j)}[k] = 0 \text{ and } \delta_{(j)}[k] = 1 \\ 0 & \text{otherwise.} \end{cases}$$

iv.
$$False\ Negative = \begin{cases} 1 & \text{if } \delta^*_{(j)}[k] = 1 \text{ and } \delta_{(j)}[k] = 0 \\ 0 & \text{otherwise.} \end{cases}$$

(f) Hence calculate the probabilities for declaring effects active/inactive either correctly or incorrectly;

i.
$$P(True\ active\ declared\ active) = \frac{\sum_{ij}(True\ Positive)}{qr}$$

ii.

$$P(True\ inactive\ declared\ inactive) \quad = \quad \frac{\sum_{ij}(True\ Negative)}{rs}$$

iii.

$$U_1 = P(Type\ I\ error) \quad = \quad \frac{\sum_{ij}(False\ Positive)}{rs}$$

iv.

$$U_2 = P(Type\ II\ error) \quad = \quad \frac{\sum_{ij}(False\ Negative)}{qr}.$$

(Note that True Actives $= 1 - U_1$ and True Inactives $= 1 - U_2$).

(g) Hence calculate the utility for each design;

$$U = \gamma U_1 + (1 - \gamma)U_2.$$

(h) Calculate mean over r samples to obtain $\widehat{E(U)}$.

2.5 Summary

This chapter has presented standard optimality criteria used to determine the efficiency of designs and also highlighted circumstances where they may not be the most appropriate criteria to use. Some utility functions have been presented which can be used in circumstances where standard optimality criteria may be lacking, namely, utility functions that can be used dependent on the purpose of the experiment, i.e. optimization or screening.

In the following chapter, the utility functions set out above will be applied with some examples to illustrate how they can be used to discriminate between designs. To begin with, a small example is presented comparing two candidate designs of the same size to illustrate the use of the utility function for optimization. An example is then presented where a comparison of designs of different sizes is made, namely a 12-run Plackett-Burman design and a 16-run design, and the utility function for screening is demonstrated. Finally a larger and more complex example will be presented where the utility function for optimization is again demonstrated. In this example, 5 factors will be investigated in two different designs, with one design involving all factors at two levels, and the other design involving all factors at three levels.

Chapter 3

Application of Utility Functions for Optimization and Screening

3.1 Introduction

The utility functions and methods described in Chapter 2 provide a way to obtain a design calculated as being most 'efficient' in terms of providing the most information by generating data sets and then observing the estimated and true optimum treatment combinations.

Some of the drawbacks of standard optimality criteria for a design have been previously highlighted (§2.2). It has also been brought to attention that standard optimality criteria may not always be the most appropriate criteria to answer questions that an experimenter may have. Thus, the candidate designs presented in this chapter are not optimal according to any of the criteria set out in the previous chapter such as A- or D-optimality.

The simulation programs used were written in R and data sets from one thousand experiments were generated from each model for each design using specified prior distributional assumptions. The program used is given in the appendix.

One thousand experiments were found to be a sufficient number of simulations to calculate the estimated utilities for optimization with a reasonable degree of confidence and reliability. The standard errors were calculated for each of the different forms of prior, for the mean of the estimated loss, gain and relative gain. As displayed in Table 3.1, the 95% confidence intervals of the estimated optimization utilities are given for prior 1 (which shall be outlined later), for each of the three different forms of prior. The 95% confidence interval of the estimated utility is $\bar{x} \pm 1.96\frac{sd}{\sqrt{n}}$, where $\bar{x}$ and sd are the mean and standard deviation of the estimated utility, $\widehat{E(Utility)}$, and n is the simulation size. To give an indication of the computing time, the average time taken per n simulations for prior 1 is displayed in Table 3.2. The computing times indicated in the table are with reference to running simulations for the optimization utility function, for both designs A and B (designs A and B with regards to the optimization utility function shall be outlined in the following section).

The same number of experiments, i.e. one thousand experiments, were also found to be a sufficient number of simulations to calculate the estimated utilities for screening within a reasonable degree of confidence. The 95% confidence intervals of the estimated screening utilities are shown in Table 3.3. This is for prior 1 in the case of the normal prior and prior 1(a) in the case of the normal mixture prior, looking at main effects along with two-factor interactions. This shall be explained in more detail later on in this chapter. In this table the utility U displayed is calculated as $\gamma U_1 + (1-\gamma)U_2$, where U_1 and U_2 are the probabilities of Type I and Type II errors respectively. γ is the weight of importance placed on a Type I error and consequently $1-\gamma$ the weight of importance on a Type II error. Where method 1 has been stated, this is referring to the utility being calculated based upon an effect being declared active if

it is greater than some constant ω. This constant ω is one that the experimenter will have set out. Method 2 indicates a utility which has been calculated based upon an effect being declared active using Lenth's method. Table 3.4 shows the time taken to run the simulations for the screening utility function for design A (where design A is the 16 run design and design B is the 12 run design as outlined in §3.3). The results in Table 3.3 and 3.4 are for when $\omega = \sigma$.

Table 3.1: 95% Confidence Interval for Estimated Optimization Utilities

	500 Simulations		1000 Simulations		10000 Simulations	
	Design A	Design B	Design A	Design B	Design A	Design B
Point Prior						
Mean Loss	4 ± 0	2 ± 0	4 ± 0	2 ± 0	4 ± 0	2 ± 0
Mean Gain	26 ± 0	28 ± 0	26 ± 0	28 ± 0	26 ± 0	28 ± 0
Relative Gain	0.867 ± 0	0.933 ± 0	0.867 ± 0	0.933 ± 0	0.867 ± 0	0.933 ± 0
Normal Prior						
Mean Loss	4.265 ± 0.306	2.286 ± 0.143	3.994 ± 0.223	1.997 ± 0.104	4.360 ± 0.069	2.362 ± 0.031
Mean Gain	26.145 ± 0.339	28.124 ± 0.294	25.974 ± 0.239	27.971 ± 0.212	25.975 ± 0.078	27.972 ± 0.067
Relative Gain	0.864 ± 0.010	0.927 ± 0.004	0.867 ± 0.006	0.934 ± 0.004	0.860 ± 0.002	0.924 ± 0.000
Normal Mixture Prior						
Mean Loss	7.421 ± 0.817	9.318 ± 0.484	5.973 ± 0.617	7.026 ± 0.353	7.320 ± 0.184	9.030 ± 0.127
Mean Gain	5.145 ± 0.858	9.789 ± 0.408	6.093 ± 0.617	5.040 ± 0.316	5.626 ± 0.186	3.916 ± 0.092
Relative Gain	0.373 ± 0.063	0.710 ± 0.029	0.510 ± 0.041	0.426 ± 0.020	0.409 ± 0.014	0.304 ± 0.006

Table 3.2: Computing Time Per n Simulations (seconds) - Optimization Utility

	n = 500 Simulations	n = 1000 Simulations	n = 10000 Simulations
Point Prior	2.98	4.53	10.90
Normal Prior	2.54	3.13	10.79
Normal Mixture Prior	4.29	6.32	24.84

Table 3.3: 95% Confidence Interval for Estimated Screening Utilities

	500 Simulations		1000 Simulations		10000 Simulations	
	Design A	Design B	Design A	Design B	Design A	Design B
Normal Prior						
Method 1	0.229 ± 0.008	0.391 ± 0.016	0.243 ± 0.006	0.381 ± 0.011	0.237 ± 0.002	0.384 ± 0.004
Method 2	0.571 ± 0.033	0.439 ± 0.010	0.603 ± 0.023	0.442 ± 0.007	0.597 ± 0.007	0.437 ± 0.002
Normal Mixture Prior						
Method 1	0.020 ± 0.001	0.170 ± 0.013	0.015 ± 0.001	0.149 ± 0.008	0.018 ± 0.000	0.164 ± 0.003
Method 2	0.048 ± 0.002	0.392 ± 0.012	0.044 ± 0.001	0.374 ± 0.007	0.044 ± 0.000	0.387 ± 0.003

Table 3.4: Computing Time Per n Simulations (seconds) - Screening Utility

	n = 500 Simulations	n = 1000 Simulations	n = 10000 Simulations
Normal Prior	7.61	10.76	284.79
Normal Mixture Prior	16.86	21.26	306.91

3.2 Optimization: Comparison of 2^{3-1} Designs

3.2.1 Introduction

Let us first consider a small example comparing two designs of the same size to illustrate the concepts of the decision theoretic approach and utility functions that we shall use, before a larger case is presented later on in this chapter. The criteria used to compare the performance of the designs will be the optimization utility function, so the mean loss, mean gain and relative gain will be looked at.

Consider a scenario where an experimenter can study three factors, all at two levels, and there are only 4 runs available in which to study them. The full 2^3 factorial design is given in Table 3.5. Thus, the decision to be made in this situation is which candidate design (Table 3.6) should be used?

Table 3.5: Full 2^3 factorial design

	Variable		
Run	**1**	**2**	**3**
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

It should first be noted that this is, as mentioned, a small example to illustrate the use of the utility functions before larger designs are presented. Studying three factors in 4 runs would typically not be carried out in practice by an experimenter, the main reason being that a 2^{3-1} design would most likely be too small for an experimenter

Table 3.6: Candidate Half-fractions of 2^3 Factorial design

Candidate Design A

	Variable		
Run	**1**	**2**	**3**
1	-	-	-
2	-	+	+
3	+	-	+
4	+	+	-

Candidate Design B

	Variable		
Run	**1**	**2**	**3**
1	-	-	0
2	+	-	0
3	-	+	0
4	+	+	0

to see anything and make any insightful inferences.

Candidate design A has the defining relation $\mathbf{I}$ = **-123** and is the *complementary half-fraction* of the design with defining relation $\mathbf{I}$ = **123**, i.e. the two half-fractions together make up the complete 2^3 factorial design. All the aliasing relations for this design are

$$\mathbf{1 = -23, 2 = -13, 3 = -12}.$$

Reasons for choosing a defining relation with a minus sign, which therefore also results in defining contrast words with a minus sign, are considered by Gilmour (2001) in the contribution to the discussion of the paper by Lewis and Dean (2001). It was stated that in circumstances when experimenters are able to specify which factor level is likely to give a higher response, therefore ensuring that all main effects are expected to be positive, then the two-factor interactions are expected to be negative due to limitations upon how much improvement can be made. Gilmour thus outlined that in defining contrasts, a plus sign should be used with even-lettered words and a minus sign with odd-lettered words. A result of doing so avoids cancellation of effects and inflates moderate sized effects.

The full model with all effects associated with variables **1**, **2** and **3** is

$$\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3,$$

where μ is the mean response, β_0 is the average at standard operating conditions, β_i is the parameter variable i, and similarly β_{ij} and β_{ijk} are effects ij and ijk.

However, there are not enough degrees of freedom to estimate all effects and another disadvantage of this design is that all main effects are aliased with two-factor interactions. Thus, none of the effects are clear and to be able to estimate the main effects a strong assumption that two-factor interactions are negligible will have to made. So in fact the model that is being estimated is

$$\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3.$$

Candidate design B is constructed by allocating minuses and pluses to variables **1** and **2** by standard ordering, and variable **3** is held fixed at 0. The treatment combination (0, 0, 0) for variables **1**, **2** and **3** is assumed to be standard operating conditions.

Candidate design B has variable **3** fixed at 0, and so the main effect of variable **3** cannot be estimated nor can any of the interactions involving the variable **3**. Due to main effect **3** and its associated interactions being inestimable, this design would typically only be chosen in cases when there is a reasonable amount of information to suggest the size of the variable **3** effect will be small or inactive. As a result of variable **3** being held fixed at 0, the effects that are estimated are done so clearly, i.e. there are no aliased effects.

The model being estimated for this design is

$$\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{12} x_1 x_2.$$

As discussed previously, careful consideration needs to be taken in formulating the priors for the sizes of effects in order to provide a correct representation. In this

example to compare the performances of designs using the loss and gain utility criteria, three different forms of priors will be used. From simulation of data sets, results of the mean loss and gain are presented and also how the choice of prior affects the data is seen.

3.2.2 Choice of Prior

The first form of prior chosen is having a known size of effect in the case where an experimenter has very strong prior knowledge about the effects being investigated. In practice, this is probably not very realistic but may be of interest to experimenters at Pfizer to see what would be the result if they did in fact know exactly how certain factors behave.

The second form of prior chosen is for all effects to be modelled by a normal distribution and for all effects to be treated independently, i.e. assuming that the size of an effect does not depend on the size of another effect. This can be shown as:

$$\begin{aligned} \beta_0 &\sim N(\mu_0, \sigma_0^2) \\ \beta_1 &\sim N(\mu_1, \sigma_1^2) \\ &\vdots \\ \beta_{123} &\sim N(\mu_{123}, \sigma_{123}^2) \end{aligned}$$

The final choice of prior will be to represent effects by normal mixture priors, in a similar fashion to that presented by Chipman, Hamada and Wu (1997). This approach

is considered for the general linear model,

$$\mathbf{y} = \mathbf{X}\beta + \epsilon.$$

A vector α, of length p, will be used to indicate the importance of main effects and interaction effects. Its values will take either 0 if the effect is inactive, or 1 if the effect is active, and is determined by some probability p_i of being active. The value of probability p_i will be somewhere in the range $(0, 0.5]$ to account for the fact that a relatively small number of effects are expected to be active, and typically a value of 0.25 would be a sensible value to choose. Taking this value appropriately represents the principle of effect sparsity.
So the values of α describing the main effects β_i will be generated from

$$\alpha_i \sim Bern(p_i).$$

Then if $\alpha_i = 0$ and therefore inactive, the size of β_i is small. Likewise when $\alpha_i = 1$ the size of β_i is large. A normal mixture prior $p(\beta_i|\alpha_i)$ for the coefficients β_i is thus specified as follows:

$$\beta_i \sim \begin{cases} N(0, \sigma_{1i}^2) & \text{if } \alpha_i = 0 \\ N(0, \sigma_{2i}^2) & \text{if } \alpha_i = 1. \end{cases} \tag{3.1}$$

In the case when $\alpha_i = 0$, i.e. β_i comes from the distribution $N(0, \sigma_{1i}^2)$, the variance σ_{1i}^2 needs to be specified such that β_i is tightly centred about 0 and hence does not have a large effect. In contrast, when $\alpha_i = 1$, the variance σ_{2i}^2 needs to be chosen carefully with size much greater than σ_{1i}^2, so that β_i comes from a distribution with high variance to allow for the possibility of it having a large effect.

Typically, the independence prior

$$p(\alpha) = \prod_{i=1}^{p+1} p_i^{\alpha_i}(1 - p_i)^{1-\alpha_i}$$

will be used to imply that the *activity* of a variable (i.e. whether a variable is active or inactive) does not depend upon whether another variable is active or not, i.e. that

the variables are independent. (Note that in reference to this independence prior, $p_i = P(\alpha_i = 1)$). However, this may not be considered to be true in light of the effect heredity principle and so to use the independence prior when modelling interaction effects is not appropriate. Hierarchical priors, where prior distributions are assumed based upon the principles of conditional independence and inheritance should be used (Chipman, 1996). This is also discussed by Wu and Hamada (2000).

Consider the candidate designs A and B mentioned previously, and all the effects involved from the three main effects **1**, **2** and **3**, to the three-factor interaction **123**. Then $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_{12}, \alpha_{13}, \alpha_{23}, \alpha_{123})$, and according to the *effect heredity principle* an interaction effect, say **12**, will be active dependent upon the activity of the parent main effects **1** and **2**. The term *parent* refers to the factors making up the interaction, thus the parents for the interaction **12** are the main effects **1** and **2**.

This idea can be represented for the two-factor interactions **ij**, where they are dependent upon the parent main effects **i** and **j**, by the following probabilities:

$$p_{ij} = \begin{cases} p_{00} & \text{if } (\alpha_i,\, \alpha_j) = (0,0) \\ p_{01} & \text{if } (\alpha_i,\, \alpha_j) = (0,1) \\ p_{10} & \text{if } (\alpha_i,\, \alpha_j) = (1,0) \\ p_{11} & \text{if } (\alpha_i,\, \alpha_j) = (1,1). \end{cases}$$

The value of p_{ij} will again be chosen to be less than 0.5 to represent the expectation that only a few terms will be truly active. The value of p_{00} is usually chosen to be small, for example 0.01, to represent the idea that an interaction effect with no active parent effects is highly unlikely to be active. Also, p_{01} and p_{10} will be slightly larger, say about 0.10, to represent that if one of the parent effects is active it is more likely the interaction effect will be active. Finally, p_{11} will typically be chosen to be somewhere in the same region as the probability of a main effect being active, which is 0.25. This illustrates the idea that if both parent effects are active then the

interaction effect is more likely to be active.

The value of α_{ij} is then generated from the distribution,

$$\alpha_{ij} \sim Bern(p_{ij})$$

where the normal mixture prior $p(\beta_{ij}|\alpha_{ij})$ for the interaction effects is specified as follows:

$$\beta_{ij} \sim \begin{cases} N(0, \sigma^2_{1ij}) & \text{if } \alpha_{ij} = 0 \\ N(0, \sigma^2_{2ij}) & \text{if } \alpha_{ij} = 1. \end{cases}$$

This similarly forms the basis for how the normal mixture prior for the three-factor interaction is set up. The idea that the three-factor interaction **ijk** is dependent upon the parent main effects **1**, **2** and **3** is represented by the following probabilities:

$$p_{ijk} = \begin{cases} p_{000} & \text{if } \alpha_i + \alpha_j + \alpha_k = 0 \\ p_{100} & \text{if } \alpha_i + \alpha_j + \alpha_k = 1 \\ p_{110} & \text{if } \alpha_i + \alpha_j + \alpha_k = 2 \\ p_{111} & \text{if } \alpha_i + \alpha_j + \alpha_k = 3. \end{cases}$$

Probability p_{000} represents an interaction effect with no active parent effects. This will typically be chosen to be very small, for example 0.001, to demonstrate the idea that it is highly unlikely for a three-factor interaction effect to be active with no active parent effects. The probability p_{100} represents the three-factor interaction effects (1,0,0), (0,1,0) and (0,0,1), i.e. a three-factor interaction effect with one active parent effect. The size of this probability will be larger than p_{000}, but slightly smaller than p_{01} or p_{10}, say about 0.05. The probability p_{110} represents the three-factor interaction effects (1,1,0), (1,0,1) and (0,1,1), i.e. a three-factor interaction with two active parent effects. This will be larger than p_{100}, say about 0.1. Finally, the probability p_{111} represents the three factor interaction effect with all parent effects active. Thus, the size of this probability will be in the same region as the probability, p, that a

main effect is active, say about 0.25. As stated, the normal mixture prior for the three-factor interaction is set up so that the interaction **ijk** is dependent upon the parent main effects **i**, **j** and **k**. It would then also be appropriate to assume that the three-factor interaction is also dependent upon the two-factor interactions. However, in this instance we are assuming that the three-factor interaction is conditionally independent of the two-factor interactions due to computational simplicity for the simulation work.

The value α_{ijk} is distributed as follows;

$$\alpha_{ijk} \sim Bern(p_{ijk}).$$

The normal mixture prior $p(\beta_{ijk}|\alpha_{ijk})$ for the three factor interaction effect is then specified as:

$$\beta_{ijk} \sim \begin{cases} N(0, \sigma^2_{1ijk}) & \text{if } \alpha_{ijk} = 0 \\ N(0, \sigma^2_{2ijk}) & \text{if } \alpha_{ijk} = 1. \end{cases}$$

Now that the different priors that have been used to generate data have been demonstrated, the prior distributions on the effect sizes used for the simulations are given in the following section.

3.2.3 Prior Distributions on the Effect Sizes

The prior distributions on the effect sizes used for the simulation work are now discussed. All tables referred to in this section are given at the end of the section.

Throughout, the models used to simulate data are for

$$\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_{12}, \beta_{13}, \beta_{23}, \beta_{123}]^{\mathrm{T}}.$$

Given a prior distribution specified for the p parameters of interest, $\boldsymbol{\beta}$, the data $\mathbf{y}$

collected from the experiment consisting of N observations which uses the design with design matrix $\mathbf{X}$, it is then assumed that

$$\mathbf{y}|\boldsymbol{\beta} \sim MVN_N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \tag{3.2}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix.

If prior knowledge about the parameters of interest $\boldsymbol{\beta}$ can be sufficiently expressed by a prior distribution of the *conjugate form*, i.e. where the posterior distribution is of the same form as the prior distribution, then the form of the posterior distribution is more straightforward to manipulate. That is, the analysis and interpretation of results that are following from the combination of prior information about the parameters contained in the observed data to give a composite picture of the final judgments about the values of the parameter, are simplified.

Thus, the model used to describe $\mathbf{y}|\boldsymbol{\beta}$ which is of the multivariate normal form as given previously, has the multivariate normal conjugate prior,

$$\boldsymbol{\beta} \sim MVN_p(\mu_p, \Sigma_p).$$

Since the priors on the parameters of interest $\boldsymbol{\beta}$ are individually represented by normal distributions, this can be extended to all p parameters of interest being represented by the multivariate normal distribution. The case of the normal mixture prior shall be considered later on in this section. It can be shown using Bayes' Theorem that the posterior distribution of $\boldsymbol{\beta}|\mathbf{y}$ is

$$\boldsymbol{\beta}|\mathbf{y} \sim MVN_p(\mu, \Sigma)$$

where

$$\mu = [\frac{1}{\sigma^2}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \Sigma_p^{-1}]^{-1}[\frac{1}{\sigma^2}\mathbf{X}^{\mathrm{T}}\mathbf{y} + \Sigma_p^{-1}\mu_p]$$

and

$$\Sigma = [\frac{1}{\sigma^2}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \Sigma_p^{-1}]^{-1}.$$

Values of μ_p and Σ are given in tables that are to follow. Where Σ is specified for the normal prior distributions, this is assumed to be the diagonal variance matrix, i.e. the diagonal variance entries for the variance-covariance matrix Σ_p, and throughout σ^2 is assumed to be known. The variance of the intercept, β_0, for all normal prior distributions shall be assumed to approach ∞. It is usual to do this since we are not generally concerned with estimating the intercept, but more so with estimating the main effects and interaction effects and also because the utility functions do not depend upon β_0.

When setting up the normal mixture prior, we shall consider (3.2) as part of a larger hierarchical model as set out by George and McCulloch (1993). As mentioned in the previous section, the key idea is that each component of $\boldsymbol{\beta}$ is modelled as coming from a mixture of two normal distributions with different variances, as seen in (3.1). Although this is similar to the normal mixture prior as set out by Mitchell and Beauchamp (1988), in the sense of being a mixture of two normal distributions, a distinct difference is that they in fact consider a "spike and slab" mixture. Their approach has β_i uniformly distributed between two limits except for some part of the probability mass of β_i which is concentrated at 0. This demonstrates the idea of an effect being 'active' if it comes from somewhere in the uniform distribution and 'inactive' if it has value 0. Although there have been previous references in this thesis to an effect being 'inactive', this is in the sense that an effect has a size that could be considered to be negligible and is therefore not large enough to be considered one of the key effects driving the process in hand. Thus, where an effect is referred to as being 'inactive' it should be considered in this sense rather than an effect having a size of 0.

The vector α, which was previously discussed, captures the importance of effects where $\alpha_i = 0$ or 1, for $i = 1, \ldots, p$. This is where p is the number of parameters being

considered. Then the normal mixture is represented as

$$\beta_i|\alpha_i \sim (1-\alpha_i)N(0,\sigma_{1i}^2) + \alpha_i N(0,\sigma_{2i}^2).$$

The prior for $\beta_i|\alpha_i$ then takes the form of the multivariate normal prior where

$$\beta|\alpha \sim N_p(0, \mathbf{D}_\alpha \mathbf{R} \mathbf{D}_\alpha).$$

$\alpha = (\alpha_1, \ldots, \alpha_p)$, $\mathbf{R}$ is the prior correlation matrix and $\mathbf{D}_\alpha \equiv diag[a_1, \ldots, a_p]$, where $a_i = \sigma_{1i}^2$ if $\alpha_i = 0$ and $a_i = \sigma_{2i}^2$ if $\alpha_i = 1$. The value of σ_{1i}^2 is set to be small so that those β_i where $\alpha_i = 0$, will come from a distribution with low variance and be clustered somewhere about 0. σ_{2i}^2 is chosen to be large ($\sigma_{2i}^2 \gg \sigma_{1i}^2$), and therefore those β_i for which $\alpha_i = 1$ will come from the distribution with high variance and be much more dispersed.

Typically, the prior on the residual variance σ^2 takes the form of an inverse gamma conjugate prior as follows,

$$\sigma^2 \sim IG(\nu/2, \nu\lambda/2).$$

This is equivalent to $\nu\lambda/\sigma^2 \sim \chi_\nu^2$. However, for the simulation work, it will be assumed that σ^2 is known and is equal to 1.

The values for μ_p and Σ_p, assuming that strong prior information is known, are given in Table 3.7. The diagonal variance matrix entries are assumed to be 0 for this strong prior information. This is a very unrealistic assumption to make, since it is highly unlikely that an experimenter will know exactly the outcome of the effect sizes. However, the scenario of assuming effect sizes are known is primarily for illustration rather than investigatory purposes, before more interesting prior distributions are displayed. Some points to note about the prior distributions for the strong prior distributions, assuming known sizes of effects, are:

1. For prior 1 there are expected to be two large main effects, **1** and **2**, a large

two-factor interaction **12**, a fairly small main effect **3** and all other effect sizes small.

2. For prior 2, it is expected that all three main effects will be large and all other effects small.

3. Priors 3 and 4 both indicate the expectation of all three main effects being large, one large two-factor interaction, and all other effects small.

4. There are expected to be three large main effects and two large two-factor interactions, **12** and **13**, resulting from prior 5.

5. Priors 6, 7 and 12 indicate the expectation of two large main effects and one large two-factor interaction.

6. For prior 8 it is expected that the two main effects **1** and **2** will be large and all other effects small.

7. Priors 9, 10 and 11 all indicate the expectation of one large main effect and one large two-factor interaction.

8. For prior 13, there are expected to be two large main effects, **2** and **3**, and two large two-factor interactions, **12** and **13**.

Some further priors, which take the form of the normal distribution, are given in Tables 3.8, 3.9 and 3.10. Each of the priors in these three tables contain useful information about the expectation of effect sizes. The sizes of the main effects, two-factor and three-factor interactions correspond with those set out in points 1-8 as given previously for the strong prior information. However, the three tables do vary according to how informative the priors are, ranging from informative priors with low prior variances (3.8), to somewhat informative priors (3.9), and finally to non-

informative priors with high prior variances (3.10).

Some additional, and perhaps more interesting, normal prior distributions are given in Table 3.11. Points to note are:

1. Prior distributions 1-3 are non-directional - they express the belief that there is no indication about the likely direction of effects. The variances differ from high prior variances to somewhat more informative variances.

2. Priors 4 and 5 are also non-directional. Looking at prior 4 we see that all main effects have high prior variances and all two-factor interactions have somewhat informative variances. The three-factor interaction has an informative variance where it is expected this interaction will be somewhere close to 0. The converse is true for prior 5 in the case of the main effects, where this time the variances are informative. The two-factor interactions for prior 5 are non-informative with high prior variance and the three-factor interaction is somewhat informative.

3. Prior 6 is not highly informative but does contain some useful information. There is expected to be one large main effect, one fairly large main effect and one fairly small main effect. All the interaction effects are small except for that which has large main effects.

4. The means for prior 7 are consistent with those of prior 6. However, there is more prior information in this case about those effects which are expected to be large and little information about other effects.

5. Prior 8 indicates the expectation of two large main effects, one with much less certainty than the other, and one small main effect. The interaction effect which has large parent main effects is expected to be fairly large with more certainty than all other interaction effects, which are expected to be small.

6. Prior 9 denotes that there is expected to be one large main effect with a high level of certainty and two fairly large main effects but with much less certainty. All two-factor interactions are expected to be somewhat small with the three-factor interaction much smaller. These are all with fairly uninformative variances.

7. Prior 10 corresponds with prior 9 in the sense that there is also expected to be one large main effect with a high level of certainty and two fairly large main effects with much less certainty. However, there is also expected to be one fairly large two-factor interaction, although this is not with much certainty. All other two- and three-factor interactions are expected to be small also with a low level of certainty.

The values set out in Table 3.12, are the probabilities of the main effects and interaction effects for variables **1**, **2** and **3** being active. The probabilities p_1, p_2 and p_3 represent the main effects being active and p_{00}, p_{01}, p_{10} and p_{11} represent the probabilities for a two-factor interaction effect being active. The probability p_{00} is for the interaction effects with no active parent effects, p_{01} and p_{10} for interaction effects with one active parent effect and p_{11} for an interaction effect with both parents active, as outlined in the previous section. The probabilities p_{000}, p_{100}, p_{110} and p_{111} similarly represent the three-factor interaction effect **123**, dependent upon whether there are respectively no, one, two or three active parents. These probabilities are then used to generate the vector α, to indicate the importance of main effects and interaction effects. These are generated using the fact that the entries of the vector α come from a Bernoulli distribution with relevant probability, depending on whether the particular entry of α is referring to a main, two-factor interaction or three-factor interaction effect.

For each of the effects, $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_{12}, \boldsymbol{\beta}_{13}, \boldsymbol{\beta}_{23}, \boldsymbol{\beta}_{123}$, i.e. for all $\boldsymbol{\beta}_i$ where $i = 1, \ldots, p$,

the normal components are

$$\beta_i \sim \begin{cases} N(0, 0.04) & \text{if } \alpha_i = 0; \\ N(0, 4) & \text{if } \alpha_i = 1. \end{cases} \tag{3.3}$$

Some points to note about the normal mixture priors given in Table 3.12, is that owing to the Bernoulli distributions set out for the values α_i:

1. Prior 1 indicates that there are expected to be three large main effects.

2. Priors 2, 3 and 4 indicate that there are expected to be two large main effects.

3. Priors 5, 6 and 7 indicate that there is expected to be one large main effect.

4. Prior 8 indicates that all main effects are expected to be small.

5. According to prior 9 all main effects are independent and all have an equal probability of being active. The probability of two-factor and three-factor interactions being active is smaller than for the main effects.

6. Looking at priors 10 and 11, we see that all effects are independent and have an equal probability of being active. This probability is smaller for prior 11 than prior 10.

The normal mixture prior is set up so that the higher order interactions are dependent upon whether the main effects are active or not, for priors 1-8. It may be expected that there will be a greater number of active two-factor interaction effects for prior 1 in comparison to some of the other scenarios considered. Conversely, it may be expected that there is the smallest number of active two-factor interaction effects for priors 5, 6 and 7. In the case of priors 9-11, we observe that the principle of effect heredity is not taken into account. For instance, with prior 9 we note that $p_{00} = p_{01} = p_{10} = p_{11}$. Hence, a two-factor interaction will have the same probability

of being active regardless of whether one or both or neither of its parent main effects are active. This is similarly observed for priors 10 and 11 where in fact we have that $p_1 = p_2 = p_3 = p_{00} = p_{01} = p_{10} = p_{11} = p_{000} = p_{100} = p_{110} = p_{111}$ in both cases. Although we would expect that effects are likely to be active according to the effect heredity principle, it should be noted that effect heredity is only an assumption and so to consider the circumstance where effect heredity does not apply is a reasonable assumption to make.

Note also that the probabilities given in Table 3.12 are much more extreme than set out in the previous section. It was stated that typically a main effect will be active with probability 0.25. However for the simulation work the probability has been set at either 0.75 or 0.8 when $\alpha_i = 1$ for variable i. More extreme values are also seen for the probabilities which have been set for an interaction effect to be active. To illustrate this for one case, when a two factor interaction **ij**, with parent main effects **i** and **j** have values of α_i and α_j both being generated to be 0, the resulting probability of p_{00} would usually be set to be very small such as 0.01 to represent the idea that an interaction effect with no active parent main effect is highly unlikely to be active. However, in the situation of the simulation work, the probability for p_{00} has been chosen to be 0.1. The reason for the probabilities being more extreme than as typically chosen in literature is due to the size of the designs being relatively small, where $n = 4$. Thus, to obtain results where some effects are being declared as active the probabilities have been chosen to be larger than perhaps would normally be chosen. However, it could also be argued that in practice an interaction effect could be active even in the circumstance when neither of the parent main effects are active, more so than the probability of $p_{00} = 0.01$ suggests and so to choose p_{00} to be as small as this would not be an accurate reflection of circumstances occurring in the real world.

Prior Distributions on the Effect Sizes

Table 3.7: Strong Prior Information - Point Prior

Prior	μ_p	Σ
1	[0, 10, 10, 2, 8, 0, 0, 0]	[0, 0, 0, 0, 0, 0, 0, 0]
2	[0, 10, 10, 10, 0, 0, 0, 0]	[0, 0, 0, 0, 0, 0, 0 ,0]
3	[0, 10, 10, 10, 0, 8, 0, 0]	[0, 0, 0, 0, 0, 0, 0 ,0]
4	[0, 10, 10, 10, 8, 0, 0, 0]	[0, 0, 0, 0, 0, 0, 0 ,0]
5	[0, 10, 10, 10, 8, 8, 0, 0]	[0, 0, 0, 0, 0, 0, 0 ,0]
6	[0, 10, 0, 10, 0, 8, 0, 0]	[0, 0, 0, 0, 0, 0, 0 ,0]
7	[0, 10, 10, 0, 0, 8, 0, 0]	[0, 0, 0, 0, 0, 0, 0, 0]
8	[0, 10, 10, 0, 0, 0, 0, 0]	[0, 0, 0, 0, 0, 0, 0 ,0]
9	[0, 0, 0, 10, 0, 8, 0, 0]	[0, 0, 0, 0, 0, 0, 0 ,0]
10	[0, 0, 0, 10, 8, 0, 0, 0]	[0, 0, 0, 0, 0, 0, 0 ,0]
11	[0, 10, 0, 0, 0, 8, 0, 0]	[0, 0, 0, 0, 0, 0, 0 ,0]
12	[0, 0, 10, 10, 8, 0, 0, 0]	[0, 0, 0, 0, 0, 0, 0 ,0]
13	[0, 0, 10, 10, 8, 8, 0, 0]	[0, 0, 0, 0, 0, 0, 0 ,0]

Table 3.8: Normal Prior Distribution on Effect Sizes: Informative Prior

Prior	μ_p	Σ
1	[0, 10, 10, 2, 8, 0, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]
2	[0, 10, 10, 10, 0, 0, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]
3	[0, 10, 10, 10, 0, 8, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]
4	[0, 10, 10, 10, 8, 0, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]
5	[0, 10, 10, 10, 8, 8, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]
6	[0, 10, 0, 10, 0, 8, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]
7	[0, 10, 10, 0, 0, 8, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]
8	[0, 10, 10, 0, 0, 0, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]
9	[0, 0, 0, 10, 0, 8, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]
10	[0, 0, 0, 10, 8, 0, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]
11	[0, 10, 0, 0, 0, 8, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]
12	[0, 0, 10, 10, 8, 0, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]
13	[0, 0, 10, 10, 8, 8, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]

Table 3.9: Normal Prior Distribution on Effect Sizes: Mildly Informative Prior

Prior	μ_p	Σ
1	[0, 10, 10, 2, 8, 0, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]
2	[0, 10, 10, 10, 0, 0, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]
3	[0, 10, 10, 10, 0, 8, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]
4	[0, 10, 10, 10, 8, 0, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]
5	[0, 10, 10, 10, 8, 8, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]
6	[0, 10, 0, 10, 0, 8, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]
7	[0, 10, 10, 0, 0, 8, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]
8	[0, 10, 10, 0, 0, 0, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]
9	[0, 0, 0, 10, 0, 8, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]
10	[0, 0, 0, 10, 8, 0, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]
11	[0, 10, 0, 0, 0, 8, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]
12	[0, 0, 10, 10, 8, 0, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]
13	[0, 0, 10, 10, 8, 8, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]

Table 3.10: Normal Prior Distribution on Effect Sizes: Weakly Informative Prior

Prior	μ_p	Σ
1	[0, 10, 10, 2, 8, 0, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]
2	[0, 10, 10, 10, 0, 0, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]
3	[0, 10, 10, 10, 0, 8, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]
4	[0, 10, 10, 10, 8, 0, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]
5	[0, 10, 10, 10, 8, 8, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]
6	[0, 10, 0, 10, 0, 8, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]
7	[0, 10, 10, 0, 0, 8, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]
8	[0, 10, 10, 0, 0, 0, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]
9	[0, 0, 0, 10, 0, 8, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]
10	[0, 0, 0, 10, 8, 0, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]
11	[0, 10, 0, 0, 0, 8, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]
12	[0, 0, 10, 10, 8, 0, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]
13	[0, 0, 10, 10, 8, 8, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]

Table 3.11: Additional Normal Prior Distributions

Prior	μ_p	Σ
1	[0, 0, 0, 0, 0, 0, 0, 0]	[∞, 4, 4, 4, 4, 4, 4, 4]
2	[0, 0, 0, 0, 0, 0, 0, 0]	[∞, 2, 2, 2, 2, 2, 2, 2]
3	[0, 0, 0, 0, 0, 0, 0, 0]	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]
4	[0, 0, 0, 0, 0, 0, 0, 0]	[∞, 4, 4, 4, 2, 2, 2, 0.25]
5	[0, 0, 0, 0, 0, 0, 0, 0]	[∞, 0.25, 0.25, 0.25, 4, 4, 4, 2]
6	[0, 10, 6, 2, 4, 1, 1, 1]	[∞, 2, 2, 2, 2, 2, 2, 2]
7	[0, 10, 6, 2, 4, 1, 1, 1]	[∞, 0.25, 0.25, 2, 0.25, 2, 2, 2]
8	[0, 10, 10, 2, 4, 1, 1, 1]	[∞, 0.25, 2, 2, 1, 2, 2, 2]
9	[0, 10, 4, 4, 2, 2, 2, 1]	[∞, 0.25, 2, 2, 2, 2, 2, 2]
10	[0, 10, 4, 4, 4, 1, 1, 1]	[∞, 0.25, 2, 2, 2, 2, 2, 2]

Table 3.12: Probabilities for Normal Mixture Prior Distribution on Effect Sizes

Prior	p_1	p_2	p_3	p_{00}	p_{01}	p_{10}	p_{11}	p_{000}	p_{100}	p_{110}	p_{111}
1	0.8	0.8	0.75	0.1	0.4	0.4	0.7	0.1	0.3	0.3	0.6
2	0.8	0.2	0.75	0.1	0.4	0.4	0.7	0.1	0.3	0.3	0.6
3	0.8	0.75	0.2	0.1	0.4	0.4	0.7	0.1	0.3	0.3	0.6
4	0.2	0.8	0.75	0.1	0.4	0.4	0.7	0.1	0.3	0.3	0.6
5	0.8	0.2	0.2	0.1	0.4	0.4	0.7	0.1	0.3	0.3	0.6
6	0.2	0.8	0.2	0.1	0.4	0.4	0.7	0.1	0.3	0.3	0.6
7	0.2	0.2	0.8	0.1	0.4	0.4	0.7	0.1	0.3	0.3	0.6
8	0.2	0.2	0.2	0.1	0.4	0.4	0.7	0.1	0.3	0.3	0.6
9	0.4	0.4	0.4	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1
10	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
11	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2

3.2.4 Results for Optimization Utility

In this section, tables are presented which display the results when applying the optimization utility function in the case of various prior assumptions. Comments on the results are also given.

Table 3.13: Results for Strong Prior Information - Point Prior

	Design A			Design B		
Prior	Mean Loss	Mean Gain	Relative Gain	Mean Loss	Mean Gain	Relative Gain
1	4.000	26.000	0.867	2.000	28.000	0.933
2	0.000	30.000	1.000	10.000	20.000	0.667
3	0.000	38.000	1.000	18.000	20.000	0.526
4	0.000	38.000	1.000	10.000	28.000	0.737
5	0.000	46.000	1.000	18.000	28.000	0.609
6	0.000	28.000	1.000	18.000	10.000	0.357
7	8.432	19.567	0.699	8.000	20.000	0.714
8	0.000	20.000	1.000	0.000	20.000	1.000
9	7.856	10.144	0.564	18.000	0.000	0.000
10	8.064	9.936	0.552	10.000	8.000	0.444
11	8.112	9.888	0.549	8.000	10.000	0.556
12	8.320	19.680	0.703	10.000	18.000	0.643
13	16.352	19.648	0.546	18.000	18.000	0.500

Table 3.14: Results for Normal Prior Distribution on Effect Sizes: Informative Prior

	Design A			Design B		
Prior	Mean Loss	Mean Gain	Relative Gain	Mean Loss	Mean Gain	Relative Gain
1	3.994	25.974	0.867	1.997	27.971	0.934
2	0.000	29.990	1.000	10.005	19.985	0.666
3	0.020	37.975	0.999	18.003	19.992	0.526
4	0.019	37.986	1.000	10.012	27.993	0.737
5	0.036	45.935	0.999	17.986	27.984	0.609
6	0.388	28.015	0.986	18.370	10.034	0.353
7	7.977	20.044	0.715	8.007	20.013	0.714
8	0.321	20.070	0.984	0.401	19.990	0.981
9	8.595	9.823	0.532	18.258	0.159	0.009
10	8.018	10.394	0.565	10.289	8.123	0.441
11	8.564	9.862	0.533	8.380	10.046	0.545
12	7.920	20.083	0.717	10.007	17.996	0.643
13	15.908	20.093	0.559	17.989	18.012	0.500

Table 3.15: Results for Normal Prior Distribution on Effect Sizes: Mildly Informative Prior

	Design A			Design B		
Prior	Mean Loss	Mean Gain	Relative Gain	Mean Loss	Mean Gain	Relative Gain
1	5.508	25.908	0.835	3.615	27.802	0.890
2	0.046	29.824	0.998	9.855	20.015	0.677
3	3.795	34.113	0.903	17.898	20.010	0.531
4	3.830	34.189	0.902	10.094	27.926	0.741
5	16.001	30.219	0.655	18.166	28.053	0.610
6	3.232	27.772	0.900	19.968	11.037	0.357
7	11.315	16.913	0.605	8.241	19.998	0.717
8	1.980	21.201	0.915	3.154	20.027	0.869
9	10.083	11.217	0.518	19.009	2.292	0.112
10	13.318	7.853	0.381	11.129	10.042	0.481
11	9.004	12.375	0.567	9.860	11.518	0.549
12	11.679	16.366	0.594	10.115	17.931	0.646
13	22.451	13.404	0.377	17.820	18.035	0.505

Table 3.16: Results for Normal Prior Distribution on Effect Sizes: Weakly Informative Prior

	Design A			Design B		
Prior	Mean Loss	Mean Gain	Relative Gain	Mean Loss	Mean Gain	Relative Gain
1	9.273	25.332	0.730	6.555	28.049	0.819
2	2.867	29.009	0.902	11.912	19.964	0.637
3	7.508	31.770	0.818	19.028	20.250	0.521
4	8.246	30.899	0.804	10.959	28.186	0.732
5	21.960	23.862	0.526	17.912	27.910	0.618
6	7.684	26.351	0.774	21.919	12.117	0.358
7	13.981	16.437	0.550	11.843	18.575	0.612
8	5.467	21.589	0.792	6.956	20.099	0.747
9	12.148	13.482	0.512	19.557	6.074	0.247
10	15.475	9.915	0.394	12.735	12.654	0.512
11	12.324	12.593	0.486	12.131	12.785	0.524
12	13.185	17.121	0.572	12.115	18.191	0.611
13	23.412	12.888	0.371	18.254	18.046	0.506

Table 3.17: Results for Additional Normal Prior Distributions

	Design A			Design B		
Prior	Mean Loss	Mean Gain	Relative Gain	Mean Loss	Mean Gain	Relative Gain
1	9.703	6.315	0.360	10.914	5.105	0.323
2	4.666	3.429	0.406	5.345	2.749	0.344
3	0.755	0.235	0.229	0.881	0.109	0.104
4	2.815	8.488	0.711	5.168	6.135	0.527
5	10.113	-0.031	-0.132	9.890	0.192	0.014
6	7.470	18.008	0.724	5.420	20.058	0.797
7	7.860	17.414	0.707	5.278	19.996	0.805
8	7.816	21.614	0.746	5.591	23.840	0.819
9	6.064	19.233	0.768	9.577	15.720	0.629
10	7.693	17.520	0.706	7.029	18.184	0.731

Table 3.18: Results for Normal Mixture Prior Distribution on Effect Sizes

	Design A			Design B		
Prior	Mean Loss	Mean Gain	Relative Gain	Mean Loss	Mean Gain	Relative Gain
1	5.973	6.093	0.510	7.026	5.040	0.426
2	4.547	4.995	0.523	6.578	2.964	0.307
3	4.870	5.149	0.517	5.014	5.005	0.501
4	4.855	4.844	0.508	6.617	3.081	0.318
5	3.958	3.431	0.442	4.295	3.093	0.405
6	3.829	3.432	0.446	4.266	2.996	0.389
7	4.297	3.196	0.426	6.194	1.299	0.131
8	3.027	1.657	0.265	3.418	1.266	0.198
9	2.272	3.615	0.561	3.400	2.488	0.376
10	4.929	3.501	0.390	5.813	2.617	0.287
11	2.717	1.627	0.277	3.240	1.103	0.200

Strong Prior Information about Effect Sizes

Looking at the results in Table 3.13 and 3.14 it is clear to see that the estimated utilities, i.e. mean loss, mean gain and relative gain are comparatively similar for most of the priors when comparing the same design. Take for example the relative gain of design A for prior 1 which is 0.867 for both the strong prior information and the normal prior. Also, looking still at prior 1 there is only a difference of 0.001 between the relative gains of design B when either assuming strong prior information or a normal prior distribution. Thus, it can be seen that whether assuming strong prior information or informative normal prior distributions on effect sizes, the conclusion is the same for all but two cases, as to which design has the maximum relative gain. This means that given the majority of assumptions that have been made in this example about the effect sizes, the conclusions given strong prior information and assuming a normal distribution on the whole concur as to which design is the 'best'.

From the results in Table 3.13, there are some interesting points to note. Except for

prior 1, prior 7 and prior 11, the relative gain is greater for design A than design B, and the relative gains for both design are equal in the case of prior 8. This implies that given the majority of assumptions that have been considered in this particular example about the effect sizes, i.e. those that are set out in 3.7, it would be more sensible to estimate the three main effects rather than two main effects and a two-factor interaction.

In the case of prior 1, where it is expected that main effects **1**, **2** and the two-factor interaction **12** are large, design B has the greater relative gain. This seems quite an obvious result, since if an experimenter had made the assumption that in fact main effect **3** would be small and two-factor interaction **12** would be large, there appears to be little benefit in choosing design A. However, the difference of the relative gain between design A and B is 0.066 and does not seem to be as large as one would expect to reflect the greater benefit in choosing design B. For prior 7, although design B also has the higher relative gain, this is only greater than the relative gain of design A by an amount of 0.015, which could be considered to be a minimal difference. In this case it is expected that main effects **1** and **2** and the two-factor interaction **13** will be large. Thus, if one suspects these effects to be large, although design B does have the greater relative gain, choosing a design to estimate all three main effects may be equally beneficial. This idea is reiterated with prior 11 where main effect **1** and two-factor interaction **13** are expected to be large. Design B has the greater relative gain than design A but only by a difference of 0.007. It appears that with design B having the greater relative gain for priors 7 and 11, more importance is being placed on estimating any potentially large interaction effects involving the assumed to be large parent main effect **1**.

It is also worth pointing out that for prior 8, which expects main effects **1** and **2** to be large and all other effects to be small, both design A and design B have a relative gain of 1. To identify large main effects may usually be of prime importance to an

experimenter, particularly in the case of screening experiments, however if two main effects are expected to be large then it would also make sense to want to estimate the interaction of these effects. Thus, when assuming strong prior knowledge that main effects **1** and **2** are large, the utilities result in a conclusion that it would be of equal benefit to either estimate main effect **3** or two-factor interaction **12**. In this situation the most sensible course of action would be for the experimenter to use their scientific knowledge of the process in hand, or any knowledge from previous experimental runs to decide if it is of more benefit to estimate a main effect or an interaction effect. Alternatively, the experimenter may wish to choose a design based on the purpose of the experiment and whether it is more worthwhile to correctly identify all active main effects, or identify potential active interaction effects along with some of the active main effects.

As mentioned, except for prior 1, prior 7 and prior 11, and prior 8 where both designs have equal maximum relative gain, all other priors result in design A having the greater relative gain in comparison to design B. Amongst these priors with design A coming out as 'best', there are some interesting points to note about prior 6 and prior 9.

