UNIVERSITY OF SOUTHAMPTON

Faculty of Engineering, Science and Mathematics
School of Mathematics

Two-Stage Group Screening

by

Anna-Jane Elizabeth Vine

Thesis submitted for the degree of Doctor of Philosophy
August 2004

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
SCHOOL OF MATHEMATICS

Doctor of Philosophy

TWO-STAGE GROUP SCREENING

by Anna-Jane Elizabeth Vine

The work in this thesis focuses on product improvement experiments in which the response or product performance can depend on the levels of a number of different factors. Factorial experiments are used which allow the joint effects of the different factors to be investigated and thus provide valuable information on possible interactions between the factors. As the number of factors to be investigated increases, the number of observations needed for such an experiment can rapidly become economically infeasible. One approach which aims to achieve a practical number of runs is to group factors together. New grouped factors are defined to represent each group and these grouped factors are investigated in a first stage experiment. The individual factors within those groups found to be important in the first stage experiment are then investigated in a second stage experiment.

This thesis presents new theory for this group screening approach in which the group sizes are unequal and individual control main effects, noise main effects, control×noise interactions and control×control interactions are assumed to be active with possibly different probabilities. Examples are presented which demonstrate the investigation of different grouping strategies and group sizes through an examination of their impact on the distribution of the predicted number of effects that require estimation in the two-stage experiment.

This theory has been used in the planning of a two-stage group screening experiment at Jaguar Cars whose aim is to identify the key or active factors that influence engine cold start performance. The experiment is described and the results are presented.

An analysis based on the Bayesian method of Stochastic Search Variable Selection (SSVS) is used to investigate an alternative screening strategy of using supersaturated designs. SSVS is first investigated in a small simulation study which compares results from a full factorial experiment, analysed using SSVS, with those from a half-replicate. The results show that if an active effect is totally aliased with a non-active effect, then the prior information determines which effect is identified as active. SSVS is compared with all-subsets regression through an investigation of the performance of a particular supersaturated design using simulation. This study indicates that the performance of SSVS is similar to that of all-subsets regression in the number of active effects that are missed. Finally, a two-stage group screening strategy and the use of a supersaturated design are also compared through simulated experiments. In this study, two-stage group screening was more successful in identifying the active effects than the supersaturated design, but at the cost of more observations.

Contents

List of Figures

List of Tables

Acknowledgements

I would firstly like to thank my supervisor, Professor Susan Lewis, for her continual support, encouragement, guidance and enthusiasm. I would like to thank Professor Angela Dean for helpful discussions and for the use of her simulation software and Dr David Woods for many helpful discussions, and help with computing.

I would like to thank EPSRC and the three industrial collaborators, Jaguar Cars, Hossiden Besson and Goodrich. In particular, I thank David Brunson from Jaguar Cars for enabling a two-stage group screening experiment to be run and for arranging financial support for me.

I would also like to thank Professor Jonathan Forster for discussions regarding the work in Chapter 5, and Professor Hugh Chipman for his advice and the use of his code for SSVS.

I also thank my boyfriend Matthew Farey for his encouragement and care, my good friend Jonathan Jackson for his help in the final stages of thesis preparation and my parents for their constant support.

Chapter 1

Background

This thesis is concerned with experiments for product improvement in which a large number of factors require investigation. Factorial designs are used for these experiments as they are able to assess, simultaneously, the joint influence of the different factors on the response and to provide valuable information on interactions. As the number of factors to be investigated increases, the number of runs needed in an experiment can rapidly become economically infeasible. The work in this thesis investigates two-stage group screening as a way of carrying out experiments to identify the most important factors in a practical number of runs. As in most screening experiments, all factors are assumed to have two levels.

In Section 1.1 an introduction to factorial experiments is given. Section 1.2 gives an introduction to screening in industrial experiments and selected literature on screening experiments is briefly reviewed in Section 1.3. Group screening is described in Section 1.4 and an overview of the work in this thesis is given in Section 1.5.