For prior 6, it is expected that main effects **1**, **3** and two-factor interaction **13** are large and design A results in a relative gain of 1 compared to 0.357 for design B. Therefore, when main effect **3** is expected to be large it is better to go with the design that can estimate this effect. This idea is reiterated with the case of prior 9 where main effect **3** and two-factor interaction **13** are expected to be large and the resulting relative gain of design A is 0.564 and for design B is 0. The relative gain of 0 indicates that it is pointless to use design B which would estimate main effects **1**, **2** and two-factor interaction **12**, and would entirely overlook identifying the true active effects.

Informative Normal Prior Distributions on Effect Sizes

As stated previously, when assuming strong prior information on the effect sizes, the conclusions for which design has the higher relative gain for each of the priors is in agreement with that when assuming informative normal prior distributions, except in two cases. The two cases are prior 7 and prior 8. Looking more closely at the results for prior 7, we see that when assuming a known size of effect, design B is declared as the better design in terms of having a greater relative gain of 0.714. However, when assuming an informative normal prior distribution for the same prior, design B also has the same relative gain of 0.714, although design A in this case is declared as better than design B by what could be considered to be a minimal amount of 0.001. Thus, it could be concluded that in the case of prior 7 whether assuming that the effect sizes are known or an informative normal prior distribution on the effect sizes is assumed, both design A and B could be considered to be equally as informative.

Also, with prior 7, it is expected that main effects **1** and **2** and two-factor interaction **13** are large. If an experimenter did have strong prior knowledge that both main effect **2** and two-factor interaction **13** may be large, then although design A has the somewhat higher relative gain this design should be chosen with careful consideration. Looking at the design structure of design A, due to the defining relation of design A being $\mathbf{I} = -\mathbf{123}$, the aliased effects are:

$$\mathbf{1} = -\mathbf{23}, \mathbf{2} = -\mathbf{13}, \mathbf{3} = -\mathbf{12}.$$

Thus, if in fact both effects **2** and **13** are large, then because these large effects are aliased the experimenter ma not be able to conclude strongly which effect is in fact the truly active effect, or that in fact both are. There may also be the possibility of *cancellation*, that is effects may cancel each other out if they have opposite signs. Thus, even if both effects are truly active, they may go undiscovered by the experimenter. In these types of cases, follow-up experiments may need to be

carried out to clearly identify the true active effects.

Despite this argument against choosing design A, although it resulted in the higher relative gain in the case of normal priors on effect sizes, it can also be argued that design B may also have its own drawbacks in the case of prior 7. Owing to the fact that design B has variable **3** held fixed at 0, i.e. standard operating conditions, main effect **3** cannot be estimated nor can any of the interactions involving variable **3**. So, if in fact the two-factor interaction **13** is truly large and therefore an active effect, this cannot be estimated and may lead to problems such as non-active effects being declared active or an overinflated variance.

Looking at prior 8, designs A and B are found to be equally efficient with relative gains of 1 when assuming known sizes of effects. We also see that when placing informative normal prior distributions on effect sizes, design A has the greater relative gain, but this is only greater than design B by an amount of 0.003. Such a strong conclusion that the relative gain will be 1 for both design A and B can be drawn since strong prior knowledge about the effect sizes are assumed. Taking into consideration that the relative gains for designs A and B when assuming informative normal priors on effect sizes are both close to 1 (0.984 and 0.981 respectively), with a minimal difference of 0.003, it does not seem surprising that both designs have a relative gain of 1 when strong prior knowledge is assumed.

Mildly Informative and Weakly Informative Normal Prior Distributions on Effect Sizes

Now, we can observe the results obtained from the utility functions when assuming mildly informative normal prior distributions on effect sizes, as displayed in Table 3.15. It can be seen that for the most part, the design which has the higher relative

gain does correspond to that when making stronger assumptions about the effect sizes as in Tables 3.13 and 3.14. However, the few occurrences where the conclusions do not agree are in the cases of priors 10, 11, 12 and 13. Observing prior 10, we see that it is expected that effects **3** and **12** are large and in this case the design estimating the main effects **1** and **2** and the interaction effect **12** is deemed to be more efficient as it has the higher relative gain. This contrasts with previous results where when stronger assumptions about the sizes of the effects could be made, the design estimating all main effects resulted in having the higher relative gain. This leads the initial judgement to be made that if one suspects these particular effects may be large but can only specify within a vague range as to where the effect sizes will lie, then more benefit will be gained from choosing a design that does not estimate variable **3** but instead estimates an interaction effect.

For prior 11, where it is that expected main effect **1** and interaction effect **13** will be large, design A has the greater relative gain by an amount of 0.018. For priors 12 and 13 design B has the higher relative gain where similarly to prior 10, design A had the higher relative gain when making stronger assumptions about the effect sizes. Prior 12 expects effects **2**, **3** and **12** to be large and for prior 13, the same effects are expected to be large along with **13**. Thus, in both these circumstances, when only a fairly informative normal prior distribution can be used to model the effect sizes, it appears more gain will be made from estimating the interaction effect **12** instead of main effect **3**. This may seem a strange result as the main effect of **3** is expected to be large in both cases and also its associated interaction effect **13** being large in the case of prior 13. However, since neither of the candidate designs can estimate any interaction effects involving **3**, and **-2** is aliased with **13** this does lead to potential drawbacks for both designs as mentioned previously.

Similar results were seen when weakly informative normal prior distributions were assumed for the effect sizes, as seen in Table 3.16. Design B also had the higher

relative gain for priors 10, 12 and 13. All other results agreed with the much more informative priors, except in the case of prior 5. Here, effects **1, 2, 3, 12** and **13** are expected to be large and where for all other more informative priors design A had been deemed to be more efficient, in this case design B had the greater relative gain. This may not seem a surprising result considering in this situation there is expected to be almost all effects active in such a small design and that there is little information incorporated into the normal prior distribution.

Additional Normal Prior Distributions on Effect Sizes

The results in Table 3.17 are for some additional normal prior distributions. The first three priors describe the situation where an experimenter has no prior information about the likely size or direction of the effect, ranging from a high prior variance (prior 1) to an informative prior variance (prior 3). Observing the first three rows of results in Table 3.17, it is clear that if an experimenter has no indication of the likely direction or size of effects, then making a judgement based on the relative gain of a design it would always be best to use the design that can estimate all main effects. This is regardless of how informative the prior variance is, since the design A, which estimates all main effects of variables **1**, **2** and **3**, has the higher relative gain for all three possible choices of variance. The worth of choosing a design, measuring this by the size of the relative gain, increases as the variance becomes less vague. This is observed where for the most vague distribution, design A is perceived as having a greater relative gain of 0.037 than design B, and this increases to being 0.125 in terms of relative gain better than design B, for the most informative distribution.

The other non-directional prior in the case of prior 4, where the prior variances are not equal for all effects, has a similar result to priors 1-3 where design A has a larger relative gain. This appears to be an understandable result owing to the

prior variances which suggest the interaction effects will be somewhere close to 0 and the main effects could be somewhat larger. This contrasts greatly with the results obtained for prior 5. Although this prior is also non-directional like priors 1-4, high prior variances for the two-factor interactions suggested these effects could be large whilst low prior variances were observed for the main effects. Relative gains obtained from the simulations are -0.132 for design A and 0.014 for design B. These results clearly show that both designs do quite badly when making these prior assumptions on effect sizes as both relative gains are extremely small. However, it is evident that design A is less efficient a design than design B, since a negative gain resulted from the simulations. This seems an obvious result, that the design estimating all main effects and no interactions is less successful because the prior assumptions make it clear that main effects are highly likely to be small and there is more chance of two-factor interactions having a larger effect size.

The other results in Table 3.17, specifically priors 6-8, show that in the cases where there is expected to be a large main effect and a fairly large interaction effect involving that main effect, the relative gain is always larger for the design which can estimate the interaction effect. This is even in the cases when the interaction effect is expected to be large but with not a great deal of certainty.

Prior 9 indicates that if there is any belief of there being a fairly large main effect, even if this is not with much certainty, and all interaction effects are believed to be fairly small, it would be better to choose the design that will estimate all main effects. However, looking at the final prior of Table 3.17, prior 10, we see that even if there is expected to be a fairly large main effect, namely **3**, the design which estimates the other two variables, **1** and **2** and its associated interaction effect, has the higher relative gain. In this case, where there is expected to be a fairly large interaction effect, which does not involve **3**, more emphasis appears to be placed on being able to estimate this interaction effect. The greater importance put on estimating an

interaction effect rather than a main effect, which typically would be considered to be of more importance, may be due to one of the associated parent effects, main effect **1**, being expected to be large with a greater degree of certainty. This is despite the fact that both the main effect **3** and interaction effect **12** are considered to be fairly large with the same amount of certainty.

Normal Mixture Prior Distributions on Effect Sizes

Now, looking at Table 3.18 which displays the results of the utilities when assuming normal mixture prior distribution on effect sizes, we can see that design A has the greater relative gain each time. The greatest benefit of choosing design A is seen in the case of prior 7. This is where only variable **3** is expected to be large, and has the greatest difference of relative gain between design A and design B of 0.295. This makes sense since if one does not expect either the main effect of **1** or **2** to be large, and consequently, neither its associated two-factor interaction effect **12**, there would be no benefit in choosing design B. It is also worth mentioning that the smallest gain of carrying out design A over design B is seen in the case of prior 3. In this situation, the hierarchical prior has been set up so that variables **1** and **2** are highly likely to be active and due to the dependent nature of the priors, the interaction effect **12** is then also highly likely to be active. Thus, with such a scenario where one would expect the effects **1**, **2** and **12** to be active in the experiment, it may be expected that design B would have the greater relative gain. However, this is not the case and as with all of the prior assumptions considered, the design that can estimate all main effects has been deemed to be the better design. To try to come to some understanding of why this is the case, it is of interest to calculate the probability that two-factor interaction effect **12** is active. This is based on the prior probabilities p_i and p_{ij}, which are respectively the probabilities that factor i and two-factor interaction ij are active. For prior 3, the probabilities were set out as $p_1 = 0.8$, $p_2 = 0.75$ and $p_3 = 0.2$.

That is, it was expected that variables **1** and **2** were likely to be active and there was a small probability of variable **3** being active. Due to the activity of interaction effects depending upon the activity of their parent effects, the probabilities p_{00}, p_{01}, p_{10} and p_{11} were also set out. For prior 3, these were $p_{00} = 0.1$, $p_{01} = 0.4$, $p_{10} = 0.4$ and $p_{11} = 0.7$. We can then obtain the probability that β_{12} is active as

$$\begin{aligned} P(\beta_{12}\ is\ active) &= p_1p_2p_{11} + p_1(1-p_2)p_{10} + (1-p_1)p_2p_{01} + (1-p_1)(1-p_2)p_{00} \\ &= (0.8)(0.75)(0.7) + (0.8)(0.25)(0.4) + (0.2)(0.75)(0.4) + \\ &\quad (0.2)(0.25)(0.1) \\ &= 0.565. \end{aligned}$$

The probability $P(\beta_{12}\ is\ active) = 0.565$ may not be considered to be that large. Thus, this may provide some explanation for why although the variables **1** and **2** have been set up to be highly likely to be active in this case, the design which is able to estimate variable **3** has been chosen as the better design, over the design which can estimate two-factor interaction **12**. Whilst considering this, it would be of some value to note that as previously stated the difference in relative gain between the two designs for prior 3 is the smallest out of all the prior assumptions that have been made.

For the final three priors of the table, to be precise priors 9, 10 and 11, where the effect heredity principle has not been assumed and the effects are considered to be independent, we see that again design A has the higher relative gain. Out of these three priors we observe that the greatest relative gain in carrying out design A as opposed to design B is seen for prior 9. This seems fairly self-explanatory since in this case the probabilities of the main effects being active are all equal, but more importantly, are greater than the probability that a two- or three-factor interaction effect is active. Thus, greater benefit would be acquired from carrying out a design estimating main effects. Where the probabilities for the effects to be active are all equal, as in priors 10 and 11, we see that the difference in relative gain between

design A and B decreases as the probability of an effect being active decreases. This is seen where the difference in relative gain between design A and B is 0.103 when $p_1 = p_2 = \ldots = p_{111} = 0.4$ but decreases to 0.077 in the case of prior 11 where $p_1 = p_2 = \ldots = p_{111} = 0.2$.

It is also of interest to note that for design A variables **1**, **2** and **3**, and for design B variables **1** and **2** can be considered to be *exchangeable*, i.e. that the labels identifying them are uninformative (Bernardo, 1996), and . Thus, it would be expected that the results obtained for the priors 2 and 4, where the main effects **1** and **3** (prior 2), and main effects **2** and **3** (prior 4), are expected to be large and interaction effects large dependent on the parent effects, will be relatively similar. Likewise, similar results would also be expected for priors 5 and 6 where either the main effect of **1** and **2** respectively is expected to be large, and similarly the interaction effects will be large dependent on parent effects. Looking at the results in 3.18 it can be seen that this is in fact the case.

3.2.5 Discussion

The use of the utility function outlined in Chapter 2, primarily for the purposes of optimization, has been demonstrated with a small example comparing two designs each consisting of 4 runs, where various prior assumptions about the effect sizes have been made. As mentioned earlier on in the chapter, this example is primarily for illustrating the rationale of the optimization utility function before a larger, and perhaps more thought-provoking, case is presented. Thus, the results observed in this section may appear to be relatively intuitive as a result of the small designs considered.

As mentioned in the previous chapter, this work stems from that of Gilmour and

Mead (1995) where they considered the estimated quantity $E\{Y(\mathbf{x}_{max}) - Y(\hat{\mathbf{x}}_{max})\}$, that refers to the gain in yield or some other similar measure that an experimenter is seeking to maximize. In their paper, they considered the question of whether or not the next stage of experimentation is worthwhile in the case of sequential experiments. They did this by simulating from the distribution of $\boldsymbol{\beta}|\mathbf{y}$ (where $\boldsymbol{\beta}$ refers to the set of parameters and $\mathbf{y}$ refers to the data) and obtaining $L(\mathbf{x})$ for each realization of $\boldsymbol{\beta}$ to obtain the posterior distribution of $L(\mathbf{x})|\mathbf{y}$. This is where, for any $\mathbf{x}$,

$$L(\mathbf{x}) = Y(\mathbf{x}_{max}) - Y(\mathbf{x}).$$

That is, $L(\mathbf{x})$ is the quantity for assessing the difference between the expected response at the true optimum and the expected response at $\mathbf{x}$.

The optimal factor combination $\mathbf{x}_{max}$ could then be estimated by the $\mathbf{x}$ which minimizes $E_{\boldsymbol{\beta}|\mathbf{y}}\{L(\mathbf{x})\}$. The posterior distribution of $L(\hat{\mathbf{x}}_{max})|\mathbf{y}$ could then be used to decide whether or not to stop at a particular stage of experimentation or continue and carry out another experiment, for example deciding to stop experimentation when

$$E_{\beta|y}\{L(\hat{\mathbf{x}}_{max})\} < \epsilon$$

where the value of ϵ is to be decided by the experimenter. Thus, it can be seen that in comparison to their approach of using this quantity to decide whether or not to stop experimentation, our approach in this thesis is concerned with selecting an optimal design from a set of candidate designs. An optimal design is chosen by identifying that which minimizes the quantity $U_{Loss}(\hat{\mathbf{x}}_{opt})$, which is the difference between the expected response at the true optimum and the expected response at the predicted optimum, i.e. $\mu_{\mathbf{x}_{opt}} - \mu_{\hat{\mathbf{x}}_{opt}}$, and is comparable to $L(\hat{\mathbf{x}}_{max})$.

It cannot be claimed that this utility function should be used exclusively by an experimenter when making a choice as to the best design to choose in their experimentation. However, it can be used as an aid along with the experimenters' judgement about

which design will potentially provide the most insight about the estimates of the effects and in locating an optimum treatment combination. The optimum treatment combination will of course be with respect to the objectives of the experiment. Thus, it is for the experimenter to use the principal aims of the experiment or investigation, alongside the utility function to come to a conclusion as to what will be the 'best' design to most accurately exploit the greatest information. From the results obtained from the simulation work and observations that have been made, some general conclusions can be drawn. Furthermore, some guidelines can then be set out with the intention of assisting an experimenter in the circumstance where they are required to choose a design, incorporating any prior information they may have about the parameters that are to be estimated.

The results obtained when assuming that the prior information about the effect sizes are described using normal mixture priors could be considered to be the most realistic representation of experimentation. This is due to the fact that the effect heredity principle, namely that an interaction effect will be active dependent upon the parent main effects, which is incorporated into the hierarchical prior, would typically be considered to accurately represent the majority of variables considered by an experimenter. A stronger justification for the use of the normal mixture prior is that the nature of this particular prior allows an effect to come from a high variance distribution, typically with a small probability, and from a low variance distribution otherwise, hence taking into account the principle of effect sparsity. Thus, to apply an independence prior, and treat effects independently without carefully considering if the effects are in fact independent, may be a perilous decision.

It is clear to see that when considering hierarchical priors, the design which would estimate all main effects, as opposed to setting one of the variables to 0 in order for the other variables' main effects and their associated interaction effects to be estimated, is deemed to be the better design in terms of having the greater relative gain. Relating

this back to the idea of the optimization utility function, this corresponds to the estimated improvement on standard operating conditions, in relation to the gain that would have been made if the true optimal treatment combination had been used compared to standard operating conditions, is always increased for the prior assumptions made in this thesis if the design estimating all main effects is used. This makes for a surprising result, particularly in the case where it was highly likely that the effects **1**, **2** and **12** would be active, as in the case of prior 3. In this situation, one may have expected that the design estimating these three effects would be more beneficial, and in actual fact the alternative candidate design had the greater relative gain, although some explanation for this interesting result was provided by the calculation of $P(\beta_{12}\ is\ active)$. This adds more weight to the argument that although there is benefit to the utility function, it should always be used in its own place and along with sensible judgement. It should never be used blindly as a substitute for, but in addition to, various other methods that an experimenter may use when deciding which design to run. It should also be noted that since the designs being considered are relatively small, only consisting of 4 runs, conclusions drawn from this simulation work should be treated with some caution.

Results also indicate that if an experimenter has carried out no previous investigative work on the variables being studied and therefore has no knowledge of how the factors will behave, then it is best to carry out a design estimating all main effects. This is assuming the experimenter has set prior means for all effects to be 0 to encompass the idea that they have no indication about the likely direction of the effect. Then regardless of whether the variance is vague, i.e. a high prior variance, or is informative, i.e. a low prior variance, it appears best to go with a design that, as mentioned, estimates all main effects.

Some interesting observations were made in the cases where an experimenter assumed effects to be independent. This is particularly in the case where the experimenter had

knowledge on the effect sizes and either assumed strong prior information, i.e. point priors, or normal distribution priors on effect sizes with prior variance ranging from low to high prior variance. It appears from results obtained that if an experimenter is in a similar situation and has knowledge about the size of effects and strong information about the range within which this effect size will lie, then it would on the whole always be best to carry out a design estimating all main effects, unless there is strong belief that some interactions will be active. When the experimenter does not have a great deal of information about the range within which the effect will lie, then suggestions for the experimenter are not as exact or clear-cut, although it is apparent that in this instance more care should be taken if any suspected active effects will be aliased with one another.

It is important to note that there is an interplay between the priors and aliasing patterns in the experiments and the resulting losses and gains observed. This therefore means there are some constraints in the investigation and the conclusions or inferences that have been drawn are dependent upon the beliefs about the effects of the factors that have been chosen to be studied, and also the aliasing structure of both the designs. An example of this is clearly seen in the priors investigated where the effects that are believed to be large and active are in fact the effects that are aliased with one another. This is the case for several of the priors such as for priors 4 and 5 where strong prior information and also normal prior distributions are assumed on the effects. In these cases, the effects **3** and **12** are both expected to be large and therefore active, and are also aliased due to the defining relation of design A being $\mathbf{I} = -\mathbf{123}$. That is, $\mathbf{3} = -\mathbf{12}$ and in this case the direction of the interaction effect does evidently matter. As a result of the prior beliefs set out where both main effect **3** and the interaction effect **12** are expected to be large, the negative direction of the interaction effect is likely to lead to cancellation where the effects cancel each other out owing to their opposite signs. This is also likely to occur for those other priors where both the main effects and interaction effects that are believed to be large are

aliased with one another. For example, when $\mathbf{2} = -\mathbf{13}$ such as in the priors 3, 5, 7 and 13. Therefore, it may have been more interesting to look at other combinations of the choice of priors and aliasing structures where pitfalls such as cancellation did not occur. It then may have been more evident whether the failure or success of the performance of the design investigated was actually down to the beliefs about the effects of the factors or the aliasing structure, rather than the interplay between these two things. Thus, it should be stated that the results displayed are dependent upon the various choices of prior distributions and experimental designs made and as such does mean that there are limitations to the study.

Hence, an experimenter can use a utility function to assess the benefits of a design and also compare designs in terms of loss and gain. However, as displayed due to potential drawbacks such as the aliasing structure which results from the defining relation chosen, and also the dependency between this and the choice of prior upon the results, an experimenter should use the utility function side by side with their own judgement as to what seems sensible.

3.3 Screening: Comparison of 5 Factors in 12 and 16 Runs

3.3.1 Introduction

The algorithm outlined in §2.4.1 is to be used when an experimenter is primarily concerned with finding those key factors driving a process and identifying the active effects that are large in comparison to noise. The algorithm mentioned can be used to aid an experimenter in a screening situation and assess the benefits of various designs.

The majority of the fractional factorial designs mentioned so far have been of the form 2^{k-p}, i.e. a power of two. However, looking at the run size of these designs as they get larger, the run size being 4, 8, 16, 32, 64, ..., we see that the size gets progressively larger. These large jumps between run sizes can prove problematic in situations where each run may be costly or time-consuming, an experiment may be destructive or there may be limited resources available. Say, for example, there were 11 factors to be studied, then the minimum 2^{k-p} design that could be used to study the factors in would be a 16 run design. However, by instead studying the 11 factors in a 12 run design then 4 runs could be saved to overcome the difficulties mentioned. Thus, time and money could be saved.

A 12 run design, is in fact a *Plackett-Burman* design since its run size is a multiple of 4 but not a power of two. Plackett-Burman designs are nongeometric and have more complicated aliasing structures than 2^{k-p} designs, which are *geometric* designs. Geometric designs are those in which the run size is a power of two. That is, where the run size is 4, 8, 16, 32, 64, etc. The complicated aliasing in nongeometric designs may be a large reason for the reluctance to use them as readily as the geometric designs, as the complex aliasing makes them much more difficult to analyse.

Looking at the 16 run design in Table 3.19, the column for variable 5 is achieved using the design generator $\mathbf{5 = 1234}$. It can therefore be inferred that all 5 main effects will be aliased with 4-factor interactions, and so are strongly clear. Two-factor interactions are aliased with three-factor interactions, and so are clear. The observation to make here is that the aliasing structure in this case is fairly straightforward in that the aliased terms have coefficients of 1, as can be seen by looking at the aliased terms for this 16 run design as presented in Table 3.20. The confounding that exists in geometric designs is complete, which means that if two effects are aliased they may be identical numerically except for their sign. That is, one may be the negative of another. Comparing this to the aliasing structure for the 12 run Plackett-Burman design (Table 3.19) and we see that this is much more complex as presented again in Table 3.20. In the case of the 12 run design we see that all the two-factor interactions not involving $\mathbf{1}$ are partial aliases of $\mathbf{1}$ with an aliasing coefficient of $\frac{1}{3}$ or $-\frac{1}{3}$, and this is similarly seen for all other effects. Two effects are said to be *partially aliased* if the absolute value of their aliasing coefficient is strictly between 0 and 1 (Hamada and Wu, 2000). Thus, Plackett-Burman designs will result in particular aliased terms that can have fractional coefficients due to being associated with more than one data contrast or column. The complete aliasing structure for both the 12 and 16 run design is displayed in Table 3.20.

Owing to this partial aliasing of the Plackett-Burman designs, these designs which were first presented by Plackett and Burman (1946), were intended to be used as an economical and useful way of detecting large main effects, but by assuming all interactions are negligible. However, it can be seen in the literature since then that considerable efforts have been made to overcome this such as by Box and Meyer (1993). By using a Bayesian method which did not assume that interactions were negligible, but in fact allowed for their possibility, posterior probabilities were able to be computed for a factor being active.

Table 3.19: Design Matrix for 5 Factors in 12 and 16 Runs

12-run Plackett-Burman Design

	Variable				
Run	**1**	**2**	**3**	**4**	**5**
1	+	-	+	-	-
2	+	+	-	+	-
3	-	+	+	-	+
4	+	-	+	+	-
5	+	+	-	+	+
6	+	+	+	-	+
7	-	+	+	+	-
8	-	-	+	+	+
9	-	-	-	+	+
10	+	-	-	-	+
11	-	+	-	-	-
12	-	-	-	-	-

2^{5-1} Design

	Variable				
Run	**1**	**2**	**3**	**4**	**5**
1	-	-	-	-	+
2	+	-	-	-	-
3	-	+	-	-	-
4	+	+	-	-	+
5	-	-	+	-	-
6	+	-	+	-	+
7	-	+	+	-	+
8	+	+	+	-	-
9	-	-	-	+	-
10	+	-	-	+	+
11	-	+	-	+	+
12	+	+	-	+	-
13	-	-	+	+	+
14	+	-	+	+	-
15	-	+	+	+	-
16	+	+	+	+	+

Table 3.20: Aliasing Structure for 5 Factors in 12 and 16 Runs

12 Run Design	16 Run Design
1 + $\frac{1}{3}$(**-23+24+25-34-35-45**)	**1 + 2345**
2 + $\frac{1}{3}$(**-13+14+15-34+35-45**)	**2 + 1345**
3 + $\frac{1}{3}$(**-12-14-15-24+25-45**)	**3 + 1245**
4 + $\frac{1}{3}$(**12-13-15-23-25-35**)	**4 + 1235**
5 + $\frac{1}{3}$(**12-13-14+23-24-34**)	**5 + 1234**
$\frac{1}{3}$(**-12+13-14+15+23-24-25+34-35-45**)	**12 + 345**
$\frac{1}{3}$(**-12-13-14+15-23+24-25+34-35+45**)	**13 + 245**
$\frac{1}{3}$(**12+13-14-15-23-24-25-34+35+45**)	**14 + 235**
$\frac{1}{3}$(**-12-13-14-15+23+24-25-34-35-45**)	**15 + 234**
$\frac{1}{3}$(**-12-13+14-15-23-24-25+34+35-45**)	**23 + 145**
$\frac{1}{3}$(**-12+13+14-15-23-24+25-34-35+45**)	**24 + 135**
	25 + 134
	34 + 125
	35 + 124
	45 + 123

Despite this, when using Plackett-Burman designs their analysis should be carried out carefully, especially in the case where interaction effects are to be considered. Typically, if an experimenter wants to estimate interaction effects as well as main effects it would usually be recommended that a geometric design of type 2^{k-p} be used, unless there is a strong reason for wanting to opt for a nongeometric design, such as for economical reasons. However, it is of interest to investigate how much benefit there is in using an extra n runs, if there is a choice of which design to use, and if so being able to quantify this value. If for example, an experimenter wants to carry out a screening experiment and therefore only wants to look at main effects, is the cost of carrying out an extra n runs to use a geometric design as opposed to a Plackett-Burman design, really worthwhile enough for the benefit of not having to make such strong assumptions that all interactions are negligible? Also, the type of design chosen as being more beneficial per run may also be affected by the amount of prior knowledge that is assumed before the experiment is carried out. Thus, prior knowledge should also be taken into consideration when looking at the choice of design.

3.3.2 Models to be Considered

To illustrate the algorithm to aid an experimenter in a screening situation, consider that there are 5 factors and only a maximum of 16 runs with which to investigate them. The suggested designs are a 12 run Plackett-Burman design and a 16 run, 2^{5-1}, design (Table 3.19).

However, there may be various scenarios where an experimenter only wants to estimate main effects as a preliminary experiment, or wants to estimate interactions as well. Depending on which of the two designs are chosen, different effects will be able to be estimated. Thus, it would be interesting to look at the following comparisons,

where the basis of the comparison is the identification of active effects:

1. Compare a main effects model in 12 and 16 runs.

2. Compare a main effects model in 12 runs with a main effects and all 2-factor interactions model in 16 runs.

3. Compare a main effects with some 2-factor interactions model in 12 runs with a main effects and all 2-factor interactions model in 16 runs.

For all of these comparisons, both fixed effect sizes and prior distributions on the effect sizes will be considered.

It is worthwhile mentioning that in 12 runs, up to a maximum of 11 contrasts can be estimated and in the case of a 16 run design, the number of contrasts that can be estimated increases to 15. Thus, if carrying out an experiment consisting of 12 experimental runs then an experimenter will have to either look at a main effects only model if studying 5 factors, as outlined in scenarios 1 and 2 above. If the experimenter wants to also estimate some two-factor interactions then these shall have to be carefully selected due to the fact that all 10 two-factor interactions cannot be estimated because of a lack of degrees of freedom. Possible ways of selecting which two-factor interactions it would be best to estimate could be to either consider suggested interaction effects the experimenter feels may have some effect on the process being studied, or alternatively, consider a 'two-step' method. The 'two-step' method would consist of initially analysing all 5 factors and after obtaining estimates of their main effects, then estimate those two-factor interactions with at least one active parent effect. If a large number of main effects have been identified as active, then the principle of strong heredity could be applied where only those interaction effects with both active parent effects are considered. Another possibility could be to estimate all those two-factor interactions associated with the three most important main effects,

where 'important' would be considered to be the three largest main effects.

The strategy taken in this thesis concerning which two-factor interactions to estimate along with the 5 main effects when investigating 5 factors in 12 experimental runs will be the latter approach as outlined above. That is, to initially estimate the 5 main effects and then identify those three main effects with the largest absolute value. Having done this we will then estimate the three two-factor interactions associated with those three most 'important' main effects. For example, if when investigating variables **1**, **2**, **3**, **4** and **5** it is found that the estimated sizes of the main effects are $1 = -1.375$, $2 = 19.5$, $3 = -0.625$, $4 = 10.75$ and $5 = -6.25$, these results taken from the reactor data example from Box, Hunter and Hunter (1978), then it would be concluded that variables **2**, **4** and **5** are the most 'important'. Thus, the three two-factor interactions associated with these variables, namely **24**, **25** and **45**, would then be estimated. This approach could be considered to be based on the effect heredity principle, which states that in order for an interaction effect to be active we would expect at least one of its parent factors to be active.

3.3.3 Prior Distributions on Effect Sizes

The priors being considered in order to generate the simulated datasets, and hence calculate the estimated utility which provides a measure of how well active and inactive effects are identified as active and inactive correctly, will take two of the same three forms of priors as set out previously in the case presented in §3.2. That is normal prior distributions will be considered ranging from vague to informative priors, and a mixture of normal prior distributions will also be used. Strong prior information cannot really be assumed about the effect sizes. If we assumed this, then normal prior distributions may be used in this case with $\sigma_i^2 = 0$ for the i^{th} effect, therefore these priors being effectively point priors. However, by definition a screening experiment

is used to eliminate the unimportant factors of the many whose importance cannot be ruled out at the beginning of a study. Thus, if at the preliminary stages of an investigation an experimenter assumes a great deal of knowledge about the sizes of the effects, this defeats the object of running a screening experiment. It therefore seems rather unnecessary to consider point priors.

Reactor data taken from Box, Hunter and Hunter (1978) is an example typical of a screening experiment. In this reactor example, there are 5 factors being studied, all at 2 levels and so the full fraction consists of 32 experimental runs. Standard analyses of the full fraction, including a normal plot and estimates of effects, found that the effects distinguishable from noise and therefore effects that could be considered to be active were main effects **2, 4** and **5**, and two-factor interactions **24** and **45**. Higher order interactions than two-factor ones have been disregarded as with the previous example, based on the hierarchical ordering principle. This reactor data example will be revisited in greater length in Chapter 4, where the design used and the analyses mentioned will be given in more detail. For now, we shall be focusing on the effect sizes of the main effects and two-factor interactions (Table 3.21) that resulted from this experiment and will discuss appropriate priors that could have been considered in this case, before outlining the priors that we shall consider ourselves to generate the simulated datasets.

Table 3.21: Analysis of 2^5 Factorial Experiment: Reactor Data

Estimates of Effects	
Average = 65.5	
1 = -1.375	12 = 1.375
2 = 19.5	13 = 0.75
3 = -0.625	14 = 0.875
4 = 10.75	15 = 0.125
5 = -6.25	23 = 0.875
	24 = 13.25
	25 = 2.0
	34 = 2.125
	35 = 0.875
	45 = -11.0

So, observing the estimates of effects we see that there are three main effects and two two-factor interactions that are strongly considered to be active. Thus, if an experimenter suspected this may be the case they could have set the prior variances for these effects to be large. All prior means for effect sizes should be set at zero since in the case of screening an experimenter will typically not know the sizes of the effects. However they may suspect some effects will be larger than others, therefore justifying the decision to have some large prior variances, although they may not know in which direction this effect size will lie. Alternatively, the experimenter may wish to be more cautious and assume equal prior variances on all effects.

Thus, the priors that we shall consider for our simulation work will be chosen with all prior means on effect sizes to be zero. This is to correctly represent the scenario that an experimenter does not have knowledge about which factors are important, therefore giving validation for the need to run a screening experiment. However, the prior variances with varying degrees of confidence will be investigated to display the different scenarios that could occur. This could be an experimenter having some prior belief about the range within which the effect sizes will lie or having very little information and not knowing accurately about this range. The priors used to generate

the simulated datasets are now set out.

In Table 3.22 are the models used to simulate data for the main effects, i.e. these are for

$$\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5]^{\mathrm{T}}.$$

As set out previously, the parameters of interest β_i can be individually represented by normal distributions, which can then be extended to the multivariate normal distribution to represent all p parameters. Where values of μ_p and Σ have been given in the tables for the normal prior distributions, μ_p represents the prior means for the parameters and Σ indicates the diagonal variance entries for the variance-covariance matrix Σ_p.

When all main effects and two-factor interactions are able to be estimated, as is in the case of the 2^{5-1} design consisting of 16 runs, then the following is to be estimated:

$$\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{15}, \beta_{23}, \beta_{24}, \beta_{25}, \beta_{34}, \beta_{35}, \beta_{45}]^{\mathrm{T}}.$$

Then the models used to simulate data for these effects are given in Table 3.23.

Table 3.22: Normal Prior Distributions for Main Effects

Prior	μ_p	Σ
1	**0**	[∞, 0.25, 0.25, 0.25, 0.25, 0.25]
2	**0**	[∞, 1, 1, 1, 1, 1]
3	**0**	[∞, 2, 2, 2, 2, 2]
4	**0**	[∞, 4, 4, 4, 4, 4]
5	**0**	[∞,10,10,10,10,10]
6	**0**	[∞,0.25,0.25,0.25,2,2]
7	**0**	[∞,0.25,0.25,2,2,2]
8	**0**	[∞,0.25,0.25,0.25,4,4]
9	**0**	[∞,0.25,0.25,4,4,4]
10	**0**	[∞,0.25,0.25,10,10,10]
11	**0**	[∞,0.25,0.25,2,2,10]
12	**0**	[∞,2,2,4,4,10]

Table 3.23: Normal Prior Distributions for Main Effects and Two-Factor Interactions

Prior	μ_p	Σ
1	**0**	[∞, 0.25, 0.25, 0.25, 0.25, 0.25, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1]
2	**0**	[∞, 1, 1, 1, 1, 1, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5]
3	**0**	[∞, 2, 2, 2, 2, 2, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5]
4	**0**	[∞, 4, 4, 4, 4, 4, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2]
5	**0**	[∞,10,10,10,10,10, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6]
6	**0**	[∞,0.25,0.25,0.25,2,2, 0.25, 0.25, 1, 1, 0.25, 1, 1, 1, 1, 1.5]
7	**0**	[∞,0.25,0.25,2,2,2, 0.25, 1, 1, 1, 1, 1, 1, 1.5, 1.5, 1.5]
8	**0**	[∞,0.25,0.25,0.25,4,4, 0.25, 0.25, 1.5, 1.5, 0.25, 1.5, 1.5, 1.5, 1.5, 2]
9	**0**	[∞,0.25,0.25,4,4,4, 0.25, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 2, 2, 2]
10	**0**	[∞,0.25,0.25,10,10,10, 0.25, 2, 2, 2, 2, 2, 2, 6, 6, 6]
11	**0**	[∞,0.25,0.25,2,2,10, 0.25, 0.25, 2, 6, 2, 2, 6, 6, 6, 6]
12	**0**	[∞,2,2,4,4,10, 1.5, 1.5, 1.5, 6, 2, 2, 6, 6, 6, 6]

For the normal mixture priors, firstly just looking at the case where we are estimating main effects, the values of p_i are set out in Table 3.24 where

$$\alpha_i \sim Bern(p_i),$$

i.e. p_i indicates the probability of α_i being 0 or 1 which denotes if an effect i is inactive or active respectively. Depending on the outcome of α_i, β_i will come from either of two normal distributions. The parameters β_{ia} and β_{ib} indicated in Table 3.25 are such that

$$\beta_i = (1 - \alpha_i)\beta_{ia} + \alpha_i\beta_{ib},$$

where

$$\beta_{ia} \sim N(\mu_{ia}, \sigma^2_{ia})$$
$$\beta_{ib} \sim N(\mu_{ib}, \sigma^2_{ib}).$$

Thus, if β_i is inactive it takes the distribution denoted by β_{ia} and if it is active it then takes the distribution denoted by β_{ib}.

When we estimate two-factor interaction effects along with the main effects, the probabilities that the two-factor interaction effects are active, i.e. $\alpha_{ij} = 1$, are given in Table 3.26. These probabilities have been chosen in order for the principle of effect heredity to be observed. This is where an interaction with both parent main effects active is more likely to be active than an interaction effect with one active parent main effect. Likewise, an interaction effect with one active parent main effect has a greater probability of being active than an effect with no active parent main effects.

For each of the different combinations of probabilities for an effect being active as set out in Tables 3.24 and 3.26, several scenarios for the two possible distributions β_{ia} and β_{ib} will be looked at. These scenarios are set out in Table 3.25. In the case where $\alpha_i = 1$ and an effect is active, possible distributions will vary from having little information about the range of the effect size, to having a large amount of information about the range. For all cases when $\alpha_i = 0$, the distribution will be $N(0, 0.01)$.

Table 3.24: Probabilities for Normal Mixture Prior Distributions on Main Effects

Prior	p_1	p_2	p_3	p_4	p_5
1	0.2	0.2	0.2	0.2	0.2
2	0.4	0.4	0.4	0.4	0.4
3	0.2	0.2	0.2	0.6	0.6

Table 3.25: Normal Mixture Prior Distributions for Main Effects

Prior Assumption	β_{ia}	βib
(a)	N(0,0.01)	N(0, 4)
(b)	N(0,0.01)	N(0,10)
(c)	N(0,0.01)	N(0,25)

Table 3.26: Probabilities for Normal Mixture Prior Distributions on Two-Factor Interactions Effects

Prior	p_{00}	p_{01}	p_{10}	p_{11}
1	0.1	0.4	0.4	0.7
2	0.1	0.4	0.4	0.7
3	0.1	0.4	0.4	0.7

The distribution β_{ib} is one with a high variance to account for β_i coming from it when it is considered to be an active effect. However, what an experimenter considers to be a high variance and consequently an active effect will depend upon the context of the experiment and what the experimenter deems to be as such. One of the methods proposed to declare an effect as active was to consider some constant ω which the experimenter will have chosen. Then, an effect will be declared as active when $|\beta_i| > \omega$. This method was set out in the algorithm outlined in Chapter 2 for the screening utility function where an alternative method considered for finding active effects was Lenth's method. The experimenter should first decide upon an appropriate value of ω, which will then determine if an effect size is active or not. Then the experimenter should decide upon the prior variances for the effects. If an experimenter considers all effects to be distributed the same, the prior standard deviations for all effects may be set to be equal to ω, i.e. $\omega = \sigma$. Thus, anything more than 1 standard deviation away from the prior mean would be considered to be active. It may also be interesting to consider the case when $\omega = 1.5\sigma$ and also $\omega = 2\sigma$, although this final case may be setting the margin with which to determine an active effect too high. In doing so, we may run the risk of true active effects not being correctly declared as active.

The normal priors set out in Table 3.22 will be considered for the cases when $\omega = \sigma, 1.5\sigma$ and 2σ, where σ^2 is the diagonal variance entry as indicated by Σ. There are some cases where the diagonal variance entries are not equal for all effects. This is

seen for priors 6 - 12 and in these cases the level of $\omega = \sigma$ will be set to 1. Thus, where $\omega = 1.5\sigma$ this will refer to $\omega = 1.5$ in these cases and similarly $\omega = 2\sigma$ will refer to $\omega = 2$. For the normal priors set out in Table 3.23 $\omega = \sigma, 1.5\sigma$ and 2σ shall be considered for all priors where the diagonal variance entries are equal for all main effects as in priors 1 - 5. Again, where the variances for the main effects are not equal (priors 6 - 12), $\omega = 1, 1.5$ and 2. The normal mixture priors are set out in Tables 3.24, 3.25 and 3.26. As displayed in Table 3.25, the high variance distribution β_{ib} follows a normal distribution where $\beta_{ib} \sim N(0, \sigma^2)$. Then for all three prior assumptions where $\sigma^2 = 4, 10$ and 25, the level of ω is taken to be $\sigma, 1.5\sigma$ and 2σ.

It should be noted that, as similarly stated for the simulation work in the previous section where we were comparing two designs of the same size, when also looking at comparing a 12 and a 16 run design in this section we shall assume that the residual variance is known and is equal to 1. For all cases looked at, γ is taken to be 0.5, therefore placing equal importance on both Type I and Type II errors.

The priors used to generate the simulated datasets have been outlined in this section. In the following section, results obtained from the utility functions using these priors will be presented. As stated at the beginning of the chapter, where method 1 is indicated this refers to a utility which has been calculated based upon an effect declared active if it is greater than some constant ω set out by the experimenter. Method 2 refers to one where an effect has been declared active using Lenth's method.

3.3.4 Results for Screening Utility

In this section tables are presented which display the results when applying the screening utility function in the case of various prior assumptions. Comments on the results are also given.