1.1 Factorial experiments

A continuous response $\mathbf{Y}$ in a factorial experiment depends on the levels of a number of different factors. In this thesis a linear model is assumed to describe the relationship between these factors and the response, namely

$$\mathbf{Y} = \boldsymbol{\tau} + \boldsymbol{\epsilon}, \tag{1.1}$$

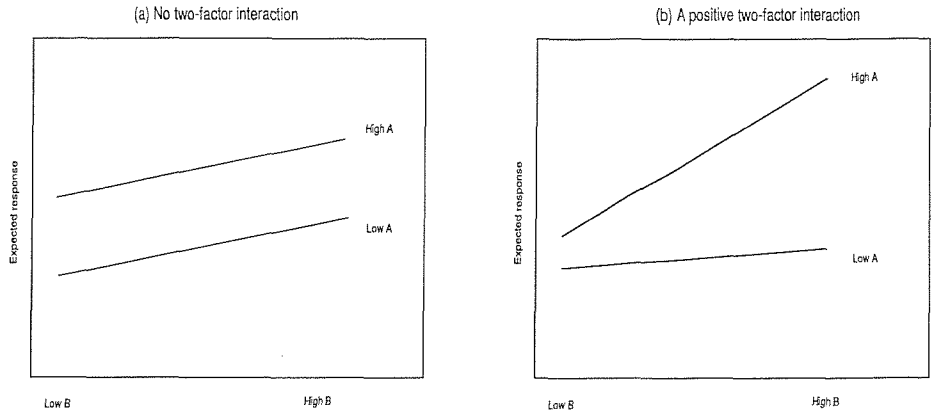


Figure 1.1: Illustration of the situation where two factors A and B do not interact together in (a) and where they do interact in (b).

where $\boldsymbol{\tau}$ is a vector whose fixed entries are the effects of the combinations of the levels of the factors and $\boldsymbol{\epsilon}$ is a vector of independent, normally distributed error variables, each with mean zero and variance σ^2.

In a factorial experiment, two types of contrasts amongst the effects are of interest: main effects and interactions. The main effect of a factor is the difference between the average of the treatment effects when the factor is set at its high level, and at its low level (where average is over the other factor levels). Factors may also interact with each other to influence the response. When factors A and B interact together, the change in the average treatment effect produced by moving from the low to the high level of factor A is different for the two levels of factor B, as illustrated in Figure 1.1(b). The case of no interaction is illustrated in Figure 1.1(a).

In a factorial experiment involving $m > 1$ two-level factors $A_1, ..., A_m$, the low and high levels of factor A_i are often denoted respectively by 0 and 1, or by -1 and $+1$. In a full factorial experiment, the treatments are formed as all 2^m combinations $a_1, a_2, ..., a_m$ of the levels of each of the factors, where a_i denotes the level of factor A_i, for $i = 1, ..., m$. Let $\tau_{a_1 a_2 ... a_m}$ denote the effect of the treatment combination $a_1 a_2 ... a_m$. Then $\boldsymbol{\tau}$, in equation (1.1), contains all the 2^m treatment effects listed, by convention, in lexicographical order, so that

$$\boldsymbol{\tau}' = (\tau_{0...00}, \tau_{0...01}, \tau_{0...10}, ..., \tau_{1...11}).$$

The main effects and interactions, known collectively as the factorial effects, can each be formulated as $C^x\boldsymbol{\tau}$, where the contrast matrix C^x is defined as follows for the main effect of factor A_i,

$$\begin{aligned} C^x &= \frac{1}{2^m} C_2^{x_1} \otimes C_2^{x_2} \otimes \cdots \otimes C_2^{x_m} \\ &= \frac{1}{2^m} \otimes_{k=1}^{m} C_2^{x_k} \end{aligned} \tag{1.2}$$

where $x = x_1 x_2 ... x_m$ is an m-digit binary number, with $x_i = 1$ and $x_k = 0$ for $k \neq i$ $(1 \leq k \leq m)$, with

$$C_2^{x_k} = \begin{cases} 2I_2 - J_2 & \text{if } x_k = 1 \\ \mathbf{1}_2' & \text{if } x_k = 0 \end{cases}$$

and where $\otimes$ denotes the Kronecker product. For factors A_i and A_j, the contrast matrix for testing the two-factor interaction A_iA_j can be expressed as

$$C^x = \frac{1}{2^m} \otimes_{k=1}^{m} C_2^{x_k}, \tag{1.3}$$

where $x_i = 1$, $x_j = 1$ and $x_k = 0$ for $k \neq i, j$ $(1 \leq k \leq m)$. Interactions between three or more factors are defined similarly; see, for example, John and Williams (1995), page 162.

The size of a full factorial experiment can become too large for use in practice when the number of factors is large. This is because a small increase in the number of factors results in a rapid increase in the number of combinations of factor levels. For example, for five factors, each having two levels, there are $2^5 = 32$ possible combinations, but for seven such factors there are 128 combinations. It can then be necessary to sacrifice some of the information on interactions in order to keep the size of the experiment feasible, by using a subset or fraction of the possible treatment combinations, i.e. a fractional factorial experiment. In choosing a design for such an experiment, interactions believed to be negligible are often deliberately made indistinguishable from main effects and two-factor interactions, as illustrated by the following small example of a regular fractional factorial design.