Table 3.27: Type I Errors: Normal Prior Distributions for 16 Run Design - Main Effects Only

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1	0.196	0.013	0.091	0.019	0.042	0.025
2	0.095	0.010	0.044	0.018	0.015	0.023
3	0.062	0.015	0.024	0.021	0.011	0.027
4	0.039	0.013	0.019	0.017	0.009	0.022
5	0.020	0.014	0.010	0.017	0.003	0.023
6	0.055	0.022	0.020	0.034	0.010	0.045
7	0.072	0.017	0.036	0.024	0.017	0.033
8	0.052	0.026	0.019	0.030	0.013	0.040
9	0.056	0.022	0.026	0.026	0.020	0.036
10	0.058	0.042	0.031	0.049	0.019	0.055
11	0.060	0.029	0.027	0.034	0.020	0.042
12	0.096	0.014	0.058	0.017	0.037	0.019

Table 3.28: Type II Errors: Normal Prior Distributions for 16 Run Design - Main Effects Only

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1	0.183	0.840	0.074	0.273	0.010	0.035
2	0.118	0.835	0.048	0.275	0.010	0.045
3	0.090	0.829	0.041	0.293	0.007	0.040
4	0.060	0.823	0.026	0.271	0.003	0.037
5	0.039	0.807	0.018	0.285	0.003	0.043
6	0.079	0.530	0.021	0.184	0.007	0.054
7	0.087	0.769	0.035	0.341	0.019	0.117
8	0.056	0.592	0.027	0.306	0.022	0.144
9	0.065	0.827	0.055	0.611	0.041	0.374
10	0.033	0.815	0.036	0.745	0.035	0.609
11	0.071	0.812	0.055	0.527	0.032	0.300
12	0.070	1.078	0.061	0.984	0.066	0.743

Table 3.29: Type I Errors: Normal Prior Distributions for 12 Run Design - Main Effects Only

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1	0.235	0.038	0.124	0.037	0.059	0.040
2	0.107	0.020	0.047	0.023	0.023	0.028
3	0.081	0.021	0.032	0.025	0.014	0.029
4	0.048	0.022	0.021	0.024	0.009	0.028
5	0.032	0.014	0.012	0.017	0.004	0.023
6	0.072	0.045	0.025	0.054	0.012	0.062
7	0.086	0.035	0.036	0.038	0.021	0.045
8	0.068	0.049	0.022	0.051	0.014	0.058
9	0.080	0.039	0.032	0.042	0.024	0.051
10	0.083	0.047	0.033	0.070	0.028	0.073
11	0.068	0.051	0.035	0.054	0.022	0.060
12	0.127	0.023	0.068	0.024	0.041	0.025

Table 3.30: Type II Errors: Normal Prior Distributions for 12 Run Design - Main Effects Only

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1	0.212	0.843	0.086	0.275	0.011	0.035
2	0.123	0.839	0.053	0.275	0.010	0.045
3	0.094	0.830	0.041	0.292	0.007	0.039
4	0.065	0.824	0.027	0.272	0.005	0.037
5	0.037	0.807	0.025	0.284	0.004	0.043
6	0.083	0.537	0.034	0.188	0.010	0.054
7	0.105	0.771	0.054	0.342	0.026	0.116
8	0.075	0.605	0.040	0.318	0.025	0.151
9	0.068	0.825	0.065	0.615	0.044	0.377
10	0.057	0.871	0.038	0.747	0.039	0.610
11	0.095	0.809	0.058	0.526	0.035	0.300
12	0.072	1.075	0.063	0.981	0.066	0.739

Table 3.31: Screening Utility Results: Normal Prior Distributions for 16 Run Design - Main Effects Only

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1	0.190	0.427	0.083	0.146	0.026	0.030
2	0.107	0.422	0.046	0.146	0.012	0.034
3	0.076	0.422	0.032	0.157	0.009	0.033
4	0.049	0.418	0.023	0.144	0.006	0.029
5	0.030	0.411	0.014	0.151	0.003	0.033
6	0.067	0.276	0.020	0.109	0.008	0.049
7	0.080	0.393	0.035	0.183	0.018	0.075
8	0.054	0.309	0.023	0.168	0.017	0.092
9	0.061	0.424	0.041	0.318	0.031	0.205
10	0.045	0.429	0.034	0.397	0.027	0.332
11	0.065	0.420	0.041	0.281	0.026	0.171
12	0.083	0.546	0.059	0.500	0.051	0.381

Table 3.32: Screening Utility Results: Normal Prior Distributions for 12 Run Design - Main Effects Only

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1	0.223	0.440	0.105	0.156	0.035	0.038
2	0.115	0.429	0.050	0.149	0.016	0.037
3	0.087	0.425	0.036	0.158	0.011	0.034
4	0.057	0.423	0.024	0.148	0.007	0.033
5	0.035	0.411	0.018	0.151	0.004	0.033
6	0.077	0.291	0.029	0.121	0.011	0.058
7	0.096	0.403	0.045	0.190	0.023	0.081
8	0.075	0.605	0.031	0.184	0.019	0.104
9	0.074	0.432	0.048	0.328	0.034	0.214
10	0.070	0.459	0.036	0.408	0.034	0.341
11	0.082	0.430	0.046	0.290	0.029	0.180
12	0.100	0.549	0.066	0.503	0.054	0.382

Table 3.33: Type I Errors: Normal Prior Distributions for 16 Run Design - Main Effects and all Two-Factor Interactions

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1	0.164	0.003	0.061	0.004	0.022	0.006
2	0.075	0.003	0.026	0.006	0.009	0.009
3	0.051	0.003	0.020	0.004	0.007	0.007
4	0.031	0.002	0.009	0.005	0.004	0.007
5	0.020	0.001	0.007	0.002	0.002	0.004
6	0.066	0.004	0.027	0.008	0.012	0.013
7	0.076	0.002	0.037	0.004	0.016	0.006
8	0.068	0.002	0.027	0.005	0.014	0.011
9	0.077	0.001	0.037	0.003	0.022	0.006
10	0.080	0.001	0.041	0.002	0.028	0.005
11	0.071	0.001	0.038	0.004	0.022	0.007
12	0.109	0.000	0.058	0.001	0.034	0.002

Table 3.34: Type II Errors: Normal Prior Distributions for 16 Run Design - Main Effects and all Two-Factor Interactions

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1	0.323	1.204	0.108	0.401	0.015	0.044
2	0.214	1.213	0.081	0.446	0.012	0.054
3	0.121	1.180	0.097	0.684	0.019	0.115
4	0.115	1.170	0.049	0.473	0.007	0.051
5	0.069	1.243	0.045	0.578	0.006	0.077
6	0.147	1.113	0.130	0.773	0.045	0.224
7	0.133	1.124	0.129	1.063	0.087	0.506
8	0.122	1.046	0.114	0.961	0.073	0.560
9	0.098	1.066	0.100	1.106	0.118	0.938
10	0.062	0.971	0.071	0.986	0.084	1.001
11	0.062	0.949	0.069	0.963	0.076	0.955
12	0.060	1.029	0.077	1.085	0.088	1.172

Table 3.35: Type I Errors: Normal Prior Distributions for 12 Run Design - Main Effects and 3 Two-Factor Interactions

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1	0.179	0.478	0.098	0.488	0.047	0.498
2	0.146	0.485	0.072	0.492	0.032	0.502
3	0.182	0.504	0.098	0.497	0.048	0.499
4	0.129	0.484	0.061	0.492	0.026	0.500
5	0.151	0.495	0.079	0.495	0.037	0.499
6	0.191	0.528	0.094	0.516	0.045	0.512
7	0.224	0.522	0.120	0.504	0.066	0.499
8	0.223	0.525	0.124	0.512	0.069	0.508
9	0.259	0.525	0.158	0.505	0.096	0.499
10	0.320	0.530	0.217	0.503	0.149	0.491
11	0.370	0.562	0.267	0.543	0.187	0.533
12	0.361	0.543	0.261	0.520	0.196	0.510

Table 3.36: Type II Errors: Normal Prior Distributions for 12 Run Design - Main Effects and 3 Two-Factor Interactions

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1	0.584	0.405	0.151	0.069	0.014	0.003
2	0.637	0.469	0.230	0.129	0.026	0.010
3	0.730	0.590	0.494	0.358	0.086	0.056
4	0.619	0.470	0.228	0.130	0.021	0.007
5	0.696	0.538	0.356	0.244	0.040	0.026
6	0.736	0.611	0.483	0.402	0.171	0.130
7	0.710	0.597	0.634	0.503	0.286	0.209
8	0.669	0.565	0.625	0.511	0.396	0.301
9	0.638	0.539	0.694	0.556	0.564	0.416
10	0.575	0.511	0.589	0.503	0.586	0.482
11	0.650	0.569	0.721	0.624	0.759	0.655
12	0.588	0.508	0.643	0.536	0.699	0.573

Table 3.37: Screening Utility Results: Normal Prior Distributions for 16 Run Design - Main Effects and all Two-Factor Interactions

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1	0.243	0.603	0.084	0.202	0.018	0.025
2	0.145	0.608	0.054	0.226	0.010	0.032
3	0.086	0.591	0.059	0.344	0.013	0.061
4	0.073	0.586	0.029	0.239	0.005	0.029
5	0.045	0.622	0.026	0.290	0.004	0.041
6	0.106	0.558	0.078	0.391	0.029	0.119
7	0.104	0.563	0.083	0.534	0.051	0.256
8	0.095	0.524	0.071	0.483	0.044	0.286
9	0.087	0.533	0.069	0.555	0.070	0.472
10	0.071	0.486	0.056	0.494	0.056	0.503
11	0.067	0.475	0.054	0.483	0.049	0.481
12	0.085	0.515	0.068	0.543	0.061	0.587

Table 3.38: Screening Utility Results: Normal Prior Distributions for 12 Run Design - Main Effects and 3 Two-Factor Interactions

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1	0.381	0.442	0.124	0.278	0.030	0.250
2	0.392	0.477	0.151	0.311	0.029	0.256
3	0.456	0.547	0.296	0.427	0.067	0.277
4	0.374	0.477	0.145	0.311	0.024	0.253
5	0.423	0.517	0.217	0.370	0.038	0.263
6	0.464	0.570	0.288	0.459	0.108	0.321
7	0.467	0.560	0.377	0.503	0.176	0.354
8	0.446	0.545	0.375	0.511	0.233	0.404
9	0.448	0.532	0.426	0.531	0.330	0.457
10	0.448	0.520	0.403	0.503	0.368	0.487
11	0.510	0.565	0.494	0.583	0.473	0.594
12	0.474	0.523	0.452	0.528	0.447	0.542

Table 3.39: Type I Errors: Normal Mixture Prior Distributions for 16 Run Design - Main Effects Only

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1(a)	0.007	0.077	0.002	0.103	0.001	0.117
1(b)	0.004	0.096	0.002	0.128	0.001	0.143
1(c)	0.002	0.127	0.001	0.156	0.000	0.167
2(a)	0.012	0.099	0.007	0.133	0.003	0.152
2(b)	0.009	0.152	0.003	0.191	0.001	0.213
2(c)	0.005	0.193	0.002	0.226	0.001	0.243
3(a)	0.010	0.100	0.005	0.139	0.002	0.156
3(b)	0.007	0.162	0.004	0.201	0.001	0.218
3(c)	0.003	0.193	0.002	0.230	0.001	0.248

Table 3.40: Type II Errors: Normal Mixture Prior Distributions for 16 Run Design - Main Effects Only

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1(a)	0.007	0.013	0.001	0.001	0.000	0.000
1(b)	0.005	0.011	0.002	0.002	0.000	0.000
1(c)	0.002	0.010	0.001	0.003	0.000	0.000
2(a)	0.012	0.079	0.005	0.011	0.001	0.001
2(b)	0.014	0.079	0.005	0.014	0.001	0.001
2(c)	0.007	0.079	0.004	0.015	0.000	0.001
3(a)	0.017	0.070	0.003	0.011	0.001	0.001
3(b)	0.011	0.047	0.002	0.010	0.000	0.001
3(c)	0.008	0.047	0.002	0.007	0.000	0.000

Table 3.41: Type I Errors: Normal Mixture Prior Distributions for 12 Run Design - Main Effects Only

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1(a)	0.007	0.104	0.004	0.128	0.002	0.142
1(b)	0.003	0.129	0.002	0.159	0.001	0.173
1(c)	0.003	0.161	0.001	0.189	0.000	0.199
2(a)	0.016	0.126	0.007	0.156	0.002	0.175
2(b)	0.010	0.181	0.003	0.217	0.002	0.238
2(c)	0.006	0.225	0.002	0.255	0.000	0.270
3(a)	0.013	0.128	0.006	0.165	0.003	0.180
3(b)	0.009	0.193	0.003	0.229	0.001	0.245
3(c)	0.004	0.228	0.003	0.262	0.000	0.279

Table 3.42: Type II Errors: Normal Mixture Prior Distributions for 12 Run Design - Main Effects Only

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1(a)	0.008	0.014	0.002	0.001	0.000	0.000
1(b)	0.006	0.011	0.001	0.002	0.000	0.000
1(c)	0.002	0.010	0.001	0.003	0.000	0.000
2(a)	0.020	0.080	0.006	0.011	0.001	0.001
2(b)	0.013	0.079	0.005	0.014	0.001	0.001
2(c)	0.008	0.079	0.003	0.015	0.000	0.001
3(a)	0.021	0.072	0.004	0.011	0.001	0.001
3(b)	0.012	0.047	0.003	0.009	0.001	0.001
3(c)	0.007	0.047	0.003	0.007	0.000	0.000

Table 3.43: Screening Utility Results: Normal Mixture Prior Distributions for 16 Run Design - Main Effects Only

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1(a)	0.007	0.045	0.002	0.052	0.000	0.059
1(b)	0.005	0.054	0.002	0.065	0.000	0.071
1(c)	0.002	0.069	0.001	0.080	0.000	0.084
2(a)	0.012	0.089	0.006	0.072	0.002	0.077
2(b)	0.011	0.115	0.004	0.102	0.001	0.107
2(c)	0.006	0.136	0.003	0.121	0.001	0.122
3(a)	0.014	0.085	0.004	0.075	0.002	0.079
3(b)	0.009	0.104	0.003	0.105	0.001	0.110
3(c)	0.006	0.120	0.002	0.119	0.000	0.124

Table 3.44: Screening Utility Results: Normal Mixture Prior Distributions for 12 Run Design - Main Effects Only

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1(a)	0.007	0.059	0.003	0.065	0.001	0.071
1(b)	0.005	0.070	0.002	0.081	0.000	0.086
1(c)	0.003	0.085	0.008	0.096	0.000	0.100
2(a)	0.018	0.103	0.006	0.084	0.001	0.088
2(b)	0.011	0.130	0.004	0.115	0.001	0.119
2(c)	0.007	0.152	0.003	0.135	0.000	0.136
3(a)	0.017	0.100	0.005	0.088	0.002	0.091
3(b)	0.010	0.120	0.003	0.119	0.001	0.123
3(c)	0.005	0.138	0.003	0.135	0.000	0.140

Table 3.45: Type I Errors: Normal Mixture Prior Distributions for 16 Run Design - Main Effects and all Two-Factor Interactions

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1(a)	0.005	0.054	0.003	0.086	0.001	0.101
1(b)	0.003	0.096	0.002	0.129	0.001	0.146
1(c)	0.002	0.119	0.001	0.153	0.000	0.169
2(a)	0.010	0.075	0.004	0.118	0.003	0.142
2(b)	0.008	0.142	0.004	0.188	0.001	0.210
2(c)	0.003	0.200	0.002	0.247	0.001	0.268
3(a)	0.008	0.080	0.005	0.127	0.002	0.150
3(b)	0.006	0.144	0.003	0.195	0.001	0.219
3(c)	0.004	0.191	0.002	0.236	0.001	0.254

Table 3.46: Type II Errors: Normal Mixture Prior Distributions for 16 Run Design - Main Effects and all Two-Factor Interactions

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1(a)	0.024	0.035	0.013	0.004	0.001	0.000
1(b)	0.016	0.016	0.009	0.002	0.001	0.000
1(c)	0.013	0.008	0.003	0.002	0.001	0.000
2(a)	0.058	0.162	0.023	0.030	0.004	0.004
2(b)	0.040	0.114	0.023	0.028	0.001	0.003
2(c)	0.024	0.096	0.015	0.026	0.003	0.002
3(a)	0.051	0.090	0.021	0.017	0.005	0.001
3(b)	0.051	0.054	0.018	0.013	0.003	0.002
3(c)	0.017	0.041	0.009	0.012	0.002	0.001

Table 3.47: Type I Errors: Normal Mixture Prior Distributions for 12 Run Design - Main Effects and 3 Two-Factor Interactions

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1(a)	0.037	0.522	0.012	0.519	0.004	0.516
1(b)	0.037	0.519	0.013	0.513	0.004	0.510
1(c)	0.036	0.495	0.011	0.489	0.005	0.488
2(a)	0.065	0.518	0.025	0.515	0.010	0.513
2(b)	0.066	0.515	0.029	0.512	0.011	0.510
2(c)	0.062	0.514	0.025	0.509	0.011	0.507
3(a)	0.056	0.521	0.021	0.518	0.008	0.516
3(b)	0.056	0.518	0.021	0.513	0.009	0.512
3(c)	0.057	0.514	0.021	0.510	0.009	0.508

Table 3.48: Type II Errors: Normal Mixture Prior Distributions for 12 Run Design - Main Effects and 3 Two-Factor Interactions

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1(a)	0.260	0.225	0.078	0.069	0.013	0.011
1(b)	0.282	0.254	0.080	0.071	0.013	0.012
1(c)	0.301	0.270	0.078	0.068	0.014	0.011
2(a)	0.478	0.399	0.186	0.154	0.029	0.026
2(b)	0.503	0.425	0.187	0.157	0.036	0.030
2(c)	0.486	0.417	0.170	0.149	0.030	0.025
3(a)	0.458	0.381	0.162	0.138	0.023	0.019
3(b)	0.435	0.380	0.142	0.121	0.024	0.020
3(c)	0.487	0.421	0.182	0.153	0.030	0.026

Table 3.49: Screening Utility Results: Normal Mixture Prior Distributions for 16 Run Design - Main Effects and all Two-Factor Interactions

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1(a)	0.015	0.044	0.008	0.045	0.001	0.051
1(b)	0.010	0.056	0.006	0.066	0.001	0.073
1(c)	0.008	0.063	0.002	0.077	0.001	0.085
2(a)	0.034	0.118	0.014	0.074	0.004	0.073
2(b)	0.024	0.128	0.013	0.108	0.001	0.107
2(c)	0.014	0.148	0.008	0.137	0.002	0.135
3(a)	0.030	0.085	0.013	0.072	0.004	0.076
3(b)	0.029	0.099	0.011	0.104	0.002	0.111
3(c)	0.011	0.117	0.005	0.124	0.001	0.128

Table 3.50: Screening Utility Results: Normal Mixture Prior Distributions for 12 Run Design - Main Effects and 3 Two-Factor Interactions

	$\omega = \sigma$		$\omega = 1.5\sigma$		$\omega = 2\sigma$	
Prior	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1(a)	0.149	0.374	0.045	0.294	0.009	0.264
1(b)	0.160	0.386	0.046	0.292	0.008	0.261
1(c)	0.169	0.382	0.045	0.279	0.009	0.249
2(a)	0.271	0.458	0.106	0.335	0.020	0.269
2(b)	0.284	0.470	0.108	0.335	0.023	0.270
2(c)	0.274	0.466	0.098	0.329	0.020	0.266
3(a)	0.257	0.451	0.092	0.328	0.016	0.267
3(b)	0.246	0.449	0.082	0.317	0.016	0.266
3(c)	0.272	0.467	0.102	0.331	0.019	0.267

Normal Prior Distributions on Effect Sizes when Estimating Main Effects Only

Looking at the results in Tables 3.27 - 3.32, when assuming normal prior distributions on effect sizes and estimating main effects only there are some interesting points to note in both the cases of the 12 and 16 run designs. The size of the Type I errors

for both the 12 and 16 run designs (Tables 3.27 and 3.29) are less in the case of method 2 (using Lenth's method to declare an effect active) in comparison to method 1 (declaring an effect active when its absolute value is greater than ω) for all priors when $\omega = \sigma$. The size of the Type I errors are also generally quite low for all levels of ω, i.e. σ, 1.5σ and 2σ. The largest Type I error is for prior 11 for the 12 run design with a size of 0.051, and the smallest is 0.010 (prior 2 for the 16 run design), when $\omega = \sigma$. Relating this back to the definition of a Type I error in the context of identifying active and inactive effects correctly, this translates to an inference that when using Lenth's method, an objective method for deciding which effects are active, inactive effects are declared as active incorrectly very infrequently.

We have previously defined P(Type I error) in this setting as

$$\frac{\sum_{ij}(False\ Positive)}{rs}$$

where i denotes an effect, $j = 1, \ldots, r$ denotes the j^{th} simulation and s represents the number of true inactives. Then with a Type I error of size 0.051 for the particular example of prior 11 for the 12 run design, and the mean number of 'true' actives per dataset for the 1000 simulations being 1.794, we can then calculate the mean number of 'true' inactives as 3.206. Therefore $rs = 3206$ and for the same number of simulations $\sum_{ij}(False\ Positive)$ is approximately 164, in order that the size of the Type I error in this particular case is 0.051. In the case of the smallest Type I error, prior 2 for the 16 run design, the mean number of 'true' actives per dataset is 1.536 and therefore the mean number of 'true' inactives is 3.464. Then, since $rs = 3464$, there are only approximately 35 inactive effects declared as active incorrectly resulting in the proportion of 0.010 of Type I errors for this particular case. It can be seen that of the various combinations of priors investigated, the worst case scenario for Type I errors has only 164 inactive effects declared active incorrectly out of the total possible 3206 true inactives, and in the best case scenario only 35 out of a total possible 3464

true inactives, are identfied as active incorrectly. Thus, it could be concluded based on this information that Lenth's method provides a relatively efficient method for declaring effects as active and inactive correctly. However, looking in more detail at the output of the simulations it can be noted that many of the entries for the vectors $\delta_{(j)}$ where $j = 1, \ldots, r$ are 0. It was previously outlined in the algorithm for the screening utility function in §2.4.3 that an entry in $\delta_{(j)}$ shall be given as 1 if the effect associated with that entry is active and 0 otherwise. In fact, for prior 2 for the 16 run design it can be noted that whilst the mean number of actives for the 'true' datasets is 1.536, for the simulated datasets where Lenth's method has been used to identify active effects there is only a mean of 0.038 actives. Therefore, a more reasonable explanation for the lack of Type I errors is actually due to almost all effects being declared as inactive, rather than Lenth's method being an efficient method. This explanation is also reflected by the Type II errors which are conversely very large in comparison. This is clear to see since almost all effects have been declared as inactive, many truly active effects amongst these effects will also have been identified inactive wrongly, increasing the proportion of Type II errors.

The choice of ω represents the minimum level an effect needs to be greater than for that effect to be declared active for the 'true' datasets, denoted by $\delta^*_{(j)}$. As ω increases from σ to 1.5σ and again to 2σ, we see that the Type I errors also increase in all cases, i.e. for both the 16 and 12 run design, methods 1 and 2 and also for all priors. Although the Type I errors do increase, the increase from σ to 2σ is fairly minimal. In contrast, the Type II errors and consequently the value of the screening utility decrease as the level of ω increases from σ to 1.5σ to 2σ, for both sets of designs and methods, except in the case of priors 10 and 12.

For the priors where the variances on the main effects are all equal, as for priors 1 - 5, it appears that the choice of ω when using method 1, has an effect on the Type I and II errors and the screening utility. In these cases, as the size of the prior

variance on the main effects increase, both Type I and II errors decrease and as a result the screening utility also decreases. It does not appear that the size of the prior variance on the main effect has as great an effect on the Type I and II errors when using Lenth's method. From observing the results of the tables 3.27 - 3.32 it can be concluded that the 16 run design outperforms the 12 run design. This is when looking at the two alternative methods for identifying active effects and also in the case of all priors. However, although the values for the Type I and II errors and the screening utility are minimized when using the 16 run design, it should be noted that the values for each of the corresponding priors for the 12 run design are only marginally worse in comparison. An example of this is seen for the first value displayed in Table 3.31 and 3.32. It can be seen that the screening utility for the 16 run design is 0.190 in comparison to 0.223 for the 12 run design. Thus, it could be concluded that the amount lost in efficiency relative to the benefit of saving 4 runs may be a compromise worth making.

Normal Prior Distributions on Effect Sizes when Estimating Main Effects and Two-Factor Interactions

The tables 3.33 - 3.38 refer to those Type I and II errors and the screening utility obtained when using both a 16 and 12 run design to estimate main effects and two-factor interactions. In the case of the 16 run design, all two-factor interactions have been estimated as there are five main effects and 10 two-factor interactions to be assigned to the 15 contrasts available. This would be seen to be a relatively good design in terms of aliasing since both main effects and two-factor interactions are clear. In the case of the 12 run design, three two-factor interactions have been estimated along with the five main effects, these three two-factor interactions being those associated with the three main effects that have the largest absolute value.

For the Type I and II errors for the 16 run design (Tables 3.33 and 3.34), there are some similarities in the results in comparison to the case of the 16 run design when estimating main effects only as outlined in the previous subsection. The Type I errors are again less when using Lenth's method as opposed to the method where one declares an effect active if it is greater than ω. The size of the Type I errors when using method 1 to identify active effects are fairly similar to those Type I errors for the 16 run design estimating main effects only. It is also seen that the Type I errors when using Lenth's method are extremely small, where the largest is 0.013 for prior 6 and when $\omega = 2\sigma$, and the size of the smallest is 0.001 seen in several cases. Also, as the size of ω increases from σ to 1.5σ to 2σ it is again seen that the Type I errors decrease for all of the priors investigated, and similarly the Type II errors decrease as ω increases, except in the case of priors 9 - 12. Looking at priors 1 - 5 for both Type I and II errors it is seen that as the prior variance on the main effects increase the size of these errors decrease. As was seen when estimating main effects only, the Type I and II errors do not appear to be affected by the varying size of the prior variance on the main effects when using Lenth's method.

When looking at the Type I and II errors in the case of the 12 run design where main effects and two-factor interactions have been estimated (Tables 3.35 and 3.36) it appears that there are fewer patterns emerging. The Type I errors are much larger when using Lenth's method in comparison to method 1; however when using either method for identifying active effects, it appears that there is little effect on the Type I and II errors when varying the size of the prior variance on the main effects. It appears that when the prior variances on the main effects are all equal as in priors 1 - 5 then however large the prior variance, the size of the errors are all fairly similar given the same method and size of ω. The Type I and II errors do again tend to decrease as the size of ω increases, except in the case of priors 9-12 for the Type II errors.

The results from the screening utility (Tables 3.37 and 3.38) indicate that when using method 1, the 16 run design has the optimum utility in comparison to the 12 run design. However, the converse is true when using method 2.

Normal Mixture Prior Distributions on Effect Sizes when Estimating Main Effects Only

Tables 3.39 - 3.44 display the results for the 16 and 12 run designs when normal mixture prior distributions are assumed on the effect sizes and only main effects are estimated.

In the case of both the 16 and 12 run design it can be seen that as the prior variance on the main effects increase, the Type I errors decrease when using method 1 but increase when using method 2. The same pattern is true when the level of ω increases. That is, as ω increases from σ to 1.5σ to 2σ, the Type I errors decrease when using method 1 and increase when using method 2. The Type I errors in the case of designs of both sizes are also minimal when using method 1, particularly when $\omega = 2\sigma$ where for both designs the largest Type I error is 0.003.

For both the 16 and 12 run design, the Type II errors are generally smaller when using method 1 and for both methods the Type II errors decrease as the size of ω increases. The Type II errors are minimal for both methods when $\omega = 2\sigma$, where the largest error is 0.001.

The screening utility indicates that method 1 is preferable for both designs as the utility gives a smaller quantity in all cases. The 16 run design also appears to be preferable in the majority of cases due to the lower utility when comparing the same priors for both designs, although it should be noted that the difference between the utilities when comparing the two designs is small.

Normal Mixture Prior Distributions on Effect Sizes when Estimating Main Effects and Two-Factor Interactions

Tables 3.45 - 3.50 display the results of the Type I and II errors and screening utility for the 16 and 12 run design. This is when normal mixture prior distributions are assumed on all main effects and two-factor interactions in the case of the 16 run design and all main effects and those three two-factor interactions associated with the three most important main effects for the 12 run design.

As was seen when estimating main effects only, for the 16 run design as the prior variance on the main effects increase and the level of ω increases from σ to 1.5σ to 2σ, the Type I errors decrease when using method 1. Conversely, the Type I errors increase when using method 2. It should also be noted that the Type I errors are fairly similar for both the 16 run designs whether estimating main effects only or estimating main effects and all two factor-interactions. However, in the case of the 12 run design where main effects and two-factor interactions have been estimated, the Type I errors are larger when comparing the same priors to that when estimating main effects only. Also, when estimating main effects and two-factor interactions for the 12 run design, the Type I errors are of a similar value despite the prior variance on the main effects increasing when using method 1. For example, the priors $1(a)$, $1(b)$ and $1(c)$ in Table 3.47 represent the scenario where all five main effects have the probability 0.2 of being active, and the two-factor interactions have probability of 0.1, 0.4 or 0.7 of being active dependent upon whether both parent effects are not active, one of the parent effects is active or both parent effects are active, respectively. If a main effect or two-factor interaction is then active, its distribution is a high variance normal distribution, either $N(0, 4)$, $N(0, 10)$ or $N(0, 25)$ (respectively in the case of $1(a)$, $1(b)$ or $1(c)$), and takes $N(0, 0.01)$ otherwise. It can be seen in Table 3.47 that even as the prior variance on the effects increase the Type I error for $1(a)$, $1(b)$ and $1(c)$ is 0.037, 0.037 and 0.036 respectively. A similar pattern is also seen for all other

priors indicating that the prior variance does not appear to be having an effect on the Type I error when using method 1. However, when using method 2, it can be seen that the Type I error decreases as the size of the prior variance on the main effects increases. This is in contrast to the 16 run design, or to the 12 run design when estimating main effects only. As the level of ω increases the Type I errors decrease when using either method.

The Type II errors for both the 16 and the 12 run designs are larger when estimating two-factor interactions in addition to the main effects, in comparison to when estimating main effects only. For almost all cases for the 16 run design, as the prior variance on the effects and the level of ω increase, the Type II errors decrease and this is true for both methods of identifying active effects. For the 12 run design, as the level of ω increases the Type II errors again decrease, although the prior variance on the effects increasing does not appear to have an impact on the Type II error.

The values of the screening utility given in Tables 3.49 and 3.50 indicate that the 16 run design is preferable for identifying active and inactive effects correctly due to the utilities being lower than for the 12 run design. Where the difference in utilities between the two designs appeared to be minimal when estimating main effects only, the difference in utilities is much greater between the 16 and 12 run design when estimating two-factor interactions along with the main effects. It can be seen that the utilities for the 12 run design are much greater when estimating two-factor interactions in comparison to those utilities obtained when only estimating main effects. The utilities for the 16 run design when estimating main effects and two-factor interactions (Table 3.49) are fairly similar to those utilities for the 16 run design when estimating main effects only (Table 3.43). The utilities are also generally lower when using method 1 for both the 12 and 16 run designs.

3.3.5 Discussion

The utility function for screening as outlined in §2.4, which assesses how well active and inactive effects are identified correctly in a particular design, has been demonstrated with an example comparing a 2^{5-1} design and a 12 run Plackett-Burman design where five variables are being investigated. Various prior assumptions about the effect sizes were considered and two different methods were also considered to decide whether an effect will be declared active, namely that an effect is declared active if it is greater than some constant ω and Lenth's method.

In many of the cases it was seen that the probability of Type I errors were much smaller when using Lenth's method and the Type II errors much larger, when comparing Lenth's method to the alternative method for identifying active and inactive effects. Due to the Type II errors being much larger in comparison to those when using the alternative method for declaring effects active, i.e. if $|\beta_{i(j)}| > \omega$, which we refer to as method 1, the quantity $\gamma U_1 + (1-\gamma)U_2$ is much larger when using Lenth's method. As we wish to minimize this quantity, it would be advisable for an experimenter to use method 1 for the purpose of deciding which effects are active, when measuring the performance of a design. This is in order to reduce the loss of efficiency in identifying active and inactive effects correctly. It should be noted that the recommendation for using this method is when an experimenter is concerned with the choice of design and not for when the experimenter is at the analysis stage of the investigation. Although at first sight the much smaller Type I errors when using Lenth's method appears to indicate that this method proves more efficient, a closer look at the simulated datasets actually highlights that many effects are being declared as inactive, which when comparing to the 'true' dataset, are effects that should have in fact been active. This seems to occur when using Lenth's method due to the PSE being relatively large, and consequently the levels of ME and SME also being large. Thus, many estimated effects are failing to exceed SME and are therefore recorded as 0, an inactive entry.

As well as highlighting an inadequacy of Lenth's method in this context, it also draws attention to the utility function being used and how it is defined and whether using solely Type I and Type II errors as a basis for deciding upon the efficiency of a design is an appropriate measure. If the efficiency of a candidate design is being based upon how well active and inactive effects are declared correctly, to achieve minimal Type I errors one may seek to falsely declare all effects as inactive to minimize this type of error. Conversely, to minimize Type II errors one may falsely declare all effects as active. Experimenters should be aware of the impact of making such false declarations and so treat results with caution in such situations. Also, the choice of the weight γ must be seriously considered by the experimenter to correctly represent the proportions of the two components U_1 and U_2 where a greater weight would be placed on that component the experimenter considers to be more important. If an experimenter feels both Type I and Type II errors are of equal importance then this would imply that $\gamma = 0.5$. However, if an experimenter feels that Type I errors are three times as important as Type II errors then $\gamma = 0.75$ will be used, placing a weight of 0.75 on U_1 and accordingly, $(1 - \gamma) = 0.25$ will be placed upon U_2. In deciding the weights, it is the responsibility of the experimenter to act as decision maker and assign importance to the criteria.

A sensitivity analysis was carried out where several different prior distributions were applied to the case where only main effects were estimated and also when both main effects and two-factor interactions were estimated. This was done to investigate the effect, if any, of changing the prior distribution and determine if this does have any impact on the value of utility obtained. Also relating this to the interpretation of the utility function, this would be to determine if changing the prior distribution has any impact on how well a design performs in terms of active and inactive effects being identified.

The mean of the prior distributions were fixed to zero for all cases investigated, for

the reasons stated earlier on in a previous subsection (§3.3.3). However, the variances of the prior distributions were changed and investigated where all variances on effects were equal and ranged from low prior variances to high prior variances. Prior distributions where the variances are not all equal were also looked at. When estimating main effects only and assuming normal prior distributions with equal variance on all main effects, the greatest impact of changing the variances was seen when using the method of assigning a constant ω, and declaring an effect active if it is greater than ω. The quantity $\gamma U_1 + (1-\gamma)U_2$, which we would seek to minimize, was reduced when a high prior variance was assumed. In comparison, when using Lenth's method there appeared to be little impact in changing the variance. An implication of these results may be that when using method 1, if an experimenter believes it is appropriate to assume high prior variances on main effects then the loss in efficiency of the design in terms of identifying active and inactive effects correctly will be reduced.

Observing the results when the variances on the main effects are not equal, it appears that there is a slight indication that when the unequal variances all tend to be smaller, the quantity $\gamma U_1 + (1-\gamma)U_2$ is somewhat reduced compared to when all unequal variances are larger. These results may highlight that the low prior variances have been understated. However, if an experimenter does believe that low prior variances do appropriately represent their beliefs about the parameters being investigated, then the loss of efficiency in correctly identifying active and inactive effects is reduced. This inference does need to be treated with some caution though as the value of $\widehat{E(U)}$ (where $\widehat{E(U)} = \gamma U_1 + (1-\gamma)U_2$) for the normal prior distribution with all relatively small variances, ($\Sigma = [\infty, 0.25, 0.25, 0.25, 2, 2]$) is 0.077 and when the normal prior distribution has all much larger variances ($\Sigma = [\infty, 2, 2, 4, 4, 10]$) it is 0.100. An experimenter may have to decide whether this difference is large enough to be considered a substantial difference and therefore whether specifying low prior variances do in fact lead to a greater efficiency in terms of identifying actives and inactives correctly. These implications apply in the case when using both methods

for declaring effects active and also for both the 12 and 16 run designs.

When assuming normal prior distributions on effect sizes and both main effects and all two-factor interactions are estimated in the case of the 16 run design, and variances on all main effects are equal and also variances on all two-factor interactions are equal when using method 1, then similar results are seen to the estimating main effects only case. That is, the quantity we seek to minimize is reduced when high prior variances are assumed. Thus, the implication for an experimenter is that if they believe it is reasonable to place high prior variances on all effects, and they consider all main effects to be exchangeable and also two-factor interactions to be exchangeable, then the efficiency of the design will be improved. In comparison, when assuming normal prior distributions and estimating all main effects and only the three two-factor interactions associated with the three largest main effects in the case of the 12 run design, differing results are observed.

There appears to be little impact in changing the prior variances on the utility when using either method to declare effects active and may in fact lead us to reconsider the way in which the two-factor interactions that have been estimated have been chosen. For both the 12 and 16 run designs when assuming normal prior distributions and prior variances on main effects and two-factor interactions are not all equal, there appears to be little connection between changing the prior variance and the value $\widehat{E(U)}$. The implication of these results are that if an experimenter believes main effects and two-factor interactions are not exchangeable, the size of the prior variance will have little impact upon the efficiency of the design.

When assuming normal mixture prior distributions, using method 1 and estimating only main effects for both designs, as the prior variance on the active effect is increased, the active effect coming from the high variance distribution β_{ib}, the value of $\widehat{E(U)}$ decreases. This implies that an experimenter should place high prior variances

such that $\sigma_{ib}^2 \gg \sigma_{ia}^2$ and also where σ_{ib}^2 is at the maximum value that the experimenter believes it can take, on the high variance component β_{ib} when the prior distribution for the effects takes the form of the normal mixture prior, i.e. when

$$\beta_{ia} \sim N(\mu_{ia}, \sigma_{ia}^2)$$
$$\beta_{ib} \sim N(\mu_{ib}, \sigma_{ib}^2).$$

This would be in order to improve the efficiency of the design in terms of correctly identifying active and inactive effects. However, in contrast when using Lenth's method the converse is true for both designs in that as the prior variances on main effects increase, $\widehat{E(U)}$ also increases. The small prior variance leading to a loss in $\gamma U_1 + (1-\gamma)U_2$ may perhaps though be indicative of the variances being understated. The above implications also hold true when looking at the 16 run design, assuming normal mixture prior distributions on effects and estimating main effects and all two-factor interactions. That is, high prior variances lead to a reduction of $\widehat{E(U)}$ when using method 1 and the opposite when using Lenth's method. However, when considering the 12 run design and estimating main effects and three two-factor interactions when placing normal mixture prior distributions on effects it appears there is little or almost no effect at all of changing the size of the prior variance. It can be seen that for all probabilities $p_1, \ldots, p_5$ of a main effect to be active and for all probabilities $p_{00}, p_{01}, p_{10}, p_{11}$ of a two-factor interaction being active considered, regardless of increasing the prior variance of the high variance distribution β_{ib}, the value of $\widehat{E(U)}$ is approximately the same. This may imply that if experimenters believed it would be appropriate to model parameter effects with prior normal mixture distributions, then they can be advised that the size of the prior variance is irrelevant. However, a more plausible explanation for the results may actually be to do with the fact that all two-factor interactions have not been estimated. Again, attention is drawn to the way in which the 3 two-factor interactions have been chosen, the method by which only some of the two-factor interaction effects have been estimated when using a 12 run design as there are not enough contrasts available to estimate them all.

Now considering the value of ω, the level of which is set by the experimenter and determines how large an effect has to be to be considered active, it appears that some care should be taken in setting this value. It appears in most cases, particularly when assuming normal mixture prior distributions, that the value of the expected utility is greatly reduced as the value of ω is increased from the smallest level investigated, σ, to the highest level investigated, 2σ. This may be indicative that a higher level of ω results in a design where effects are more correctly identified as active and inactive. However, a more credible explanation may be that the level of ω is set to such a high level that very few effects are declared as active, thus resulting in misleading values of Type I errors as mentioned earlier leading to incorrect inferences. Therefore, it is important that if using this method an experimenter uses any historical data or prior knowledge about the process being investigated to correctly set the level of ω. This is to ensure that not too many or too few effects are declared as active and to therefore enable the key few active factors having an impact on the process in hand to be correctly distinguished.

The values of $\widehat{E(U)}$ calculated as displayed in Tables 3.31, 3.32, 3.37, 3.38, 3.43, 3.44, 3.49 and 3.50 are a measure of how well active and inactive effects are correctly identified for a particular design. To be able to draw some conclusions regarding whether it is worthwhile for an experimenter to carry out extra runs, which would then typically result in the experiment being more costly and time-consuming, it will be necessary to judge the design by the worth of each particular run. That is, the benefit of carrying out extra runs has to be weighed against the run. In the context of this example we are comparing a 16 run design which we shall denote design 1, and the run size as $n_1 = 16$, and a 12 run design which we shall denote design 2 and its run size as $n_2 = 12$. Thus, to try to quantify whether there is any benefit of carrying out an extra 4 runs, we shall need to judge the value of $\widehat{E(U)}$ against n_1 and n_2. In almost all cases the value of $\widehat{E(U)}$ is in fact lower for the 16 run design than the 12 run design when comparing the same prior assumptions. However, we need some

measure by which to compare the two quantities on the same scale. A measure which could be of some meaning would be to look at $n_1\widehat{E(U)}$ and $n_2\widehat{E(U)}$. Although the value of $\widehat{E(U)}$ is lower for the 16 run design than the 12 run design when looking at the same prior assumptions, the 12 run design actually appears to give the optimal result when comparing $n_1\widehat{E(U)}$ and $n_2\widehat{E(U)}$ for the majority of assumptions. Since $\widehat{E(U)}$ is a quantity that we wish to minimize, when calculating $n_2\widehat{E(U)}$ as being less than $n_1\widehat{E(U)}$, the implication of a smaller value is that the 12 run design can be concluded as bring more beneficial a design when measuring the worth of the design per run. Thus, given the majority of prior assumptions that have been made one could suggest to an experimenter that it may be worthwhile carrying out a 12 run design as opposed to the 16 run design. However, in this particular example this would mean that strong assumptions would have to be made that 7 of the two-factor interactions are negligible, an assumption that would not have to be made if carrying out a 16 run experiment. Thus, although it appears it may be more beneficial to carry out only 12 experimental runs given the prior assumptions that we have investigated here, this advice should be treated with caution if for some reason an experimenter has a reason for not wanting to make this strong assumption or wanted to estimate more two-factor interactions. It should be noted that as mentioned previously, cost considerations can also be considered rather than looking at the worth per run in the arbitrary manner that has been considered here. Although an approach for quantifying the benefits of carrying out an additional number of runs and cost considerations has not been discussed here in great detail, this is of much importance and is an area that would be valuable to investigate further. It would also be worthwhile considering costs in the situation where both types of mistake are made by an experimenter. That is, both falsely declaring an inactive effect as active and also failing to declare a truly active effect as active.

The limitations of the study should be discussed, and as was highlighted with the research in the previous section looking at the optimization utility function, in this

section there are again dependencies of the results upon the various choices of prior distributions and experimental design made and the interplay between the prior beliefs and aliasing structures of the designs considered. It may also be thought that the probabilities chosen for an effect to be active in the case of the normal mixture priors were quite high and potentially violate the notion of effect sparsity. Although this may be the case and perhaps is not a wholly accurate depiction of a situation that would occur in industry, for the purposes of this study these probabilities were considered in order for there to be a high chance for some effects to be active in order to then be able to illustrate whether these active effects were then correctly identified as active using the various methods outlined.

The potential drawbacks of Lenth's method were discussed in Chapter 1 and may provide some explanation for the high errors seen in the results. It was observed when using Lenth's method that much less was declared active, even when several of the effects were found to be truly active resulting in high Type II errors. This may be due to the inaccuracies in the computation of Lenth's critical values for the SME and ME. It has been shown that the margin of error ME is very conservative for small experiments and the inaccuracy even larger for the simultaneous margin of error SME (Olguin and Fearn, 1997). Besides the lack of precision of the approximations, when calculating the SME the assumption is made that the test statistics are independent which is only approximately true, in order to obtain the critical values (Lenth, 1989). It was also observed that although the Type II errors were approximately the same when comparing the 16 and 12 run design in the case of a main effects only model, when interaction effects were included in the model the Type II errors were much lower for the 12 run design when using Lenth's method to identify active effects. This again potentially demonstrates some inadequacies of Lenth's method. One may expect that as fewer contrasts are included in the Plackett-Burman design (since only three of the interaction effects are estimated along with all main effects) that the ME and SME would be inaccurately approximated leading to higher Type II errors due to

the failure of active effects being declared active. However, the 16 run design in fact has higher errors in this case and may highlight that there is some interplay between the choice of method in identifying significant effects, and the aliasing structure of the experimental design. The partial aliasing involved in the Plackett-Burman design may lead to some inaccuracies when calculating the critical values. The disadvantages of Lenth's method that have been highlighted lead to some question surrounding the strategies used to identify active effects, and it may have been worthwhile to consider some alternative such as using Bayesian methods.

There are also some limitations in the study which arise from the same value of ω that is used in both the data generation and the analysis, and this may not be entirely appropriate. It was also previously stated that the recommendation of using the method, $|\beta_{i(j)}| > \omega$, for declaring effects active is when an experimenter is concerned with the choice of design and not at the analysis stage of an investigation. However, this statement may not be strictly true considering that the design choice is based upon the analysis in this case.

3.4 Optimization: Comparison of 5 Factors in a 2-level or 3-level Design

3.4.1 Introduction

The example presented in §3.2 was used to demonstrate how a utility function could aid an experimenter in assessing the benefits of various designs. This utility measured the gain that could be obtained in terms of using the estimated optimal treatment combination from a particular design, in place of the standard operating conditions. Alternatively, one could think of the benefits of a design based on the loss that would be incurred by running it, and thus one would want to try to minimise this loss. This loss being referred to can be considered to be a measure of what one would obtain from the using the estimated optimal treatment combination, instead of the true optimal treatment combination.

As mentioned previously, this example was a relatively small one, where in practice it would be very rare for an experimenter to carry out a design consisting of 4 runs. Thus, an example which more accurately represents problems that an experimenter would face is presented in this section. This example will also not only be more challenging in terms of the run size being larger, but also more complex since the two designs being compared will involve one design with all factors at two levels, and the other design with all factors at three levels. An experimenter may be faced with such a situation if they wanted to investigate a number of factors, but only had enough resources to check either main effects and interaction effects (two-level design) or main effects and quadratic effects (three-level design) but not both.

3.4.2 Possible 2-level and 3-level Designs to use

Consider the case where an experimenter wants to investigate 5 factors and is concerned with locating the optimal treatment combination, but only has enough resources for approximately 18 experimental runs. Hence, the experimenter is confronted with the situation of choosing the design which will best exploit maximum information about the factors and determine the treatment combination which produces the optimum response relative to the objectives of their investigation.

In the case where one would want to estimate main effects and two-factor interaction effects, it would clearly make sense to consider two levels of each of the factors, in order for these effects to be estimable. The most straightforward design to use would appear to be one where each of the factors has runs allocated according to the conventional standard ordering. For 5 factors to be investigated in such a design, a standard half-fraction of a 2^5 design is most appropriate where all main effects and two-factor interaction effects can be estimated clearly. However, effectively two experimental runs are being thrown away due to the size of such a design consisting of 16 experimental runs. If resources do allow that up to 18 experimental runs can be carried out, the half-fraction which is in fact a 2_V^{5-1} design, could be carried out along with the last two runs being used as *centrepoints*. Centrepoints are runs that are at the centre of the design region. This would ensure that the two extra runs are not discarded and it is also typically good practice to use centrepoints. Benefits of using centrepoints are that repeatibility and lack of fit are able to be assessed. However, in such a situation where an experimenter is primarily concerned with optimization, it is not necessary to check for such things and would be sufficient to carry out a standard half-fraction of the 2^5 design, although this is provided that the correct model is being assumed. The design that would be used in this scenario is displayed in Table 3.51. As mentioned, the first 4 factors have their runs allocated to standard ordering and factor **5** has being assigned the column **1234** as a result of the design generator which

is $\mathbf{5 = 1234}$.

Table 3.51: Possible design for 5 2-level Factors in 16 Runs: 2^{5-1} Design

	Variable				
Run	**1**	**2**	**3**	**4**	**5**
1	-	-	-	-	+
2	+	-	-	-	-
3	-	+	-	-	-
4	+	+	-	-	+
5	-	-	+	-	-
6	+	-	+	-	+
7	-	+	+	-	+
8	+	+	+	-	-
9	-	-	-	+	-
10	+	-	-	+	+
11	-	+	-	+	+
12	+	+	-	+	-
13	-	-	+	+	+
14	+	-	+	+	-
15	-	+	+	+	-
16	+	+	+	+	+

In the circumstance where an experimenter wants to estimate main effects and quadratic effects, it would be best to consider a three-level design. Tsai, Gilmour and Mead (2000, 2004) outlined orthogonal main effects plans for three-level designs for 4, 5 and 6 factors in 18 runs. These were found to have better projection properties and provided better parameter estimates for a range of possible models in comparison to designs obtained from the existing L_{18} orthogonal array. They also introduced a criterion (Tsai, Gilmour and Mead, 2000), denoted by $Q(\Gamma^k)$ for a k-factor design, which can be used to explore the projection efficiencies of the design. It was found that designs with lower values of $Q(\Gamma^k)$ are more likely to have efficient projections and on average can provide better parameter estimates over a range of models than designs with higher $Q(\Gamma^k)$. Plans for designs for 5 factors in 18 runs were presented by Tsai, Gilmour and Mead (2004), some of which were found to have low values of

$Q(\Gamma^5)$, which therefore means these designs have better projection properties than other designs. They were also found to be ranked higher than projected 5-factor design from the existing L_{18} in terms of the $Q(\Gamma^5)$ criterion. Thus, a potential design an experimenter could consider to estimate main effects and quadratic effects could be one of these main effects plans. The best three designs according to the $Q(\Gamma^k)$ criterion for 5 factors in 18 runs are presented in Table 3.52.

The main effects plan for Design 3 as indicated in Table 3.52 shall be used in the simulation work later on in this chapter to assess the benefits of a three-level design. The standard half-fraction of a 2^5 design as indicated in Table 3.51 shall be used as the main effects plan for a two-level design. These main effects plans for both two- and three-level designs shall be used to assess whether a two- or three-level design is better to use when resources dictate that only a maximum of 18 experimental runs are available in which to investigate 5 factors.

3.4.3 The Nature of the Second-Order Surface

Using the least squares method to estimate the parameters in the model, the fitted second-order model is

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i + \sum \sum_{i<j}^{k} \hat{\beta}_{ij} x_i x_j + \sum_{i=1}^{k} \hat{\beta}_{ii} x_i^2.$$

In this second-order model k denotes the number of variables, β_i represents the linear effect of x_i, β_{ij} represents the linear-by-linear interaction effect of x_i and x_j and β_{ii} represents the quadratic effect of x_i. This fitted second-order model can also be expressed as

$$\hat{y} = \hat{\beta}_0 + \mathbf{x}^{\mathrm{T}}\mathbf{b} + \mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x},$$

Table 3.52: Designs for 5 Factors in 18 Runs

Design 1

	Variable				
Run	**1**	**2**	**3**	**4**	**5**
1	-	-	-	-	0
2	-	-	0	+	0
3	-	0	-	0	-
4	-	0	+	-	+
5	-	+	0	0	+
6	-	+	+	+	-
7	0	-	0	0	+
8	0	-	+	-	-
9	0	0	-	+	+
10	0	0	+	0	0
11	0	+	-	+	0
12	0	+	0	-	-
13	+	-	-	0	-
14	+	-	+	+	+
15	+	0	0	-	0
16	+	0	0	+	-
17	+	+	-	-	+
18	+	+	+	0	0

Design 2

	Variable				
Run	**1**	**2**	**3**	**4**	**5**
1	-	-	-	-	-
2	-	-	+	0	0
3	-	0	-	+	+
4	-	0	0	-	0
5	-	+	0	+	-
6	-	+	+	0	+
7	0	-	0	0	-
8	0	-	+	+	+
9	0	0	-	0	0
10	0	0	+	-	-
11	0	+	-	-	+
12	0	+	0	+	0
13	+	-	-	+	0
14	+	-	0	-	+
15	+	0	0	0	+
16	+	0	+	+	-
17	+	+	-	0	-
18	+	+	+	-	0

Design 3

	Variable				
Run	**1**	**2**	**3**	**4**	**5**
1	-	-	-	-	0
2	-	-	0	0	-
3	-	0	-	+	+
4	-	0	+	+	-
5	-	+	0	-	+
6	-	+	+	0	0
7	0	-	+	-	+
8	0	-	+	+	0
9	0	0	-	0	+
10	0	0	0	0	0
11	0	+	-	-	-
12	0	+	0	+	-
13	+	-	-	0	-
14	+	-	0	+	+
15	+	0	0	-	0
16	+	0	+	-	-
17	+	+	-	+	0
18	+	+	+	0	+

i.e. in matrix form where $\mathbf{x}^{\mathrm{T}} = (x_1, \ldots, x_k)$, $\mathbf{b}^{\mathrm{T}} = (\hat{\beta}_1, \ldots, \hat{\beta}_k)$ and $\mathbf{B}$ is the following $k \times k$ symmetric matrix,

$$\mathbf{B} = \begin{bmatrix} \hat{\beta}_{11} & \frac{1}{2}\hat{\beta}_{12} & \cdots & \frac{1}{2}\hat{\beta}_{1k} \\ \frac{1}{2}\hat{\beta}_{12} & \hat{\beta}_{22} & \cdots & \frac{1}{2}\hat{\beta}_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{1}{2}\hat{\beta}_{1k} & \frac{1}{2}\hat{\beta}_{2k} & \cdots & \hat{\beta}_{kk} \end{bmatrix}.$$

Then the *stationary point* of the quadratic surface can be obtained from

$$\mathbf{x}_s = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{b}.$$

Due to there not being enough degrees of freedom to estimate both linear-by-linear effects and quadratic effects of 5 factors in 18 runs, obtaining the stationary point when estimating only linear and quadratic effects becomes

$$\mathbf{x}_s = -\frac{1}{2}\begin{bmatrix} \frac{1}{\hat{\beta}_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\hat{\beta}_{22}} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\hat{\beta}_{kk}} \end{bmatrix}\begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{2}\frac{\hat{\beta}_1}{\hat{\beta}_{11}} \\ \vdots \\ -\frac{1}{2}\frac{\hat{\beta}_k}{\hat{\beta}_{kk}} \end{bmatrix}.$$

Now that the location of the stationary point has been outlined, we shall consider the nature of the stationary point. Let $\mathbf{P}$ be some $k \times k$ matrix. Then

$$\mathbf{P}^{\mathrm{T}}\mathbf{B}\mathbf{P} = \mathbf{\Lambda},$$

where the columns of $\mathbf{P}$ are the standardized eigenvectors of $\mathbf{B}$ and $\mathbf{\Lambda} = diag(\lambda_1, \ldots, \lambda_k)$ is a diagonal matrix whose entries λ_i are the eigenvalues of $\mathbf{B}$. The nature of the stationary point is then determined by the signs of the eigenvalues of the matrix $\mathbf{B}$ and can be summarized as follows:

1. If $\lambda_1, \ldots, \lambda_k$ are all negative, then the stationary point is a point of *maximum* response.

2. If $\lambda_1, \ldots, \lambda_k$ are all positive, then the stationary point is a point of *minimum* response.

3. If $\lambda_1, \ldots, \lambda_k$ are of mixed signs, then the stationary point is neither a maximum nor a minimum point and is called a *saddle point*.

Since it is often the case in industrial experimentation that one seeks to maximize the response, such as maximizing the yield of a drug being produced in the pharmaceutical setting, we shall focus on the case of obtaining the point of maximum response. That is, we wish to determine the optimal treatment combination that will produce the maximum response and this will occur when the eigenvalues for the matrix $\mathbf{B}$ are all negative and the maximum response is then at the stationary point. We can denote the stationary point for $\mathbf{x}_j$ as $\mathbf{x}_{j0}$. We will consider the region of experimentation to be between the levels -1 and $+1$ for each factor, since we are observing the 5 factors at the three levels $\{-1, 0, +1\}$. It then follows that the optimal level of $\mathbf{x}_j$, denoted by $\mathbf{x}_{j_{opt}}$ is as follows:

$$\mathbf{x}_{j_{opt}} = \begin{cases} -\frac{1}{2}\frac{\hat{\beta}_j}{\hat{\beta}_{jj}} & \text{if } \hat{\beta}_{jj} < 0 \text{ and } -1 \leq \hat{\mathbf{x}_{j0}} \leq +1 \\ -1 & \text{if } \hat{\beta}_{jj} \geq 0 \text{ and } \hat{\beta}_j < 0 \\ & \text{or } \hat{\beta}_{jj} < 0 \text{ and } \hat{\mathbf{x}_{j0}} < -1 \\ +1 & \text{if } \hat{\beta}_{jj} \geq 0 \text{ and } \hat{\beta}_j > 0 \\ & \text{or } \hat{\beta}_{jj} < 0 \text{ and } \hat{\mathbf{x}_{j0}} > +1. \end{cases}$$

As it is assumed that the true optimal treatment combination is at a point of maximum response, it then follows that this optimal treatment combination is

$$\mathbf{x}_{opt} = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{b}.$$

The response at this point of maximum response $\mathbf{x}_{opt}$ is then

$$\begin{aligned}
\mu_{\mathbf{x}_{opt}} &= \beta_0 + \mathbf{x}_{opt}^{\mathrm{T}}\mathbf{b} + \mathbf{x}_{opt}^{\mathrm{T}}\mathbf{B}\mathbf{x}_{opt} \\
&= \beta_0 - \frac{1}{2}\mathbf{b}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b} + \frac{1}{4}\mathbf{b}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{B}\mathbf{B}^{-1}\mathbf{b} \\
&= \beta_0 - \frac{1}{4}\mathbf{b}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b}.
\end{aligned}$$

It is also assumed that the response at standard operating conditions $\mu_{\mathbf{x}_0}$, is β_0. It then follows that the maximum gain is

$$\mu_{\mathbf{x}_{opt}} - \mu_{\mathbf{x}_0} = -\frac{1}{4}\mathbf{b}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b}.$$

The response at the estimated optimal treatment combination $\mu_{\hat{\mathbf{x}}_{opt}}$ is

$$\begin{aligned}
\mu_{\hat{\mathbf{x}}_{opt}} &= \beta_0 + \hat{\mathbf{x}}_{opt}^{\mathrm{T}}\mathbf{b} + \hat{\mathbf{x}}_{opt}^{\mathrm{T}}\hat{\mathbf{B}}\hat{\mathbf{x}}_{opt} \\
&= \beta_0 - \frac{1}{2}\hat{\mathbf{b}}^{\mathrm{T}}\hat{\mathbf{B}}^{-1}\mathbf{b} + \frac{1}{4}\hat{\mathbf{b}}^{\mathrm{T}}\hat{\mathbf{B}}^{-1}\mathbf{B}\hat{\mathbf{B}}^{-1}\hat{\mathbf{b}}.
\end{aligned}$$

Similar expressions can then be obtained for the loss, which is

$$\mu_{\mathbf{x}_{opt}} - \mu_{\hat{\mathbf{x}}_{opt}} = -\frac{1}{4}\mathbf{b}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b} + \frac{1}{2}\hat{\mathbf{b}}^{\mathrm{T}}\hat{\mathbf{B}}^{-1}\mathbf{b} - \frac{1}{4}\hat{\mathbf{b}}^{\mathrm{T}}\hat{\mathbf{B}}^{-1}\mathbf{B}\hat{\mathbf{B}}^{-1}\hat{\mathbf{b}}$$

and the gain, which is

$$\mu_{\hat{\mathbf{x}}_{opt}} - \mu_{\mathbf{x}_0} = -\frac{1}{2}\hat{\mathbf{b}}^{\mathrm{T}}\hat{\mathbf{B}}^{-1}\mathbf{b} + \frac{1}{4}\hat{\mathbf{b}}^{\mathrm{T}}\hat{\mathbf{B}}^{-1}\mathbf{B}\hat{\mathbf{B}}^{-1}\hat{\mathbf{b}}.$$

In the following section we shall set out the prior specifications for the simulation work when applying the optimization utility function in order to determine whether a two-level or three-level design is more efficient to use when resources dictate that only a maximum of 18 experimental runs can be made. As we shall be looking at the case when an experimenter wishes to maximize a response we shall set out the priors so that the resulting eigenvalues are all negative.

3.4.4 Prior Specification

The prior assumptions that are to be made for the linear, interaction and quadratic effects shall now be set out. Priors for the linear, interaction and quadratic effects shall be specified for both the two- and three-level design, although only the linear and interaction effects shall be estimated from the two-level design and only linear and quadratic effects shall be estimated in the case of the three-level design. The specified priors are assumed when carrying out the simulation work, the results for which will be given in the following section.

As in previous examples presented, the intercept shall be assumed to be distributed normally with mean 0 and variance approaching ∞.

Normal independent priors shall be considered on the linear and interaction effects. These shall be

$$\beta_i, \beta_{ij} \sim N(0, 4)$$

and

$$\beta_i, \beta_{ij} \sim N(0, 100)$$

to assess the impact, if any, of varying the size of the variance.

Normal independent priors shall also be considered on the quadratic effects, and the two cases of prior assumptions which shall be looked at are:

$$\beta_{ii} \sim N(-30, 4)$$

and

$$\beta_{ii} \sim N(-30, 100).$$

That is, the priors for the quadratic effects shall be assumed to be large and negative, however we shall investigate this where the prior variance is relatively small, i.e. when $\sigma^2 = 4$ and also when the variance is much larger, i.e. when $\sigma^2 = 100$.

As well as normal independent priors on the effects, hierarchical normal mixture priors shall also be investigated for the linear and interaction effects. For both the two-level and three-level design, hierarchical prior assumptions on the linear effects shall be considered, which will be:

$$\beta_i \sim \begin{cases} N(0,1) & \text{if } \alpha_i = 0 \\ N(0,100) & \text{if } \alpha_i = 1 \end{cases}$$

where $p_i = 0.5$. As previously outlined in this chapter, this is where α_i is represented by the Bernoulli distribution with parameter p_i and determines the importance of main effect i. When $\alpha_i = 0$, the main effect i is inactive, and similarly, when $\alpha_i = 1$ the main effect i is active.

The interaction effects β_{ij} will also be assumed to be modelled by hierarchical priors and will be represented by:

$$\beta_{ij} \sim \begin{cases} N(0,1) & \text{if } \alpha_{ij} = 0 \\ N(0,100) & \text{if } \alpha_{ij} = 1 \end{cases}$$

where

$$p_{ij} = \begin{cases} 0.1 & \text{if } (\alpha_i, \alpha_j) = (0,0) \\ 0.25 & \text{if } (\alpha_i, \alpha_j) = (0,1) \\ 0.25 & \text{if } (\alpha_i, \alpha_j) = (1,0) \\ 0.75 & \text{if } (\alpha_i, \alpha_j) = (1,1). \end{cases}$$

The priors to be investigated can be summarized in Table 3.53 which follows, where priors 1-4 indicate normal independent prior assumptions on all effects. Priors 5 and 6 indicate hierarchical normal mixture prior on linear and interaction effects, and normal independent priors on the quadratic effects.

Table 3.53: Prior Distributional Assumptions

	Effect		
Prior	Linear	Interaction	Quadratic
1	$\beta_i \sim N(0, 4)$	$\beta_{ij} \sim N(0, 4)$	$\beta_{ii} \sim N(-30, 4)$
2	$\beta_i \sim N(0, 100)$	$\beta_{ij} \sim N(0, 100)$	$\beta_{ii} \sim N(-30, 100)$
3	$\beta_i \sim N(0, 100)$	$\beta_{ij} \sim N(0, 4)$	$\beta_{ii} \sim N(-30, 100)$
4	$\beta_i \sim N(0, 100)$	$\beta_{ij} \sim N(0, 100)$	$\beta_{ii} \sim N(-30, 4)$
5	$\beta_i \sim N(0, 1)$ if effect is inactive	$\beta_{ij} \sim N(0, 1)$ if effect is inactive	$\beta_{ii} \sim N(-30, 4)$
	$\beta_i \sim N(0, 100)$ if effect is active	$\beta_{ij} \sim N(0, 100)$ if effect is active	
6	$\beta_i \sim N(0, 1)$ if effect is inactive	$\beta_{ij} \sim N(0, 1)$ if effect is inactive	$\beta_{ii} \sim N(-30, 100)$
	$\beta_i \sim N(0, 100)$ if effect is active	$\beta_{ij} \sim N(0, 100)$ if effect is active	

3.4.5 Results

The results when applying the optimization utility function in order to assess the benefits of the two-level and three-level designs based upon the loss, gain and relative gain of each of the designs are now presented. 1000 simulations were found to be a sufficient number of simulations in order for the estimated utilities to be calculated within a reasonable degree of confidence. Results for the mean values of the loss, gain and relative gain are given in the following table along with the mean value of the maximum gain, which is defined as

$$\mu_{\mathbf{x}_{opt}} - \mu_{\mathbf{x}_0},$$

i.e. the gain that would have been made if the true optimal treatment combination had been used rather than standard operating conditions.

Table 3.54: Results for Two- and Three-Level Design

	Loss	Gain	Maximum Gain	Relative Gain
Two-Level Design:				
Prior				
1	0.0924	0.0801	0.1724	0.3958
2	0.0192	6.5211	6.5403	0.9860
3	0.0381	5.1791	5.2172	0.9966
4	0.0405	4.5413	4.5817	0.9814
5	0.0613	2.1225	2.1838	0.9674
6	0.0471	3.4652	3.5123	0.9786
Three-Level Design:				
Prior				
1	0.0351	0.1373	0.1724	0.6756
2	125.5413	-119.0010	6.5403	-17.8303
3	161.3855	-156.1683	5.2172	-13.1544
4	99.1967	-94.6150	4.5817	-23.4287
5	3.5801	-1.3963	2.1838	-2.0103
6	13.0477	-9.5355	3.5123	-17.5148

Looking at the results for the two-level design it can be seen that the mean loss is fairly small for all priors investigated, where the smallest mean loss is in the case of prior 2 with a value of 0.0192, and the largest mean loss is in the case of prior 1 with a value of 0.0924. This indicates that regardless of the variety of prior assumptions investigated, the loss is small when using the estimated optimal treatment combination instead of the true optimal treatment combination. It can also be seen that the mean loss is relatively small when comparing these with the corresponding priors for the three-level design. The mean losses for the three-level design are much larger than those for the two-level design, except in the case of prior 1. This is the only prior where the loss for the two-level design is larger than that of the three-level design. However, it should be noted that this difference of 0.0573 is much smaller than the difference in loss between the two- and three-level design for priors 2-6. For example, the difference

in loss between the two- and three-level design for prior 2 is a value of 125.5221.

The mean gain for the two-level design is larger and positive than the mean gain when looking at the corresponding priors for the three-level design in all cases, with the exception of prior 1. This implies that an improvement is to be made from using the estimated optimal treatment combination in place of the standard operating conditions. Due to the mean loss for all of the six priors investigated being fairly small, the maximum gain in most cases is approximately the same as the mean gain. It can be shown that,

$$(\mu_{\mathbf{x}_{opt}} - \mu_{\hat{\mathbf{x}}_{opt}}) + (\mu_{\hat{\mathbf{x}}_{opt}} - \mu_{\mathbf{x}_0}) = \mu_{\mathbf{x}_{opt}} - \mu_{\mathbf{x}_0}.$$

That is, the addition of the loss and gain results in the maximum gain. A large value for the maximum gain suggests that the response obtained at the true optimal treatment combination is much greater than the response obtained at standard operating conditions. Since the values for the maximum gain are calculated from

$$\mu_{\mathbf{x}_{opt}} - \mu_{\mathbf{x}_0} \quad = \quad -\frac{1}{4}\mathbf{b}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b},$$

this quantity is dependent only on $\mathbf{b}$ and $\mathbf{B}$, and not on the design or simulated data, thus resulting in the same maximum gain for both designs.

When looking at the results for the three-level design, it is clear to see that there is a much greater range in values for the mean loss than was seen for the two-level design, and for many of the priors investigated the mean loss is very large. The smallest mean loss is seen for prior 1 with a value of 0.0351, and the largest mean loss for the three-level design is for prior 3 with a value of 161.3855, although the mean loss for prior 2 is also extremely large. Some interesting results are also seen for the mean gain for the three-level design where for five out of the six priors investigated, the mean gain is large and negative. This implies that the expected response at standard operating conditions is greater, and therefore better, than the expected response at the estimated optimal treatment combination.

For prior 1, normal independent priors are placed on linear, interaction and quadratic effects and the variance σ^2, is equal to 4, the smallest variance size investigated. For prior 2, normal independent priors are again assumed except that the variances on all effect parameters are increased to 100. For the three-level design, the results appear to indicate that mean loss is greatly increased when the variances on all effect parameters are also increased. However, the converse is true for the two-level design; when the variances on all effect parameters are increased to 100, the mean loss is subsequently reduced. This is also reiterated in the case of priors 5 and 6, where normal mixture priors are assumed on linear and interactions effects and normal independent priors are assumed on the quadratic effects. When the variances on the quadratic effects are increased from $\sigma^2 = 4$, to $\sigma^2 = 100$, the mean loss again decreases for the two-level design but increases for the three-level design.

Looking at the results for prior 3 for the two-level design, the mean loss is smaller when the quadratic effects have a large variance, and the interaction effects have a smaller variance, and for prior 4 the mean loss increases somewhat when the interaction effects have a large variance and the quadratic effects have a smaller variance. The converse is true when only linear and quadratic effects can be estimated, i.e. with the three-level design. That is, when the quadratic effects have a large variance, and the interaction effects have a smaller variance, the mean loss is large, and is reduced greatly when the interaction effects have a large variance and the quadratic effects have a smaller variance.

There does appear to be some difference in the results depending on whether normal mixture priors, or normal independent priors are assumed on the parameters, and also the size of the variance placed on the effect parameter. For the three-level design, when normal mixture priors are assumed on the linear and interaction effects as with priors 5 and 6, the mean loss is much smaller and the mean gain much larger than for most of the other priors for the three-level designs where the effects are assumed

to come from independent normal distributions.

There appears to be a clear indication that a two-level design is more favourable to use in terms of being able to obtain the optimal treatment combination, for the various prior assumptions that have been investigated. This is due to the mean gain being positive for all of these priors, and so the resulting relative gain is also positive in all cases. This means that there is always some gain to be made from using the estimated optimal treatment combination rather than standard operating conditions. This is in relation to the maximum gain that would have been made if the true optimal treatment combination had been used compared to standard operating conditions. In contrast when looking at the three-level design, since the gain in many of the cases is large and negative, and the maximum gain is comparatively small, the resulting relative gain for most of the priors investigated is also fairly large and negative. This implies that that the estimated optimal treatment combination is in fact worse than standard operating conditions, in relation to the maximum gain that would have been made if the true optimal treatment combination had been used compared to standard operating conditions.

Thus, if an experimenter was to make a decision about which design to use based upon the loss that will be incurred if the estimated optimal treatment combination is used instead of the true optimal treatment combination, and in terms of gain to be made from using the estimated optimal treatment combination in place of standard operating conditions, the inference can be drawn that the two-level design should be used. This is because the loss is reduced and the gain increased for the majority of priors assumed in this thesis, in comparison to the loss and gain for the three-level design.

3.4.6 Discussion

The application of the optimization utility function has been demonstrated where two designs of different sizes have been compared. In addition to the designs having a different number of experimental runs, the two designs compared in this section were also different in terms of what effects were able to be estimated. That is, for the two-level design, linear and interaction effects have been able to be estimated whereas in the case of the three-level design, linear and quadratic effects are estimated. This does mean that a completely fair comparison perhaps cannot be made between the two- and three-level designs and imposes some restrictions on the comparisons and conclusions that can be drawn.

For many of the priors investigated for the three-level design, a large and negative value for the gain was seen. Although this is possible, this is an unusual result as one would expect the optimal treatment combination to result in the best response, which in this case would be the maximum response. Thus, it would be hoped that the expected response obtained at the estimated optimal treatment combination would be larger, and therefore improve on the response obtained when the operating conditions currently in place are used. However, this was not the case for many of the prior assumptions considered. This may lead one to suspect that perhaps the maximum response, which in this case should be at the location of the stationary point, does not actually lie within the region of experimentation. However, checks were carried out to confirm that the location of the maximum response was in fact between the levels -1 and $+1$. If checks found that the maximum response did not actually lie within the region of experimentation, which was not the case in this thesis, it would be of interest to perhaps consider *ridge analysis* which can be performed to determine the direction in which additional experimentation should be carried out.

In this study, the choice of experimental design and prior beliefs about the effects have

been examined under the assumption of searching for the combination of various factors that gives the maximum response. That is, the true surface is assumed to have elliptical contours and the optimum point on this surface is a maximum not a minimum. This problem was similarly considered by Curnow (1972) where the question of whether one-variable experimentation, two-variable experimentation or no experimentation was preferable in terms of obtaining the estimated optimum was discussed, where the estimated optimal treatment combination was assumed to be the maximum response. However, it may not be the case that the optimum is a maximum. Due to errors of estimation the turning point on the fitted surface may be a minimum or a saddle point. If this was to transpire then it would be advisable for further experimentation to take place rather than merely locating the estimated position of the turning point and using this as the best estimate of the position of the maximum on the true fitted surface.

It would also be of interest to investigate further and assess whether the form of the prior, and size of the mean and variance assumed for the effect parameters has any influence on the gain being negative. As has been discussed with the other studies in this thesis, there is a dependence of the results on the various prior distributions and experimental design choices and so the results are limited to the few scenarios that have been selected to be investigated. The size of the probabilities for an effect to be active in the hierarchical prior set-up may also be considered to be too large and that it violates the principle of effect sparsity. However, it was felt that these probabilities should be large enough in order for some effects to result in being active given that the run sizes of these designs may be considered to be fairly small. Also, there is an argument for the case that these probabilities are a more accurate representation of what occurs in industry and that the probabilities typically considered in literature, such as Chipman, Hamada and Wu (1997) are not totally realistic. In their paper p_{00} is chosen to be small, e.g. 0.01, p_{01} and p_{10} somewhat larger, e.g. 0.10 and p_{11} largest, e.g. 0.25. However, in practice there may be a greater than a one in 100 chance that

an effect could be active even if its parent main effects are not active.

The results may have been more thought-provoking if designs had been presented where not all important effects were able to be estimated. In this situation, all main effects and two-factor interactions were able to be estimated when using the two-level design and all main effects and quadratic effects in the instance of the three-level design. Examples of perhaps more interesting circumstances to investigate would be estimating 5 factors in less than 16 experimental runs for a two-level design so that not all two-factor interactions can be estimated, or perhaps a fractional factorial experiment in the case of the three-level design.

There was some indication that the form of the prior had some impact upon the loss, gain and relative gain. Therefore, it may be of interest to investigate further, and perhaps more informative, prior assumptions to assess whether this does give any benefits in terms of choice of design.

3.5 Summary

Various approaches for discriminating between a selection of candidate designs have been presented. Methods have been outlined for choosing between designs which are to be used in the context of optimization, when a large number of variables have been narrowed down to the key few variables truly driving the process under consideration and an optimum treatment combination is sought. Methods have also been set out for the selection of designs that are to be used in a screening context at the initial stages of experimentation where the large number of variables being investigated are to be screened in order for further experimentation at a later stage.

Some benefits have been shown in giving experimenters a more formal manner in which to choose the design that they shall use in experimentation. Although the methods presented have not been directly compared to standard optimality criteria, the methods in this thesis could be used as opposed to standard optimality design criteria such as D-optimality or utility functions such as those making use of Shannon information or the Kullback-Leibler distance. These standard optimality criteria perhaps do not translate naturally to criteria that will answer questions that an industrial experimenter may have although no claim is being made that the methods laid out in this thesis are better than these traditional criteria or utility functions but perhaps more appropriate in this context of industrial experimentation. On the other hand, some words of caution should be asserted that the inferences drawn from the results have been based on the set of priors that have been chosen and therefore relate to the various prior assumptions that have been made about the parameters being investigated. Also, it should be stressed that the methods presented in chapter 2 and the applications of these that have been displayed in this chapter do set out guidelines in choosing a design. However, in practice they should be applied more informally. The methods are not to be used as a blind substitute for selecting a design but rather

in conjunction with any ad hoc methods that an experimenter may already have in place. These methods can be deemed to be helpful in providing experimenters with guidance particularly in the situation when they may need to consult statisticians for such assistance.

Chapter 4

Bayesian Analysis of Fractional Factorial Experiments

4.1 Introduction

Fully Bayesian methods are rarely used in multifactor experiments, particularly in the case of fractional factorial designs, although a general set-up for the analysis has been proposed by Nobile and Green (2000). Since an experimenter will typically have some prior knowledge about the sizes of effects, to use in this a fully Bayesian data analysis would seem a natural approach. In this chapter, the Bayesian analysis of fractional factorial experiments will be presented where both vague and more informative priors will be explored.

It is usually the case that several alternative probability models can reasonably fit the same data, where the alternative models being specified are in reference to either the prior distribution, the likelihood (sampling distribution) or both. Thus, it would be necessary for a *sensitivity analysis* to be performed to assess how much the poste-

rior inference is affected by the use of alternative reasonable models, to ensure that prior beliefs or likelihood assumptions made do not unduly affect the results. Several probability models will be fitted to the same problem in this chapter to gauge the sensitivity of posterior inferences to different model assumptions and assess whether there is variation in posterior inferences for estimands and predictive quantities of interest. The data will be modelled assuming a normal distribution and a scaled-t distribution. Prior distributions on the effect parameters that will be considered are a normal distribution and a mixture of normal distributions. Alternative prior distributions that will be considered for the variance are a uniform prior and more typically used, an inverse-gamma prior. When considering the uniform distribution, the prior is placed on $\log \sigma$. That is, we work with the logarithm of the standard deviation and assume that $\log \sigma$ is distributed uniformly. When assuming an inverse-gamma prior distribution we instead place a gamma prior distribution on the *precision*, τ, which is the inverse of the variance.

Before the Bayesian analyses of fractional factorial experiments are examined, the most notable works in this area are considered and discussed in the following section.

4.2 Literature Review

A prominent paper that introduced the use of Bayesian analysis methods in the field of industrial experimentation was that of Box and Meyer (1986a), as outlined in Chapter 1. They presented a formal analysis of fractional factorial designs to be used in conjunction with the graphical analysis due to Daniel (1959), to be used most appropriately in the circumstance of factor sparsity. Their method allowed posterior probabilities for active effects to be calculated and made use of Box and Tiao (1968).

Box and Meyer (1993) also proposed an alternative Bayesian approach to model identification for the analysis of data where there is a high amount of aliasing involved, where they focussed on factors rather than specific effects. They were able to develop a Bayesian method that more completely considered the various hypotheses of which factors are active when analysing the results of a screening experiment. This is even when the effects are thoroughly confounded, as is typically the case in Plackett-Burman designs where the alias structure can be complicated. They considered the set of possible models $M_0, \ldots, M_m$ where each model M_i has an associated vector of parameters θ_i. Then the sampling distribution of data $\mathbf{y}$, given the model M_i is described by the probability density $f(\mathbf{y}|M_i, \theta_i)$. The prior probability of the model M_i is $p(M_i)$ and the prior probability density of θ_i is $f(\theta_i|M_i)$. Then the predictive density of $\mathbf{y}$, given model M_i, is written $f(\mathbf{y}|M_i)$ and can be expressed as

$$f(\mathbf{y}|M_i) = \int_{\mathbf{R_i}} f(\mathbf{y}|M_i, \theta_i) f(\theta_i|M_i)\, d\theta_i$$

where $\mathbf{R_i}$ is the set of possible values of θ_i. The posterior probability of the model M_i, given the data $\mathbf{y}$ is then

$$p(M_i|\mathbf{y}) = \frac{p(M_i) f(\mathbf{y}|M_i)}{\sum_{h=0}^{m} p(M_h) f(\mathbf{y}|M_h)}.$$

These posterior probabilities $p(M_i|\mathbf{y})$ for each possible model M_i can then be used as a

basis for model identification where those potentially likely models can be identified by their large posterior probability. As was similarly calculated in Box and Meyer (1986a) for the marginal probabilities p_i that an effect i is active, in this case the probabilities $p(M_i|\mathbf{y})$ can be accumulated to compute the marginal posterior probabilities P_j that factor j is active. That is

$$P_j = \sum_{M_i:\, factor\, j\, active} p(M_i|\mathbf{y}).$$

Therefore, the probabilities $\{P_j\}$ are calculated by adding up the posterior probabilities over all possible models M_i where factor j is active. Thus, a large value of P_j will indicate that factor j is active whereas a small value of P_j or a value close to 0 will indicate that j is inactive (or passive).

Chipman, Hamada and Wu (1997) also proposed a methodology for a Bayesian approach to model selection, however focussed on effects rather than factors. Their method also differed where in addition to marginal posteriors, joint posterior probabilities were also considered, namely posterior probabilities of the models. In comparison to an all-subsets approach or exhaustive search as used by Box and Meyer, in this paper it was shown that the methodology required less computation due to the search through the model space being done stochastically. The search also focussed on a class of reasonable models as a result of the specification of flexible hierarchical priors. The keys aspects set out in this paper, namely the stochastic search and use of hierarchical priors adopted those methods as set out by George and McCulloch (1993).

The general set-up that Nobile and Green (2000) proposed in their paper provides a framework for the approach that shall be taken in this thesis in terms of the Bayesian analysis carried out, although there are some significant differences. They presented a Bayesian analysis for factorial experiments using finite mixture distributions to model the main effects and interactions. The Bayesian analysis allowed both estimation and

an analogue of hypothesis testing in a posterior analysis using a single prior specification. The Bayesian methods used were restricted to a two-way, 'row-plus-column' model with replications, possibly unequal and/or missing and allowing interactions although they did state that the approach was intended to be extendible to more complicated designs and experiments using covariates. Computations carried out in the paper were done by *Markov chain Monte Carlo* (MCMC) methods. MCMC is a method by which we can draw values of θ from an approximate distribution and then correcting those draws to better approximate the target distribution, $p(\theta|y)$, i.e. the target posterior distrbution as set out previously (§1.4.5). Samples are drawn in a sequential manner, and the distribution of the sampled draws depend on the previous value drawn, therefore the draws forming a *Markov chain*. From probability theory, a Markov chain is defined as a sequence of random variables $\theta_1, \theta_2, \ldots$ for which for any t, the distribution of θ_t given all previous $\theta's$ depends only on the most recent value, θ_{t-1}. At each step in the simulation the approximate distributions are improved, in that they are converging to the target distribution. In order to achieve this and start the chain, a starting value of θ must be specified and in the context of this thesis this would be chosen by the experimenter. Once the chain has reached an equilibrium, we can say that it has converged and values in the chain prior to convergence should not be included in calculation of the target distribution summaries due to the possibility of incorrect inferences being made. Instead, the values in the chain prior to convergence are discarded, these values being termed 'burn-in'.

In Nobile and Green's paper it is assumed that

$$y_{ijk} = \theta_{ij} + \epsilon_{ijk} \, (i = 1, \ldots, m; \; j = 1, \ldots, n; \; k = 1, \ldots, r_{ij})$$

for a two-way layout model. There are r_{ij} replicates in cell (i, j) corresponding to the i^{th} level of factor 1 and the j^{th} level of factor 2. θ_{ij} is the sum of the overall level μ, the main effects α_i and β_j and the interaction γ_{ij} so that it follows that

$$\theta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}. \tag{4.1}$$

The error terms are normally distributed $\epsilon_{ijk} \sim N(0, \sigma_{ij})$ and the overall level μ has normal prior distribution $\mu \sim N(\eta, \sigma^{\mu})$. The remaining terms in (4.1) are assumed to proceed from finite mixtures of unknown numbers of normal component distributions subject to the following classical constraints

$$\sum_i \alpha_i = 0, \sum_j \beta_j = 0, \sum_i \gamma_{ij} = 0, \sum_j \gamma_{ij} = 0.$$

It is also considered that

$$\alpha_i \sim \sum_{t=1}^{k^{\alpha}} w_t^{\alpha} N(\mu_t^{\alpha}, \sigma_t^{\alpha})$$

independently for all i. For the number of components k^{α} the prior is uniform on the integers from 1 to some maximum value $k_{\max}^{\alpha}$. The mixture weights follow a Dirichlet distribution as follows

$$w^{\alpha} \sim Dir(d_1^{\alpha}, \ldots, d_{k^{\alpha}}^{\alpha})$$

and independent normal and inverse gamma distributions are the priors placed on the means and variances:

$$\mu_t^{\alpha} \sim N(\xi_t^{\alpha}, 1/\tau^{\alpha}),\ (\sigma_t^{\alpha})^{-1} \sim Ga(a_t^{\alpha}, b_t^{\alpha}).$$

Similarly, the other parameter effects are distributed as follows

$$\beta_j \sim \sum_{s=1}^{k^{\beta}} w_s^{\beta} N(\mu_s^{\beta}, \sigma_s^{\beta}),\ \gamma_{ij} \sim \sum_{u=1}^{k^{\gamma}} w_u^{\gamma} N(\mu_u^{\gamma}, \sigma_u^{\gamma})$$

and a similar structure is assumed for the $\beta' s$ and $\gamma' s$ as for the $\alpha' s$, where the mixture weights follow a Dirichlet distribution and the priors for the mean and variances are independent normal and inverse gamma distributions respectively. Rather than i and j levels of the factors as laid out in this paper, the experiments that shall be considered in this chapter consist of two-level factors and thus, this set up would then be reduced to β_j for the j effect parameters in this instance.

Thus, the approach of Bayesian analysis of factorial experiments with a single prior specification suitable for both estimation and testing and computations being done using MCMC methods is similar to the approach that shall be taken in this thesis and demonstrated in the subsequent sections. Nobile and Green's approach of explicitly-specified mixtures of normals, building on Richardson and Green (1997), and features of the approach of George and McCulloch (1993) shall also be used in this thesis.

Another relevant work is that of Baba and Gilmour, chapter 12 of Colosimo and del Castillo (2006). They demonstrated the applications of Bayesian methods. However, in this case the methods were applied to data from saturated designs. Similarly to the approach that is to be taken in this thesis, they extended the work of Box and Meyer (1993). In extending the work of Box and Meyer, rather than just using a mixture of two normal distributions for each effect (both having mean zero and one with a small variance and one with a large variance) to obtain the posterior probability of each effect of being active, Baba and Gilmour used the full prior and posterior distributions and concentrated on estimating the effects rather than just identifying them as being active or inactive. This approach will also be demonstrated in the following sections.

4.3 Methodology

As has been previously mentioned, Bayesian methods are infrequently used in industry, such as in the analysis of data from fractional factorial experiments. Therefore an experimenters' prior knowledge fails to be explicitly exploited and this is despite a typical assumption being made *a priori* that some effects will be negligible. In the case of fractional factorial designs effects considered to be negligible will usually be higher order interactions, and this leads to not all effects being estimated. Thus, although Bayesian methods are not standardly applied in a formal manner, it could be reasoned that making assumptions such as some effects are negligible does in fact represent an extreme form of prior. This is then one reason for justifying the use of a Bayesian analysis in the context of industrial experimentation.

An experimenter can specify their belief about the size of each effect for the parameters of interest $\boldsymbol{\beta}$, or at least give the range within which they believe it will lie. In doing so they can place a joint prior distribution on the set of parameters of interest $\boldsymbol{\beta}$ and then upon collection of the data $\mathbf{y}$, the joint posterior distribution of $\boldsymbol{\beta}|\mathbf{y}$ can be obtained by applying Bayesian methods. The procedure of obtaining the posterior distribution via a Bayesian analysis will be illustrated theoretically in the following section. This will be shown when assuming a conjugate prior and in both the cases where σ^2 is known and unknown.

The priors on the parameters of interest $\boldsymbol{\beta}$ are typically individually represented by normal distributions which can then be extended to all p parameters of interest being represented by the multivariate normal distribution. The normal distribution is typically used by experimenters to represent their prior beliefs about the parameter effects due to the normal distribution being fairly flexible and various other properties. The prior mean can take any value along the real line and the normal distribution can closely represent the experimenters' beliefs if they expect the prior to be symmetric

and uni-modal about some point. Another property of the normal distribution which makes it a reasonable choice of prior is that it is conjugate, thus the posterior distribution will also be normal and therefore be of a form that is easier to manipulate. The same properties will also apply when extending the normal distribution to the multivariate normal distribution when dealing with p parameters of interest and assuming a joint prior distribution on the parameters. When carrying out a Bayesian analysis of a dataset obtained from a typical industrial experiment, the results of which shall be presented in a later section (§4.5), normal conjugate priors shall be assumed for the parameters of interest as well as scaled-t prior distributions. Scaled-t prior distributions shall also be considered to illustrate the impact, if any, on the analysis resulting from the choice of prior distribution. It will also demonstrate that an experimenter may not need to make typical assumptions such as assuming normal distributions. Choosing a different prior distribution such as the scaled-t distribution which has the property of having longer tails than that of the normal, may actually prove to be a better choice of model if the experimenter suspects their dataset of containing outliers, which will be investigated further in Chapter 5.

When assuming that σ^2 is unknown, the choice of prior distribution for the variance component is commonly the inverse gamma distribution. This is the conjugate of the normal distribution (Gelman et al. 2003) and is typically used when some information about the variance component is available. However, an experimenter may possess little or no information about run-to-run variation and so a vague or noninformative prior distribution will be desired. The choice of prior distribution for the variance in this situation was discussed by Daniels (1999) and some recommendations given. In this thesis the inverse gamma prior distribution for the variance shall be considered in the situation where some information about σ^2 is available. A uniform prior distribution shall also be considered when little or no information about σ^2 is to hand, and the uniform prior distribution shall be put on $\log \sigma$, defining the prior distribution on a compact set $[-A, A]$ for some large value of A. Taking the logarithm of σ is

appropriate in this instance since the variance parameter must be positive, although this does lead to an improper posterior distribution. However, defining the prior distribution on the compact set $[-A, A]$ rather than on the range $(0, A)$ is a possible way of trying to overcome this problem. Alternatively, a uniform prior distribution on the parameter σ itself can be considered although some issues also arise in this situation where there is bias, or *miscalibration* when speaking in the Bayesian rather than the classical sense, towards positive values. Further description on the specification of a uniform prior distribution on either σ or $\log \sigma$ are provided in Gelman (2006) as well as other details on non-informative and weakly-informative prior distributions for hierarchical variance parameters.

4.3.1 Procedure for Bayesian Analysis when assuming a Conjugate Prior

The procedure of obtaining the posterior distribution when carrying out a Bayesian analysis shall now be set out mathematically for both cases where σ^2 is known and unknown.

σ^2 known

A multivariate normal likelihood is assumed, i.e. that

$$\mathbf{y}|\boldsymbol{\beta} \sim MVN_N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}).$$

This is where a joint prior distribution is specified for the p parameters of interest, $\boldsymbol{\beta}$. Data $\mathbf{y}$ are collected consisting of N observations from an experiment which uses the design matrix $\mathbf{X}$ and $\mathbf{I}$ is the $N \times N$ identity matrix. A multivariate normal conjugate

prior is then placed on the parameters of interest $\boldsymbol{\beta}$ where

$$\boldsymbol{\beta} \sim MVN_p(\mu_p, \Sigma_p)$$

and

$$p(\boldsymbol{\beta}) = (2\pi)^{-\frac{p}{2}} \left|\Sigma_p\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\mu_p)^{\mathrm{T}}\Sigma_p^{-1}(\boldsymbol{\beta}-\mu_p)\right\}.$$

Also,

$$p(\mathbf{y}|\boldsymbol{\beta}) = (2\pi)^{-\frac{N}{2}}\sigma^{-N} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}.$$

We can then obtain the posterior distribution using Bayes Theorem which implies

$$\begin{aligned}
p(\boldsymbol{\beta}|\mathbf{y}) &\propto p(\boldsymbol{\beta})p(\mathbf{y}|\boldsymbol{\beta}) \\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\mu_p)^{\mathrm{T}}\Sigma_p^{-1}(\boldsymbol{\beta}-\mu_p) - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\} \\
&= \exp\left\{-\frac{1}{2}\left[(\boldsymbol{\beta}-\mu_p)^{\mathrm{T}}\Sigma_p^{-1}(\boldsymbol{\beta}-\mu_p) + \frac{1}{\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}^{\mathrm{T}}\Sigma_p^{-1}\boldsymbol{\beta} - \boldsymbol{\beta}^{\mathrm{T}}\Sigma_p^{-1}\mu_p - \mu_p^{\mathrm{T}}\Sigma_p^{-1}\boldsymbol{\beta} + \mu_p^{\mathrm{T}}\Sigma_p^{-1}\mu_p\right.\right. \\
&\qquad \left.\left. + \frac{1}{\sigma^2}\mathbf{y}^{\mathrm{T}}\mathbf{y} - \frac{1}{\sigma^2}\mathbf{y}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} - \frac{1}{\sigma^2}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} + \frac{1}{\sigma^2}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}^{\mathrm{T}}(\frac{1}{\sigma^2}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \Sigma_p^{-1})\boldsymbol{\beta} - \boldsymbol{\beta}^{\mathrm{T}}(\frac{1}{\sigma^2}\mathbf{X}^{\mathrm{T}}\mathbf{y} + \Sigma_p^{-1}\mu_p)\right.\right. \\
&\qquad \left.\left. - (\frac{1}{\sigma^2}\mathbf{y}^{\mathrm{T}}\mathbf{X} + \mu_p^{\mathrm{T}}\Sigma_p^{-1})\boldsymbol{\beta} + (\frac{1}{\sigma^2}\mathbf{y}^{\mathrm{T}}\mathbf{y} + \mu_p^{\mathrm{T}}\Sigma_p^{-1}\mu_p)\right]\right\}.
\end{aligned}$$

Then only considering the terms involving $\boldsymbol{\beta}$, it can be found that

$$p(\boldsymbol{\beta}|\mathbf{y}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\mu)^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{\beta}-\mu)\right\}$$

where

$$\mu = \left[\frac{1}{\sigma^2}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \Sigma_p^{-1}\right]^{-1}\left[\frac{1}{\sigma^2}\mathbf{X}^{\mathrm{T}}\mathbf{y} + \Sigma_p^{-1}\mu_p\right]$$

and

$$\Sigma = \left[\frac{1}{\sigma^2}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \Sigma_p^{-1}\right]^{-1}.$$

Thus, the form of the posterior distribution is also multivariate normal where

$$\boldsymbol{\beta}|\mathbf{y} \sim MVN_p(\mu, \Sigma).$$

σ^2 unknown

A multivariate normal likelihood is again assumed, as in the case where σ^2 was known. Throughout this thesis it has generally been assumed that an experimenter will have some prior knowledge about run-to-run variation. This will usually be based upon past experience or due to the fact that there is not expected to be much variability between *experimental units*, the smallest unit to which a treatment can be applied (Bailey, 2008). In the pharmaceutical industry setting the units will typically be runs of pilot-scale or full-scale equipment depending on the stage of the investigation, such as high shear mixers and blenders. However, if an experimenter does not possess much information about the variation in experimental units we can instead assume that σ^2 is unknown and put a prior distribution on this parameter.

Let a multivariate normal prior be placed on the parameters of interest $\boldsymbol{\beta}$ and an inverse gamma prior be placed on the variance σ^2. Then the multivariate normal-inverse gamma prior distribution is conjugate to the normal likelihood and is given by

$$[\boldsymbol{\beta}, \sigma^2] \sim NIG(\mu_p, \mathbf{A}, a, b)$$

which can also be expressed by

$$\boldsymbol{\beta}|\sigma^2 \sim N_p(\mu_p, \mathbf{A}\sigma^2)$$

and

$$\sigma^2 \sim IG(a, b).$$

Thus, the prior distribution for σ and $\boldsymbol{\beta}$ is

$$p(\sigma) \propto (\sigma)^{-(a+1)} \exp\left\{-\frac{ab}{2\sigma^2}\right\}$$

and

$$p(\boldsymbol{\beta}|\sigma) \propto \frac{|\mathbf{A}^{-1}|^{\frac{1}{2}}}{\sigma^p} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \mu_p)^{\mathrm{T}}\mathbf{A}^{-1}(\boldsymbol{\beta} - \mu_p)\right\}.$$

Then the resulting posterior is also of the form normal-inverse gamma with parameters

$$\begin{aligned}
\tilde{\mu} &= \tilde{\Sigma}(\mathbf{X}^{\mathrm{T}}\mathbf{y} + \mathbf{A}^{-1}\mu_p) \\
\tilde{\Sigma} &= (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \mathbf{A}^{-1})^{-1} \\
\tilde{a} &= \frac{N}{2} + a \\
\tilde{b} &= \frac{SS}{2} + b \\
SS &= \mathbf{y}^{\mathrm{T}}\mathbf{y} - \mu^{\mathrm{T}}\tilde{\Sigma}^{-1}\mu + \mu_p^{\mathrm{T}}\mathbf{A}^{-1}\mu_p
\end{aligned}$$

and the marginal posterior distribution of $\boldsymbol{\beta}|\mathbf{y}$ integrated over σ has p.d.f

$$p(\boldsymbol{\beta}|\mathbf{y}) \propto \Big\{ ab + \mathbf{y}^{\mathrm{T}}\mathbf{y} + \mu_p^{\mathrm{T}}\mathbf{A}^{-1}\mu_p - \mu^{\mathrm{T}}(\mathbf{A}^{-1} + \mathbf{X}^{\mathrm{T}}\mathbf{X})\mu + (\boldsymbol{\beta} - \mu)^{\mathrm{T}}(\mathbf{A}^{-1} + \mathbf{X}^{\mathrm{T}}\mathbf{X})(\boldsymbol{\beta} - \mu) \Big\}^{-\frac{N+a+p}{2}}.$$

That is, the marginal posterior distribution of $\boldsymbol{\beta}$ is a multivariate scaled-t distribution with mean μ_p.

4.3.2 The Gibbs Sampler

In the previous subsection (§4.3.1), the mechanics of a Bayesian analysis were presented in order to obtain the posterior distribution. Obtaining the posterior distribution often requires the integration of high-dimensional functions which can be very difficult to achieve analytically, particularly when non-conjugate prior distributions have been assumed.

MCMC techniques have enabled highly complicated models to be used and their corresponding posterior distributions to be estimated with accuracy. The WinBUGS software (Spiegelhalter et al., 2003), which is based upon the BUGS project (Spiegelhalter et al., 1996) that started in 1989, is the Windows version of BUGS and is a statistical software for Bayesian analysis that uses MCMC methods. BUGS is so

called since it is a Bayesian inference Using Gibbs Sampling. The *Gibbs Sampler* is a particular Markov chain algorithm which was developed by Geman and Geman (1984) and has its origins in image processing and it should be noted that the Gibbs sampler is a special case of the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970). It is perhaps the rediscovery of MCMC methods in the early 1990s (Gelfand et al., 1990; Gelfand and Smith, 1990) which has lead to the spark of interest in the use of Bayesian methods in applied statistics and has resulted in MCMC methods and the Gibbs sampler becoming one of the principal computational tools for carrying out a Bayesian inference.

The Gibbs sampling procedure will now be outlined. Let the parameter vector θ be partitioned into J subvectors or components so that $\theta = (\theta_1, \ldots, \theta_J)$ and θ_j is the j^{th} component of parameters. Each component contains one or more parameters and θ_{-j} denotes the set of all parameters not in component j. Relating this to Bayes Theorem we have

$$p(\theta_1, \ldots, \theta_J | y_1, \ldots, y_N) \propto p(\theta_1, \ldots, \theta_J) p(y_1, \ldots, y_N | \theta_1, \ldots, \theta_J),$$

which gives the shape of the joint posterior density of all the parameters. Gibbs sampling requires that we know the full conditional distribution of all the components θ_j, given all other parameters θ_{-j} and observed data $\mathbf{y} = (y_1, \ldots, y_N)^{\mathrm{T}}$. Then let the full conditional distribution of component θ_j be denoted as

$$p(\theta_j | \theta_{-j}, \mathbf{y}) = p(\theta_j | \theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_J, \mathbf{y}).$$

The Gibbs sampler cycles through each of the parameter components in turn, drawing each one from its full conditional distribution given the most recent values of all the other parameter components and the observed data.

The steps for the algorithm of the Gibbs sampler can be summarized as follows:

1. Start from some arbitrary point in the parameter space $\theta^0 = (\theta_1^{(0)}, \ldots, \theta_J^{(0)})$.

2. For each iteration $t = 1, \ldots, T$

   (a) Set $\theta = \theta^{t-1}$.

   (b) For $j = 1, \ldots, J$ update θ_j by drawing $\theta_j^{(t)}$ from
   $\theta_j \sim p(\theta_j|\theta_1^{(t)}, \ldots, \theta_{j-1}^{(t)}, \theta_{j+1}^{(t-1)}, \ldots, \theta_J^{(t-1)}, \mathbf{y})$.

   (c) Set $\theta^t = \theta$ and save this as the generated set of values at the $t+1$ iteration of the algorithm.

Thus, given a particular state of the chain $\theta^{(t)}$, we generate the new parameter values by

$$\begin{aligned}
\theta_1^{(t)} \quad & from \quad p(\theta_1|\theta_2^{(t-1)}, \theta_3^{(t-1)}, \ldots, \theta_J^{(t-1)}, \mathbf{y}), \\
\theta_2^{(t)} \quad & from \quad p(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_J^{(t-1)}, \mathbf{y}), \\
\vdots \quad & from \quad \vdots \\
\theta_J^{(t)} \quad & from \quad p(\theta_J|\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_J^{(t)}, \mathbf{y}).
\end{aligned}$$

The long-run distribution of $\theta^{(T)} = (\theta_1^{(T)}, \ldots, \theta_J^{(T)})$ is then the true posterior $p(\theta_1, \ldots, \theta_J|\mathbf{y})$ so that for a large T the value $\theta^{(T)}$ will approximately be a random draw from the true posterior.

As mentioned the Gibbs sampler is a special case of the Metropolis-Hastings algorithm, where the conditional candidate density for each of the parameter components is the conditional density of that component, given all the parameters in the other components and the observed data. Then, due to the candidates being drawn from the correct full conditional distribution, we will always accept the draw of each proposed move. Further details of the Gibbs sampler can be found in Casella and George (1992) and its application in Smith and Roberts (1993).

WinBUGS makes great use of this algorithm in order to perform a Bayesian analysis.

The results presented in the following section will be obtained using WinBUGS in order to perform a Bayesian analysis. Hence, data will be analysed based upon the Bayesian methodology that has been discussed thus far in this chapter. The data that are to be analysed in order to achieve these results shall also be given in the following section.

4.4 Bayesian Analysis of Reactor Data

The Bayesian methods set out thus far shall now be illustrated with a practical example where some reactor data, which shall be discussed in some detail in the following subsection, shall be analysed in a Bayesian way. The model and prior assumptions to be made shall then be set out before finally the results are presented and discussed.

4.4.1 Reactor Data

The reactor data which are to be analysed in a Bayesian fashion are taken from Box, Hunter and Hunter (1978). This particular experiment and the resulting dataset is being considered due to its feature of being typical of an industrial experiment. There are 5 factors being investigated all at two levels and these variables and the particular settings for their low and high levels are given in Table 4.1. Due to the design being a 2^5 factorial design, the design consists of 32 runs as displayed in Table 4.2.