Example 1.1.1 Consider three factors A_1, A_2 and A_3, each taking two levels. A full factorial experiment for these factors requires $2^3 = 8$ runs:

A_1	A_2	A_3
0	0	0
0	0	1
0	1	0
1	0	0
0	1	1
1	0	1
1	1	0
1	1	1

If it was infeasible to use all eight runs, a $\frac{1}{2}$-replicate could be used, for example:

A_1	A_2	A_3
0	0	0
1	1	0
1	0	1
0	1	1

The main effect of A_1 is given by the contrast

$$1/4[(\tau_{111} + \tau_{100} + \tau_{110} + \tau_{101}) - (\tau_{011} + \tau_{000} + \tau_{010} + \tau_{001})]$$

where τ_{ijk} is the effect of factor A_1 at its ith level, factor A_2 at its jth level and factor A_3 at its kth level (from (1.2)). The main effect of A_1 cannot be estimated from the four runs used as observations are not available on four of the treatments; for example, there is no observation on 010. The interaction between factors A_2 and A_3 is, from (1.3), the contrast

$$A_2A_3 = 1/4[(\tau_{111} + \tau_{011} + \tau_{100} + \tau_{000}) - (\tau_{010} + \tau_{001} + \tau_{101} + \tau_{110})].$$

This interaction also cannot be estimated from the four runs used. However, the treatment contrast $A_1 - A_2A_3$ is

$$1/4[\tau_{110} + \tau_{101} - \tau_{011} - \tau_{000} - \tau_{011} - \tau_{000} + \tau_{101} + \tau_{110}] = 1/2[\tau_{110} + \tau_{101} - \tau_{000} - \tau_{011}],$$

which can be estimated from the four runs in the fraction. The same is true for the contrasts $A_3 - A_1A_2$ and $A_2 - A_1A_3$. The main effect A_1 is said to be **aliased** with A_2A_3. This relationship is denoted $A_1 = A_2A_3$. Similarly, A_3 is aliased with A_1A_2 and A_2 is aliased with A_1A_3. The main effects A_1, A_2 and A_3 can be estimated from the experiment provided that the two-factor interactions can be assumed to be negligible.

There is no way of estimating the three-factor interaction

$$A_1A_2A_3 = 1/4[(\tau_{100} + \tau_{010} + \tau_{001} + \tau_{111}) - (\tau_{000} + \tau_{110} + \tau_{101} + \tau_{011})],$$

because the effects of the four treatments in the fraction all have the same sign in the contrast $A_1A_2A_3$. Thus $A_1A_2A_3$ evaluated for the runs in the fraction is indistinguishable from the mean. $A_1A_2A_3$ is called the defining contrast for the fraction and the fraction is identified by, and can be constructed from, its defining relation $I = A_1A_2A_3$.

If two factorial effects are aliased, then the correlation between their least squares estimators is 1 or -1 and these estimators are identical, as in the above example. In some (non-regular) fractions, estimators of two factorial effects may be non-zero and have value strictly between 1 and -1. Such factorial effects are said to be partially aliased. If two factorial effects are not aliased, either partially or fully, then their estimators are uncorrelated.

The number of runs needed for a fractional factorial experiment to estimate all main effects and interactions of interest can still be too large when the number of factors is large. In this situation, pilot studies or engineering knowledge in the problem area can sometimes be used to reduce the number of factors in the experiment to a manageable size. In the absence of such information, an initial screening experiment can be used to identify the important or **active** factors. The next step in an investigation is to approximate the relationship (response surface) between the response and the levels of the factors using one or more further experiments. The fitted response surface can then be used to find, for example, combinations of the levels of the factors that give an optimum response.

1.2 Screening in industrial experiments

When the aim of an experiment is to improve the quality of a product, particularly in an industrial setting, the factors involved may be of two types: **control** or **noise**. A **control factor** is assigned particular values in the specification of the product or manufacturing process and can be well controlled. **Noise factors** typically arise in the manufacturing process or the environment where the product

is used, where they cannot be controlled. They can, however, be well controlled or simulated in an experiment in order to investigate the variation they cause in product performance. Experiments investigating both control and noise factors are used to identify the levels of the control factors which improve the quality of a product through attaining both the required mean performance and insensitivity, or robustness, in the performance when the noise factors vary (see, for example, Deming (1986), Taguchi (1986) and Wu and Hamada (2000)).