From the standard analyses of the complete factorial arrangement, estimates of effects and a normal plot of effects as displayed in Table 4.3 and Figure 4.1, allow us to conclude that the only effects distinguishable from noise are the main effects **2, 4** and **5** and two-factor interactions **24** and **45**. The normal plot has the notation of the variables identified by capital letters rather than notated by numbers, where A denotes main effect **1**, B denotes main effect **2** and so on. BD and DE are the only two-factor interaction effects that are labelled on the normal plot as they fall away from the line and are clearly distinguishable from noise in comparison to the other interactions. These represent **24** and **45** respectively.

Table 4.1: Variables for 2^5 factorial design, reactor example

	Variable	-	+
1	Feed Rate (litres/min)	10	15
2	Catalyst (%)	1	2
3	Agitation Rate (rpm)	100	120
4	Temperature (°C)	140	180
5	Concentration (%)	3	6

Figure 4.1: Normal Plot for Full Factorial Design of Reactor Data

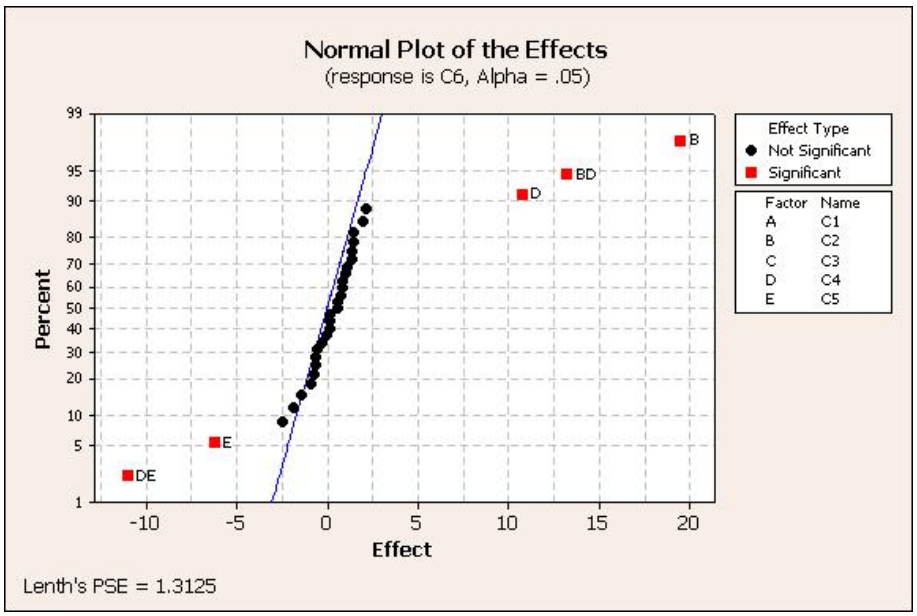


Table 4.2: Results from 2^5 factorial design, reactor example

	variable					response (% reacted)
Run	**1**	**2**	**3**	**4**	**5**	**y**
1	-	-	-	-	-	61
∗2	+	-	-	-	-	53
∗3	-	+	-	-	-	63
4	+	+	-	-	-	61
∗5	-	-	+	-	-	53
6	+	-	+	-	-	56
7	-	+	+	-	-	54
∗8	+	+	+	-	-	61
∗9	-	-	-	+	-	69
10	+	-	-	+	-	61
11	-	+	-	+	-	94
∗12	+	+	-	+	-	93
13	-	-	+	+	-	66
∗14	+	-	+	+	-	60
∗15	-	+	+	+	-	95
16	+	+	+	+	-	98
∗17	-	-	-	-	+	56
18	+	-	-	-	+	63
19	-	+	-	-	+	70
∗20	+	+	-	-	+	65
21	-	-	+	-	+	59
∗22	+	-	+	-	+	55
∗23	-	+	+	-	+	67
24	+	+	+	-	+	65
25	-	-	-	+	+	44
∗26	+	-	-	+	+	45
∗27	-	+	-	+	+	78
28	+	+	-	+	+	77
∗29	-	-	+	+	+	49
30	+	-	+	+	+	42
31	-	+	+	+	+	81
∗32	+	+	+	+	+	82

Table 4.3: Analysis of 2^5 factorial design, reactor example

estimates of effects	
average = 65.5	
1 = -1.375	123 = 1.50
2 = 19.5	124 = 1.375
3 = -0.625	125 = -1.875
4 = 10.75	134 = -0.75
5 = -6.25	135 = -2.50
	145 = 0.625
12 = 1.375	235 = 0.125
13 = 0.75	234 = 1.125
14 = 0.875	245 = -0.250
15 = 0.125	345 = 0.125
23 = 0.875	
24 = 13.25	1234 = 0.0
25 = 2.0	1245 = 0.625
34 = 2.125	2345 = -0.625
35 = 0.875	1235 = 1.5
45 = -11.0	1345 = 1.0
	12345 = -0.25

As previously stated, the run size for a full 2^k factorial design increases geometrically as the number of variables k increases, and for the full 2^5 factorial, the following effects can be estimated as has just been demonstrated:

Table 4.4: Analysis of 2^5 factorial design, reactor example

Type of Effect	Number
Average	1
Main Effects	5
2-Factor Interactions	10
3-Factor Interactions	10
4-Factor Interactions	5
5-Factor Interactions	1

Although all the effects can be estimated, they will not all be of considerable size.

Thus, higher order interactions than two-factor ones can be disregarded based on the hierarchical ordering principle which states that lower order effects are more likely to be important than higher order effects. In this way, information that an experimenter wishes to gain can be obtained by carrying out an experiment where only a fraction of the full factorial design is used. Applying this in the context of the reactor example, only 16 runs of the original design are now used where previously the full factorial design was considered. This half-fraction, which is a Resolution V design, is constructed using the design generator $\mathbf{5 = 1234}$, the design for which is presented in Table 4.5. The 16 runs which make up this design are marked with asterisks where the full factorial design has been given in Table 4.2.

By using the design generator $\mathbf{5 = 1234}$, we assign the column that equals the product of the columns for factors **1, 2, 3** and **4** to factor **5**. As a result, the data from such a design will not be able to distinguish the estimate of main effect **5** from the estimate of the 4-factor interaction **1234**, and we say main effect **5** is aliased with the **1234** interaction. This is a price that we pay for using a fraction of the full factorial design. In the case of this Resolution V design, all main effects are strongly clear (they are not aliased with any other main effects, two-factor or three-factor interactions) and the two-factor interactions are clear (they are not aliased with main effects or other two-factor interactions).

The analysis of the half-fraction produces estimates of effects and a normal plot, displayed in Table 4.6 and Figure 4.2, that draw attention to precisely the same effects as the analysis of the full factorial. Thus, it can be observed that although aliasing is introduced as a result of carrying out only half of the total number of runs, the most essential information from this experiment can still be gained, saving the experimenter considerable efforts in terms of both time and cost.

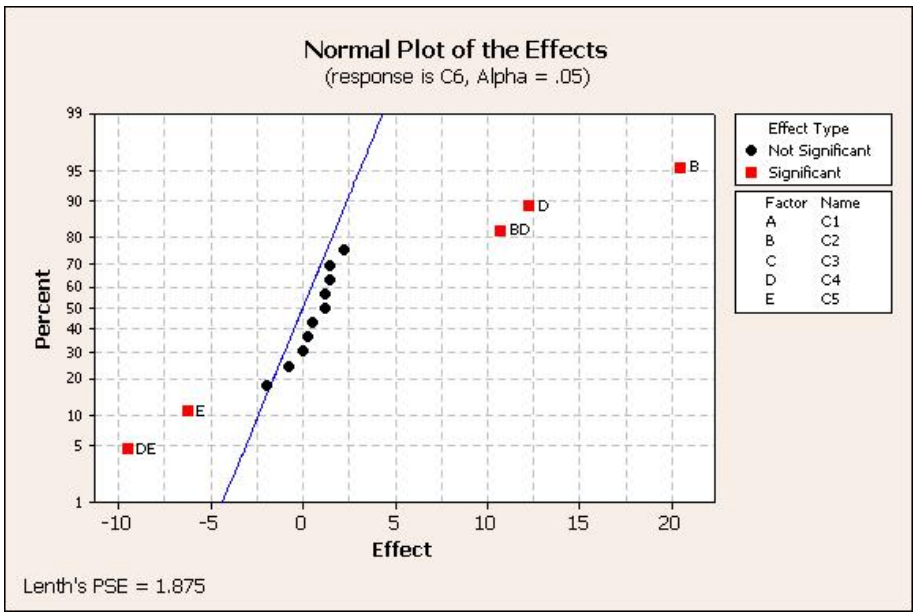
Table 4.5: Design for half-fraction of the full 2^5 factorial design, reactor example

	Design					response (% reacted)
Run	**1**	**2**	**3**	**4**	**5 = 1234**	**y**
17	-	-	-	-	+	56
2	+	-	-	-	-	53
3	-	+	-	-	-	63
20	+	+	-	-	+	65
5	-	-	+	-	-	53
22	+	-	+	-	+	55
23	-	+	+	-	+	67
8	+	+	+	-	-	61
9	-	-	-	+	-	69
26	+	-	-	+	+	45
27	-	+	-	+	+	78
12	+	+	-	+	-	93
29	-	-	+	+	+	49
14	+	-	+	+	-	60
15	-	+	+	+	-	95
32	+	+	+	+	+	82

Table 4.6: Analysis of a Half-fraction of the full 2^5 factorial design, reactor example

estimates of effects	
average = 65.25	
1 = -2.0	12 = 1.5
2 = 20.5	13 = 0.5
3 = 0.0	14 = -0.75
4 = 12.25	15 = 1.25
5 = -6.25	23 = 1.50
	24 = 10.75
	25 = 1.25
	34 = 0.25
	35 = 2.25
	45 = -9.50

Figure 4.2: Normal Plot for Half-fraction of Reactor Data



Now that the reactor data and the standard analyses of these data have been presented, the dataset will now be re-analysed in the following section using Bayesian methods. The Bayesian analyses will be compared with the standard analyses to assess whether any differing conclusions are drawn when making various prior distributional assumptions on the model and parameter effects based upon prior knowledge that is available. It will also be of interest to observe if using prior knowledge in a fully Bayesian data analysis improves upon standard analyses and provides an experimenter with any useful information or enables them to gain greater insight of the scientific process they are investigating.

4.4.2 Prior Specification

The various likelihood model assumptions on the data and prior assumptions on the effect parameters and variance being investigated will now be set out.

Assumptions about how the data are modelled will be with respect to the observed data $\mathbf{y} = [y_1, \ldots, y_N]^{\mathrm{T}}$ where

$$\mathbf{y} = [56, 53, 63, 65, 53, 55, 67, 61, 69, 45, 78, 93, 49, 60, 95, 82]^{\mathrm{T}}.$$

That is, the observed data being considered shall be the data when a half-fraction of the complete 2^5 factorial design is used. This is due to reasons set out in the previous subsection (§4.4.1) which are that the most important information can still be gained from carrying out only half the number of runs of the complete factorial design.

The distributions that shall be assumed about how the data are modelled will be the normal and the scaled-t distribution. The observed data shall take the following form,

$$y_i = \mu_i + \epsilon_i$$

where

$$\mu_i = \beta_0 + \sum_{i=1}^{5} \beta_i x_i + \sum \sum_{i<j}^{5} \beta_{ij} x_i x_j.$$

When assuming the normal distribution,

$$\epsilon_i \sim N(\mu, \sigma^2)$$

and when assuming the scaled-t distribution,

$$\epsilon_i \sim t_\nu(\mu, \sigma^2)$$

That is, it is the errors that are assumed to have a normal or scaled-t distribution. The scaled-t distribution has three parameters which it is determined by, the location, or centre μ, the scale σ, and the degrees of freedom, or shape parameter ν.

When assuming that the errors are modelled by the scaled-t distribution, we shall take the degrees of freedom parameter ν, which determines the shape of the distribution,

to be 4. When taking the extreme case of $\nu = 1$, this is in fact the same as the Cauchy distribution and thus the shape of the distribution is so long-tailed that the mean is undefined and the variance are infinite. When looking at the opposite extreme, i.e. when $\nu \to \infty$, then the scaled-t distribution can then be shown to tend to the normal distribution. This gives some explanation for the choice of $\nu = 4$. That is, in the case when the degrees of freedom parameter is set to be extremely small, for example at one or two, then the prior distribution will have infinite variance which will typically not be truly representative particularly when looking at the far tails. Conversely, when taking the degrees of freedom to be extremely large or tending to infinity, then we will be looking at the normal distribution and we are seeking to investigate an alternative distribution to the normal. In fact, the choice of scaled-t distribution chosen as a robust alternative to the normal distribution, particularly when an experimenter suspects their dataset of containing outliers, will be discussed further in the following chapter.

Note that in the cases of the normal distribution given above, the variance has been denoted by its standard notation of σ^2. However, throughout this section for all other distributions that are to be specified, the variance shall now be represented in terms of the precision τ rather than being represented by its typical notation of σ^2. This is due to the fact that WinBUGS works with the precision rather than the variance. For the normal distribution this will be represented as

$$\epsilon_i \sim N(\mu, \tau),$$

and for the scaled-t distribution this will be represented as

$$\epsilon_i \sim t_\nu(\mu, \tau).$$

Also note that the effects being estimated from the analysis of these data are main effects and two-factor interactions, due to the hierarchical ordering principle which

states that lower order effects are more likely to be important than higher order effects. The aliasing structure also means that both the main effects and two-factor interactions are clear since none of their aliases are any other main effects or two-factor interactions. The prior distributions that shall be assumed for the effect parameters, i.e. the main effects and two-factor interaction effects are normal distributions and also a mixture of normal distributions. In all cases the intercept shall be distributed normally with a large variance that shall be assumed to approach ∞. This is to represent the belief that we are generally not concerned with estimating the intercept parameter.

The prior distributions that shall be placed on the residual variance shall be the inverse gamma distribution and a uniform distribution placed on $\log \sigma$. Since the precision, τ, is the inverse of the variance then when considering an inverse gamma distribution on the variance parameter, i.e. that

$$\sigma^2 \sim IG(a, b)$$

for some values a and b, then it follows that the prior distribution on the precision, τ can be represented as

$$\tau \sim Gamma(a, b).$$

When considering the uniform prior distribution on $\log \sigma$, this can be represented as

$$\log \sigma \sim Unif[-c, c]$$

for some value c.

A summary of the prior assumptions that shall be investigated, and also the parameter values that shall be assumed for each of the distributions investigated are displayed in the following table. As previously stated, where the value for the variance parameter is stated, such as with 10^{-3} in the case of the normal distribution for the main effects and two-factor interactions, this is in fact the value for the precision, τ.

Table 4.7: Prior Distributional Assumptions

	Distribution	Parameter Values
Model for errors	Normal	$N(\mu, \tau)$
	Scaled-t	$t_4(\mu, \tau)$
Intercept	Normal	$N(0, 10^{-\infty})$
Main Effects and two-factor interactions	Normal	$N(0, 10^{-3})$
	Mixture of normals	$N(0, \tau_1)$ if effect is inactive
		$N(0, \tau_2)$ if effect is active
		where $\tau_1 = \frac{1}{\sigma_1^2}$, $\tau_2 = \frac{1}{\sigma_2^2}$
		and $\sigma_1 \sim$ Unif $(0, 1)$, $\sigma_2 \sim$ Unif $(0, 10)$
		$p_0 = 0.85$ (probability effect is inactive)
		$p_1 = 0.15$ (probability effect is active)
Residual Variance	Inverse gamma	$\tau \sim$ Gamma $(0.001, 0.001)$
		$\tau \sim$ Gamma $(1, 1)$
		$\tau \sim$ Gamma $(5, 5)$
	Uniform	$\log \sigma \sim$ Unif $[-10, 10]$

4.4.3 MCMC Diagnostics

Before the summary statistics obtained when using WinBUGS to analyse the reactor data are presented, some comments on the MCMC diagnostics are given. Using MCMC methods, by way of an iterative simulation method posterior distributions were computed for each of the models fitted in order to achieve the summary statistics displayed in the following subsection (§4.4.4). As described previously, at each step in the simulation the approximate distributions are improved so that the chain is converging to the target distribution. Thus, before any inferences can be made based upon these posterior estimates, it is necessary to look at MCMC diagnostics to assess whether the computation has 'worked' and that the chain has in fact converged to equilibrium.

For each of the models fitted and prior assumptions made, 20000 iterative simulations were run for two chains, where different initial values for the two chains were used.

The purpose of running more than one chain is to check that, given that each chain is started at a different place, the two chains eventually converge so that both are representative of the same target distribution. The iterations were also *thinned* by only storing every 10^{th} simulation, which proves useful when running the chain for a long time and when there are a large number of parameters to store simulations for. 20000 simulations were found to be a sufficient number of simulations for convergence to be reached for all parameters, assessed by looking at trace plots, autocorrelation function plots and Brooks-Gelman-Rubin plots produced in WinBUGS. Trace plots give the value of the parameter monitored against the iteration number and can be assessed visually to check for convergence. Autocorrelation function plots can be used to detect if there is correlation within values in the chain and the Brooks-Gelman-Rubin diagnostic is a method for assessing the convergence of parallel chains. Assuming that the chains start from widely different starting values, one would expect that as the chains come closer into agreement the variability of the pooled chains should be similar to the average variability of the individual chains. Selected plots for the main effects, σ and τ, produced in WinBUGS when analysing the reactor data assuming a normal model on the data, normal mixture distribution on effect parameters and Gamma$(1, 1)$ distribution on the precision are given in Figures 4.3 - 4.7. Looking at the trace plots it can be seen that the chains have converged and are well 'mixed'. All the autocorrelation function plots are tailing off, becoming indistinguishable from zero very quickly for the main effects and within less than 10 lags for σ and τ. This indicates that values within the chain are not highly correlated. The Brooks-Gelman-Rubin plots also confirm that convergence of parallel chains is achieved. The red line displays the ratio of the variability of the pooled chains to the average variability of the individual chains and confirms that convergence is satisfactory since in all plots, this ratio converges to one.

Carrying out checks of the convergence and correlation by looking at these plots for various prior assumptions made when analysing the reactor data in a Bayesian

fashion in WinBUGS, it was found that it was suitable to run the chain for 20000 simulations, as mentioned previously. A rule-of-thumb for checking that the number of iterations is suitable is by looking at the MC error, which is an adjusted standard error and is obtained for each parameter monitored. A suitable number of further iterations are required once the chain has converged, in order to obtain accurate posterior estimates and the simulations should be run for further iterations until the MC error is approximately no more than 5% of the sample standard deviation for all parameters. Values for the MC error of all parameters are given in the results table in the following subsection. A burn-in of the first 5000 simulations was also found to be sufficient. Thus, 'throwing away' the first 5000 simulations ensures that any results and inferences will only be based upon those sample draws after convergence has been reached.

Figure 4.3: Trace Plots for $\beta_1, \ldots, \beta_5$, σ and τ

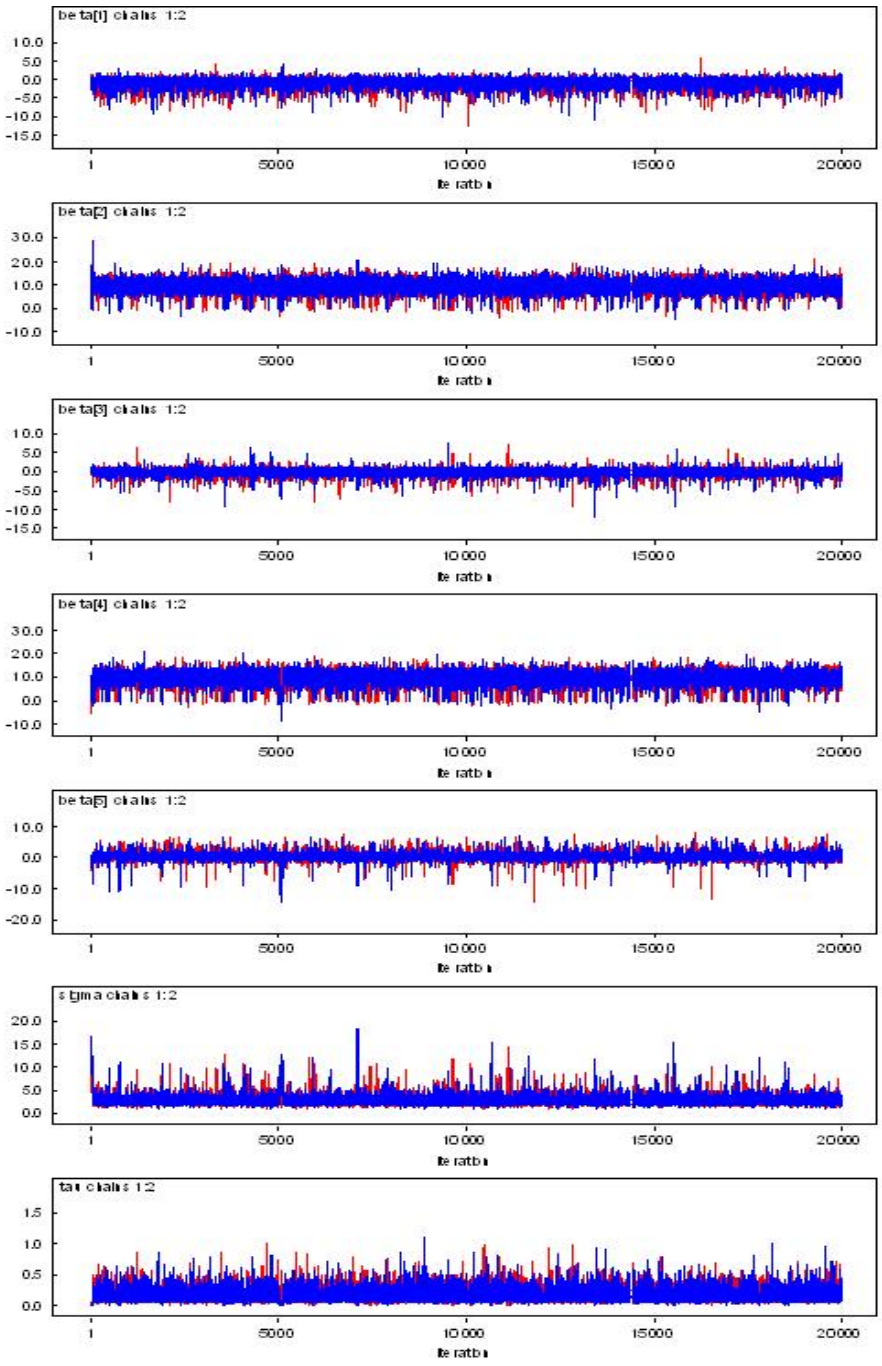


Figure 4.4: Autocorrelation Function Plots for $\beta_1, \ldots, \beta_5$

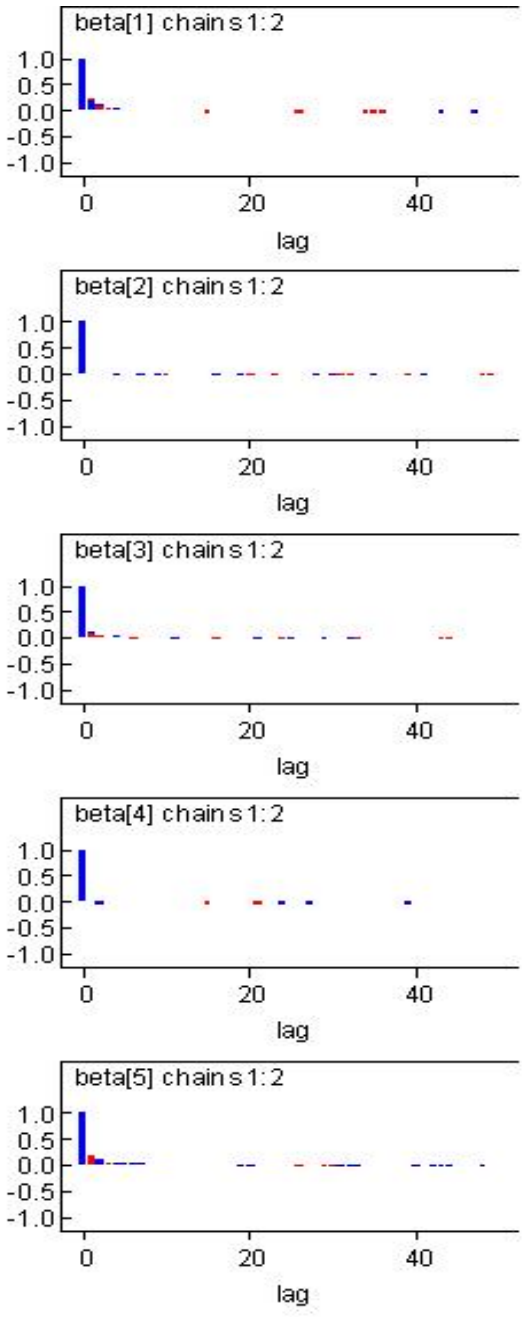


Figure 4.5: Brooks-Gelman-Rubin Plots for $\beta_1, \ldots, \beta_5$

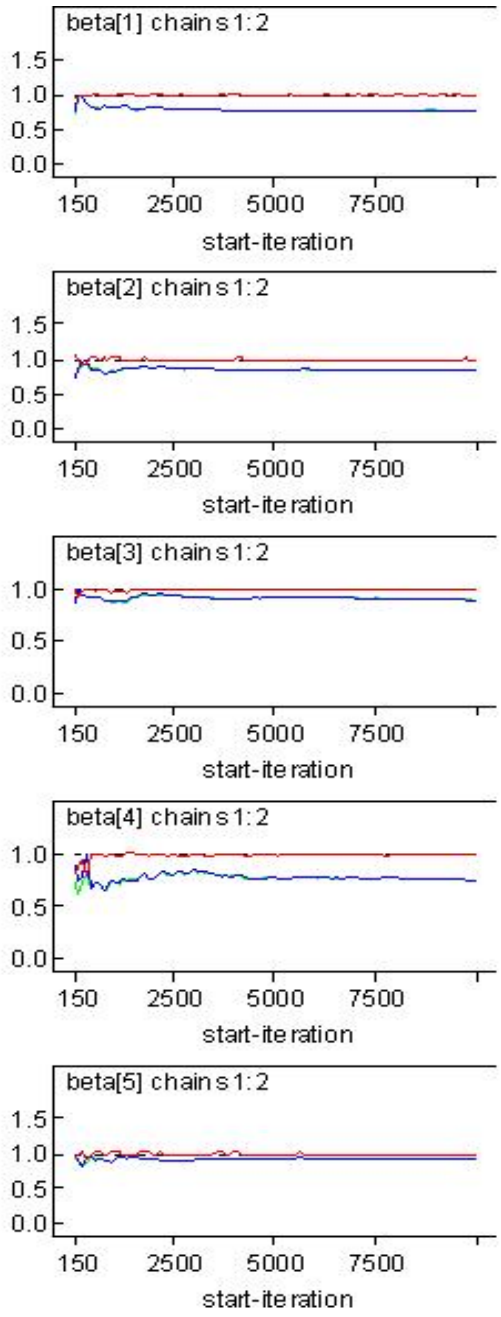


Figure 4.6: Autocorrelation Function Plots for σ and τ

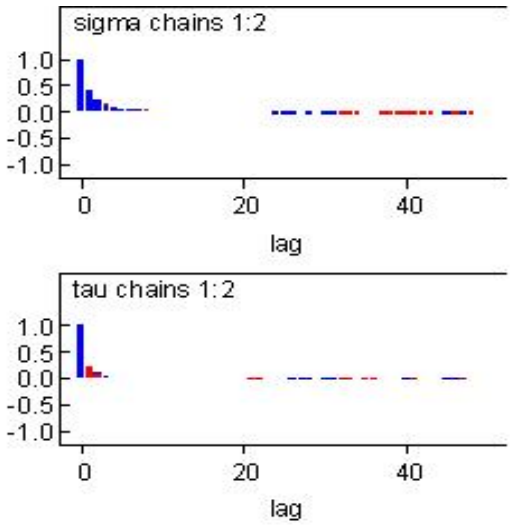
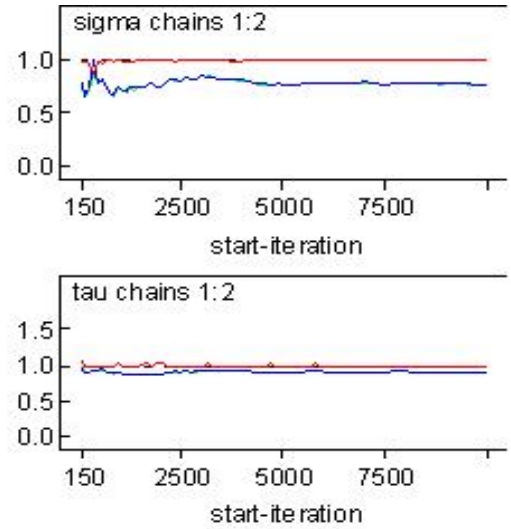


Figure 4.7: Brooks-Gelman-Rubin Plots for σ and τ



4.4.4 Results

The summary statistics obtained when using WinBUGS to analyse the reactor data are now presented. This is the analysis when estimating all main effects and two-factor interactions when experimenting using a half-fraction (Table 4.5) of the original 2^5 experiment. After the tables displaying the summary statistics are presented, comments on the results are given.

Table 4.8: Summary statistics from WinBUGS after fitting normal model to reactor data, with normal distribution on effect parameters and Gamma$(0.001, 0.001)$ prior on precision

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	65.27	3.492	0.01965	58.63	65.25	71.9
β_1	−0.995	3.194	0.01713	−7.261	−1.0	5.448
β_2	10.12	3.143	0.01761	3.182	10.25	15.9
β_3	−0.1377	3.201	0.01808	−6.499	−0.0573	6.307
β_4	6.055	3.24	0.0186	−0.9433	6.125	12.27
β_5	−3.102	3.236	0.01847	−9.621	−3.124	3.588
β_{12}	0.7467	3.169	0.01645	−5.562	0.7506	7.036
β_{13}	0.2552	3.181	0.01868	−6.205	0.2502	6.586
β_{14}	−0.4176	3.177	0.01865	−6.97	−0.3758	5.872
β_{15}	0.6146	3.189	0.01788	−5.897	0.6254	7.019
β_{23}	0.75	3.204	0.01689	−5.605	0.7509	7.096
β_{24}	5.331	3.303	0.01846	−1.285	5.375	11.57
β_{25}	0.6041	3.223	0.01905	−6.144	0.624	7.166
β_{34}	0.117	3.208	0.02013	−6.408	0.1254	6.649
β_{35}	1.142	3.157	0.01617	−5.179	1.126	7.401
β_{45}	−4.736	3.185	0.01779	−11.1	−4.749	1.772
σ	6.415	12.13	0.2476	0.0357	1.118	43.89
τ	79.26	268.7	3.955	0.000521	0.801	785.4

Table 4.9: Summary statistics from WinBUGS after fitting normal model to reactor data, with normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	65.25	0.6356	0.003984	64.19	65.25	66.31
β_1	-1.003	0.6619	0.003903	-2.042	-1.002	0.03457
β_2	10.25	0.6342	0.004096	9.178	10.25	11.29
β_3	0.00129	0.657	0.003687	-1.057	0.002397	1.087
β_4	6.122	0.6144	0.003484	5.059	6.125	7.182
β_5	-3.133	0.6196	0.003368	-4.2	-3.128	-2.106
β_{12}	0.7525	0.6234	0.003619	-0.3042	0.7533	1.827
β_{13}	0.2535	0.629	0.003502	-0.8051	0.2557	1.301
β_{14}	-0.3823	0.6413	0.00376	-1.478	-0.3807	0.6814
β_{15}	0.625	0.6495	0.003875	-0.4409	0.6282	1.701
β_{23}	0.7493	0.6313	0.003982	-0.3068	0.7497	1.819
β_{24}	5.373	0.6247	0.003555	4.286	5.372	6.437
β_{25}	0.6226	0.6814	0.004071	-0.4383	0.6233	1.671
β_{34}	0.1211	0.6287	0.003626	-0.9284	0.124	1.154
β_{35}	1.125	0.6251	0.003446	0.05736	1.125	2.187
β_{45}	-4.744	0.638	0.003664	-5.814	-4.746	-3.67
σ	1.713	1.906	0.02265	0.518	1.205	6.086
τ	1.001	1.004	0.008064	0.0271	0.6882	3.73

Table 4.10: Summary statistics from WinBUGS after fitting normal model to reactor data, with normal distribution on effect parameters and Gamma$(5, 5)$ prior on precision

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	65.25	0.2807	0.001612	64.69	65.25	65.81
β_1	−0.9987	0.2794	0.001557	−1.552	−0.9997	−0.4417
β_2	10.25	0.2824	0.001504	9.687	10.25	10.81
β_3	0.0275	0.2793	0.001712	−0.5599	0.0270	0.5533
β_4	6.124	0.2791	0.001699	5.566	6.126	6.688
β_5	−3.126	0.2801	0.001552	−3.683	−3.126	−2.563
β_{12}	0.7502	0.2807	0.001702	0.1936	0.7501	1.306
β_{13}	0.2506	0.2812	0.00148	−0.3142	0.2539	0.8052
β_{14}	−0.3879	0.2794	0.001612	−0.9365	−0.3779	0.7124
β_{15}	0.6263	0.2792	0.001488	0.06774	0.626	1.181
β_{23}	0.749	0.2789	0.001424	0.1882	0.7506	1.299
β_{24}	5.373	0.28	0.001617	4.818	5.37	5.937
β_{25}	0.6287	0.2773	0.001655	0.07382	0.6276	1.18
β_{34}	0.1252	0.2779	0.001618	−0.4284	0.1246	0.6777
β_{35}	1.125	0.2783	0.001513	0.5729	1.124	1.683
β_{45}	−4.749	0.2804	0.001521	−5.314	−4.749	−4.191
σ	1.083	0.2751	0.001596	0.7009	1.033	1.761
τ	1.001	0.446	0.002472	0.3225	0.9366	2.036

Table 4.11: Summary statistics from WinBUGS after fitting normal model to reactor data, with normal distribution on effect parameters and Unif$[-10, 10]$ prior on $\log \sigma$

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	65.25	2.327	0.01421	62.0	65.25	68.54
β_1	-0.9987	2.067	0.01168	-4.112	-1.0	2.219
β_2	10.22	2.066	0.01253	6.895	10.25	13.29
β_3	0.00172	2.09	0.01118	-3.167	-0.0033	3.185
β_4	6.095	2.106	0.01283	2.711	6.125	9.164
β_5	-3.121	2.132	0.01345	-6.376	-3.125	0.2634
β_{12}	0.7382	2.207	0.01292	-2.609	0.75	3.876
β_{13}	0.2388	2.17	0.01193	-3.12	0.25	3.303
β_{14}	-0.3574	2.155	0.01111	-3.426	-0.375	2.896
β_{15}	0.6328	2.124	0.01276	-2.686	0.625	3.751
β_{23}	0.7533	2.108	0.01351	-2.411	0.75	3.922
β_{24}	5.364	2.049	0.01195	2.168	5.375	8.51
β_{25}	0.6168	2.149	0.01147	-2.669	0.625	3.813
β_{34}	0.1132	2.14	0.01218	-3.162	0.125	3.258
β_{35}	1.121	2.128	0.01247	-2.125	1.125	4.342
β_{45}	-4.728	2.114	0.0117	-8.022	-4.75	-1.354
σ	2.961	8.605	0.2409	0.0425	0.03601	29.69
τ	2001.356	6974.587	593.0403	0.001139	771.5	7482.205

Table 4.12: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with normal distribution on effect parameters and Gamma$(0.001, 0.001)$ prior on precision

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	65.25	3.287	0.01877	58.9	65.25	71.6
β_1	−0.9843	3.139	0.01755	−6.988	−0.9995	5.352
β_2	10.2	3.092	0.02013	3.995	10.25	16.14
β_3	−0.04449	3.06	0.0182	−6.391	−0.1019	5.838
β_4	6.072	3.125	0.01773	−0.3598	6.124	12.14
β_5	−3.095	3.102	0.01797	−9.014	−3.125	3.222
β_{12}	0.7027	3.122	0.01817	−5.596	0.7494	6.841
β_{13}	0.2593	3.081	0.01885	−5.778	0.2497	6.443
β_{14}	−0.3894	3.026	0.0195	−6.578	−0.3751	5.609
β_{15}	0.5829	3.04	0.01933	−5.802	0.6242	6.446
β_{23}	0.7298	3.084	0.01672	−5.528	0.7503	7.081
β_{24}	5.321	3.061	0.01942	−1.087	5.375	11.2
β_{25}	0.6247	3.163	0.01744	−5.474	0.6252	6.941
β_{34}	0.1318	3.076	0.01967	−6.046	0.1255	6.105
β_{35}	1.08	3.072	0.01769	−5.195	1.125	7.06
β_{45}	−4.727	3.095	0.01802	−10.75	−4.749	1.406
σ	4.749	9.386	0.2082	0.0344	0.8726	32.24
τ	88.58	294.7	4.217	0.1765	1.313	845.4

Table 4.13: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	65.24	0.8434	0.004973	63.79	65.25	66.66
β_1	−0.9985	0.7977	0.005004	−2.45	−0.9976	0.4418
β_2	10.24	0.811	0.004766	8.778	10.25	11.68
β_3	−0.001304	0.8173	0.00459	−1.441	−0.1238	1.463
β_4	6.119	0.8364	0.004897	4.654	6.122	7.589
β_5	−3.124	0.8096	0.004995	−4.565	−3.125	−1.663
β_{12}	0.7449	0.8198	0.004598	−0.683	0.7455	2.2
β_{13}	0.2381	0.8194	0.004672	−1.245	0.246	1.664
β_{14}	−0.3714	0.8344	0.005078	−1.806	−0.3767	1.117
β_{15}	0.6271	0.836	0.005213	−0.843	0.626	2.055
β_{23}	0.7516	0.8216	0.004378	−0.7032	0.7511	2.216
β_{24}	5.365	0.8145	0.004467	3.922	5.371	6.814
β_{25}	0.6298	0.8154	0.004622	−0.7799	0.6256	2.051
β_{34}	0.1284	0.8301	0.004623	−1.285	0.1228	1.572
β_{35}	1.12	0.8736	0.005259	−0.3813	1.123	2.557
β_{45}	−4.746	0.8255	0.004805	−6.198	−4.747	−3.259
σ	1.672	1.773	0.01915	0.5141	1.182	5.952
τ	1.019	1.011	0.007989	0.02823	0.7164	3.784

Table 4.14: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with normal distribution on effect parameters and Gamma$(5, 5)$ prior on precision

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	65.25	0.394	0.002171	64.46	65.25	66.04
β_1	-1.0	0.3928	0.002265	-1.788	-0.9979	-0.2072
β_2	10.25	0.3947	0.002203	9.455	10.25	11.04
β_3	-0.00166	0.396	0.00241	-0.8	-0.1010	0.7871
β_4	6.125	0.3946	0.002175	5.334	6.124	6.913
β_5	-3.125	0.3961	0.002144	-3.92	-3.126	-2.322
β_{12}	0.7489	0.3932	0.002151	-0.04765	0.7519	1.534
β_{13}	0.2541	0.3931	0.00215	-0.5319	0.2545	1.042
β_{14}	-0.375	0.393	0.002174	-1.165	-0.3707	0.3989
β_{15}	0.6234	0.3938	0.002211	-0.1604	0.6204	1.414
β_{23}	0.7497	0.3947	0.002304	-0.04447	0.7523	1.536
β_{24}	5.375	0.3941	0.002205	4.59	5.376	6.171
β_{25}	0.6265	0.3904	0.002198	-0.1558	0.6259	1.423
β_{34}	0.124	0.3951	0.002254	-0.6739	0.1252	0.912
β_{35}	1.126	0.3963	0.002259	0.3346	1.126	1.915
β_{45}	-4.749	0.3932	0.002295	-5.531	-4.752	-3.958
σ	1.084	0.2744	0.001509	0.6964	1.034	1.762
τ	0.9999	0.4485	0.002587	0.3221	0.9357	2.062

Table 4.15: Summary statistics from WinBUGS after fitting normal model to reactor data, with normal mixture distribution on effect parameters and Gamma$(0.001, 0.001)$ prior on precision

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	65.25	0.5298	0.003245	64.21	65.25	66.29
β_1	−0.755	0.4063	0.006768	−1.219	−0.9565	0.1747
β_2	10.14	0.7976	0.03183	8.846	10.25	11.04
β_3	−0.00181	0.2383	0.001291	−0.533	−0.0510	0.5221
β_4	6.056	0.6456	0.02105	4.836	6.122	6.99
β_5	−2.966	0.697	0.01247	−3.861	−3.12	−0.3239
β_{12}	0.566	0.3446	0.004952	−0.2464	0.7173	1.011
β_{13}	0.1869	0.2488	0.002119	−0.4129	0.2375	0.6617
β_{14}	−0.2833	0.2679	0.002828	−0.7468	−0.03577	0.3487
β_{15}	0.4721	0.3144	0.004313	−0.2782	0.598	0.9263
β_{23}	0.7456	0.4641	0.003241	−0.2871	0.7495	1.751
β_{24}	5.314	0.5855	0.01739	4.111	5.373	6.25
β_{25}	0.6173	0.4541	0.003682	−0.4135	0.6247	1.599
β_{34}	0.09468	0.2428	0.00172	−0.4611	0.1179	0.599
β_{35}	0.8481	0.4441	0.007369	−0.1644	1.047	1.319
β_{45}	−4.693	0.5714	0.01645	−5.637	−4.748	−3.526
σ	1.265	1.729	0.05647	0.03407	0.6061	5.212
τ	91.51	290.1	3.808	0.03682	2.723	861.6

Table 4.16: Summary statistics from WinBUGS after fitting normal model to reactor data, with normal mixture distribution on effect parameters and Gamma$(1, 1)$ prior on precision

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	65.25	0.4373	0.002493	64.37	65.25	66.16
β_1	−0.6774	0.3981	0.003744	−1.319	−0.7535	0.1275
β_2	10.19	0.4425	0.002871	9.208	10.22	11.03
β_3	0.002537	0.2752	0.001499	−0.5639	0.001111	0.5698
β_4	6.089	0.4466	0.002931	5.14	6.103	6.95
β_5	−3.046	0.5385	0.004394	−3.945	−3.097	−1.749
β_{12}	0.508	0.3506	0.002852	−0.2052	0.5566	1.11
β_{13}	0.1674	0.2869	0.001855	−0.4238	0.1707	0.7238
β_{14}	−0.2544	0.2972	0.002004	−0.8286	−0.2673	0.3527
β_{15}	0.4259	0.3275	0.002449	−0.2343	0.4611	1.019
β_{23}	0.7462	0.4292	0.002383	−0.1621	0.7478	1.633
β_{24}	5.344	0.4356	0.002426	4.404	5.36	6.203
β_{25}	0.619	0.4289	0.002516	−0.2811	0.6235	1.486
β_{34}	0.08649	0.281	0.001622	−0.4871	0.08426	0.6507
β_{35}	0.7637	0.4261	0.003857	−0.1035	0.8528	1.422
β_{45}	−4.722	0.4342	0.002644	−5.583	−4.737	−3.791
σ	1.503	0.8624	0.008583	0.5325	1.275	3.622
τ	0.9346	0.9402	0.007312	0.07637	0.6156	3.527

Table 4.17: Summary statistics from WinBUGS after fitting normal model to reactor data, with normal mixture distribution on effect parameters and Gamma$(5, 5)$ prior on precision

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	65.25	0.2855	0.001559	64.68	65.25	65.81
β_1	−0.8134	0.2876	0.001832	−1.333	−0.8327	−0.1642
β_2	10.22	0.286	0.001562	9.65	10.22	10.79
β_3	0.001718	0.2478	0.001469	−0.486	0.001089	0.5009
β_4	6.108	0.2878	0.001619	5.524	6.112	6.671
β_5	−3.105	0.2938	0.001577	−3.68	−3.106	−2.511
β_{12}	0.6116	0.2722	0.001728	0.02755	0.6236	1.125
β_{13}	0.2022	0.2507	0.001483	−0.3035	0.2034	0.6934
β_{14}	−0.305	0.2545	0.001577	−0.8033	−0.3092	0.2057
β_{15}	0.5074	0.2647	0.001492	−0.03352	0.5167	1.016
β_{23}	0.7469	0.2861	0.001692	−0.1768	0.7457	1.322
β_{24}	5.361	0.2863	0.001591	4.788	5.363	5.933
β_{25}	0.6225	0.2858	0.001609	0.04716	0.6231	1.191
β_{34}	0.1013	0.2497	0.001406	−0.3952	0.1005	0.5968
β_{35}	0.9195	0.2991	0.002034	0.2304	0.9443	1.447
β_{45}	−4.737	0.2871	0.001618	−5.305	−4.739	−4.164
σ	1.11	0.292	0.001864	0.7072	1.056	1.823
τ	0.9647	0.4452	0.002575	0.301	0.8969	1.999

The node column heading in the tables represent the parameters that have been monitored and that posterior estimates such as the mean and median have been calculated for using WinBUGS. In all cases the intercept β_0, main effects $\beta_1, \ldots, \beta_5$, two-factor interactions $\beta_{12}, \beta_{13}, \beta_{14}, \beta_{15}, \beta_{23}, \beta_{24}, \beta_{25}, \beta_{34}, \beta_{35}, \beta_{45}$, residual standard deviation σ and precision τ have all been monitored. sd denotes the standard deviation for each of the individual parameters and the MC error is an adjusted standard error as mentioned previously. Observing the values of the MC error in all tables it can be seen that the MC error is less than, or approximately 5% of the sample standard deviation which indicates that we can conclude that an appropriate number of iterations have been carried out. The values given in the tables under the column headings 2.5% and 97.5% are the end-points of the 95% Bayesian *credible intervals*. The interpretation of this interval differs greatly from the confidence interval considered in the classical, or frequentist, sense. A classical 95% confidence interval is interpreted such that if we were to draw repeatedly from the population being considered, then 95% of our confidence intervals would contain the population parameter. This contrasts to a 95% Bayesian credible interval which can be interpreted as after observing our data, there is a 95% chance that the parameter will fall in the interval.

Looking at the results it can be seen that there is some impact on the posterior estimated mean for σ and τ when varying the prior on the precision from non-informative, such as in the case of Gamma$(0.001, 0.001)$, to informative such as Gamma$(5, 5)$. As the prior on the precision becomes more informative, the posterior estimated mean of σ and τ is reduced in both cases when assuming the data are modelled normally or by a scaled-t distribution and in both cases this is when assuming normal distributions on the effects parameters. The size of τ is extremely large when a Gamma$(0.001, 0.001)$ distribution is assumed. When assuming a normal mixture prior distribution on the effect parameters it can be seen that the posterior estimate of σ is relatively small when assuming a Gamma$(0.001, 0.001)$ distribution, in comparison to the other models. However, it is not clear whether making the Gamma distribution on the precision

more informative has an impact on the posterior estimate of σ, as when a Gamma$(1, 1)$ distribution is assumed the posterior estimate is increased to 1.503, but then reduced again to 1.11 when placing a Gamma$(5, 5)$ distribution on τ. Conversely, the posterior estimate of τ is extremely large with a value of 91.51 when assuming a weakly informative prior on the precision, similarly to the other models. However, as the prior becomes more informative with the Gamma$(1, 1)$ distribution the posterior estimate is reduced to 0.9346, but then again increased slightly to 0.9647 when assuming a Gamma$(5, 5)$ distribution. Table 4.11 displays the results when assuming a weakly informative Uniform prior on $\log \sigma$ and normal distribution assumed on data and effect parameters. Again, it can be seen that when assuming a weakly informative prior, which could be said to be comparable to the Gamma$(0.001, 0.001)$ distribution, the posterior estimate of τ is extremely large with a value of 2001.356.