A common approach to investigating a large number of factors is that of **classical screening**. This strategy typically involves two phases of experimentation. First, only main effects are investigated. Then, in a second experiment, interactions between those factors whose main effects are identified as substantial at the first stage are investigated. This method is useful when interactions exist only between factors with large main effects. In practice, however, interactions may involve one or more factors that have negligible main effects. A control factor having a small main effect but involved in an important interaction could fail to be identified during the first experiment. The interaction would then not be investigated in the second experiment. Hence, information on important interactions may be unavailable from classical screening methods. This loss can be a serious problem for industrial experimentation where identification of interactions between control and noise factors plays a fundamental role in product improvement by allowing a reduction in variation in the response (product performance) due to the changing values of the noise factors (see, for example, Shoemaker et al. (1991) and Wu and Hamada (2000)).

1.3 Literature

Although an experimenter might identify a long list of factors for experimentation, it is not unusual to find that the number of important factorial effects is small. This was called the **effect sparsity principle** by Box and Meyer (1986) and is based on the idea that a large proportion of variation in a response can be explained by a small proportion of the factors. This principle has had an important influence

in the literature and in the development of screening experiments whose aim is to identify these few crucial factors. In this section, various approaches to screening are discussed, and the relevant literature is briefly reviewed.

1.3.1 Supersaturated designs

A **supersaturated design** is a fractional factorial design with m factors and n_0 runs, where $n_0 < m + 1$ so that there are not enough runs to estimate all main effects contrasts as well as the mean. Such designs have been proposed to identify large main effects using a small number of runs, when a model with main effects only is assumed and the effect sparsity principle is also assumed to hold. There has been a great deal of work on these designs in the literature in recent years, which is now reviewed. A major difficulty has been the analysis of data from such designs and this issue is also considered.

Construction of designs

Satterthwaite (1959) suggested constructing supersaturated designs by random construction of the design matrix, that is the matrix whose (k, l)th entry is the level of the lth factor in the kth run, $1 \leq k \leq n_0$, $1 \leq l \leq m$. That is, the combinations of factor levels to include in the design are selected at random. In these designs, and many of the later designs, only designs where each factor had an equal number of '+1' and '-1' values were considered. These designs were called random balance designs.

The first systematic construction of supersaturated designs was given by Booth and Cox (1962) who used the criterion of the minimisation of

$$max|\boldsymbol{c_i}'\boldsymbol{c_j}|,$$

where $\boldsymbol{c_i}'\boldsymbol{c_j}$ is the inner product of two columns $\boldsymbol{c_i}$ and $\boldsymbol{c_j}$ of the design matrix, and the maximum is evaluated over $1 \leq i < j \leq m$. Note that when $\boldsymbol{c_i}'\boldsymbol{c_j} = 0$, columns $\boldsymbol{c_i}$ and $\mathbf{c_j}$ are said to be **orthogonal**. Seven supersaturated designs were found ranging from 16 factors in 12 runs to 36 factors in 18 runs using a computer

search. Booth and Cox (1962) also proposed the criterion of the minimisation of

$$E(s^2) = \sum_{i<j} s_{ij}^2 / \binom{m}{2}, \tag{1.4}$$

where $s_{ij} = \boldsymbol{c}_i' \boldsymbol{c}_j$. This criterion gives an intuitive measure of non-orthogonality of a design and has been used in many papers on the construction of supersaturated designs.

Interest in the design construction problem was renewed by Lin (1993) and Wu (1993). Lin proposed a class of supersaturated designs constructed using half fractions of Hadamard matrices. A Hadamard matrix H_N is a square matrix of order N, whose first column consists of 1s and whose remaining $N-1$ columns are orthogonal to the first column. Each of these latter columns have half of their entries '+1' and the remaining entries '-1'. These matrices have the property that $H_N' H_N = H_N H_N' = N I_N$ (where I_N is the identity matrix of order N). Lin (1993) used a 12-run Plackett and Burman design for 11 factors, which is a particular type of Hadamard matrix, to illustrate the construction method for supersaturated designs. In this method a 'branching column' is selected to provide the basis for splitting the Plackett and Burman design into two parts. The rows of the design matrix can then be split into two groups: the first group contains rows with a '+1' in the branching column and the second contains rows with a '-1' in the branching column. Removal of the branching column results in the two groups both being supersaturated designs which can examine up to $N-2$ factors using $N/2$ runs. Lin constructed supersaturated designs for experiments where $N \leq 60$ and investigated all possible choices of branching columns. Comparisons with designs given by Satterthwaite (1959) and Booth and Cox (1962) were made, and Lin's designs were found to be more efficient when judged by the criterion of minimising $E(s^2)$, as given in (1.4).