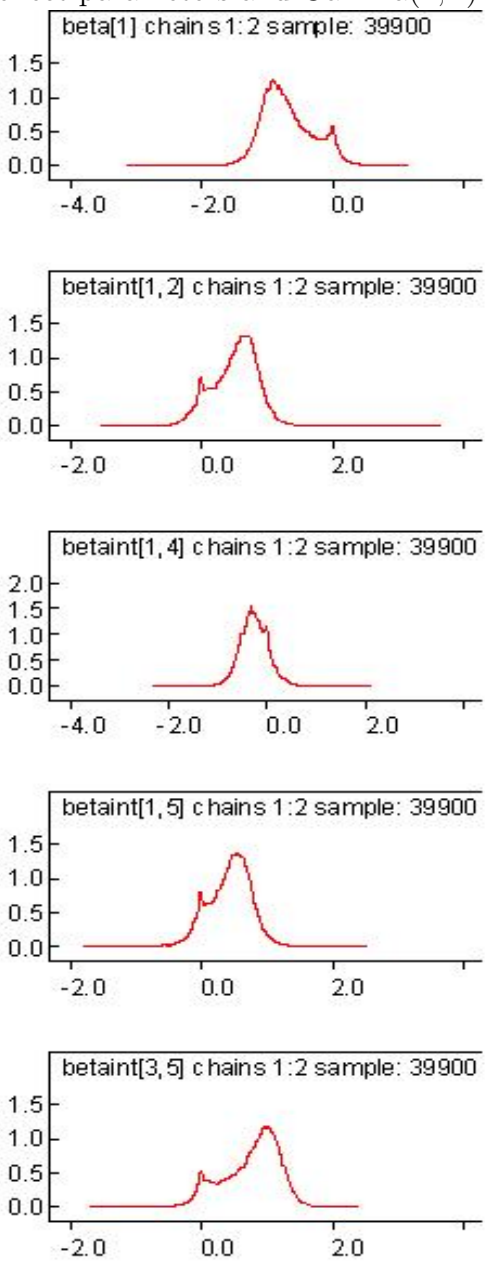
It should be noted that although a weakly informative prior such as the Unif$[-10, 10]$ distribution on $\log \sigma$ could be considered to be comparable to an alternative weakly informative prior such as the Gamma$(0.001, 0.001)$ distribution, the value of τ is much larger when assuming the Uniform distribution and also when assuming the Gamma$(0.001, 0.001)$ prior, the credible intervals are wider. When keeping the distribution that the data are modelled by, and the prior distribution on the effect parameters the same, but varying the prior on the precision from a weakly informative prior to an informative prior, the 95% Bayesian credible interval becomes narrower for all parameters monitored. This is also reiterated by the values of the standard deviation for each of the individual parameters which are reduced as the prior on the precision becomes more informative. It can also be seen that when assuming a mixture of normal distributions on the effect parameters and a Gamma$(0.001, 0.001)$ distribution on the precision (Table 4.15), the 95% Bayesian credible interval is much narrower in comparison to when assuming normal distributions on effect parameters (Table 4.8).

We can compare Tables 4.12 - 4.14 with Tables 4.8 - 4.10 to assess the impact of using a scaled-t distribution to model the data in place of the normal distribution which is perhaps a more typical assumption to make. Comparing the corresponding tables where the same prior assumptions have been made on the effect parameters and the precision, i.e. comparing Table 4.8 with Table 4.12, and so on, we observe that similar results are obtained for the posterior mean and median for each of the parameters, and also for the 95% Bayesian credible interval. Although very similar, the credible interval is slightly wider for two of the priors on the precision (Gamma$(1, 1)$ and Gamma$(5, 5)$) when assuming a scaled-t distribution on the data.

For all prior assumptions made, the corresponding Bayesian analysis indicates that the true active main effects **2**, **4** and **5**, and the true active two-factor interactions **24** and **45** should be identified as active due to the large values of their parameter effects β_{24}, β_4, β_5, β_{24} and β_{45}. This is also indicated by the 95% Bayesian credible intervals where all those truly inactive effects have credible limits close to 0. By definition, this means that 95% of the time the parameter will fall into this interval that is close to 0, and so it is highly likely these effects are inactive. The posterior estimated means for the parameter effects correspond to those calculated using standard methods for estimating effect sizes, but the posterior means are in fact approximately half of those calculated using standard methods. Density plots produced in WinBUGS when analysing the reactor data and assuming normal mixture prior distributions on the effect parameters indicate that there is some evidence of bimodal distributions in the case of some of the effects, and appears to occur more so in the case of inactive effects. Density plots for some of those effects displaying a bimodal distribution are given in Figure 4.8, namely β_1, β_{12}, β_{14}, β_{15} and β_{35}. This is when a Gamma$(1, 1)$ prior is placed on the precision. The density plot labelled beta[1,2] corresponds to that of β_{12} and similarly for the other two-factor interactions. For all density plots displayed, there can be seen a large peak away from 0.0, in the case of β_1 this is at approximately -0.6, and so on for the other effects, and a much smaller peak about

0.0 in all cases. This is to be expected since the priors assumed on the effects indicate there is some belief they follow a bimodal distribution.

Figure 4.8: Density Plots for β_1, β_{12}, β_{14}, β_{15} and β_{35} with normal mixture distribution on effect parameters and Gamma$(1, 1)$ prior on precision



4.4.5 Discussion

The results given in the previous subsection illustrate the impact of assuming different prior assumptions upon the analysis when analysing data in a Bayesian manner. It appears that more accurate conclusions can be drawn when assuming a more informative prior on the precision. This is due to the 95% Bayesian credible interval for each of the parameters becoming narrower as the prior on the precision becomes more informative. However, it can be seen that when a weakly informative prior is used in the case Gamma$(0.001, 0.001)$ when assuming normal or scaled-t distribution on the data and normal distribution on the effect parameters, or when a Unif$[-10, 10]$ distribution is placed on the precision, this results in a large value of the precision τ. This suggests an extremely small posterior estimate for the residual variance σ^2 since by definition the precision is the reciprocal of the variance. This leads to some questions around the choice of prior on the precision and variance, as contradicting conclusions appear to be drawn where the residual standard deviation σ is reduced as the prior on the precision become more informative, yet the precision is also reduced in the same instance.

Results obtained when making the assumption that the data are modelled by a longer tailed distribution in the case of the scaled-t distribution instead of the normal distribution, illustrate that typical assumptions such as using the normal distribution to model data do not always have to be made. This is particularly true from observing the summary statistics, given that similar results were observed whether using the scaled-t or normal distribution. It would also be expected that the 95% Bayesian credible intervals would be wider for the scaled-t distribution given that it has longer tails, and this was observed from looking at the results in two out of the three instances for the prior on the precision investigated. However, for these two instances the credible intervals were not particularly wider than for the normal distribution and so may prove reassuring to an experimenter especially when they are seeking a robust

analysis and wish to use a distribution with longer tails to account for extreme observations if they suspect their dataset of containing outliers. This shall be considered in more detail, as mentioned previously, in the following chapter.

For all prior assumptions made, the true active main effects and two-factor interactions would be identified as active due to their posterior means being larger in comparison to the other effects and so corresponds to those conclusions that would be drawn if using standard analysis such as calculating estimates of effects or using normal plots which proves reassuring to an experimenter. However, given that a credible interval can also be obtained therefore being able to state within what interval that parameter will lie, this appears to add additional information to that obtained using standard analysis. This also extends further the work of Box and Meyer (1993), where they demonstrated a method for obtaining the posterior probability of a factor being active but did not state an interval within which a particular effect would lie.

4.5 Summary

The ability to quantify the prior knowledge that an experimenter may have about effect sizes and the residual variance and incorporating this into a Bayesian analysis when analysing fractional factorial experiments has been illustrated in this chapter. It has been shown that there is some benefit to be gained in terms of not only being able to obtain a posterior mean for effect parameters monitored, and also for the variance and precision, but also being able to calculate a credible interval and thus being able to infer the 95% chance say, or any other percentage chance, that an observed parameter will lie within that interval.

In the following chapter, this method of incorporating prior knowledge into the Bayesian analysis of industrial fractional factorial experiments shall be extended further to the scenario of analysing a dataset that an experimenter may suspect of containing outliers.

Chapter 5

Outliers in Fractional Factorial Experiments

5.1 Introduction

The analysis of unreplicated fractional factorial experiments can be impacted in a disastrous way in the presence of outliers. An outlier is typically considered to be an observation for which its associated residual is significantly larger than anticipated and from the vast majority of other observations. As defined by Daniel (1960) "an outlier in a factorial experiment is an observation whose value is not in the pattern of values produced by the rest of the data". Thus we can consider this type of observation to be suspected of not being generated from the mechanism which produced the majority of observations.

Outliers are an unavoidable circumstance, they may occur as a result of temporary changes in experimental conditions and perhaps be indicated by a relatively large interaction, be due to excessive random noise, or even result from human error such

as measurement error or recording data incorrectly. There are ways to safeguard against anomalies or outliers that may result from mistakes such as incorrect data entries. It is good practice to scrutinize data in order to pick up on any obvious outliers. However, unlike fully replicated experiments where there are at least two observations for each treatment combination which provides an additional safeguard against outliers, when using small unreplicated fractional factorial experiments, it becomes difficult for an experimenter to locate any bad values, or even be able to determine the possible cause of this outlier.

Therefore, given the possibility that outliers may occur is always present and that the analysis of a dataset containing them may be distorted as a result, an analysis that is robust to outliers is sought to ensure that inferences based upon this analysis, about the parameters of interest are also robust. If an experimenter suspects their dataset of containing 'non-genuine' outliers such as those arising from human error, then one would expect the analysis to be distorted a great deal. One would expect the analysis to also be impacted upon in the case of so-called 'genuine' outliers, i.e. those resulting from extreme random noise or changes in experimental conditions. In this circumstance it may be possible to guard oneself against outliers by using longer-tailed distributions to account for the possibility of such extreme observations. Thus, in this chapter we intend to find whether any particular choice of prior assumptions about how the data or effect parameters are distributed has any impact upon the analysis, and ultimately makes the inferences more robust to outliers. This will be done by applying different models to the data, namely the normal and scaled-t, to compare the effect of using a longer-tailed distribution. Prior assumptions made about the effect parameters will also be considered where a mixture of two normal distribution will be assumed, with a high-variance and low variance component so that any extreme observations may be treated as arising from the high-variance component rather than being treated as an outlying observation.

5.2 Linear Model when outliers are considered

Before the various likelihood model assumptions on the data and prior assumptions on the effect parameters and variance being investigated are set out, the linear model assumed when accounting for outliers shall be described.

Consider the general linear model previously set out in this thesis (§2.2), i.e.

$$\mathbf{y} = \mathbf{X}\beta + \epsilon.$$

In the case when considering outliers, we shall rewrite this as

$$\mathbf{y}^* = \mathbf{X}\beta + \epsilon$$

where

$$\begin{aligned} y_1 &= y_1^* + \delta \\ y_i &= y_i^*;\ i = 2, \ldots, N. \end{aligned}$$

That is, the above case assuming that observation 1, y_1 is an outlier and all other $N-1$ observations are believed to be data that are generated from the 'true' model.

Then, given that the least squares estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y},$$

we can then find

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\boldsymbol{\beta} + \begin{bmatrix} \delta \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right) \\ &= \boldsymbol{\beta} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\begin{bmatrix} \delta \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \end{aligned}$$

This can be generalised to more than one suspected outlier, as was outlined by Abraham and Box (1978). That is, when some of the observations are suspected to have not been generated in the same manner as the majority of observations then the following linear model is considered,

$$\mathbf{y}^* = \mathbf{X}\beta + \delta\mathbf{Z} + \epsilon,$$

where $\mathbf{Z}$ is an $N \times 1$ vector whose entries are either 1 with probability α and 0 with probability $1 - \alpha$, and the amount of 'contamination' is described by δ.

Then if analysing a dataset suspected of containing outliers in a Bayesian manner and assuming the above adjusted linear model, the posterior distribution is

$$P(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{Z}) = \int P(\boldsymbol{\beta}, \delta|\boldsymbol{Z})P(\boldsymbol{y}|\boldsymbol{\beta}, \delta, \boldsymbol{Z})\, d\delta\, dy.$$

5.3 Reactor Data Revisited

5.3.1 Design fractions to be analysed

The reactor data which has previously been introduced in this thesis and analysed using Bayesian methodology in Chapter 4 will now be revisited. The half-fraction consisting of 16 runs will again be analysed, however as we are now looking at the robustness of particular model and prior assumptions to outliers, we shall adjust some of the response values to fit the criteria of having an outlier in the dataset and assess the impact of this. This half-fraction is a Resolution V design and as previously stated, has all main effects and two-factor interactions clear as none are aliased with any other main effects or two-factor interactions.

A quarter-fraction of the original 2^5 factorial design shall also be considered. Using standard ordering to allocate the first three columns of the design matrix, and the design generator $\mathbf{4 = 12}$ and $\mathbf{5 = 13}$ to allocate the final two columns, the 2^{5-2}_{III} design consisting of eight runs is achieved as displayed in Table 5.1.

Table 5.1: Design for quarter-fraction of the full 2^5 factorial design, reactor example

	Design					response (% reacted)
Run	**1**	**2**	**3**	**4 = 12**	**5 = 13**	**y**
1 (25)	-	-	-	+	+	44
2 (2)	+	-	-	-	-	53
3 (19)	-	+	-	-	+	70
4 (12)	+	+	-	+	-	93
5 (13)	-	-	+	+	-	66
6 (22)	+	-	+	-	+	55
7 (7)	-	+	+	-	-	54
8 (32)	+	+	+	+	+	82

The run numbers indicated in brackets correspond to those runs of the original 2^5

factorial design, i.e. the first run of the quarter-fraction which has all three variables at the lower level is actually run 25 of the original full factorial design. This quarter-fraction is a Resolution III design and so has no effects clear due to an even smaller fraction than the half-fraction being used. We ignore three-factor and higher order interactions as a result of the hierarchical ordering principle. Then the aliases for this design when considering only main effects and two-factor interactions are given in Table 5.2.

Table 5.2: Aliasing Structure for quarter-fraction

8 Run Design
1 = 24 = 35
2 = 14
3 = 15
4 = 12
5 = 13
23 = 45
25 = 34

Carrying out a Bayesian analysis of these eight runs and adjusting the dataset so that it contains an observation believed not to have been generated from the mechanism producing the majority of other observations, will be interesting given that this design has a high level of aliasing. An experimenter may wish to assess whether the aliasing structure of a design will have any impact upon the analysis, particularly when the aliasing structure is such that lower order, and subsequently, more important effects are aliased with one another and also if they suspect the dataset of containing some extreme observations.

5.3.2 Prior Specification

The likelihood model distributions to be assumed on the data and the prior assumptions on the effect parameters and variance to be investigated shall now be outlined. These assumptions will be made when carrying out a Bayesian analysis of the half-fraction and quarter-fraction of the reactor data and adjusting the respective 16 run and 8 run dataset in such a way that the experimenter may suspect one of the observations of being an outlier.

Before the dataset is adjusted however, we shall analyse the original datasets for the half-fraction and the quarter-fraction when assuming minimal prior knowledge. The datasets are

$$\mathbf{y} = [56, 53, 63, 65, 53, 55, 67, 61, 69, 45, 78, 93, 49, 60, 95, 82]^{\mathrm{T}}$$

and

$$\mathbf{y} = [44, 53, 70, 93, 66, 55, 54, 82]^{\mathrm{T}}$$

respectively. This is considering the situation where an experimenter may have little or no prior information about the process in hand or how the factors will behave upon the process. It is highly unlikely that an experimenter would have no knowledge at all of the scientific process that they are intending to study but we shall look at this situation in order to compare with the standard analyses. When we assume little or no prior information, the prior distribution plays a minimal role in the resulting posterior distribution and so we would expect the size of effects to correspond to those calculated using standard methods as presented in §4.4.1. We shall use this analysis of the 16 and 8 run dataset to compare with when we do adjust the dataset in order to assess the impact on the analysis when there is a suspected outlier. When we are analysing both the original half-fraction and quarter-fraction datasets and assuming minimal prior knowledge, we shall assume that the data are modelled by a

normal distribution, that the effect parameters are represented by independent prior normal distributions and that the precision is represented by a mildly informative Gamma$(1, 1)$ distribution. This will be with the same parameter values as set out in the previous chapter (§4.4.2). That is, we shall assume that the data takes the following form,

$$y_i = \mu_i + \epsilon_i$$

where

$$\mu_i = \beta_0 + \sum_{i=1}^{5} \beta_i x_i + \sum \sum_{i<j}^{5} \beta_{ij} x_i x_j,$$

the errors are modelled by a normal distribution as follows,

$$\epsilon_i \sim N(\mu, \tau)$$

and the intercept is represented by the following prior

$$\beta_0 \sim N(0, 10^{-\infty}).$$

Finally, all the main effects and two-factor interactions are represented by

$$\beta_i, \beta_{ij} \sim N(0, 10^{-3})$$

and the prior distribution on the precision is

$$\tau \sim Gamma(1, 1).$$

Thus, in the case of analysing the 16 run dataset the resulting summary statistics will just be those as displayed in Chapter 4 in Table 4.9. When analysing the 8 run dataset, for both the original dataset and then when considering an outlier, only the intercept and main effects shall be estimated due to there not being enough degrees of freedom to estimate all two-factor interactions. Thus, the prior specifications set out for the 8 run case shall just be the prior distributions on the intercept, main effects and variance.

When analysing the datasets when they are suspected of containing an outlier, the prior assumptions to be investigated for the model on the errors will be both a normal distribution and scaled-t distribution. This will be to assess whether using a distribution from a longer-tailed family of distributions, such as the scaled-t, has any impact on the resulting analysis when an experimenter is concerned the dataset contains an extreme observation. In the case of the scaled-t distribution, by varying the degrees of freedom from small to large we can also assess sensitivity to the normal model assumption. Taking the degrees of freedom to be small, would be tending to a much heavier tailed distribution and in the extreme case of t_1 would in fact be the Cauchy distribution, whereas if the degrees of freedom are at the opposite extreme and much larger, say t_{20}, we would expect the distribution to tend towards the normal distribution. It will be of interest to assess how the analysis is impacted upon as the degrees of freedom vary. The case where $\nu = 1$ will not be investigated in this instance since the assumption that is to be made in this case is that the prior distribution will have infinite variance which is not representative of what would happen in practice in experimentation. Also, the Cauchy distribution is not able to be implemented in WinBUGS. Varying the prior distribution on the effect parameters will also be investigated, looking at both independent normal distributions on the parameters, and then focusing on a mixture of normal distribution to assess whether any difference in the analysis is found. Throughout, a mildly informative Gamma$(1, 1)$ distribution is to be assumed on the precision and a normal distribution used to represent the intercept, β_0. A summary of the prior specifications just mentioned is set out in Table 5.3.

Table 5.3: Prior Distributional Assumptions

	Distribution	Parameter Values
Model for errors	Normal	$N(\mu_i, \tau)$
	Scaled-t	$t_\nu(\mu_i, \tau)$; $\nu = 2, 5, 10, 20, 50$
Intercept	Normal	$N(0, 10^{-\infty})$
Main Effects and two-factor interactions	Normal	$N(0, 10^{-3})$
	Mixture of normals	$N(0, \tau_1)$ if effect is inactive
		$N(0, \tau_2)$ if effect is active
		where $\tau_1 = \frac{1}{\sigma_1^2}$, $\tau_2 = \frac{1}{\sigma_2^2}$
		and $\sigma_1 \sim$ Unif $(0, 1)$, $\sigma_2 \sim$ Unif $(0, 10)$
		$p_0 = 0.85$ (probability effect is inactive)
		$p_1 = 0.15$ (probability effect is active)
Residual Variance		$\tau \sim$ Gamma $(1, 1)$

As mentioned, when looking at both the half-fraction and quarter-fraction, the dataset will be adjusted in order to fit the criterion that an experimenter would suspect the dataset of containing an outlier. This will be achieved firstly, by changing one of the observations in a very obvious and rudimentary way and treating one of the observations as if it had been affected by human error and the observation was the result of measurement error on an incorrect data entry. For example, in the case of the 16 run dataset, the first observation is 56, however we will adjust this value to 156. Again, with the 8 run dataset the first observation 44 shall be adjusted to 144. This rather basic way of adjusting the dataset in order to satisfy the criteria that the dataset contains an outlier means that the adjusted observation does stand out quite clearly from the rest just from looking at the dataset. One would hope that by scrutinizing the dataset before any analysis methods are carried out, an experimenter would pick up on this obvious outlier. However, we shall initially consider this scenario for illustration purposes and in order to observe the impact, if any, on the analysis and whether any differences are seen in the analysis depending on the prior assumptions made.

We shall then consider a more subtle sensitivity analysis where we shall adjust a particular observation in the dataset by increasing or decreasing the response by slightly greater than 3σ and assessing whether any impact is seen in the resulting analysis. In this case we shall assume that the variance and therefore the standard deviation σ is known. The value of 3σ is to be considered due to the *empirical rule* (Wackerly et al., 2002). This states that given some population which has population mean μ, and population standard deviation σ for a distribution that is approximately normal, it follows that the interval with endpoints

$\mu \pm \sigma$ contains approximately 68% of the observations,
$\mu \pm 2\sigma$ contains approximately 95% of the observations,
$\mu \pm 3\sigma$ contains almost all of the observations.

Thus, it shall be of interest to note whether the analysis differs greatly when assuming the data are normally distributed and changing an observation from y to the new adjusted response y_{adj} where y_{adj} is increased or decreased by slightly more than 3σ. This is given that assuming the distribution is approximately normal, it would be considered that nearly all observations would lie within $\pm 3\sigma$ of the population mean. In the case when assuming the data are modelled by the scaled-t distribution, it will be of interest to note whether changing the degrees of freedom from small to large also has any impact on the analysis when adjusting a response y to $y \pm 3\sigma$, given that as the degrees of freedom approach infinity, the distribution tends to the normal distribution.

5.3.3 Results

The summary statistics obtained when using WinBUGS to analyse a half-fraction and quarter-fraction of the reactor data are now presented. 20000 iterations were

run for two chains, thinning by storing only every 10^{th} simulation and consequently this was found to be a sufficient number of simulations in order for convergence to be achieved. MCMC diagnostic checks such as looking at autocorrelation function, Brooks-Gelman-Rubin and trace plots also confirmed that convergence had been reached and that the two chains were well mixed.

Results from analysis of original data

Table 5.4: Summary statistics from WinBUGS after fitting normal model to reactor data, with normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	65.25	0.6356	0.003984	64.19	65.25	66.31
β_1	−1.003	0.6619	0.003903	−2.042	−1.002	0.03457
β_2	10.25	0.6342	0.004096	9.178	10.25	11.29
β_3	0.00129	0.657	0.003687	−1.057	0.002397	1.087
β_4	6.122	0.6144	0.003484	5.059	6.125	7.182
β_5	−3.133	0.6196	0.003368	−4.2	−3.128	−2.106
β_{12}	0.7525	0.6234	0.003619	−0.3042	0.7533	1.827
β_{13}	0.2535	0.629	0.003502	−0.8051	0.2557	1.301
β_{14}	−0.3823	0.6413	0.00376	−1.478	−0.3807	0.6814
β_{15}	0.625	0.6495	0.003875	−0.4409	0.6282	1.701
β_{23}	0.7493	0.6313	0.003982	−0.3068	0.7497	1.819
β_{24}	5.373	0.6247	0.003555	4.286	5.372	6.437
β_{25}	0.6226	0.6814	0.004071	−0.4383	0.6233	1.671
β_{34}	0.1211	0.6287	0.003626	−0.9284	0.124	1.154
β_{35}	1.125	0.6251	0.003446	0.05736	1.125	2.187
β_{45}	−4.744	0.638	0.003664	−5.814	−4.746	−3.67
σ	1.713	1.906	0.02265	0.518	1.205	6.086
τ	1.001	1.004	0.008064	0.0271	0.6882	3.73

Table 5.5: Summary statistics from WinBUGS after fitting normal model to reactor data, with normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-2} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	64.63	4.716	0.02861	55.14	64.63	74.14
β_1	6.025	4.613	0.02537	−3.443	6.067	15.19
β_2	9.899	4.6	0.02658	0.528	9.959	19.07
β_3	−0.3851	4.586	0.02781	−9.657	−0.3819	8.913
β_4	6.469	4.647	0.02778	−2.918	6.511	15.83
β_5	−1.847	4.586	0.0256	−11.04	−1.861	7.601
σ	12.19	5.498	0.03003	6.025	10.85	26.59
τ	0.01008	0.006932	0.02601	0.001416	0.008497	0.02755

Results from analysis of data with obvious outlier

Table 5.6: Summary statistics from WinBUGS after fitting normal model to reactor data, with normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	71.5	0.6357	0.003985	70.44	71.5	72.56
β_1	−7.25	0.6639	0.003903	−8.283	−7.251	−6.21
β_2	4.003	0.63484	0.004084	2.936	4.004	5.039
β_3	−6.246	0.6568	0.003674	−7.296	−6.247	−5.156
β_4	−0.1258	0.615	0.003463	−1.183	−0.1243	0.943
β_5	3.114	0.6199	0.003364	2.044	3.121	4.136
β_{12}	7.0	0.6239	0.003627	5.939	7.003	8.068
β_{13}	6.501	0.6311	0.003538	5.435	6.505	7.548
β_{14}	5.865	0.6441	0.00381	4.759	5.869	6.921
β_{15}	−5.622	0.6465	0.003824	−6.681	−5.621	−4.54
β_{23}	6.997	0.6315	0.00397	5.938	6.999	8.063
β_{24}	11.62	0.6245	0.003556	10.52	11.62	12.68
β_{25}	−5.625	0.6804	0.004043	−6.681	−5.626	−4.574
β_{34}	6.369	0.6286	0.003651	5.312	6.373	7.397
β_{35}	−5.122	0.6253	0.003457	−6.188	−5.124	−4.055
β_{45}	−10.99	0.6388	0.003677	−12.06	−10.99	−9.909
σ	1.714	1.906	0.02271	0.518	1.205	6.055
τ	1.001	1.004	0.008063	0.02734	0.6886	3.729

Table 5.7: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 2$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	71.5	1.322	0.007329	68.99	71.51	73.95
β_1	−7.229	1.283	0.006677	−9.604	−7.242	−4.715
β_2	3.984	1.282	0.007395	1.512	3.994	6.379
β_3	−6.238	1.33	0.0075744	−8.685	−6.245	−3.742
β_4	−0.1305	1.307	0.007918	−2.635	−0.1275	2.321
β_5	3.111	1.3	0.007752	0.6177	3.123	5.525
β_{12}	6.986	1.318	0.007834	4.491	6.996	9.462
β_{13}	6.466	1.309	0.007607	3.881	6.488	8.841
β_{14}	5.856	1.284	0.007286	3.328	5.864	8.271
β_{15}	−5.602	1.304	0.007372	−8.042	−5.62	−3.108
β_{23}	6.975	1.307	0.007373	4.373	6.999	9.437
β_{24}	11.61	1.278	0.007791	9.024	11.62	13.97
β_{25}	−5.62	1.329	0.007729	−8.088	−5.629	−3.125
β_{34}	6.361	1.321	0.007806	3.83	6.375	8.821
β_{35}	−5.121	1.295	0.007702	−7.562	−5.122	−2.64
β_{45}	−10.98	1.292	0.007416	−13.44	−11.0	−8.432
σ	1.623	1.549	0.01713	0.5173	1.186	5.545
τ	1.025	1.015	0.00789	0.03258	0.7116	3.738

Table 5.8: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 5$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	71.5	0.7671	0.004424	70.17	71.5	72.84
β_1	−7.245	0.7685	0.00452	−8.557	−7.248	−5.898
β_2	3.997	0.7854	0.004378	2.624	3.998	5.319
β_3	−6.238	0.7735	0.00462	−7.51	−6.248	−4.881
β_4	−0.1281	0.7765	0.004619	−1.479	−0.1233	1.207
β_5	3.125	0.7641	0.004454	1.822	3.125	4.469
β_{12}	6.989	0.8245	0.004962	5.624	6.997	8.334
β_{13}	6.505	0.7624	0.004398	5.166	6.503	7.832
β_{14}	5.868	0.7784	0.004241	4.507	5.875	7.179
β_{15}	−5.612	0.7794	0.004126	−6.923	−5.62	−4.264
β_{23}	6.992	0.8104	0.004474	5.592	6.997	8.323
β_{24}	11.62	0.8002	0.004401	10.25	11.63	12.99
β_{25}	−5.617	0.7894	0.004497	−6.973	−5.623	−4.26
β_{34}	6.375	0.7736	0.004396	5.048	6.378	7.706
β_{35}	−5.128	0.7702	0.004578	−6.471	−5.127	−3.781
β_{45}	−10.99	0.7916	0.004745	−12.31	−11.0	−9.601
σ	1.68	1.839	0.01984	0.5198	1.189	5.828
τ	1.017	1.014	0.007969	0.02951	0.7068	3.703

Table 5.9: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 10$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	71.5	0.6619	0.003833	70.33	71.5	72.66
β_1	−7.255	0.6563	0.004232	−8.386	−7.251	−6.148
β_2	3.999	0.6679	0.003828	2.827	4.003	5.148
β_3	−6.249	0.6662	0.003704	−7.379	−6.248	−5.12
β_4	−0.1256	0.6821	0.004119	−1.278	−0.1281	1.059
β_5	3.125	0.662	0.003913	2.027	3.121	4.271
β_{12}	6.993	0.6649	0.003801	5.794	6.997	8.136
β_{13}	6.503	0.6753	0.004068	5.344	6.501	7.662
β_{14}	5.873	0.6623	0.003911	4.738	5.873	7.019
β_{15}	−5.624	0.6602	0.003645	−6.789	−5.623	−4.495
β_{23}	6.996	0.6829	0.003745	5.848	6.999	8.139
β_{24}	11.62	0.6665	0.003975	10.47	11.62	12.77
β_{25}	−5.626	0.6518	0.004004	−6.758	−5.627	−4.489
β_{34}	6.369	0.6762	0.003411	5.221	6.374	7.506
β_{35}	−5.11	0.6517	0.003823	−6.236	−5.119	−3.949
β_{45}	−10.99	0.7161	0.004327	−12.14	−11.0	−9.846
σ	1.662	1.779	0.01938	0.5202	1.19	5.653
τ	1.013	1.004	0.007988	0.03133	0.706	3.696

Table 5.10: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 20$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	71.51	0.6635	0.003914	70.38	71.5	72.67
β_1	−7.247	0.6763	0.003731	−8.349	−7.249	−6.125
β_2	3.997	0.6653	0.003725	2.882	3.999	5.117
β_3	−6.246	0.6473	0.003808	−7.351	−6.253	−5.103
β_4	−0.1197	0.6351	0.003708	−1.221	−0.1239	0.9746
β_5	3.128	0.7023	0.003821	2.001	3.126	4.271
β_{12}	6.994	0.652	0.004149	5.866	6.998	8.109
β_{13}	6.495	0.652	0.003841	5.396	6.499	7.59
β_{14}	5.868	0.6565	0.003848	4.712	5.873	6.962
β_{15}	−5.622	0.6627	0.004088	−6.721	−5.626	−4.498
β_{23}	6.99	0.6513	0.003387	5.875	6.996	8.095
β_{24}	11.62	0.6748	0.003894	10.47	11.62	12.71
β_{25}	−5.629	0.6489	0.004001	−6.736	−5.628	−4.546
β_{34}	6.369	0.645	0.003562	5.257	6.374	7.451
β_{35}	−5.125	0.6753	0.003696	−6.24	−5.128	−4.002
β_{45}	−10.99	0.6698	0.003468	−12.11	−11.0	−9.853
σ	1.703	1.904	0.02102	0.5182	1.195	6.075
τ	1.01	1.003	0.007961	0.02711	0.7003	3.726

Table 5.11: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 50$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	71.5	0.7144	0.003882	70.41	71.5	72.57
β_1	−7.247	0.6698	0.00369	−8.324	−7.251	−6.148
β_2	3.998	0.6867	0.004195	2.927	3.999	5.034
β_3	−6.25	0.7241	0.004277	−7.333	−6.251	−5.156
β_4	−0.1151	0.6988	0.003634	−1.182	−0.1227	0.9919
β_5	3.123	0.6896	0.003897	2.022	3.123	4.196
β_{12}	6.989	0.6927	0.00373	5.911	6.996	8.061
β_{13}	6.486	0.7416	0.003974	5.365	6.498	7.556
β_{14}	5.87	0.7217	0.004264	4.772	5.873	6.953
β_{15}	−5.62	0.6653	0.003569	−6.691	−5.623	−4.542
β_{23}	6.994	0.6676	0.003957	5.884	7.0	8.061
β_{24}	11.62	0.6977	0.004393	10.53	11.62	12.69
β_{25}	−5.62	0.7183	0.003946	−6.695	−5.622	−4.539
β_{34}	6.378	0.7448	0.004002	5.277	6.379	7.487
β_{35}	−5.127	0.6973	0.003966	−6.26	−5.126	−4.064
β_{45}	−11.0	0.7311	0.004196	−12.05	−11.0	−9.899
σ	1.725	2.194	0.02476	0.5187	1.186	6.275
τ	1.016	1.01	0.007723	0.02543	0.7113	3.717

Table 5.12: Summary statistics from WinBUGS after fitting normal model to reactor data, with normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-2} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	77.13	7.171	0.04348	62.72	77.13	91.65
β_1	−6.025	6.848	0.03766	−19.69	−6.083	7.954
β_2	−2.281	6.826	0.03922	−15.87	−2.308	11.67
β_3	−12.3	6.836	0.04139	−25.75	−12.42	1.916
β_4	18.2	6.929	0.04161	3.726	18.4	31.77
β_5	10.11	6.831	0.03826	−4.09	10.22	23.66
σ	18.71	7.948	0.04288	9.416	16.82	39.26
τ	0.004163	0.002802	0.01044	0.1189	0.003536	0.01128

Table 5.13: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 2$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-2} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	76.7	7.553	0.06147	64.26	76.21	89.53
β_1	−5.864	7.413	0.05761	−18.39	−5.381	6.421
β_2	−1.379	7.301	0.05819	−14.04	−0.2779	10.33
β_3	−11.72	7.32	0.0546	−24.46	−10.63	0.135
β_4	17.8	7.381	0.05794	5.733	16.57	30.57
β_5	9.509	7.348	0.0554	−2.334	8.399	22.28
σ	6.565	4.917	0.03259	1.465	5.231	19.32
τ	0.08458	0.1498	0.001023	0.00268	0.03654	0.4673

Table 5.14: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 5$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-2} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	77.04	7.246	0.04068	62.97	77.02	91.13
β_1	−6.056	7.04	0.04	−19.47	−6.132	7.746
β_2	−1.995	7.009	0.04343	−15.73	−1.979	11.77
β_3	−12.09	6.991	0.04099	−25.53	−12.11	1.616
β_4	18.18	6.982	0.04103	4.072	18.24	31.4
β_5	9.963	6.985	0.03907	−3.674	9.956	23.39
σ	13.78	7.04	0.0394	5.159	12.22	31.73
τ	0.009805	0.01138	0.04359	0.1823	0.006698	0.03759

Table 5.15: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 10$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-2} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	77.09	7.198	0.04102	62.83	77.07	91.31
β_1	−6.078	6.903	0.04141	−19.75	−6.135	8.039
β_2	−2.176	6.83	0.0383	−15.6	−2.205	11.64
β_3	−12.22	6.816	0.03874	−25.53	−12.33	1.749
β_4	18.27	6.942	0.03555	3.787	18.44	31.58
β_5	9.991	6.822	0.04134	−3.817	10.05	23.39
σ	16.47	7.617	0.04387	7.602	14.67	36.33
τ	0.005786	0.00449	0.01716	0.1388	0.004646	0.01731

Table 5.16: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 20$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-2} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	77.05	7.215	0.04046	62.47	77.03	91.49
β_1	−6.101	6.826	0.0424	−19.56	−6.195	7.791
β_2	−2.268	6.895	0.04231	−16.21	−2.28	11.84
β_3	−12.2	6.896	0.03907	−25.63	−12.38	2.145
β_4	18.16	6.982	0.04137	3.721	18.33	31.55
β_5	10.1	6.918	0.04048	−4.199	10.16	23.67
σ	17.66	7.769	0.04804	8.621	15.76	38.01
τ	0.004808	0.003393	0.01310	0.127	0.004025	0.01346

Table 5.17: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 50$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-2} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	77.03	7.296	0.04321	62.29	77.01	91.54
β_1	−6.042	6.872	0.03963	−19.74	−6.097	8.05
β_2	−2.159	6.884	0.03986	−16.08	−2.182	11.97
β_3	−12.29	6.822	0.03944	−25.48	−12.42	1.829
β_4	18.24	6.863	0.03844	3.994	18.41	31.56
β_5	10.1	6.802	0.03775	−3.858	10.23	23.46
σ	18.27	7.915	0.04696	9.003	16.37	39.01
τ	0.004429	0.003065	0.01265	0.120	0.00373	0.01234

Table 5.18: Summary statistics from WinBUGS after fitting normal model to reactor data, with mixture of normal distributions on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	71.49	6.445	0.03634	58.66	71.49	84.36
β_1	−0.5093	1.951	0.04756	−7.239	−0.05273	1.396
β_2	0.3596	1.816	0.02642	−2.103	0.03524	5.459
β_3	−0.4208	1.785	0.03884	−6.264	−0.04346	1.45
β_4	−0.01606	1.591	0.008909	−3.501	−0.001923	3.315
β_5	0.1929	1.468	0.01622	−1.825	0.01907	3.769
β_{12}	0.4844	1.852	0.04536	−1.355	0.05081	6.967
β_{13}	0.4621	1.827	0.04274	−1.389	0.04652	6.549
β_{14}	0.4052	1.719	0.03844	−1.485	0.04288	5.93
β_{15}	−0.3874	1.679	0.03799	−5.733	−0.03683	1.509
β_{23}	0.6615	2.171	0.04565	−1.533	0.06742	7.367
β_{24}	1.107	2.987	0.07355	−1.19	0.111	11.5
β_{25}	−0.5453	2.046	0.03743	−6.648	−0.0498	1.702
β_{34}	0.4318	1.762	0.03948	−1.429	0.04136	6.375
β_{35}	−0.3427	1.621	0.03223	−5.228	−0.03534	1.544
β_{45}	−0.9772	2.828	0.0648	−10.92	−0.1001	1.265
σ	25.22	5.929	0.1457	14.97	24.96	37.22
τ	0.01696	0.1794	0.004837	0.132	0.001605	0.004479

Table 5.19: Summary statistics from WinBUGS after fitting normal model to reactor data, with mixture of normal distributions on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-2} design

Node	Mean	sd	MC Error	2.5%	Median	97.5%
β_0	77.04	10.38	0.05573	56.34	77.01	97.66
β_1	−0.242	1.907	0.01131	−5.453	−0.01761	2.055
β_2	−0.07575	1.73	0.01099	−3.849	−0.006283	2.706
β_3	−0.6794	2.677	0.01731	−9.718	−0.04969	1.63
β_4	1.655	4.301	0.03229	−1.439	0.1164	15.68
β_5	0.4813	2.297	0.01458	−1.723	0.03841	7.843
σ	28.35	8.075	0.04247	16.65	27.0	47.98
τ	0.001539	0.157	0.01233	0.0796	0.001371	0.003608

Results from analysis of data with $y_{adj} > y + 3\sigma$

Table 5.20: Summary statistics from WinBUGS after fitting normal model to reactor data, with normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	Median	95% Credible Interval
β_0	68.31	68.31	$(67.25, 69.38)$
β_1	-4.064	-4.064	$(-5.1, 3.025)$
β_2	7.189	7.191	$(6.122, 8.224)$
β_3	-3.06	-3.06	$(-4.111, -1.972)$
β_4	3.06	3.063	$(2.001, 4.127)$
β_5	-0.07196	-0.0662	$(-1.14, 0.951)$
β_{12}	3.814	3.815	$(2.755, 4.888)$
β_{13}	3.315	3.318	$(2.252, 4.361)$
β_{14}	2.679	2.681	$(1.578, 3.738)$
β_{15}	-2.436	-2.434	$(-3.496, -1.357)$
β_{23}	3.811	3.812	$(2.753, 4.88)$
β_{24}	8.435	8.433	$(7.343, 9.498)$
β_{25}	-2.439	-2.439	$(-3.497, -1.391)$
β_{34}	3.182	3.186	$(2.13, 4.214)$
β_{35}	-1.936	-1.937	$(-3.002, -0.8729)$
β_{45}	-7.805	-7.808	$(-8.873, -6.725)$
σ	1.713	1.205	$(0.5178, 6.069)$
τ	1.001	0.6884	$(0.02716, 3.73)$

Table 5.21: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 2$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	Median	95% Credible Interval
β_0	68.31	68.32	$(65.83, 70.76)$
β_1	-4.048	-4.056	$(-6.428, -1.438)$
β_2	7.165	7.18	$(4.686, 9.543)$
β_3	-3.057	-3.059	$(-5.508, -0.5996)$
β_4	3.051	3.058	$(0.5496, 5.499)$
β_5	-0.07096	-0.06294	$(-2.524, 2.358)$
β_{12}	3.805	3.81	$(1.332, 6.301)$
β_{13}	3.284	3.302	$(0.7294, 5.66)$
β_{14}	2.675	2.678	$(0.1628, 5.108)$
β_{15}	-2.42	-2.435	$(-4.871, 0.04949)$
β_{23}	3.793	3.814	$(1.219, 6.258)$
β_{24}	8.427	8.431	$(5.867, 10.79)$
β_{25}	-2.438	-2.443	$(-4.92, 0.0462)$
β_{34}	3.18	3.19	$(0.6606, 5.648)$
β_{35}	-1.939	-1.936	$(-4.387, 0.5058)$
β_{45}	-7.797	-7.812	$(-10.27, -5.268)$
σ	1.621	1.185	$(0.5173, 5.547)$
τ	1.026	0.7118	$(0.03261, 3.737)$

Table 5.22: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 5$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	Median	95% Credible Interval
β_0	68.32	68.31	$(66.98, 69.65)$
β_1	-4.059	-4.061	$(-5.373, -2.715)$
β_2	7.183	7.185	$(5.808, 8.499)$
β_3	-3.053	-3.062	$(-4.323, -1.698)$
β_4	3.058	3.064	$(1.706, 4.389)$
β_5	-0.06057	-0.06212	$(-1.364, 1.29)$
β_{12}	3.803	3.811	$(2.445, 5.154)$
β_{13}	3.32	3.316	$(1.984, 4.651)$
β_{14}	2.682	2.688	$(1.325, 3.994)$
β_{15}	-2.427	-2.433	$(-3.734, -1.083)$
β_{23}	3.806	3.81	$(2.409, 5.137)$
β_{24}	8.438	8.439	$(7.07, 9.814)$
β_{25}	-2.432	-2.436	$(-3.797, -1.08)$
β_{34}	3.19	3.191	$(1.865, 4.521)$
β_{35}	-1.942	-1.94	$(-3.297, 0.6028)$
β_{45}	-7.802	-7.814	$(-9.132, -6.414)$
σ	1.679	1.189	$(0.5196, 5.831)$
τ	1.017	0.7069	$(0.02942, 3.706)$

Table 5.23: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 10$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	Median	95% Credible Interval
β_0	68.31	68.31	$(67.14, 69.47)$
β_1	-4.069	-4.064	$(-5.202, -2.964)$
β_2	7.185	7.19	$(6.012, 8.326)$
β_3	-3.063	-3.061	$(-4.194, -1.938)$
β_4	3.06	3.059	$(1.906, 4.239)$
β_5	-0.06069	-0.06541	$(-1.157, 1.088)$
β_{12}	3.807	3.81	$(2.609, 4.951)$
β_{13}	3.317	3.314	$(2.16, 4.477)$
β_{14}	2.687	2.686	$(1.555, 3.834)$
β_{15}	-2.438	-2.436	$(-3.605, -1.313)$
β_{23}	3.81	3.812	$(2.664, 4.952)$
β_{24}	8.436	8.436	$(7.291, 9.591)$
β_{25}	-2.44	-2.439	$(-3.574, -1.303)$
β_{34}	3.183	3.186	$(2.041, 4.322)$
β_{35}	-1.924	-1.932	$(-3.052, -0.7659)$
β_{45}	-7.809	-7.811	$(-8.955, -6.663)$
σ	1.661	1.19	$(1.773, 5.64)$
τ	1.013	0.706	$(1.005, 3.696)$

Table 5.24: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 20$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	Median	95% Credible Interval
β_0	68.32	68.31	$(67.19, 69.49)$
β_1	-4.061	-4.062	$(-5.164, -2.939)$
β_2	7.183	7.186	$(6.066, 8.297)$
β_3	-3.06	-3.066	$(-4.166, -1.918)$
β_4	3.066	3.063	$(1.964, 4.159)$
β_5	-0.05841	-0.06143	$(-1.184, 1.089)$
β_{12}	3.808	3.811	$(2.684, 4.925)$
β_{13}	3.309	3.311	$(2.213, 4.407)$
β_{14}	2.681	2.686	$(1.53, 3.779)$
β_{15}	-2.436	-2.439	$(-3.536, -1.314)$
β_{23}	3.804	3.809	$(2.691, 4.914)$
β_{24}	8.429	8.434	$(7.292, 9.532)$
β_{25}	-2.443	-2.441	$(-3.554, -1.363)$
β_{34}	3.183	3.187	$(2.077, 4.266)$
β_{35}	-1.939	-1.941	$(-3.057, -0.8163)$
β_{45}	-7.807	-7.81	$(-8.922, -6.667)$
σ	1.702	1.195	$(0.5183, 6.067)$
τ	1.01	0.7007	$(0.02718, 3.723)$

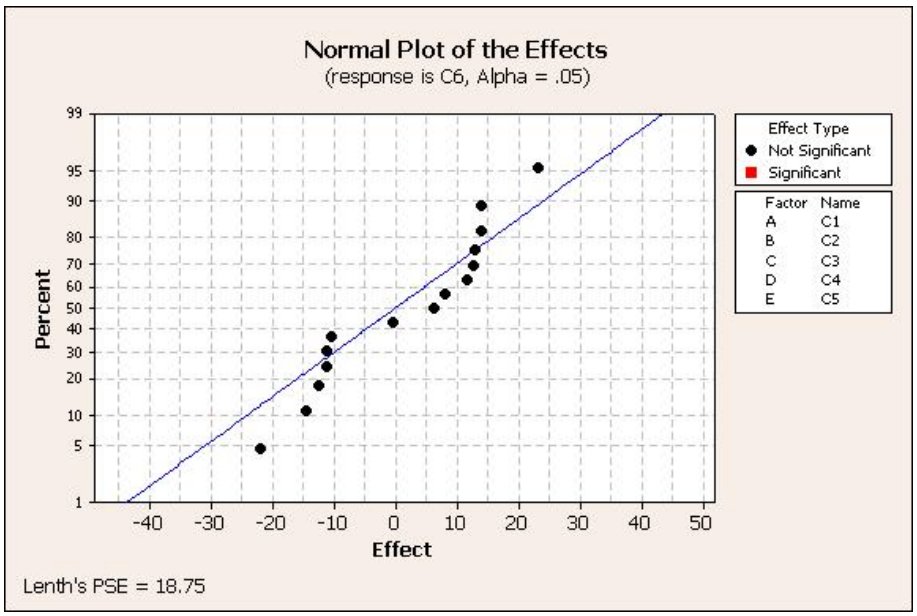
Table 5.25: Summary statistics from WinBUGS after fitting scaled-t model to reactor data, with $\nu = 50$, normal distribution on effect parameters and Gamma$(1, 1)$ prior on precision for 2^{5-1} design

Node	Mean	Median	95% Credible Interval
β_0	68.31	68.31	$(67.22, 69.39)$
β_1	-4.061	-4.063	$(-5.138, -2.965)$
β_2	7.184	7.186	$(6.113, 8.217)$
β_3	-3.064	-3.064	$(-4.149, -1.972)$
β_4	3.071	3.064	$(2.002, 4.173)$
β_5	-0.06302	-0.06389	$(-1.162, 1.014)$
β_{12}	3.803	3.809	$(2.73, 4.881)$
β_{13}	3.3	3.311	$(2.183, 4.373)$
β_{14}	2.685	2.686	$(1.589, 3.771)$
β_{15}	-2.434	-2.436	$(-3.506, -1.355)$
β_{23}	3.808	3.813	$(2.705, 4.876)$
β_{24}	8.436	8.437	$(7.344, 9.505)$
β_{25}	-2.434	-2.435	$(-3.511, -1.354)$
β_{34}	3.193	3.192	$(2.094, 4.303)$
β_{35}	-1.941	-1.938	$(-3.075, -0.8789)$
β_{45}	-7.813	-7.813	$(-8.872, -6.714)$
σ	1.724	1.186	$(0.5187, 6.274)$
τ	1.016	0.7109	$(0.02542, 3.718)$

The results in Table 5.4 are the same as those given previously in Chapter 4, in Table 4.9. That is, those results when carrying out a Bayesian analysis of a half-fraction of the reactor data assuming normal prior distributions on the model and effect parameters and a Gamma$(1, 1)$ distribution on the precision. The results correspond with those results as given by the standard analysis, which is that the true active effects **2**, **4**, **5**, **24** and **45** have much larger posterior means in comparison to the other effects and so would be identified as active. The 95% Bayesian credible intervals also correctly indicate that the other effects are inactive due to the intervals being close to 0. Table 5.5 gives the results when carrying out a Bayesian analysis with the same prior assumptions as specified for the 16 run experiment, however this time when analysing a quarter-fraction of the reactor data. The summary statistics indicate some interesting results. The posterior mean and median for the intercept are very close to the true mean as calculated from the standard analysis, however despite the summary statistics resulting from the analysis of the original data, i.e. there are no outliers, the posterior means and medians do differ quite greatly for some of the main effects. It can be seen that along with main effects **2** and **4**, main effect **1** also has a large posterior estimated mean although this effect is not considered to be active from previous analyses. Conversely, where main effect **5** has been identified as active from previous analyses, in this case it only has a posterior mean of -1.847. It is interesting to observe that the 95% Bayesian credible intervals have greatly increased for all effects, even in the case of those true inactive effects. For example, the credible interval for the parameter β_3 is $(-1.057, 1.087)$ when analysing the original data from the 16 run experiment. However, when analysing the data from the 8 run experiment instead, the credible interval widely increases to $(-9.657, 8.913)$. Thus, it is not clear which effects are truly active or inactive and demonstrates the perils of using such a small fraction for experimentation due to the high level of aliasing involved. It should also be noted that the value of the precision τ is very small with a value of 0.01008 which may be indicative of an overinflated variance.

The summary statistics presented in Table 5.6 are those from the analysis when analysing the half-fraction with one of the observations changed to an obvious outlier and assuming normal distributions on both the model for the data and the effect parameters. It can be seen that the posterior estimated mean for the intercept has been greatly affected by the outlier which has increased from 65.25 in the previous analysis, to a much greater value of 71.5. This is perhaps unsurprising as models based upon the normal distribution are known to be particularly nonrobust to outliers, where a single outlying observation can strongly impact on the inference. It is also known that the population mean can also be unduly affected by an outlier and so the posterior median may represent the data more accurately. However, it can be seen that for these particular summary statistics the posterior median also has a much larger value of 71.5. It is also worth noting that all the effects have relatively large posterior means except for the main effect **4**, which is actually one of the true active effects. The 95% Bayesian credible interval has also shifted greatly, where from the analysis of the original data the credible interval is $(5.059, 7.182)$, in the case of this dataset with an obvious outlier the credible interval is $(-1.183, 0.943)$ and now contains 0. Therefore, one may conclude incorrectly based upon this credible interval that this parameter is actually inactive. Figure 5.1 displays the normal plot for this data with the obvious outlier and it can clearly be seen that no effects are distinguished as significant.

Figure 5.1: Normal Plot for half-fraction of reactor data with obvious outlier



When replacing the normal distributional assumption for the model on the data by the scaled-t and varying the degrees of freedom from small to large it can be seen that the summary statistics tend to those when assuming a normal distribution on the data. This appears to indicate that as the degrees of freedom become larger and approach infinity, the distribution is actually tending to the normal distribution because the posterior means, medians and credible intervals become closer to those for the normal distribution on the data, i.e. those results displayed in Table 5.6. It also appears that as the degrees of freedom become larger the 95% Bayesian credible intervals tend to become narrower for the intercept and effect parameters, and the precision generally decreases indicating that the variance becomes more inflated. For all of the various degrees of freedom investigated, again it is seen that the posterior estimated mean and median for the intercept is greatly affected by the outlying observation and also that as before all effects have large posterior estimates except for main effect **4**.

When analysing the quarter-fraction with one of the observations changed to an extreme value and assuming normal distributions on the data and the effect parameters (Table 5.12) it is again seen that the posterior estimated mean and median for the intercept has been impacted upon greatly. The 95% Bayesian credible intervals have also shifted greatly in comparison to those from the analysis of the original data. The results also indicate, as similarly noted for the half-fraction, that when replacing the normal distribution model for the data with a scaled-t model and increasing the degrees of freedom from small to large, that is looking at Tables 5.13 - 5.17, the summary statistics appear to tend towards those for the normal model on the data. As with the analysis of the original quarter-fraction without an outlier, all the various prior distributions investigated when analysing the quarter-fraction with an outlier in the dataset indicate that the precision is very small.

When assuming a mixture of normal distributions on the effect parameters, and analysing the datasets with an obvious outlier, for both the half-fraction and the quarter-fraction (Tables 5.18 and 5.19) the posterior mean and median for the intercept is still affected as with previous cases where it has increased, however now all posterior estimates for effects are reduced greatly. This is for both main effects and two-factor interactions in the case of the half-fraction, and all main effects for the quarter-fraction. Also, the precision τ is reduced and the standard deviation has increased somewhat in comparison to when assuming normal distributions on the effect parameters. The 95% Bayesian credible intervals are also much wider for the intercept and effect parameters.

Only the half-fraction was considered when adjusting the dataset so that an observation was increased by slightly more than 3σ. The summary statistics for the various prior distributional assumptions made for this scenario are presented in Tables 5.20 - 5.25. For all cases the estimated posterior mean and median for the intercept has not been unduly affected by the observation y_{adj}, considered to be an outlier based

upon the empirical rule. It should be noted that the mean and median are either identical or very nearly the same when comparing the same effect parameter, for all prior assumptions made. However, it should be noted that the effect parameters have been affected by the adjusted value y_{adj} as all effect parameters appear to be large or have increased from what their true effect is, even those which are not truly active. The only effect which is not large in all cases is the main effect **5**, which is actually one of the true active main effects. All of the credible intervals have also shifted and none of these intervals contains the true posterior mean which clearly demonstrates the impact of the outlier upon the analysis. It is also seen that when assuming the data are modelled by a scaled-t distribution, when increasing the degrees of freedom from small to large, as has been seen for previous cases, the summary statistics appear to tend to those when assuming a normal model on the data demonstrating the sensitivity and approximation of the family of t distributions to the normal distribution as the degrees of freedom approach infinity.

5.3.4 Discussion

There have been some interesting results from carrying out a Bayesian analysis of a fractional factorial experiment in the presence of an outlier. Before even adjusting the dataset so that it contained what would be considered to be an extreme observation, some surprising results were also noted.

When analysing the original quarter-fraction and estimating only the intercept and main effects it was seen that the main effect of **1** was increased greatly and the main effect of **5** was reduced. Some explanation for the large estimated effect of **1** could be due to it being aliased with the true active interaction effect **24**. This leads to questions surrounding the choice of fraction chosen to be analysed, as a result of the high level of aliasing for this quarter-fraction where main effects are aliased with

two-factor interactions which would both typically be considered to be important by experimenters. It is unlikely that an experimenter would choose such a fraction due to the aliasing pattern which means no effects are clear, however the results demonstrated do serve as a warning to experimenters in carefully choosing their fraction and being aware of any possible aliasing which may lead to incorrect conclusions. In this case incorrect conclusions such as main effect **1** could have been drawn even before the added problem of outliers have even been considered. The extremely small or underestimated value of the precision τ in the case of all analyses of the quarter-fraction is also indicative of an overinflated variance, as mentioned previously. A possible suggestion for the cause of this may be due to a highly fractionated design of the original 32 run experiment being used which means that not all contrasts are included in the 8 run design. Thus, only main effects are estimated and other contrasts such as the two-factor interaction **23** is not estimated which is actually aliased with the true active two-factor interaction effect **45**. This may lead to some explanation for the overinflated variance.

As mentioned, only the main effects were estimated in the case of the quarter-fraction, however there were two additional degrees of freedom which could have been used to estimate some of the two-factor interactions. If a strategy had been considered in order to decide which two-factor interactions to estimate such as initially estimating all main effects and then estimating the 2 two-factor interactions associated with the largest main effects, then some caution should be heeded by the experimenter. Considering the example where there were no outliers and the quarter-fraction was analysed, the main effects **2**, **4** and **1** had the largest estimated effects in that order. Thus, an experimenter would then have decided to estimate **24** and **12**, however the true active effect **45** would not have been identified.

When an obvious outlier was included in the dataset for both the half-fraction and the quarter-fraction it was seen that the inferences were affected. In this case it would

be hoped that an experimenter would be able to scrutinize the data beforehand and pick up on any obvious outliers. However, without doing so it did appear that using the scaled-t distribution to model the data did result in wider 95% Bayesian credible intervals when using smaller degrees of freedom, in comparison to using the normal distribution. This result is perhaps intuitive given that one may expect wider intervals for a longer tailed distribution, however the intervals were not that much wider than for the normal distribution and so may prove reassuring to an experimenter that inferences are not overly influenced by the model assumption for the data. Despite this, it did not appear that given the prior assumptions investigated, that in this instance the scaled-t provided a much more robust inference in comparison to using the more typical assumption that the data follow a normal distribution. Very different results were seen for the summary statistics when assuming a mixture of normal distributions on the effect parameters rather than independent normal distributions, although these results did not appear to indicate more robust inferences. It may be of interest therefore to investigate these distributions further and the case of more informative priors to assess whether more robust inferences can be achieved.

The impact of not only the outliers, but also of the choice of experimental design upon the posterior means, medians and 95% Bayesian credible intervals should be noted. The intervals were affected greatly even in the case when the dataset did not contain an outlier but the aliasing structure of the quarter-fraction resulted in low order, and therefore important effects, being aliased. Although none of the truly active effects were actually aliased with one another due to the defining contrast subgroup being $\mathbf{1} = \mathbf{24} = \mathbf{35}$ which means the problem of cancellation may not occur, it appears that since main effects are aliased with two-factor interactions, all of the parameter effects have been greatly inflated whether that particular contrast has an active effect associated with it or not. This would lead to an experimenter being unable to conclude which effects are truly significant and require the need for further experimentation.

The idea of accounting for outliers in a linear model and the amount of 'contamination' by this outlier has been previously discussed, although not at great length. It would have been of interest to investigate this further by placing a prior on this contamination δ, thus considering an alternative approach to that considered in this chapter. This would be in order to use a Bayesian approach to make inferences about the parameters in the model when the dataset is suspected of containing spurious observations.

5.4 Summary

In this chapter various Bayesian analyses have been presented when making a variety of prior assumptions and in the case of different fractions of a design. It has been shown that the type of fraction chosen can impact greatly upon the analysis, particularly when there is a high level of aliasing involved, and also when the dataset is suspected of containing an outlier. Thus, an experimenter must carefully consider all design fractions that could be used and although a robust inference could be sought when carrying out a Bayesian analysis, this should not be a substitute for careful data scrutiny checks in order to identify any potential extreme observations.

Chapter 6

Conclusions and Further Work

6.1 Conclusions

This thesis has investigated various research avenues in using Bayesian methodology and decision theory ideas in both the design and analysis of small multifactor industrial experiments. This thesis has specifically focused on this with regards to the manufacturing and process improvement aspects of experimentation in the pharmaceutical industry.

In chapter 2 some utility function concepts were introduced and then later applied in chapter 3 with respect to both screening and optimization. It was shown that the choice of prior information did typically have some impact on the resulting relative gain, when considering optimization at the later stages of experimentation, and also on the utility which encompasses the weight of Type I and II errors when looking at screening at the initial stages of experimentation. Despite this, it was also seen that the type of design being investigated could also be impacted upon by other issues such as the level of aliasing involved and so it was noted that the utility function

ideas developed should not be used as a blind substitute for the choice of design but perhaps more to be used in conjunction with other methods. It was noted that the results presented do have a great dependency upon the various choices of prior distributions and experimental designs made during the study, and also the interplay between the aliasing structure in the experiments and the beliefs about the effects of the factors. Therefore, there are limitations of the studies that should be taken into consideration.

Chapter 4 presented Bayesian methodology and the application of this when analysing fractional factorial experiments, and in chapter 5 this work was further extended where datasets suspected of containing outliers were considered and a robust analysis sought in order that inferences made were not affected by any outlying observations. It was demonstrated that some benefit can be gained from incorporating prior knowledge in the analysis, 95% Bayesian credible intervals were able to be obtained and although the posterior estimated means were not as greatly impacted upon by the choice of prior, it was shown that the size of the interval did change when varying the prior assumptions made, particularly when varying the prior on the precision from a weakly informative prior to a more informative prior. The credible intervals were also found to be greatly impacted by the choice of experimental design, where they were seen to shift a great deal.

In carrying out this research, some indication of which design to use has been made when an experimenter is faced with the decision of choosing from a candidate set and limited on resources, and so facing commercial pressure to minimize the size of the experiment. An outline of how to set up a Bayesian analysis in the case of fractional factorial experiments has also been presented and demonstrated that additional information is to be gained in comparison to simply carrying out standard analyses.

6.2 Further Work

Some further research ideas stem from work carried out in this thesis. It would be interesting to further look at determining which design to choose particularly in the case of three-level designs where central composite and Box-Behnken designs could be examined. Other design types could also be focused on such as non-regular and mixed-level designs. In this thesis the response surface was explored and the optimal treatment combination sought which was assumed to be a point of maximum response. However, this is not always the case and so it would be of interest to determine the maximum response when the stationary point is a minimum or a saddle point.

Another interesting issue for further research is concerned with extending the ideas presented in chapter 5, where rather than merely seeking a robust analysis in the presence of outliers, it may be of interest to be able to actually identify which observations are potential outliers. If the potential outliers could be found, then it would be possible in some circumstances to go on and remove them. This could be investigated by using a 'Bayesian residual analysis' where differences between the observed and posterior values for the response variables could be analysed and some limit used to identify which of these residuals are large and thus outliers. Also previously discussed was the concept of describing the outlier in the linear model by the amount of 'contamination', and it would be of interest to further investigate this by placing a prior on the contamination δ. Then a Bayesian approach could be adopted in order to make inferences about the parameters in the model in the possible presence of outlying observations.

Appendix

The following appendices contain examples of the programs used to simulate the results presented in Chapter 3, which were written in R and Chapter 4, which were written in WinBUGS.

Appendix A contains the code for the optimization utility function which was written in R. A.1, and A.2 display the code for the set up of point priors and normal prior distributions on the parameters respectively. These are for the intercept, main effects **1**, **2**, **3**, two-factor interactions **12**, **13**, **23** and the three-factor interaction **123**. A.3 then gives code for the two candidate design matrices which are both half-fractions and 4 runs in size and the full factorial 2^3 design matrix with all contrasts for all effects included. Code for the algorithm to calculate the expected loss, gain and relative gain for each of the candidate designs is presented. A.4 gives the code to obtain the values of the loss, gain, maximum gain and relative gain for a three-level design when assuming independent normal priors on the main, two-factor interaction and quadratic effects.

Appendix B is again for code written in R, however this time for the screening utility function. B.1 gives the prior specifications and code to simulate the R datasets and determine active effects whereas B.2 contains code to evaluate the expected utility. B.1 displays the code when estimating main effects and two-factor interactions, as-

suming normal mixture prior distributions on these effects for both the 12 and 16 run design. Design matrices are given along with the prior specifications and code to generate r simulated datasets. Code to fit the model and estimate the effects is also given and then to determine which effects are active according to the two different methods. B.2.1 gives the code in the case of the 12 run design, where the main effects and 3 two-factor interactions are estimated. Additional code is given where initially the main effects are estimated and then code given to find the three largest main effects for each of the r simulations. The contrasts for the 3 two-factor interactions associated with these three main effects are then added to the original design matrix with main effects only. B.1.2 again gives code for the prior specification and design matrix, but is for the 16 run design where estimating main effects and all two-factor interactions. Finally, B.2 contains code to calculate the Type I and II errors and subsequently the utility, for the r simulations.

Appendix C displays code written in WinBUGS in order to carry out a Bayesian analysis of the reactor data. For each of the subsections in C, code for the model for the data is set out, and prior assumptions for the intercept and effect parameters given. Code to list the data is presented, and then vectors of $-1's$ and $+1's$ given for each of the variables to determine the design matrix. Finally, initial values for each of the two chains run is given. C.1 gives the code for this information in the case of a normal model assumed on the data, normal prior distributions assumed on the effect parameters and a uniform distribution assumed on $\log \sigma$. C.2 gives code when assuming that data are modelled by a scaled-t distribution and C.3 displays code for normal mixture prior distributions on the effect parameters.

Appendix A

Optimization Utility Function

A.1 Point Priors on Effects

```
##Fixed Size for Intercept and Effects##
b0<-0
b1<-10
b2<-0
b3<-0
b12<-0
b13<-8
b23<-0
b123<-0

##Intercept##
beta0<-matrix(b0,1,r)
```

```
##Main Effects##
beta1<-matrix(b1,1,r)
beta2<-matrix(b2,1,r)
beta3<-matrix(b3,1,r)

##Two-Factor Interactions##
beta12<-matrix(b12,1,r)
beta13<-matrix(b13,1,r)
beta23<-matrix(b23,1,r)

##Three-Factor Interaction##
beta123<-matrix(b123,1,r)
```

A.2 Normal Prior Distributions On Effects

```
##Prior Means for Parameters##
b0<-0
b1<-10
b2<-10
b3<-2
b12<-8
b13<-0
b23<-0
b123<-0

##Prior Standard Deviations for Parameters##
b0.sd<-1000
b1.sd<-2
b2.sd<-2
b3.sd<-2
b12.sd<-2
b13.sd<-2
b23.sd<-2
b123.sd<-2

##Intercept##
beta0<-rnorm(r,b0,b0.sd)

##Main Effects##
beta1<-rnorm(r,b1,b1.sd)
beta2<-rnorm(r,b2,b2.sd)
```

```
beta3<-rnorm(r,b3,b3.sd)

##Two-Factor Interactions##
beta12<-rnorm(r,b12,b12.sd)
beta13<-rnorm(r,b13,b13.sd)
beta23<-rnorm(r,b23,b23.sd)

##Three-Factor Interaction##
beta123<-rnorm(r,b123,b123.sd)
```

A.3 Code for Expected Loss, Gain and Relative Gain

```
r<-1000    # number of simulations
n<-4    # number of runs in fractional factorial design
t<-8    # number of runs in full factorial design

############################################################

##Design Matrix 1##
X1<-matrix(c(
1, -1, -1, -1,  1,  1,  1, -1,
1, -1,  1,  1, -1, -1,  1, -1,
1,  1, -1,  1, -1,  1, -1, -1,
1,  1,  1, -1,  1, -1, -1, -1),
nrow=4,ncol=8,byrow=TRUE,dimnames=list
(c("run1","run2","run3","run4"),
c("beta0","x.1","x.2","x.3","x.12","x.13","x.23","x.123")))

##Design Matrix 2##
X2<-matrix(c(
1, -1, -1,  0,  1,  0,  0,  0,
1, -1,  1,  0, -1,  0,  0,  0,
1,  1, -1,  0, -1,  0,  0,  0,
1,  1,  1,  0,  1,  0,  0,  0),
nrow=4,ncol=8,byrow=TRUE,dimnames=list
(c("run1","run2","run3","run4"),
```

```
c("beta0","x.1","x.2","x.3","x.12","x.13","x.23","x.123")))

##Matrix of Full Factorial Design##
X.full<-matrix(c(
1, -1, -1, -1,  1,  1,  1, -1,
1, -1,  1,  1, -1, -1,  1, -1,
1,  1, -1,  1, -1,  1, -1, -1,
1,  1,  1, -1,  1, -1, -1, -1,
1, -1, -1,  1,  1, -1, -1,  1,
1, -1,  1, -1, -1,  1, -1,  1,
1,  1, -1, -1, -1, -1,  1,  1,
1,  1,  1,  1,  1,  1,  1,  1),
nrow=8,ncol=8,byrow=TRUE,dimnames=list
(c("run1","run2","run3","run4","run5","run6","run7","run8"),
c("beta0","x.1","x.2","x.3","x.12","x.13","x.23","x.123")))

############################################################

##Matrix of Simulated Priors##
prior.beta<-c(beta0,beta1,beta2,beta3,beta12,beta13,beta23,beta123)
beta<-matrix(prior.beta,t,r,byrow=T)

##Errors##
sim.error<-rnorm(r*n,0,1)
E<-matrix(sim.error,n,r)

##r Sets of Simulated Data Y1##
Y1<-X1%*%beta + E
```

```
##r Sets of Simulated Data Y2##
Y2<-X2%*%beta + E

##Linear Model for Y1##
x1<-X1[,2]
x2<-X1[,3]
x3<-X1[,4]
lm(Y1~x1+x2+x3)
model1<-lm(Y1~x1+x2+x3)
coef(model1)
f1<-rbind(coef(model1),matrix(0,4,r))

##Linear Model for Y2##
x4<-X2[,2]
x5<-X2[,3]
x6<-X2[,5]
lm(Y2~x4+x5+x6)
model2<-lm(Y2~x4+x5+x6)
coef(model2)
f2<-rbind(coef(model2)[1:3,],0,coef(model2)[4],matrix(0,3,r))

##Estimates for Simulated Data for X1##
mu.hat1<-X.full%*%f1
##Estimates for Simulated Data for X2##
mu.hat2<-X.full%*%f2
##Estimates for 'True' Data##
mu<-X.full%*%beta
```

```
############################################################

##Expected Loss for X1##
Loss1 <- matrix(0,1,r)
Gain1 <- matrix(0,1,r)
Truth<-matrix(0,1,r)
for (i in 1:r){
loss1<-max(mu[,i])-mu[mu.hat1==max(mu.hat1[,i])]
gain1 <- mu[mu.hat1==max(mu.hat1[,i])]-beta[1,i]
truth<- max(mu[,i]) - beta[1,i]
Loss1[i] <- loss1
Gain1[i] <- gain1
Truth[i] <- truth
}
Loss1
Gain1
Truth

Rel.Gain1<-matrix(0,1,r)
for(i in 1:r){
Rel.Gain1[i]<-cbind(Gain1[,i]/Truth[,i])
}
Rel.Gain1

##Expected Loss for X2##
Loss2 <- matrix(0,1,r)
Gain2 <- matrix(0,1,r)
Truth <- matrix(0,1,r)
```

```
for (i in 1:r){
loss2<-max(mu[,i])-mean(mu[mu.hat2==max(mu.hat2[,i])])
gain2 <- mean(mu[mu.hat2==max(mu.hat2[,i])])-beta[1,i]
truth<- max(mu[,i]) - beta[1,i]
Loss2[i] <- loss2
Gain2[i] <- gain2
Truth[i] <- truth
}
Loss2
Gain2

Rel.Gain2<-matrix(0,1,r)
for(i in 1:r){
Rel.Gain2[i]<-cbind(Gain2[,i]/Truth1[,i])
}
Rel.Gain2

##Expected Loss, Gain and Relative Gain##
mean(Loss1)
mean(Loss2)
mean(Gain1)
mean(Gain2)
mean(Rel.Gain1)
mean(Rel.Gain2)
```

A.4 Three-Level Design

```
r<-1000    # number of simulations
n1<-18     # run size
e<-11      # number of parameters to be estimated
t<-10      # number of effects to be estimated
sd<-1      # standard deviation for errors

############################################################

##Prior Standard Deviations for Linear, Interaction##
##and Quadratic Effects##
beta.i.sd <- 2
beta.ii.sd<- 2
beta.ij.sd<- 2

##Intercept##
beta0<-rnorm(r, 0, 1000)

##Main Effects##
beta1<-rnorm(r, 0, beta.i.sd)
beta2<-rnorm(r, 0, beta.i.sd)
beta3<-rnorm(r, 0, beta.i.sd)
beta4<-rnorm(r, 0, beta.i.sd)
beta5<-rnorm(r, 0, beta.i.sd)

##Quadratic Effects ##
beta11<-rnorm(r, -30, beta.ii.sd)
```

```
beta22<-rnorm(r, -30, beta.ii.sd)
beta33<-rnorm(r, -30, beta.ii.sd)
beta44<-rnorm(r, -30, beta.ii.sd)
beta55<-rnorm(r, -30, beta.ii.sd)

##Two-Factor Interactions##
beta12<-rnorm(r, 0, beta.ij.sd)
beta13<-rnorm(r, 0, beta.ij.sd)
beta14<-rnorm(r, 0, beta.ij.sd)
beta15<-rnorm(r, 0, beta.ij.sd)
beta23<-rnorm(r, 0, beta.ij.sd)
beta24<-rnorm(r, 0, beta.ij.sd)
beta25<-rnorm(r, 0, beta.ij.sd)
beta34<-rnorm(r, 0, beta.ij.sd)
beta35<-rnorm(r, 0, beta.ij.sd)
beta45<-rnorm(r, 0, beta.ij.sd)

##Matrix of Simulated Priors##
prior.beta <-c(beta0,beta1,beta2,beta3,beta4,beta5,
beta11,beta22,beta33,beta44,beta55,
beta12,beta13,beta14,beta15,beta23,beta24,beta25,beta34,beta35,beta45)
beta <-matrix(prior.beta,21,r,byrow=TRUE)
beta

############################################################

#Design Matrix
```

```
X1<-matrix(c(
1,-1,-1,-1,-1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0,
1,-1,-1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0,-1, 0, 0,-1, 0, 0, 0, 0,
1,-1, 0,-1, 0,-1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0,
1,-1, 0, 1,-1, 1, 1, 0, 1, 1, 1, 0,-1, 1,-1, 0, 0, 0,-1, 1,-1,
1,-1, 1, 0, 0, 1, 1, 1, 0, 0, 1,-1, 0, 0,-1, 0, 0, 1, 0, 0, 0,
1,-1, 1, 1, 1,-1, 1, 1, 1, 1, 1,-1,-1,-1, 1, 1, 1,-1, 1,-1,-1,
1, 0,-1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,-1, 0, 0, 0,
1, 0,-1, 1,-1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0,-1, 1, 1,-1,-1, 1,
1, 0, 0,-1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0,-1,-1, 1,
1, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 1,-1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0,-1, 1, 0,-1, 0, 0,
1, 0, 1, 0,-1,-1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0,-1,-1, 0, 0, 1,
1, 1,-1,-1, 0,-1, 1, 1, 1, 0, 1,-1,-1, 0,-1, 1, 0,-1, 0, 1, 0,
1, 1,-1, 1, 1, 1, 1, 1, 1, 1, 1,-1, 1, 1, 1,-1,-1,-1, 1, 1, 1,
1, 1, 0, 0,-1, 0, 1, 0, 0, 1, 0, 0, 0,-1, 0, 0, 0, 0, 0, 0, 0,
1, 1, 0, 0, 1,-1, 1, 0, 0, 1, 1, 0, 0, 1,-1, 0, 0, 0, 0, 0,-1,
1, 1, 1,-1,-1, 1, 1, 1, 1, 1, 1, 1,-1,-1, 1,-1,-1, 1, 1,-1,-1,
1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0),
nrow=18, ncol=21, byrow=TRUE,dimnames=list(c
("run1","run2","run3","run4","run5","run6","run7","run8",
"run9","run10","run11","run12","run13","run14","run15",
"run16","run17","run18"),
c("beta0","x1","x2","x3","x4","x5", "x11","x22","x33","x44",
"x55","x12","x13","x14","x15","x23","x24","x25","x34","x35",
"x45")))
X1
```

```
############################################################

##Errors##
sim.error<-rnorm(r*n1,0,1)
E1<-matrix(sim.error,n1,r)
E1

##r Sets of Simulated Data Y1##
Y1<-X1%*%beta + E1
Y1

##Linear Model for Y1##
x1<-X1[,2]
x2<-X1[,3]
x3<-X1[,4]
x4<-X1[,5]
x5<-X1[,6]
x11<-X1[,7]
x22<-X1[,8]
x33<-X1[,9]
x44<-X1[,10]
x55<-X1[,11]
lm(Y1~x1+x2+x3+x4+x5+x11+x22+x33+x44+x55)
model1<-lm(Y1~x1+x2+x3+x4+x5+x11+x22+x33+x44+x55)
design1coef<-matrix(print(coef(model1)),e,r)
design1coef

############################################################
```

```
##Matrix of Quadratic Effects##
MatrixB<- matrix(0,5,r)
for (i in 1:r){
matrixB<-design1coef[7:11,i]
MatrixB[,i]<-matrixB
}
MatrixB

##k x k symmetric matrix B##
n1<-5
c1<-5
B.hat<-array(rep(0,r*n1*c1), dim=c(n1,c1,r))
for (i in 1:r){
B.hat[,,i]<-cbind(
MatrixB[1,i], 0, 0, 0, 0,
0, MatrixB[2,i], 0, 0, 0,
0, 0, MatrixB[3,i], 0, 0,
0, 0, 0, MatrixB[4,i], 0,
0, 0, 0, 0, MatrixB[5,i])
}
B.hat

##Matrix of eigenvalues##
Lambda <- matrix(0,5,r)
for (i in 1:r){
E<-eigen(B.hat[,,i])
lambda <- E$values
```

```
Lambda[,i]<-lambda
}
Lambda

##Check that eigenvalues are negative##
check.Lambda<-matrix(0,5,r)
for (j in 1:r){
for (i in 1:5){
CHECK.lambda<- if (Lambda[i,j]<0) 1 else 0
check.Lambda[i,j]<-CHECK.lambda
}
}
check.Lambda
sum(check.Lambda == 1)

Bhat<-array(rep(0,r*n1*c1), dim=c(n1,c1,r))
for (i in 1:r){
for (j in 1:5){
for (k in 1:5){
Bhat[j,k,i]<- cbind( if (B.hat[j,k,i]>0) -B.hat[j,k,i] else
B.hat[j,k,i])
}
}
}
Bhat

Lambda. <- matrix(0,5,r)
for (i in 1:r){
```

```
E.<-eigen(Bhat[,,i])
lambda. <- E.$values
Lambda.[,i]<-lambda.
}
Lambda.

check.Lambda.<-matrix(0,5,r)
for (j in 1:r){
for (i in 1:5){
CHECK.lambda.<- if (Lambda.[i,j]<0) 1 else 0
check.Lambda.[i,j]<-CHECK.lambda.
}
}
check.Lambda.
sum(check.Lambda. == 1)

##Matrix b of main effects##
bhat<- matrix(0,5,r)
for (i in 1:r){
maineffects<-design1coef[2:6,i]
bhat[,i]<-maineffects
}
bhat

##To obtain location of stationary point##
x.s<-matrix(0, 5, r)
for (i in 1:r){
x.s[,i]<- -0.5*(solve(Bhat[,,i])%*%bhat[,i])
```

```
}
x.s

##To obtain predicted response at the stationary point##
y_x.s<-matrix(0, 1, r)
for (i in 1:r){
y_x.s[,i]<- cbind(design1coef[1,i]+0.5*(t(x.s[,i])%*%bhat[,i]))
}
y_x.s

#############################################################

b1<-beta
##Matrix of Quadratic Effects##
MatrixB.1<- matrix(0,5,r)
for (i in 1:r){
matrixB.1<-beta[7:11,i]
MatrixB.1[,i]<-matrixB.1
}
MatrixB.1

##k x k symmetric matrix B##
n1<-5
c1<-5
B.1<-array(rep(0,r*n1*c1), dim=c(n1,c1,r))
for (i in 1:r){
B.1[,,i]<-cbind(
MatrixB.1[1,i],0.5*b1[12,i],0.5*b1[13,i],0.5*b1[14,i],0.5*b1[15,i],
```

```
0.5*b1[12,i],MatrixB.1[2,i],0.5*b1[16,i],0.5*b1[17,i],0.5*b1[18,i],
0.5*b1[13,i],0.5*b1[16,i],MatrixB.1[3,i],0.5*b1[19,i],0.5*b1[20,i],
0.5*b1[14,i],0.5*b1[17,i],0.5*b1[19,i],MatrixB.1[4,i],0.5*b1[21,i],
0.5*b1[15,i],0.5*b1[18,i],0.5*b1[20,i],0.5*b1[21,i],MatrixB.1[5,i])
}
B.1

B.2<-array(rep(0,r*n1*c1), dim=c(n1,c1,r))
for (i in 1:r){
for (j in 1:5){
for (k in 1:5){
B.2[j,k,i]<- cbind( if (B.1[j,k,i]>0) -B.1[j,k,i] else B.1[j,k,i])
}
}
}
B.2

##Matrix of eigenvalues##
Lambda.1 <- matrix(0,5,r)
for (i in 1:r){
E.1<-eigen(B.2[,,i])
lambda.1 <- E.1$values
Lambda.1[,i]<-lambda.1
}
Lambda.1

##Check that eigenvalues are negative##
check.Lambda.1<-matrix(0,5,r)
```

```
for (j in 1:r){
for (i in 1:5){
CHECK.lambda.1<- if (Lambda.1[i,j]<0) 1 else 0
check.Lambda.1[i,j]<-CHECK.lambda.1
}
}
check.Lambda.1
sum(check.Lambda.1 == 1)

Lambda.1.matrix<- matrix(0,1,r)
for (i in 1:r){
LAMBDA.1.matrix<- if (sum(check.Lambda.1[,i]) != 5) 1 else 0
Lambda.1.matrix[,i]<- LAMBDA.1.matrix
}
Lambda.1.matrix
sum(Lambda.1.matrix)

newB<-array(rep(0,r*n1*c1), dim=c(n1,c1,r))
for (i in 1:r){
for (j in 1:5){
for (k in 1:5){
newB[j,k,i]<-cbind(B.2[j,k,i])
newB[j,j,i]<- cbind(B.2[j,j,i]-10)
}
}
}
newB
```

```
Lambda.2 <- matrix(0,5,r)
for (i in 1:r){
E.2<-eigen(newB[,,i])
lambda.2 <- E.2$values
Lambda.2[,i]<-lambda.2
}
Lambda.2

check.Lambda.2<-matrix(0,5,r)
for (j in 1:r){
for (i in 1:5){
CHECK.lambda.2<- if (Lambda.2[i,j]<0) 1 else 0
check.Lambda.2[i,j]<-CHECK.lambda.2
}
}
check.Lambda.2
sum(check.Lambda.2 == 1)

##New k x k symmetric matrix B##
B<-array(rep(0,r*n1*c1), dim=c(n1,c1,r))
for (i in 1:r){
for (j in 1:5){
for (k in 1:5){
B[j,k,i]<- cbind(if (Lambda.1.matrix[,i]==1) newB[j,k,i]
else B.2[j,k,i])
}
}
}
```

```
B

##matrix b##
b<- matrix(0,5,r)
for (i in 1:r){
maineffects.1<-beta[2:6,i]
b[,i]<-maineffects.1
}
b

##To obtain location of stationary point##
x.s.1<-matrix(0, 5, r)
for (i in 1:r){
x.s.1[,i]<- -0.5*(solve(B[,,i])%*%b[,i])
}
x.s.1

##To obtain predicted response at the stationary point##
y_x.s.1<-matrix(0, 1, r)
for (i in 1:r){
y_x.s.1[,i]<- cbind(b1[1,i]+0.5*(t(x.s.1[,i])%*%b[,i]))
}
y_x.s.1

#############################################################

##Maximum Gain##
max.gain<-matrix(0,1,r)
```

```
for (i in 1:r){
max.gain[i]<- cbind(-0.25*(t(b[,i]%*%solve(B[,,i])%*%b[,i])))
}
max.gain

##Gain##
gain <- matrix(0,1,r)
for (i in 1:r){
gain[i]<- cbind(-0.5*t(bhat[,i])%*%solve(Bhat[,,i])%*% b[,i]
+0.25*t(bhat[,i])%*%solve(Bhat[,,i])%*%B[,,i]%*%solve
(Bhat[,,i])%*%bhat[,i])
}
gain

##Loss##
loss <- matrix(0,1,r)
for (i in 1:r){
loss[i]<- cbind(-0.25*(t(b[,i]%*%solve(B[,,i])%*%b[,i]))
+0.5*t(bhat[,i])%*%solve(Bhat[,,i])%*%b[,i]
-0.25*t(bhat[,i])%*%solve(Bhat[,,i])%*%B[,,i]%*%solve
(Bhat[,,i])%*%bhat[,i])
}
loss

##Relative Gain##
relative.gain<-matrix(0,1,r)
for(i in 1:r){
relative.gain[i]<-cbind(gain[,i]/max.gain[,i])
```

```
}
relative.gain


mean(max.gain)
mean(gain)
mean(loss)
mean(relative.gain)
```

Appendix B

Screening Utility Function

B.1 Normal Mixture Prior Distributions

B.1.1 12 Run Design: Estimating Main Effects and 3 Two-Factor Interactions

```
r<-1000          # number of simulations
n1<-12           # run size
e<-16            # number of parameters
t<-15            # number of effects
sd<-1            # standard deviation for errors
omega<-sqrt(4)   # declare effect active if > than omega
f<-6             # number of main effects and intercept
g<-5             # number of main effects

##########################################################
```

```
##Design Matrix##
X1<-matrix(c(
1,  1, -1,  1, -1, -1, -1,  1, -1, -1, -1,  1,  1, -1, -1,  1,
1,  1,  1, -1,  1, -1,  1, -1,  1, -1, -1,  1, -1, -1,  1, -1,
1, -1,  1,  1, -1,  1, -1, -1,  1, -1,  1, -1,  1, -1,  1, -1,
1,  1, -1,  1,  1, -1, -1,  1,  1, -1, -1, -1,  1,  1, -1, -1,
1,  1,  1, -1,  1,  1,  1, -1,  1,  1, -1,  1,  1, -1, -1,  1,
1,  1,  1,  1, -1,  1,  1,  1, -1,  1,  1, -1,  1, -1,  1, -1,
1, -1,  1,  1,  1, -1, -1, -1, -1,  1,  1,  1, -1,  1, -1, -1,
1, -1, -1,  1,  1,  1,  1, -1, -1, -1, -1, -1, -1,  1,  1,  1,
1, -1, -1, -1,  1,  1,  1,  1, -1, -1,  1, -1, -1, -1, -1,  1,
1,  1, -1, -1, -1,  1, -1, -1, -1,  1,  1,  1, -1,  1, -1, -1,
1, -1,  1, -1, -1, -1, -1,  1,  1,  1, -1, -1, -1,  1,  1,  1,
1, -1, -1, -1, -1, -1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1),
nrow=12,ncol=16,byrow=TRUE,dimnames=list(c
("run1","run2","run3","run4","run5","run6","run7","run8",
"run9","run10","run11","run12"),
c("beta0","x.1","x.2","x.3","x.4","x.5",
"x.12","x.13","x.14","x.15","x.23","x.24",
"x.25","x.34","x.35","x.45")))
X1
X1.full<-X1[,-c(1)]
X1.full
X1.design<-X1[,-c(7,8,9,10,11,12,13,14,15,16)]
X1.design

############################################################
```

```
b<-sqrt(4)      # prior standard deviation for active effect
a<-sqrt(0.01)   # prior standard deviation for inactive effect

##Probability that main effect i is active (i = 1,...,5)##
p1<-0.2
p2<-0.2
p3<-0.2
p4<-0.2
p5<-0.2

##Probability P_{ij}: Two-factor interaction effect is active with##
##parent main effects active if i or j = 1; inactive if i or j = 0##
p11<-0.7
p10<-0.4
p01<-0.4
p00<-0.1

##Intercept##
beta0<-matrix(0,1,r)

##Simulate alpha_{i} from Bernoulli distribution:##
##1 if i = active, 0 is i = inactive##
alpha1<-rbinom(r,1,p1)
alpha2<-rbinom(r,1,p2)
alpha3<-rbinom(r,1,p3)
alpha4<-rbinom(r,1,p4)
alpha5<-rbinom(r,1,p5)
```

```
##Prior Distributions on Main Effects##
beta1a<-rnorm(r,0,a)
beta1b<-rnorm(r,0,b)
beta1<-(1-alpha1)*beta1a + alpha1*beta1b
beta2a<-rnorm(r,0,a)
beta2b<-rnorm(r,0,b)
beta2<-(1-alpha2)*beta2a + alpha2*beta2b
beta3a<-rnorm(r,0,a)
beta3b<-rnorm(r,0,b)
beta3<-(1-alpha3)*beta3a + alpha3*beta3b
beta4a<-rnorm(r,0,a)
beta4b<-rnorm(r,0,b)
beta4<-(1-alpha4)*beta4a + alpha4*beta4b
beta5a<-rnorm(r,0,a)
beta5b<-rnorm(r,0,b)
beta5<-(1-alpha5)*beta5a + alpha5*beta5b

##Effects for Interaction 12##
alpha12.a<- matrix(0,r,1)
alpha12.b<- matrix(0,r,1)
alpha12.c<- matrix(0,r,1)
alpha12.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA12.a <- if (alpha1[i] == 1 && alpha2[i] == 1) p11 else 0
alpha12.a[i] <- ALPHA12.a
ALPHA12.b <- if (alpha1[i] == 1 && alpha2[i] == 0) p10 else 0
alpha12.b[i] <- ALPHA12.b
```

```
ALPHA12.c <- if (alpha1[i] == 0 && alpha2[i] == 1) p01 else 0
alpha12.c[i] <- ALPHA12.c
ALPHA12.d <- if (alpha1[i] == 0 && alpha2[i] == 0) p00 else 0
alpha12.d[i] <- ALPHA12.d
}
Alpha12<-alpha12.a + alpha12.b + alpha12.c + alpha12.d
alpha12<-matrix(0,r,1)
for (i in 1:r){
ALPHA12<- rbinom(1,1,Alpha12[i,])
alpha12[i]<-ALPHA12
}
beta12a<-rnorm(r,0,a)
beta12b<-rnorm(r,0,b)
beta12<- (1-alpha12)*beta12a + alpha12*beta12b

##Effects for Interaction 13##
alpha13.a<- matrix(0,r,1)
alpha13.b<- matrix(0,r,1)
alpha13.c<- matrix(0,r,1)
alpha13.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA13.a <- if (alpha1[i] == 1 && alpha3[i] == 1) p11 else 0
alpha13.a[i] <- ALPHA13.a
ALPHA13.b <- if (alpha1[i] == 1 && alpha3[i] == 0) p10 else 0
alpha13.b[i] <- ALPHA13.b
ALPHA13.c <- if (alpha1[i] == 0 && alpha3[i] == 1) p01 else 0
alpha13.c[i] <- ALPHA13.c
ALPHA13.d <- if (alpha1[i] == 0 && alpha3[i] == 0) p00 else 0
```

```
alpha13.d[i] <- ALPHA13.d
}
Alpha13<-alpha13.a + alpha13.b + alpha13.c + alpha13.d
alpha13<-matrix(0,r,1)
for (i in 1:r){
ALPHA13<- rbinom(1,1,Alpha13[i,])
alpha13[i]<-ALPHA13
}
beta13a<-rnorm(r,0,a)
beta13b<-rnorm(r,0,b)
beta13 <- (1-alpha13)*beta13a + alpha13*beta13b

##Effects for Interaction 14##
alpha14.a<- matrix(0,r,1)
alpha14.b<- matrix(0,r,1)
alpha14.c<- matrix(0,r,1)
alpha14.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA14.a <- if (alpha1[i] == 1 && alpha4[i] == 1) p11 else 0
alpha14.a[i] <- ALPHA14.a
ALPHA14.b <- if (alpha1[i] == 1 && alpha4[i] == 0) p10 else 0
alpha14.b[i] <- ALPHA14.b
ALPHA14.c <- if (alpha1[i] == 0 && alpha4[i] == 1) p01 else 0
alpha14.c[i] <- ALPHA14.c
ALPHA14.d <- if (alpha1[i] == 0 && alpha4[i] == 0) p00 else 0
alpha14.d[i] <- ALPHA14.d
}
Alpha14<-alpha14.a + alpha14.b + alpha14.c + alpha14.d
```

```
alpha14<-matrix(0,r,1)
for (i in 1:r){
ALPHA14<- rbinom(1,1,Alpha14[i,])
alpha14[i]<-ALPHA14
}
beta14a<-rnorm(r,0,a)
beta14b<-rnorm(r,0,b)
beta14 <- (1-alpha14)*beta14a + alpha14*beta14b

##Effects for Interaction 15##
alpha15.a<- matrix(0,r,1)
alpha15.b<- matrix(0,r,1)
alpha15.c<- matrix(0,r,1)
alpha15.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA15.a <- if (alpha1[i] == 1 && alpha5[i] == 1) p11 else 0
alpha15.a[i] <- ALPHA15.a
ALPHA15.b <- if (alpha1[i] == 1 && alpha5[i] == 0) p10 else 0
alpha15.b[i] <- ALPHA15.b
ALPHA15.c <- if (alpha1[i] == 0 && alpha5[i] == 1) p01 else 0
alpha15.c[i] <- ALPHA15.c
ALPHA15.d <- if (alpha1[i] == 0 && alpha5[i] == 0) p00 else 0
alpha15.d[i] <- ALPHA15.d
}
Alpha15<-alpha15.a + alpha15.b + alpha15.c + alpha15.d
alpha15<-matrix(0,r,1)
for (i in 1:r){
ALPHA15<- rbinom(1,1,Alpha15[i,])
```

```
alpha15[i]<-ALPHA15
}
beta15a<-rnorm(r,0,a)
beta15b<-rnorm(r,0,b)
beta15 <- (1-alpha15)*beta15a + alpha15*beta15b

##Effects for Interaction 23##
alpha23.a<- matrix(0,r,1)
alpha23.b<- matrix(0,r,1)
alpha23.c<- matrix(0,r,1)
alpha23.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA23.a <- if (alpha2[i] == 1 && alpha3[i] == 1) p11 else 0
alpha23.a[i] <- ALPHA23.a
ALPHA23.b <- if (alpha2[i] == 1 && alpha3[i] == 0) p10 else 0
alpha23.b[i] <- ALPHA23.b
ALPHA23.c <- if (alpha2[i] == 0 && alpha3[i] == 1) p01 else 0
alpha23.c[i] <- ALPHA23.c
ALPHA23.d <- if (alpha2[i] == 0 && alpha3[i] == 0) p00 else 0
alpha23.d[i] <- ALPHA23.d
}
Alpha23<-alpha23.a + alpha23.b + alpha23.c + alpha23.d
alpha23<-matrix(0,r,1)
for (i in 1:r){
ALPHA23<- rbinom(1,1,Alpha23[i,])
alpha23[i]<-ALPHA23
}
beta23a<-rnorm(r,0,a)
```

```
beta23b<-rnorm(r,0,b)
beta23 <- (1-alpha23)*beta23a + alpha23*beta23b

##Effects for Interaction 24##
alpha24.a<- matrix(0,r,1)
alpha24.b<- matrix(0,r,1)
alpha24.c<- matrix(0,r,1)
alpha24.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA24.a <- if (alpha2[i] == 1 && alpha4[i] == 1) p11 else 0
alpha24.a[i] <- ALPHA24.a
ALPHA24.b <- if (alpha2[i] == 1 && alpha4[i] == 0) p10 else 0
alpha24.b[i] <- ALPHA24.b
ALPHA24.c <- if (alpha2[i] == 0 && alpha4[i] == 1) p01 else 0
alpha24.c[i] <- ALPHA24.c
ALPHA24.d <- if (alpha2[i] == 0 && alpha4[i] == 0) p00 else 0
alpha24.d[i] <- ALPHA24.d
}
Alpha24<-alpha24.a + alpha24.b + alpha24.c + alpha24.d
alpha24<-matrix(0,r,1)
for (i in 1:r){
ALPHA24<- rbinom(1,1,Alpha24[i,])
alpha24[i]<-ALPHA24
}
beta24a<-rnorm(r,0,a)
beta24b<-rnorm(r,0,b)
beta24 <- (1-alpha24)*beta24a + alpha24*beta24b
```

```
##Effects for Interaction 25##
alpha25.a<- matrix(0,r,1)
alpha25.b<- matrix(0,r,1)
alpha25.c<- matrix(0,r,1)
alpha25.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA25.a <- if (alpha2[i] == 1 && alpha5[i] == 1) p11 else 0
alpha25.a[i] <- ALPHA25.a
ALPHA25.b <- if (alpha2[i] == 1 && alpha5[i] == 0) p10 else 0
alpha25.b[i] <- ALPHA25.b
ALPHA25.c <- if (alpha2[i] == 0 && alpha5[i] == 1) p01 else 0
alpha25.c[i] <- ALPHA25.c
ALPHA25.d <- if (alpha2[i] == 0 && alpha5[i] == 0) p00 else 0
alpha25.d[i] <- ALPHA25.d
}
Alpha25<-alpha25.a + alpha25.b + alpha25.c + alpha25.d
alpha25<-matrix(0,r,1)
for (i in 1:r){
ALPHA25<- rbinom(1,1,Alpha25[i,])
alpha25[i]<-ALPHA25
}
beta25a<-rnorm(r,0,a)
beta25b<-rnorm(r,0,b)
beta25 <- (1-alpha25)*beta25a + alpha25*beta25b

##Effects for Interaction 34##
alpha34.a<- matrix(0,r,1)
alpha34.b<- matrix(0,r,1)
```

```
alpha34.c<- matrix(0,r,1)
alpha34.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA34.a <- if (alpha3[i] == 1 && alpha4[i] == 1) p11 else 0
alpha34.a[i] <- ALPHA34.a
ALPHA34.b <- if (alpha3[i] == 1 && alpha4[i] == 0) p10 else 0
alpha34.b[i] <- ALPHA34.b
ALPHA34.c <- if (alpha3[i] == 0 && alpha4[i] == 1) p01 else 0
alpha34.c[i] <- ALPHA34.c
ALPHA34.d <- if (alpha3[i] == 0 && alpha4[i] == 0) p00 else 0
alpha34.d[i] <- ALPHA34.d
}
Alpha34<-alpha34.a + alpha34.b + alpha34.c + alpha34.d
alpha34<-matrix(0,r,1)
for (i in 1:r){
ALPHA34<- rbinom(1,1,Alpha34[i,])
alpha34[i]<-ALPHA34
}
beta34a<-rnorm(r,0,a)
beta34b<-rnorm(r,0,b)
beta34 <- (1-alpha34)*beta34a + alpha34*beta34b

##Effects for Interaction 35##
alpha35.a<- matrix(0,r,1)
alpha35.b<- matrix(0,r,1)
alpha35.c<- matrix(0,r,1)
alpha35.d<- matrix(0,r,1)
for (i in 1:r){
```

```
ALPHA35.a <- if (alpha3[i] == 1 && alpha5[i] == 1) p11 else 0
alpha35.a[i] <- ALPHA35.a
ALPHA35.b <- if (alpha3[i] == 1 && alpha5[i] == 0) p10 else 0
alpha35.b[i] <- ALPHA35.b
ALPHA35.c <- if (alpha3[i] == 0 && alpha5[i] == 1) p01 else 0
alpha35.c[i] <- ALPHA35.c
ALPHA35.d <- if (alpha3[i] == 0 && alpha5[i] == 0) p00 else 0
alpha35.d[i] <- ALPHA35.d
}
Alpha35<-alpha35.a + alpha35.b + alpha35.c + alpha35.d
alpha35<-matrix(0,r,1)
for (i in 1:r){
ALPHA35<- rbinom(1,1,Alpha35[i,])
alpha35[i]<-ALPHA35
}
beta35a<-rnorm(r,0,a)
beta35b<-rnorm(r,0,b)
beta35 <- (1-alpha35)*beta35a + alpha35*beta35b

##Effects for Interaction 45##
alpha45.a<- matrix(0,r,1)
alpha45.b<- matrix(0,r,1)
alpha45.c<- matrix(0,r,1)
alpha45.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA45.a <- if (alpha4[i] == 1 && alpha5[i] == 1) p11 else 0
alpha45.a[i] <- ALPHA45.a
ALPHA45.b <- if (alpha4[i] == 1 && alpha5[i] == 0) p10 else 0
```

```
alpha45.b[i] <- ALPHA45.b
ALPHA45.c <- if (alpha4[i] == 0 && alpha5[i] == 1) p01 else 0
alpha45.c[i] <- ALPHA45.c
ALPHA45.d <- if (alpha4[i] == 0 && alpha5[i] == 0) p00 else 0
alpha45.d[i] <- ALPHA45.d
}
Alpha45<-alpha45.a + alpha45.b + alpha45.c + alpha45.d
alpha45<-matrix(0,r,1)
for (i in 1:r){
ALPHA45<- rbinom(1,1,Alpha45[i,])
alpha45[i]<-ALPHA45
}
beta45a<-rnorm(r,0,a)
beta45b<-rnorm(r,0,b)
beta45 <- (1-alpha45)*beta45a + alpha45*beta45b

##Matrix of Simulated Priors##
dummy.prior.beta <-c(beta0,beta1,beta2,beta3,beta4,beta5,
beta12,beta13,beta14,beta15,beta23,beta24,beta25,beta34,
beta35,beta45)
beta.r <-matrix(dummy.prior.beta,e,r,byrow=TRUE)
Beta.r <- beta.r[-c(1),]

##Errors##
sim.error1<-rnorm(n1*r,0,sd)
E1<-matrix(sim.error1,n1,r)
E1
```

```
##r Sets of Simulated Data Y1##
Y1<-X1%*%beta.r + E1

##Linear Model for Y1##
x1<-X1[,2]
x2<-X1[,3]
x3<-X1[,4]
x4<-X1[,5]
x5<-X1[,6]
lm(Y1~x1+x2+x3+x4+x5)
model1<-lm(Y1~x1+x2+x3+x4+x5)
design1coef<-matrix(print(coef(model1)),f,r)
effects1<-design1coef[-c(1),]

############################################################
##To obtain new design matrix with all main effects and three       ##
##two-factor interactions associated with three largest main effects##

third<-matrix(0,1,r)
for (i in 1:r){
Third<-sort(abs(effects1[,i]))[3]
third[i]<-Third
}
third
large.effects<-matrix(0,g,r)
for (i in 1:g){
for (j in 1:r){
large.effects[i,j]<-as.numeric(abs(effects1[i,j])>=third[,j])
```

```
}
}
large.effects
X.expand<-array(rep(0,r*n1*t), dim=c(n1,t,r))
for (i in 1:r){
X.expand[,,i]<-cbind(X1[,2:6],
X1[,7]* large.effects[1,i]*large.effects[2,i],
X1[,8]* large.effects[1,i]*large.effects[3,i],
X1[,9]* large.effects[1,i]*large.effects[4,i],
X1[,10]*large.effects[1,i]*large.effects[5,i],
X1[,11]*large.effects[2,i]*large.effects[3,i],
X1[,12]*large.effects[2,i]*large.effects[4,i],
X1[,13]*large.effects[2,i]*large.effects[5,i],
X1[,14]*large.effects[3,i]*large.effects[4,i],
X1[,15]*large.effects[3,i]*large.effects[5,i],
X1[,16]*large.effects[4,i]*large.effects[5,i])
}
X.expand
estimated.effects<-matrix(0,16,r)
estimated.effects
for (i in 1:r){
estimated.effects[,i]<-coef(lm(Y1[,i]~(X.expand[,,i])))
}
estimated.effects
new.effects<-estimated.effects[-c(1),]
new.effects
new.effects[is.na(new.effects)]<-0
new.effects
```

```
############################################################

D <- as.numeric(abs(Beta.r)>omega)
D
Delta.star<- matrix(D,t,r)
Delta.star

##Declaring True Active Effects##
no.actives<-matrix(0,r,1)
for (i in 1:r){
no.actives[i]<-length(Delta.star[,i][Delta.star[,i]==1])
}
no.actives

############################################################

##Method 1: if effects > omega, then declare active##

D1<- as.numeric(abs(new.effects)>omega)
Delta1<- matrix(D1,t,r)

############################################################

##Method 2: Lenth's method##

Median1 <- matrix(0,1,r)
for (i in 1:r){
```

```
med1 <- median(abs(new.effects[,i]))
Median1[i] <- med1
}
s0.1 <- 1.5*Median1
T1<- matrix(0,t,r)
for (i in 1:t){
for (j in 1:r){
t.1 <- ifelse((abs(new.effects)[i,j] < 2.5*s0.1[,j]),1,0)
T1[i,j] <- t.1
}}
N1<- T1*new.effects
PSE<-matrix(0,1,r)
for (i in 1:r){
PSE[i]<-1.5*median(abs(N1[,i]))
}
t.975.d<-2.57
t.gamma.d<- 5.22
ME1<-t.975.d*PSE
SME1<- t.gamma.d * PSE
ME<-matrix(ME1,1,r)
SME<-matrix(SME1,1,r)
D1.lenth<-matrix(0,t,r)
for (i in 1:t){
for (j in 1:r){
D1.lenth[i,j]<- as.numeric(abs(new.effects[i,j])>ME1[,j])
}
}
Delta.1.lenth<- matrix(D1.lenth,t,r)
```