Wu (1993) also considered the use of $N \times N$ Hadamard matrices in the construction of supersaturated designs. His method removed the first column from the $N \times N$ Hadamard matrix to construct a design capable of estimating $N-1$ factors in N runs (a saturated design). Then he added interaction columns, formed as the products of two of the columns of the saturated design. These additional columns can then be used for studying extra factors. This allowed the construction

of supersaturated designs in which there are many more factors than runs.

A more flexible, algorithmic approach to design construction was developed by Lin (1995). This involved an algorithm which searches through possible, or candidate, columns of a design and identifies the maximum number of factors that can be accommodated in a particular number of runs, n_0, specified by the user. The degree of orthogonality between two columns $\boldsymbol{c_i}$ and $\boldsymbol{c_j}$ of the design matrix was measured by $r_{ij} = s_{ij}/n$. A design can be assessed, or two designs compared, by considering the largest absolute value of r_{ij}, denoted by r, among all pairs of columns for a given design. In Lin's algorithm, through specification of a maximum value for r, the user is able to specify an upper bound on the degree of non-orthogonality in the columns retained for a design. If n_0 is even, the algorithm generates all possible columns which contain $(n_0/2)$ occurrences of the factor at level '+1' and $(n_0/2)$ occurrences of '-1'; for example, if $n_0 = 10$, then $10!/(5! \times 5!) = 252$ columns are generated. If n_0 is odd, then the algorithm generates all possible columns with $(n_0 + 1)/2$ occurrences of a factor at its high level and $(n_0 - 1)/2$ at its low level; for example, $n_0 = 15$ gives $15!/(7! \times 8!) = 6435$ possible columns. The algorithm then selects a subset of these columns to include in a design by considering each of the columns, in turn, as follows. A candidate column is entered into the design and its inner product is calculated with each of the other columns currently in the design. If the maximum of these r_{ij} values is greater than the specified r-value, then the column is not added to the design and the search continues. Once the design has been constructed, the algorithm reorders the columns so that the columns 'nearest to orthogonality' are presented first. This is achieved by minimising the average of the r_{ij}^2, called the mean squared correlation,

$$\rho^2 = \sum_{i<j} r_{ij}^2 / \binom{m}{2}.$$

This criterion is equivalent to the $E(s^2)$ criterion given by Booth and Cox (1962), since $\rho^2 = E(s^2)/n_0^2$.

Lin (1995) showed that designs could be obtained for $3 \leq n_0 \leq 25$ and $0 \leq r \leq 1/3$ which allowed up to 276 factors to be included in the design. He also showed that designs found by his algorithm were better than those given by Booth and Cox (1962) and by Satterthwaite (1959), under the criterion of minimising $E(s^2)$.

Nguyen (1996) described a method of constructing supersaturated designs from balanced incomplete block designs and gave a lower bound for $E(s^2)$ which can be used when m is divisible by n_0. He also gave a general algorithm for finding supersaturated designs. This algorithm constructs a starting design by assigning at random a value of '+1' to half of the entries of each column and '-1' to the remaining entries. The value of $E(s^2)$ is then calculated. The impact on the value of $E(s^2)$ of exchanging the signs of pairs of entries (having opposite signs) within each column is considered. The signs of those pairs whose exchanges result in the largest reduction in the value of $E(s^2)$ are exchanged. This step is repeated until $E(s^2) = 0$ or $E(s^2)$ reaches Nguyen's lower bound or $E(s^2)$ cannot be reduced by further sign exchanges. The designs obtained were found to have similar performance to those of Booth and Cox (1962), Lin (1993, 1995) and Wu (1993), under the criterion of minimising $E(s^2)$.

A further method of constructing designs that uses a search algorithm was described by Allen and Bernshteyn (2003). The algorithm used was a genetic, rather than an exchange algorithm. A different criterion for constructing supersaturated designs was proposed. The authors assigned probabilities of factors being active and non-active and chose designs to maximise the probability of correctly identifying the active factors. Use of this criterion allowed them to identify some $E(s^2)$-optimal designs.

Butler, Mead, Eskridge and Gilmour (2001) observed that methods of constructing designs that minimise $E(s^2)$ (called $E(s^2)$-optimal designs) had only been determined for n_0-run experiments when the number of factors m is a multiple of $n_0 - 1$, and when $m = q(n_0 - 1) + t$ ($|t| \leq 2$ and q is an integer). In order to extend the range of $E(s^2)$-optimal designs, they derived a lower bound for $E(s^2)$ for $m = q(n_0 - 1) + t$ factors and gave a method of constructing supersaturated designs that attain this lower bound. Their approach was to use $E(s^2)$-optimal designs for small numbers of runs as 'building blocks' to construct designs for larger numbers of runs and factors.

A class of supersaturated designs called 'k-circulant' supersaturated designs were defined by Liu and Dean (2004). These designs are constructed from a set of one or more cyclic generators (which are rows of the design) by a cyclic shift

of k elements within each row. They gave generators which allow the construction of designs that either achieve the minimum value of $E(s^2)$ and are close to the minimum value of r_{max} (where r_{max} is the maximum value of r_{ij} over all pairs of columns), or achieved the minimum value of r_{max} and are close to the minimum value of $E(s^2)$. They also gave conditions on the columns of such a design that allow additional columns to be added which can accommodate a specific set of interactions. A consequence of the conditions is that the run size of the resulting supersaturated design must be a multiple of 4. This paper, and that of Wu (1993), appear to be the only published work that give instances of the construction of supersaturated designs for examining interactions.

Analysis of supersaturated designs

The complicated partial aliasing amongst the factorial effects in supersaturated designs and the insufficient degrees of freedom do not allow the traditional regression or analysis of variance techniques to be used. Several methods of analysis, that is, of identifying the important or active factors, have been proposed. The literature on frequentist methods is reviewed below. An alternative Bayesian approach is reviewed in Chapter 5.

Wu (1993) suggested using forward selection or all-subsets regression. Forward stepwise regression was proposed by Lin (1993) who argued that this method would identify the active factors by detecting large main effects, provided that any interaction effects are comparatively small. In this method, after each variable is entered into the model through an F-test, all variables already in the model are checked to see whether they should be removed. Lin used this procedure to identify the important factors in a data set for one of his designs, obtained by selecting the observations for the appropriate runs from the 28-run design of Williams (1968) for 24 factors, and their conclusions were very similar. For screening experiments, Lin (1995) recommended the use of conservative significance levels on the grounds that failing to detect an active factor can have more serious consequences than incorrectly identifying a non-active factor as active.

Lin also investigated the use of ridge regression in which the problem of the information matrix being singular is overcome by the addition of a multiple of the

identity matrix before matrix inversion (Draper and Smith (1998), Chapter 17). However, a simulation study by Lin showed that ridge regression did not perform well for supersaturated designs.

Some risks in the use, and particularly in the analysis of supersaturated designs were highlighted by Abraham, Chipman and Vijayan (1999). In particular, they focussed on forward selection and all-subsets regression procedures. They extended the study of Lin (1993) by considering 8 different supersaturated designs for a design of 23 factors in 14 runs, all of which were constructed by Lin's branching column method. Again the data were obtained as the relevant subset of the data from Williams' experiment. Forward regression was performed on the data for all eight designs and the results compared with Lin's conclusions. The main finding was that different designs can lead to identification of different factors as being active. The authors also described two simulation studies which are relevant to the work in Chapter 5. The first study was of a 14-run design for 23 factors given by Lin (1993). Data from experiments using this design were simulated by generating values for 14 random variables $\epsilon_i (i = 1, ..., 14)$ from a $N(0, 1)$ distribution, and then using the linear model

$$y_i = \beta_0 + \sum_{j=1}^{23} \beta_j x_{ij} + \epsilon_i, \qquad i = 1, 2, ..., 14,$$

to generate responses for specific chosen values of the vector of model parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_{14})'$. These data were then analysed using forward selection (halted after five steps) and the factors selected for the model were recorded. This process was repeated 200 times. The results showed that, even when the active effects were very large, the forward selection procedure could mislead by identifying the wrong factors as active (that is, as being included in the final fitted model). The simulation was also performed using all-subsets regression for model selection, where the search was restricted to models with 3, 4, and 5 factors. Unlike stepwise regression, all-subsets regression is able to give a choice of models of a given size that fitted the simulated data well. This allowed an assessment of this method of analysis to be made by calculating the proportion of times that the highest ranked model contained the correct model, and the proportion of times that the correct model was in the top five best fitting models in the simulated experiments. The authors found

that this method identified the correct model more often than forward stepwise regression.

The second study showed that the results of the analysis depend on which columns of the design matrix correspond to the active factors. This study used the 12-run design for 60 factors of Lin (1995). Five active factors were assigned to a different set of columns in the design matrix from those used in the simulation study given by Lin (1995). Lin's study had indicated that the five active factors were always selected in the first five steps of a forward selection procedure. Abraham et al. (1999) simulated experiments on this design, as in the first study above. Data were generated for each of the twelve runs and a forward selection procedure was used. In 200 simulated experiments, only one of the active factors was detected in the first five steps of the selection procedure.