B.1.2 16 Run Design: Estimating Main Effects and all Two-Factor Interactions

```
r<-1000             # number of simulations
n1<-16              # run size
e<-16               # number of parameters
t<-15               # number of effects
sd<-1               # standard deviation for errors
omega <- sqrt(10)   # declare effect active if > than omega

############################################################

##Design Matrix 1##
X1<-matrix(c(
1, -1, -1, -1, -1,  1,  1,  1,  1, -1,  1,  1, -1,  1, -1, -1,
1,  1, -1, -1, -1, -1, -1, -1, -1, -1,  1,  1,  1,  1,  1,  1,
1, -1,  1, -1, -1, -1, -1,  1,  1,  1, -1, -1, -1,  1,  1,  1,
1,  1,  1, -1, -1,  1,  1, -1, -1,  1, -1, -1,  1,  1, -1, -1,
1, -1, -1,  1, -1, -1,  1, -1,  1,  1, -1,  1,  1, -1, -1,  1,
1,  1, -1,  1, -1,  1, -1,  1, -1,  1, -1,  1, -1, -1,  1, -1,
1, -1,  1,  1, -1,  1, -1, -1,  1, -1,  1, -1,  1, -1,  1, -1,
1,  1,  1,  1, -1, -1,  1,  1, -1, -1,  1, -1, -1, -1, -1,  1,
1, -1, -1, -1,  1, -1,  1,  1, -1,  1,  1, -1,  1, -1,  1, -1,
1,  1, -1, -1,  1,  1, -1, -1,  1,  1,  1, -1, -1, -1, -1,  1,
1, -1,  1, -1,  1,  1, -1,  1, -1, -1, -1,  1,  1, -1, -1,  1,
```

```
1,  1,  1, -1,  1, -1,  1, -1,  1, -1, -1,  1, -1, -1,  1, -1,
1, -1, -1,  1,  1,  1,  1, -1, -1, -1, -1, -1, -1,  1,  1,  1,
1,  1, -1,  1,  1, -1, -1,  1,  1, -1, -1, -1,  1,  1, -1, -1,
1, -1,  1,  1,  1, -1, -1, -1, -1,  1,  1,  1, -1,  1, -1, -1,
1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1),
nrow=16,ncol=16,byrow=TRUE,dimnames=list
(c("run1","run2","run3","run4","run5","run6","run7","run8",
"run9","run10","run11","run12","run13","run14","run15","run16"),
c("beta0","x.1","x.2","x.3","x.4","x.5","x.12", "x.13", "x.14",
"x.15", "x.23", "x.24", "x.25", "x.34", "x.35", "x.45")))
X1

############################################################

b<-sqrt(10)        # prior standard deviation for active effect
a<-sqrt(0.01)      # prior standard deviation for inactive effect

##Probability p_{i} that main effect i is active (i = 1,...,5)##
p1<-0.2
p2<-0.2
p3<-0.2
p4<-0.6
p5<-0.6

##Probability p_{ij}: Two-factor interaction effect is active with##
##parent main effects active if i or j = 1; inactive if i or j = 0##
p11<-0.7
p10<-0.4
```

```
p01<-0.4
p00<-0.1

############################################################

##Intercept##
beta0<-matrix(0,1,r)

##Simulate alpha_{i} from Bernoulli distribution:##
##1 if i = active, 0 is i = inactive##
alpha1<-rbinom(r,1,p1)
alpha2<-rbinom(r,1,p2)
alpha3<-rbinom(r,1,p3)
alpha4<-rbinom(r,1,p4)
alpha5<-rbinom(r,1,p5)

##Prior Distributions on Main Effects##
beta1a<-rnorm(r,0,a)
beta1b<-rnorm(r,0,b)
beta1<-(1-alpha1)*beta1a + alpha1*beta1b
beta2a<-rnorm(r,0,a)
beta2b<-rnorm(r,0,b)
beta2<- (1-alpha2)*beta2a + alpha2*beta2b
beta3a<-rnorm(r,0,a)
beta3b<-rnorm(r,0,b)
beta3<- (1-alpha3)*beta3a + alpha3*beta3b
beta4a<-rnorm(r,0,a)
beta4b<-rnorm(r,0,b)
```

```
beta4<- (1-alpha4)*beta4a + alpha4*beta4b
beta5a<-rnorm(r,0,a)
beta5b<-rnorm(r,0,b)
beta5<- (1-alpha5)*beta5a + alpha5*beta5b

##Effects for Interaction 12##
alpha12.a<- matrix(0,r,1)
alpha12.b<- matrix(0,r,1)
alpha12.c<- matrix(0,r,1)
alpha12.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA12.a <- if (alpha1[i] == 1 && alpha2[i] == 1) p11 else 0
alpha12.a[i] <- ALPHA12.a
ALPHA12.b <- if (alpha1[i] == 1 && alpha2[i] == 0) p10 else 0
alpha12.b[i] <- ALPHA12.b
ALPHA12.c <- if (alpha1[i] == 0 && alpha2[i] == 1) p01 else 0
alpha12.c[i] <- ALPHA12.c
ALPHA12.d <- if (alpha1[i] == 0 && alpha2[i] == 0) p00 else 0
alpha12.d[i] <- ALPHA12.d
}
Alpha12<-alpha12.a + alpha12.b + alpha12.c + alpha12.d
alpha12<-matrix(0,r,1)
for (i in 1:r){
ALPHA12<- rbinom(1,1,Alpha12[i,])
alpha12[i]<-ALPHA12
}
beta12a<-rnorm(r,0,a)
beta12b<-rnorm(r,0,b)
```

```
beta12<- (1-alpha12)*beta12a + alpha12*beta12b

##Effects for Interaction 13##
alpha13.a<- matrix(0,r,1)
alpha13.b<- matrix(0,r,1)
alpha13.c<- matrix(0,r,1)
alpha13.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA13.a <- if (alpha1[i] == 1 && alpha3[i] == 1) p11 else 0
alpha13.a[i] <- ALPHA13.a
ALPHA13.b <- if (alpha1[i] == 1 && alpha3[i] == 0) p10 else 0
alpha13.b[i] <- ALPHA13.b
ALPHA13.c <- if (alpha1[i] == 0 && alpha3[i] == 1) p01 else 0
alpha13.c[i] <- ALPHA13.c
ALPHA13.d <- if (alpha1[i] == 0 && alpha3[i] == 0) p00 else 0
alpha13.d[i] <- ALPHA13.d
}
Alpha13<-alpha13.a + alpha13.b + alpha13.c + alpha13.d
alpha13<-matrix(0,r,1)
for (i in 1:r){
ALPHA13<- rbinom(1,1,Alpha13[i,])
alpha13[i]<-ALPHA13
}
beta13a<-rnorm(r,0,a)
beta13b<-rnorm(r,0,b)
beta13 <- (1-alpha13)*beta13a + alpha13*beta13b

##Effects for Interaction 14##
```

```
alpha14.a<- matrix(0,r,1)
alpha14.b<- matrix(0,r,1)
alpha14.c<- matrix(0,r,1)
alpha14.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA14.a <- if (alpha1[i] == 1 && alpha4[i] == 1) p11 else 0
alpha14.a[i] <- ALPHA14.a
ALPHA14.b <- if (alpha1[i] == 1 && alpha4[i] == 0) p10 else 0
alpha14.b[i] <- ALPHA14.b
ALPHA14.c <- if (alpha1[i] == 0 && alpha4[i] == 1) p01 else 0
alpha14.c[i] <- ALPHA14.c
ALPHA14.d <- if (alpha1[i] == 0 && alpha4[i] == 0) p00 else 0
alpha14.d[i] <- ALPHA14.d
}
Alpha14<-alpha14.a + alpha14.b + alpha14.c + alpha14.d
alpha14<-matrix(0,r,1)
for (i in 1:r){
ALPHA14<- rbinom(1,1,Alpha14[i,])
alpha14[i]<-ALPHA14
}
beta14a<-rnorm(r,0,a)
beta14b<-rnorm(r,0,b)
beta14 <- (1-alpha14)*beta14a + alpha14*beta14b

##Effects for Interaction 15##
alpha15.a<- matrix(0,r,1)
alpha15.b<- matrix(0,r,1)
alpha15.c<- matrix(0,r,1)
```

```
alpha15.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA15.a <- if (alpha1[i] == 1 && alpha5[i] == 1) p11 else 0
alpha15.a[i] <- ALPHA15.a
ALPHA15.b <- if (alpha1[i] == 1 && alpha5[i] == 0) p10 else 0
alpha15.b[i] <- ALPHA15.b
ALPHA15.c <- if (alpha1[i] == 0 && alpha5[i] == 1) p01 else 0
alpha15.c[i] <- ALPHA15.c
ALPHA15.d <- if (alpha1[i] == 0 && alpha5[i] == 0) p00 else 0
alpha15.d[i] <- ALPHA15.d
}
Alpha15<-alpha15.a + alpha15.b + alpha15.c + alpha15.d
alpha15<-matrix(0,r,1)
for (i in 1:r){
ALPHA15<- rbinom(1,1,Alpha15[i,])
alpha15[i]<-ALPHA15
}
beta15a<-rnorm(r,0,a)
beta15b<-rnorm(r,0,b)
beta15 <- (1-alpha15)*beta15a + alpha15*beta15b

##Effects for Interaction 23##
alpha23.a<- matrix(0,r,1)
alpha23.b<- matrix(0,r,1)
alpha23.c<- matrix(0,r,1)
alpha23.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA23.a <- if (alpha2[i] == 1 && alpha3[i] == 1) p11 else 0
```

```
alpha23.a[i] <- ALPHA23.a
ALPHA23.b <- if (alpha2[i] == 1 && alpha3[i] == 0) p10 else 0
alpha23.b[i] <- ALPHA23.b
ALPHA23.c <- if (alpha2[i] == 0 && alpha3[i] == 1) p01 else 0
alpha23.c[i] <- ALPHA23.c
ALPHA23.d <- if (alpha2[i] == 0 && alpha3[i] == 0) p00 else 0
alpha23.d[i] <- ALPHA23.d
}
Alpha23<-alpha23.a + alpha23.b + alpha23.c + alpha23.d
alpha23<-matrix(0,r,1)
for (i in 1:r){
ALPHA23<- rbinom(1,1,Alpha23[i,])
alpha23[i]<-ALPHA23
}
beta23a<-rnorm(r,0,a)
beta23b<-rnorm(r,0,b)
beta23 <- (1-alpha23)*beta23a + alpha23*beta23b

##Effects for Interaction 24##
alpha24.a<- matrix(0,r,1)
alpha24.b<- matrix(0,r,1)
alpha24.c<- matrix(0,r,1)
alpha24.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA24.a <- if (alpha2[i] == 1 && alpha4[i] == 1) p11 else 0
alpha24.a[i] <- ALPHA24.a
ALPHA24.b <- if (alpha2[i] == 1 && alpha4[i] == 0) p10 else 0
alpha24.b[i] <- ALPHA24.b
```

```
ALPHA24.c <- if (alpha2[i] == 0 && alpha4[i] == 1) p01 else 0
alpha24.c[i] <- ALPHA24.c
ALPHA24.d <- if (alpha2[i] == 0 && alpha4[i] == 0) p00 else 0
alpha24.d[i] <- ALPHA24.d
}
Alpha24<-alpha24.a + alpha24.b + alpha24.c + alpha24.d
alpha24<-matrix(0,r,1)
for (i in 1:r){
ALPHA24<- rbinom(1,1,Alpha24[i,])
alpha24[i]<-ALPHA24
}
beta24a<-rnorm(r,0,a)
beta24b<-rnorm(r,0,b)
beta24 <- (1-alpha24)*beta24a + alpha24*beta24b

##Effects for Interaction 25##
alpha25.a<- matrix(0,r,1)
alpha25.b<- matrix(0,r,1)
alpha25.c<- matrix(0,r,1)
alpha25.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA25.a <- if (alpha2[i] == 1 && alpha5[i] == 1) p11 else 0
alpha25.a[i] <- ALPHA25.a
ALPHA25.b <- if (alpha2[i] == 1 && alpha5[i] == 0) p10 else 0
alpha25.b[i] <- ALPHA25.b
ALPHA25.c <- if (alpha2[i] == 0 && alpha5[i] == 1) p01 else 0
alpha25.c[i] <- ALPHA25.c
ALPHA25.d <- if (alpha2[i] == 0 && alpha5[i] == 0) p00 else 0
```

```
alpha25.d[i] <- ALPHA25.d
}
Alpha25<-alpha25.a + alpha25.b + alpha25.c + alpha25.d
alpha25<-matrix(0,r,1)
for (i in 1:r){
ALPHA25<- rbinom(1,1,Alpha25[i,])
alpha25[i]<-ALPHA25
}
beta25a<-rnorm(r,0,a)
beta25b<-rnorm(r,0,b)
beta25 <- (1-alpha25)*beta25a + alpha25*beta25b

##Effects for Interaction 34##
alpha34.a<- matrix(0,r,1)
alpha34.b<- matrix(0,r,1)
alpha34.c<- matrix(0,r,1)
alpha34.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA34.a <- if (alpha3[i] == 1 && alpha4[i] == 1) p11 else 0
alpha34.a[i] <- ALPHA34.a
ALPHA34.b <- if (alpha3[i] == 1 && alpha4[i] == 0) p10 else 0
alpha34.b[i] <- ALPHA34.b
ALPHA34.c <- if (alpha3[i] == 0 && alpha4[i] == 1) p01 else 0
alpha34.c[i] <- ALPHA34.c
ALPHA34.d <- if (alpha3[i] == 0 && alpha4[i] == 0) p00 else 0
alpha34.d[i] <- ALPHA34.d
}
Alpha34<-alpha34.a + alpha34.b + alpha34.c + alpha34.d
```

```
alpha34<-matrix(0,r,1)
for (i in 1:r){
ALPHA34<- rbinom(1,1,Alpha34[i,])
alpha34[i]<-ALPHA34
}
beta34a<-rnorm(r,0,a)
beta34b<-rnorm(r,0,b)
beta34 <- (1-alpha34)*beta34a + alpha34*beta34b

##Effects for Interaction 35##
alpha35.a<- matrix(0,r,1)
alpha35.b<- matrix(0,r,1)
alpha35.c<- matrix(0,r,1)
alpha35.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA35.a <- if (alpha3[i] == 1 && alpha5[i] == 1) p11 else 0
alpha35.a[i] <- ALPHA35.a
ALPHA35.b <- if (alpha3[i] == 1 && alpha5[i] == 0) p10 else 0
alpha35.b[i] <- ALPHA35.b
ALPHA35.c <- if (alpha3[i] == 0 && alpha5[i] == 1) p01 else 0
alpha35.c[i] <- ALPHA35.c
ALPHA35.d <- if (alpha3[i] == 0 && alpha5[i] == 0) p00 else 0
alpha35.d[i] <- ALPHA35.d
}
Alpha35<-alpha35.a + alpha35.b + alpha35.c + alpha35.d
alpha35<-matrix(0,r,1)
for (i in 1:r){
ALPHA35<- rbinom(1,1,Alpha35[i,])
```

```
alpha35[i]<-ALPHA35
}
beta35a<-rnorm(r,0,a)
beta35b<-rnorm(r,0,b)
beta35 <- (1-alpha35)*beta35a + alpha35*beta35b

##Effects for Interaction 45##
alpha45.a<- matrix(0,r,1)
alpha45.b<- matrix(0,r,1)
alpha45.c<- matrix(0,r,1)
alpha45.d<- matrix(0,r,1)
for (i in 1:r){
ALPHA45.a <- if (alpha4[i] == 1 && alpha5[i] == 1) p11 else 0
alpha45.a[i] <- ALPHA45.a
ALPHA45.b <- if (alpha4[i] == 1 && alpha5[i] == 0) p10 else 0
alpha45.b[i] <- ALPHA45.b
ALPHA45.c <- if (alpha4[i] == 0 && alpha5[i] == 1) p01 else 0
alpha45.c[i] <- ALPHA45.c
ALPHA45.d <- if (alpha4[i] == 0 && alpha5[i] == 0) p00 else 0
alpha45.d[i] <- ALPHA45.d
}
Alpha45<-alpha45.a + alpha45.b + alpha45.c + alpha45.d
alpha45<-matrix(0,r,1)
for (i in 1:r){
ALPHA45<- rbinom(1,1,Alpha45[i,])
alpha45[i]<-ALPHA45
}
beta45a<-rnorm(r,0,a)
```

```
beta45b<-rnorm(r,0,b)
beta45 <- (1-alpha45)*beta45a + alpha45*beta45b

##Matrix of Simulated Priors##
prior.beta<-c(beta0,beta1,beta2,beta3,beta4,beta5,beta12,
beta13,beta14,beta15,beta23,beta24,beta25,beta34,beta35,beta45)
beta.r<-matrix(prior.beta,e,r,byrow=T)
Beta.r <- beta.r[-c(1),]

##Errors##
sim.error1<-rnorm(n1*r,0,sd)
E1<-matrix(sim.error1,n1,r)

##r Sets of Simulated Data Y1##
Y1<-X1%*%beta.r + E1

##Linear Model for Y##
x1<-X1[,2]
x2<-X1[,3]
x3<-X1[,4]
x4<-X1[,5]
x5<-X1[,6]
x12<-X1[,7]
x13<-X1[,8]
x14<-X1[,9]
x15<-X1[,10]
x23<-X1[,11]
x24<-X1[,12]
```

```
x25<-X1[,13]
x34<-X1[,14]
x35<-X1[,15]
x45<-X1[,16]
lm(Y1~x1+x2+x3+x4+x5+x12+x13+x14+x15+x23+x24+x25+x34+x35+x45)
model1<-lm(Y1~x1+x2+x3+x4+x5+x12+x13+x14+x15+x23+x24+x25+x34+x35+x45)
design1coef<-matrix(print(coef(model1)),e,r)
effects1<-design1coef[-c(1),]

############################################################

D <- as.numeric(abs(Beta.r)>omega)
D
Delta.star<- matrix(D,e-1,r)
Delta.star

##Declaring True Active Effects##
no.actives<-matrix(0,r,1)
for (i in 1:r){
no.actives[i]<-length(Delta.star[,i][Delta.star[,i]==1])
}
no.actives

############################################################

##Method 1: if effects > omega, then declare active##

##Design 1##
```

```
D1 <- as.numeric(abs(effects1)>omega)
Delta1<- matrix(D1,e-1,r)

############################################################

##Method 2: Lenth's method##

Median1 <- matrix(0,1,r)
for (i in 1:r){
med1 <- median(abs(effects1[,i]))
Median1[i] <- med1
}
s0.1 <- 1.5*Median1
T1<- matrix(0,e-1,r)
for (i in 1:e-1){
for (j in 1:r){
t.1 <- ifelse((abs(effects1)[i,j] < 2.5*s0.1[,j]),1,0)
T1[i,j] <- t.1
}}
N1<- T1*effects1
PSE<-matrix(0,1,r)
for (i in 1:r){
PSE[i]<- 1.5 * median(abs(N1[,i]))
}
t.975.d<-2.57
t.gamma.d<- 5.22
ME1<-t.975.d*PSE
SME1<- t.gamma.d * PSE
```

```
ME<-matrix(ME1,1,r)
SME<-matrix(SME1,1,r)
Delta.1.lenth<-matrix(0,t,r)
for (i in 1:t){
for (j in 1:r){
Delta.1.lenth[i,j]<-as.numeric(abs(effects1[i,j])>SME[,j])
}
}
```

B.2 Code for Expected Utility

```
##Utilities using Method 1##
actives.decl.active1<-matrix(0,t,r)
actives.decl.inactive1<-matrix(0,t,r)
inactives.decl.active1<-matrix(0,t,r)
inactives.decl.inactive1<-matrix(0,t,r)
for (j in 1:r)
  {
  correct1 <- as.numeric(Delta.star[,j]==Delta1[,j])
  actives.decl.active1[,j] <- Delta.star[,j]*correct1
  actives.decl.inactive1[,j] <- Delta.star[,j]*(1-correct1)
  inactives.decl.active1[,j] <- (1-Delta.star[,j])*(1-correct1)
  inactives.decl.inactive1[,j] <- (1-Delta.star[,j])*correct1
  }
actives.decl.active1
actives.decl.inactive1
inactives.decl.active1
inactives.decl.inactive1

true.pos1 <- ifelse(no.actives>0, sum(actives.decl.active1)/
(no.actives*r), 0)
type2.error1<-ifelse(no.actives>0, sum(actives.decl.inactive1)/
(no.actives*r), 0)
type1.error1<-ifelse(t-no.actives>0, sum(inactives.decl.active1)/
((t-no.actives)*r), 0)
true.neg1<-ifelse(t-no.actives>0, sum(inactives.decl.inactive1)/
((t-no.actives)*r), 0)
```

```
true.pos1
type2.error1
type1.error1
true.neg1
u2.1.1<-type2.error1
u1.1.1<-type1.error1
gamma<-0.5
U.1.1<-gamma*u1.1.1 + (1-gamma)*u2.1.1
U.1.1

############################################################

##Utilities using Method 2##
l1.1<-matrix(0,t,r)
l2.1<-matrix(0,t,r)
l3.1<-matrix(0,t,r)
l4.1<-matrix(0,t,r)
for (j in 1:r)
  {
  correct3 <- as.numeric(Delta.star[,j]==Delta.1.lenth[,j])
  l1.1[,j] <- Delta.star[,j]*correct3
  l2.1[,j] <- Delta.star[,j]*(1-correct3)
  l3.1[,j] <- (1-Delta.star[,j])*(1-correct3)
  l4.1[,j] <- (1-Delta.star[,j])*correct3
  }
l1.1
l2.1
```

```
l3.1
l4.1

true.pos3<-ifelse(no.actives>0, sum(l1.1)/(no.actives*r), 0)
type2.error3<-ifelse(no.actives>0, sum(l2.1)/(no.actives*r), 0)
type1.error3<-ifelse(t-no.actives>0, sum(l3.1)/((t-no.actives)*r), 0)
true.neg3<-ifelse(t-no.actives>0, sum(l4.1)/((t-no.actives)*r), 0)

true.pos3
type2.error3
type1.error3
true.neg3

u2.1.2<-type2.error3
u1.1.2<-type1.error3
gamma<-0.5
U.1.2<-gamma*u1.1.2 + (1-gamma)*u2.1.2
U.1.2

############################################################

##Type I and II errors##
mean(u1.1.1) # type1 error - Design 1 method 1
mean(u1.1.2) # type1.error - Design 1 method 2
mean(u2.1.1) # type2 error - Design 1 method 1
mean(u2.1.2) # type2.error - Design 1 method 2

##Expected Utilities##
```

```
mean(U.1.1) # design 1 method 1
mean(U.1.2) # design 1 method 2
```

Appendix C

WinBUGS Code for Bayesian Analysis

C.1 Normal Model on Data, Normal Prior Distribution on Effect Parameters and Uniform Prior on Log σ

```
model
{
for( i in 1 : N ) {
y[i] ~ dnorm(mu[i],tau)            # model for data

mu[i]<- beta0 + beta[1]*x1[i] + beta[2]*x2[i] + beta[3]*x3[i]
+ beta[4]*x4[i] + beta[5]*x5[i] + betaint[1,2]*x1[i]*x2[i]
+ betaint[1,3]*x1[i]*x3[i] + betaint[1,4]*x1[i]*x4[i]
```

```
+ betaint[1,5]*x1[i]*x5[i] + betaint[2,3]*x2[i]*x3[i]
+ betaint[2,4]*x2[i]*x4[i] + betaint[2,5]*x2[i]*x5[i]
+ betaint[3,4]*x3[i]*x4[i] + betaint[3,5]*x3[i]*x5[i]
+ betaint[4,5]*x4[i]*x5[i]
}
beta0 ~ dnorm(0.0,0.000001)        # prior distribution for intercept
for (j in 1:5) {
beta[j] ~ dnorm(0.0,0.001)         # prior distribution for main effects
}
for (j in 1:4){
for (k in (j+1):5){
betaint[j,k] ~ dnorm(0.0,0.001)    # prior distribution for
}                                  # two-factor interactions
}

log.sigma ~ dunif(-10, 10)         # prior for log sigma
sigma <- exp(log.sigma)
sigma.sq <- pow(sigma, 2)
tau <- 1/sigma.sq
}

##Data##
list(y = c(56, 53, 63, 65, 53, 55, 67, 61, 69, 45, 78, 93,
49, 60, 95, 82),
x1 = c(-1,  1, -1,  1, -1,  1, -1,  1, -1,  1, -1,  1, -1,  1, -1,  1),
x2 = c(-1, -1,  1,  1, -1, -1,  1,  1, -1, -1,  1,  1, -1, -1,  1,  1),
x3 = c(-1, -1, -1, -1,  1,  1,  1,  1, -1, -1, -1, -1,  1,  1,  1,  1),
x4 = c(-1, -1, -1, -1, -1, -1, -1, -1,  1,  1,  1,  1,  1,  1,  1,  1),
```

```
x5 = c( 1, -1, -1,  1, -1,  1,  1, -1, -1,  1,  1, -1,  1, -1, -1,  1),
N = 16)

##Initial values for chain 1##
list(beta0 = 0, beta=c(0, 0, 0, 0, 0) , log.sigma = 1)
betaint=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
##Initial values for chain 2##
list(beta0 = 0, beta=c(0, 0, 0, 0, 0) , log.sigma = 0.1)
betaint=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
```

C.2 Scaled-t Model on Data, Normal Prior Distribution on Effect Parameters and Gamma Prior on Precision

```
model
{
for( i in 1 : N ) {
y[i] ~ dt(mu[i], tau, d)          # model for data

mu[i]<- beta0 + beta[1]*x1[i] + beta[2]*x2[i] + beta[3]*x3[i]
+ beta[4]*x4[i] + beta[5]*x5[i] + betaint[1,2]*x1[i]*x2[i]
+ betaint[1,3]*x1[i]*x3[i] + betaint[1,4]*x1[i]*x4[i]
+ betaint[1,5]*x1[i]*x5[i] + betaint[2,3]*x2[i]*x3[i]
+ betaint[2,4]*x2[i]*x4[i] + betaint[2,5]*x2[i]*x5[i]
+ betaint[3,4]*x3[i]*x4[i] + betaint[3,5]*x3[i]*x5[i]
```

```
+ betaint[4,5]*x4[i]*x5[i]
}
beta0 ~ dnorm(0.0,0.000001)       # prior distribution for intercept
for (j in 1:5) {
beta[j] ~ dnorm(0.0,0.001)        # prior distribution for main effects
}
for (j in 1:4){
for (k in (j+1):5){
betaint[j,k] ~ dnorm(0.0,0.001)   # prior distribution for
}                                 # two-factor interactions
}

tau ~ dgamma(0.001,0.001)         # prior on precision
sigma<- 1/sqrt(tau)
d <- 4                            # degrees of freedom
}

##Data##
list(y = c(56, 53, 63, 65, 53, 55, 67, 61, 69, 45, 78, 93,
49, 60, 95, 82),
x1 = c(-1,  1, -1,  1, -1,  1, -1,  1, -1,  1, -1,  1, -1,  1, -1,  1),
x2 = c(-1, -1,  1,  1, -1, -1,  1,  1, -1, -1,  1,  1, -1, -1,  1,  1),
x3 = c(-1, -1, -1, -1,  1,  1,  1,  1, -1, -1, -1, -1,  1,  1,  1,  1),
x4 = c(-1, -1, -1, -1, -1, -1, -1, -1,  1,  1,  1,  1,  1,  1,  1,  1),
x5 = c( 1, -1, -1,  1, -1,  1,  1, -1, -1,  1,  1, -1,  1, -1, -1,  1),
N = 16)

##Initial values for chain 1##
```

```
list(beta0 = 0, beta=c(0, 0, 0, 0, 0) , tau = 0)
betaint=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
##Initial values for chain 2##
list(beta0 = 0,  beta=c(0, 0, 0, 0, 0) , tau =1)
betaint=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
```

C.3 Normal Model on Data, Normal Mixture Prior Distribution on Effect Parameters and Gamma Prior on Precision

```
model
{
for (i in 1:N) {
y[i] ~ dnorm(mu[i], tau)          # model for data

mu[i]<- beta0 + beta[1]*x1[i] + beta[2]*x2[i] + beta[3]*x3[i]
+ beta[4]*x4[i] + beta[5]*x5[i] + betaint[1,2]*x1[i]*x2[i]
+ betaint[1,3]*x1[i]*x3[i] + betaint[1,4]*x1[i]*x4[i]
+ betaint[1,5]*x1[i]*x5[i] + betaint[2,3]*x2[i]*x3[i]
+ betaint[2,4]*x2[i]*x4[i] + betaint[2,5]*x2[i]*x5[i]
+ betaint[3,4]*x3[i]*x4[i] + betaint[3,5]*x3[i]*x5[i]
+ betaint[4,5]*x4[i]*x5[i]
}
beta0 ~ dnorm(0.0,0.0001)              # prior for intercept
```

```
for (j in 1:15){
beta[j] ~ dnorm(0, prec[j])            # prior distribution for main effects
prec[j] <- T[j]*(tau2 - tau1) + tau1
u[j] ~ dcat(P[])                       # indicator variable:
                                       # 1 if effect inactive
                                       # 2 if effect active
T[j] <- u[j] - 1
}
for (j in 1:4){
for (k in (j+1):5){
betaint[j,k] ~ dnorm(0.0,prec[j])      # prior distribution for
}                                      # two-factor interactions
}

P[1] <-0.85                        # probability effect inactive
P[2] <-0.15                        # probability effect active
tau1<- 1/(sigma1*sigma1)           # precision for inactive effect
tau2<- 1/(sigma2*sigma2)           # precision for active effect
sigma1 ~ dunif(0, 1)               # standard deviation for inactive effect
sigma2 ~ dunif(0, 10)              # standard deviation for active effect

tau ~ dgamma(1, 1)                 # prior on precision
sigma<- 1/sqrt(tau)
}

##Data##
list(y = c(56, 53, 63, 65, 53, 55, 67, 61, 69, 45, 78, 93,
49, 60, 95, 82),
```

```
x1 = c(-1,  1, -1,  1, -1,  1, -1,  1, -1,  1, -1,  1, -1,  1, -1,  1),
x2 = c(-1, -1,  1,  1, -1, -1,  1,  1, -1, -1,  1,  1, -1, -1,  1,  1),
x3 = c(-1, -1, -1, -1,  1,  1,  1,  1, -1, -1, -1, -1,  1,  1,  1,  1),
x4 = c(-1, -1, -1, -1, -1, -1, -1, -1,  1,  1,  1,  1,  1,  1,  1,  1),
x5 = c( 1, -1, -1,  1, -1,  1,  1, -1, -1,  1,  1, -1,  1, -1, -1,  1),
N = 16)

##Initial values for chain 1##
list(beta0 = 0,  beta=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0), tau = 0)
##Initial values for chain 2##
list(beta0 = 0,  beta=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0), tau = 1)
```

Bibliography

[1] Abraham, B. and Box, G. E. P. (1978) Linear models and spurious observations. *Applied Statistics.* Vol. 27, No. 2, pp. 131-138.

[2] Allen, T. T. and Bernshteyn, M. (2003) Supersaturated designs that maximize the probability of identifying active factors. *Technometrics.* Vol. 45, No. 1, pp. 90-97.

[3] Anderson-Cook, C. (2005) How to choose the appropriate design. *Quality Progress.* Vol. 38, No. 10, pp. 80-82.

[4] Atkinson, A. C., Donev, A. N. and R. D. Tobias (2007) *Optimum Experimental Designs, with SAS.* Oxford University Press.

[5] Bailey, R. A., (2008) *Design of Comparative Experiments.* Cambridge University Press.

[6] Bernardo, J. M. (1996). The concept of exchangeability and its applications. *Far East Journal of Mathematical Sciences.* Vol.4, pp. 111-121.

[7] Box, G. E. P. and Draper, N. R. (1987) *Empirical Model-Building and Response Surfaces.* New York: Wiley.

[8] Box, G. E. P. and Hunter, J. S. (1961a) The 2^{k-p} fractional factorial designs Part I. *Technometrics.* Vol. 3, No. 3, pp. 311-351.

[9] Box, G. E. P. and Hunter, J. S. (1961b) The 2^{k-p} fractional factorial designs Part II. *Technometrics.* Vol. 3, No. 4 pp. 449-458.

[10] Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978) *Statistics for Experimenters.* New York: Wiley.

[11] Box, G. E. P. and Meyer, R. D. (1986a) An analysis for unreplicated fractional factorials. *Technometrics.* Vol. 28, No. 1, pp. 11-18.

[12] Box, G. E. P. and Meyer, R. D. (1986b) Dispersion effects from factorial designs. *Technometrics.* Vol. 28, No. 1, pp. 19-27.

[13] Box, G. E. P. and Meyer, R. D. (1993) Finding the active factors in fractionated screening experiments. *Journal of Quality Technology.* Vol. 25, No. 2, pp. 94-105.

[14] Box, G. E. P. and Tiao, G. C. (1968) A Bayesian approach to some outlier problems. *Biometrika.* Vol. 55, pp. 119-129.

[15] Casella, G. and George, E. (1992) Explaining the Gibbs sampler. *The American Statistician.* Vol. 46, pp. 167-174.

[16] Chaloner, K. and Verdinelli, I. (1995) Bayesian experimental design: A Review. *Statistical Science.* Vol. 10, No. 3, pp. 273-304.

[17] Chipman, H. (1996) Bayesian variable selection with related predictors. *The Canadian Journal of Statistics.* Vol. 24, pp. 17-36.

[18] Chipman, H., Hamada, M. and Wu, C. F. J. (1997) A Bayesian variable selection approach for analyzing designed experiments with complex aliasing. *Technomet-*

rics. Vol. 39, pp. 372-381.

[19] Colosimo, B. M. and del Castillo, E. (Ed.) (2006) *Bayesian Process, Monitoring, Control and Optimization.* Chapman and Hall/CRC.

[20] Curnow, R. N. (1972) The number of variables when searching for an optimum. *Journal of the Royal Statistical Society, Series B.* Vol. 34, No. 3, pp. 461-476.

[21] Daniel, C. (1959) Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics.* Vol. 1, No. 4, pp. 311-341.

[22] Daniel, C. (1960) Locating outliers in factorial experiments. *Technometrics.* Vol. 2, No. 2, pp.149-156.

[23] Daniels, M. J. (1999) A prior for the variance in hierarchical models. *Canadian Journal of Statistics.* Vol. 27, No. 3, pp. 567-578.

[24] Davies, O. L. (Ed.) (1960) *The Design and Analysis of Industrial Experiments.* Oliver and Boyd.

[25] DeGroot, M. H. (1970) *Optimal Statistical Decisions.* McGraw-Hill, New York.

[26] Finney, D. J. (1945) The fractional replication of factorial arrangements. *Annals of Eugenics* Vol. 12, pp. 291-301.

[27] Fisher, R. A. (1925) *Statistical Methods for Research Workers.* Oliver and Boyd.

[28] Fisher, R. A. (1935) *The Design of Experiments.* Oliver and Boyd.

[29] French, S. (1986) *Decision Theory: An introduction to the mathematics of rationality.* Ellis Horwood, Chichester.

[30] Fries, A. and Hunter, W. G. (1980) Minimum aberration 2^{k-p} designs. *Technometrics.* Vol. 22, pp. 601-608.

[31] Gelfand, A., Hills, S., Racine-Poon, A. and Smith, A. (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association.* Vol. 85, pp. 972-985.

[32] Gelfand, A. and Smith, A. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association.* Vol. 85, pp. 398-409.

[33] Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis.* Vol. 1, No. 3, pp. 515-533.

[34] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003) *Bayesian Data Analysis.* Second Edition. Chapman and Hall, London.

[35] Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Vol. 6, pp. 721-741.

[36] George, E. I. and McCulloch, R. E. (1993) Variable selection via Gibbs sampling. *Journal of the American Statistical Association.* Vol. 88, pp. 881-889.

[37] Gilmour, S. G. (2001) Contribution to the discussion of 'Detection of interactions in experiments on large numbers of factors' by S. M. Lewis and A. M. Dean. *Journal of the Royal Statistical Society, Series B.* Vol. 63, Part 4, pp. 662-663.

[38] Gilmour, S. G. and Mead, R. (1995) Stopping rules for sequences of factorial designs. *Applied Statistics.* Vol. 44, No. 3, pp. 343-355.

[39] Gilmour, S. G. and Mead, R. (2003) A Bayesian design criterion for locating the optimum point on a response surface. *Statistics & Probability Letters.* Vol. 64, No. 3, pp. 235-242.

[40] Hamada, M. and Wu, C. F. J. (1992) Analysis of designed experiments with complex aliasing. *Journal of Quality Technology.* Vol. 24, 130-137.

[41] Hastings, W. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika.* Vol. 57, pp. 97-109.

[42] Lenth, R. V. (1989) Quick and easy analysis of unreplicated factorials. *Technometrics.* Vol. 31, No. 4, pp. 469-473.

[43] Lewis, S. M. and Dean, A. M. (2001) Detection of interactions in experiments on large numbers of factors. *Journal of the Royal Statistical Society, Series B.* Vol. 63, Part 4, pp. 633-672.

[44] Lindley, D. V. (1956) On the measure of information provided by an experiment. *Annals of Mathematical Statistics.* Vol. 27, No. 4, pp. 986-1005.

[45] Lindley, D. V. (1972) *Bayesian Statistics - A Review.* SIAM, Philadelphia.

[46] Marley, C. J. and Woods, D. C. (2010) A comparison of design and model selection methods for supersaturated experiments. *Computational Statistics and Data Analysis.* Vol. 54, Issue 12, pp. 3158-3167.

[47] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953) Equations of state calculations by fast computing machine. *Journal of Chemical Physics.* Vol. 21, pp. 1087-1092.

[48] Meyer, R. D., Steinberg, D. M. and Box, G. E. P. (1996) Follow-up designs to

resolve confounding in multifactor experiments. *Technometrics.* Vol. 38, No. 4, pp. 303-313.

[49] Mitchell, T. J. and Beauchamp, J. J. (1988) Bayesian variable selection in linear regression. *Journal of the American Statistical Association.* Vol. 83, pp.1023-1036.

[50] Müller, P., Berry, D. A., Grieve, A. P. and Krams, M. (2006) A Bayesian decision-theoretic dose finding trial. *Decision Analysis.* Vol. 3, No. 4, pp. 197-207.

[51] Nobile, A. and Green, P. J. (2000) Bayesian analysis of factorial experiments by mixture modelling. *Biometrika.* Vol. 87, pp. 15-35.

[52] Olguin, J. and Fearn, T. (1997) A new look at half-normal plots for assessing the significance of contrasts for unreplicated factorials. *Applied Statistics.* Vol. 46, No. 4, pp. 449-462.

[53] Plackett, R. L. and Burman, J. P. (1946) The design of optimum multifactorial experiments. *Biometrika.* Vol. 33, pp. 305-325.

[54] Raiffa, H. and Schlaifer, R. (1961) *Applied Statistical Decision Theory.* Division of Research, Harvard Business School, Boston.

[55] Richardson, S. and Green, P. J. (1997) On the Bayesian analysis of mixtures with an unknown number of components (with Discussion). *Journal of the Royal Statistical Society B.* Vol. 59, pp. 731-792.

[56] Shannon, C. E. (1948) A mathematical theory of communication. *Bell System Technical Journal.* Vol. 27, pp. 379-423, 623-656.

[57] Smith, A. and Roberts, G. (1993) Bayesian computation via the Gibbs sampler

and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B.* Vol. 55, pp. 3-23.

[58] Spiegelhalter, D., Thomas, A., Best, N. and Gilks, W. (1996) *BUGS 0.5: Bayesian Inference Using Gibbs Sampling Manual.* Version 1.4, MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK.

[59] Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003) *WinBUGS User Manual.* Version 1.4, MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology and Department of Epidemiology and Public Health, Imperial College School of Medicine, UK. Available at http://www.mrc-bsu.cam.ac.uk/bugs.

[60] Taguchi, G. and Wu, Y. (1980) *Introduction to Off-Line Quality Control.* Central Japan Quality Control Association, Nagoya, Japan.

[61] Taguchi, G. (1986) *Introduction to Quality Engineering.* Asian Productivity Organization, UNIPUB, White Plains, New York.

[62] Tsai, P.-W., Gilmour, S.G. and Mead, R. (2000) Projective three-level main effects designs robust to model uncertainty. *Biometrika* Vol. 87, pp. 467-475.

[63] Tsai, P.-W., Gilmour, S.G. and Mead, R. (2004) Some new three-level orthogonal main effects designs robust to model uncertainty. *Statistica Sinica* Vol. 14, pp. 1075-1084.

[64] Wackerly, D. D., Mendenhall, W. and Scheaffer, R. L. (2002) *Mathematical Statistics with Applications.* Sixth Edition. Duxbury Press.

[65] Wu, C. F. J. and Chen, Y. (1992) A graph-aided method for planning two-level experiments when certain interactions are important. *Technometrics.* Vol. 34,

pp. 162-175.

[66] Wu, C. F. J. and Hamada, M. (2000) *Experiments: Planning, Analysis, and Parameter Design Optimization.* John Wiley & Sons.

[67] Yates, F. (1937) The design of factorial experiments. *Imperial Bureau of Soil Science.* Tech. Communication. 35.

[68] Zahn, D. A. (1975a) Modifications of and revised critical values for the half-normal plot. *Technometrics.* Vol. 17, No. 2, pp. 189-200.

[69] Zahn, D. A. (1975b) An empirical study of the half-normal plot. *Technometrics.* Vol. 17, No. 2, pp. 201-211.