In order to overcome the shortcomings in the above methods of analysing data from supersaturated experiments, a Bayesian approach was developed by Chipman, Hamada and Wu (1997), which is based on the Stochastic Search Variable Selection (SSVS) of George and McCulloch (1993) and incorporates the hierarchical priors of Chipman (1996). Discussion of these papers is postponed until Chapter 5 where work is described that uses the SSVS technique.

1.3.2 Other screening techniques

In this subsection, some screening techniques other than group screening and supersaturated designs, are briefly described.

Cheng and Wu (2001) argued that factor screening could be performed in conjunction with response surface exploration in a single experiment. This would save time and resources, compared with carrying out two separate experiments sequentially. This strategy involves performing two separate analyses of the data from the single experiment. The experiment has factors at three levels and, in the first analysis, is used to investigate a moderately large number of factors (up to 13) assuming a main effects only model. Those factors found to be active in this analysis then become the focus of a second analysis of the data set, in which the experiment is viewed as one in which the main effects of these active factors and the two-factor

interactions between them are studied. Hence a key aspect in their method is the projection of a larger factor space onto a smaller factor space. An example is given where the effects of nine factors are studied in a 3^{9-6} design. After the analysis at the first stage, consisting of an examination of main effects only, five factors were identified as being important. The projected design onto three of these factors was used to fit a second-order model, and a two-factor interaction was found to be significant. It was further concluded that two other factors, out of the five, were also likely to have been found to be important on the grounds of their partial aliasing with the significant two-factor interaction. Further analysis showed these two factors were not important once the interaction was included in the model. It was suggested that the original findings gave faulty conclusions due to the omission of interactions in the model building at the first stage.

Trocine and Malone (2001) described the Trocine Screening Procedure (TSP) which is a screening technique for a large number of factors (typically more than 20). This approach used a genetic algorithm to generate runs for the experiment and to build experiments sequentially, using the results from analysing observations on the earlier runs to choose the next set of four runs in the experiment. Those factors which are not found to be active are discarded and the next iteration is made. The approach could require many iterations and appears to be used primarily for simulation experiments, which are not considered here.

1.4 Group screening

1.4.1 Background

The idea of grouping in the context of screening began with Dorfman (1943) who considered the biological problem of the detection of a rare defect among the members of a large population by testing blood samples. He suggested that pooled blood samples should be tested and that the individual samples that formed a pool (or group) should only be tested when the pool gave a positive result. He showed that a substantial saving in the total number of tests could be made by choice of the number of samples to pool together. After this early work, the pool-

ing technique was developed for a variety of applications, for example, genetic and medical screening. Watson (1961) was the first to apply this approach to factorial experiments, where only the main effects were considered and all interactions were assumed to be negligible. For a review of group screening see Kleijnen (1987) and Ankenman and Dean (2001). The idea of grouping has also been incorporated into techniques for simulation experiments, see, for example, Cheng (1997), but these are not considered in this thesis.

In group screening for factorial experiments, individual factors are placed into groups and a new grouped factor is defined to represent each group (see Section 2.2). An experiment using these grouped factors forms the first stage of the full group screening experiment and its aim is to find the important grouped factors. The individual factors within these important groups are then investigated in a second stage experiment.

Two different two-stage group screening strategies were investigated by Lewis and Dean (2001) and Dean and Lewis (2002) under the assumption that all groups of factors have the same size and that individual factorial effects of the same type are assumed to be active with the same probability, these types being control main effects, noise main effects, control×noise interactions and control×control interactions. The motivation for their work was industrial experimentation in which both control and noise factors were involved. The first strategy investigated was **classical group screening (CGS)** in which only main effects of the grouped factors are estimated at the first stage of the experiment. The second strategy was **interaction group screening (IGS)**, where both the main effects of the grouped factors and the interactions between pairs of grouped factors are estimated at the first stage of the experiment. For both strategies, the second stage experiment is used to investigate main effects and interactions of interest among the individual factors within those grouped factors found to be active at the first stage.

Lewis and Dean (2001) defined an active factorial effect in terms of a quantity Δ that is regarded as a substantive difference in the responses between two combinations of levels of the grouped factors. A particular grouped factorial effect $\boldsymbol{c}^x\boldsymbol{\tau}$ was defined as **active** if $|\boldsymbol{c}^x\boldsymbol{\tau}| > \Delta$ and inactive if $|\boldsymbol{c}^x\boldsymbol{\tau}| \leq \Delta$, where $\boldsymbol{c}^x$ is a row of the matrix C^x, and where the contrasts are scaled to have the same variance as

the estimator of a treatment difference. The data from the first stage experiment is used to decide whether or not a factorial effect is active, for example, using t- or F-tests. This definition of active includes the definitions of other authors when $\Delta = 0$, but can also be used with $\Delta \gg 0$, when only larger improvements are of interest.

In classical group screening, a grouped factor is considered active at the first stage if it has an active main effect. At the second stage, the grouping of factors found to be active at the first stage is dismantled. All individual control×control and control×noise interactions amongst these factors are then examined.

In interaction group screening, a grouped control factor is active if it is in at least one active grouped control×noise or control×control interaction, or if it has an active main effect. A grouped noise factor is only active if it is involved in at least one active interaction with a grouped control factor. Noise main effects and noise×noise interactions are not of interest for the purpose of further screening. At the second stage of experimentation, the individual factors in the grouped factors declared active at the first stage, are examined. Within each grouped control factor, the main effects of the individual factors and the individual control×control interactions are examined. Individual control×noise and control×control interactions are only examined if their corresponding grouped interactions were declared active in the analysis of the data from the first stage experiment.

The simulation software described in Lewis and Dean (2001) and Dean and Lewis (2002) assessed the risk of failing to detect active effects by simulating a two-stage group screening experiment for CGS and IGS. The assumption is again made of equal group sizes and equal probabilities of individual factorial effects of the same type being active, as in Lewis and Dean (2001). The user specifies the numbers of grouped control and noise factors and the fixed group sizes. The user also specifies the proportions of individual control main effects, noise main effects, control×control, control×noise and noise×noise interactions that are to be made active in the simulated experiment. The user can also input their own design for the first stage experiment, or choose to let the algorithm select one appropriate to the number of grouped factors from the table of Russell, Lewis and Dean (2004). The algorithm selects at random, and as close as possible to the

specified proportions, the individual main effects and two-factor interactions to be made active. The sizes of the non-active effects are determined by a random sample from a $N(0, (\Delta/3)^2)$ distribution with probability q (user-specified) and from a $N(0, 1)$ distribution with probability $(1 - q)$. The values of the effects that are to be made active are determined by sampling from either a $N(\mu, \delta^2)$ or a $N(-\mu, \delta^2)$ distribution with equal probability, where μ and δ are specified by the user. Proportions of effects correctly identified by, hypothesis testing, as active are calculated for the simulated data sets using both the CGS and IGS strategies, so that the risk of failing to detect active effects can be assessed against the average total number of observations needed for the entire two-stage experiment. The user can then compare CGS and IGS under different choices of groupings, first stage designs, proportions of effects that are to be made active in the simulated experiment, Δ and active effect distributions. In all of the examples in Dean and Lewis (2002), classical group screening required fewer observations but performed less well than interaction group screening in terms of missing active effects. This type of simulation is used in Chapter 4.

The theory of Lewis and Dean (2001) applied to equally sized groups of factors and equal probabilities of individual factorial effects of the same type being active. However, these assumptions are too restrictive for many practical applications. In many industrial settings, it is not always possible to partition the individual factors involved into groups of equal sizes. Further, the number of individual factors may not factorise conveniently. Also, the requirement of equal probabilities does not allow detailed prior information from subject specialists on the relative likely importance of factors to be used. An overall aim of this thesis is to develop an approach which does not have these restrictions.

1.5 Outline of this thesis

In this thesis the theoretical results of Lewis and Dean (2001) for the situation where the group sizes are unequal and individual control main effects, noise main effects, control×noise interactions and control×control interactions are assumed to be active with possibly different probabilities. This increased flexibility makes

two-stage group screening a more versatile screening strategy. The generalised theory is presented in Chapter 2. This theory has been implemented in software, which is described in Chapter 3, together with a number of studies illustrating its capabilities. These studies show how strategies for group screening and group sizes can be assessed. The theory and software described in Chapters 2 and 3 have been used in the planning of a two-stage group screening experiment at Jaguar Cars. The planning of this experiment and results from the analysis are presented in Chapter 4. Chapter 5 considers the alternative screening strategy of using a supersaturated design. An investigation of these designs under a Bayesian analysis is presented using simulation. A critical comparison of a group screening design and a supersaturated design is also made, again using a Bayesian variable selection technique. Finally, in Chapter 6, conclusions are presented and some possible future work arising from this thesis is briefly described.

Chapter 2

Group screening with flexible group sizes and probabilities

2.1 Overview

In this chapter, theoretical results are obtained for the predicted distribution of the number of effects requiring estimation in a two-stage group screening experiment where there may be unequal group sizes and individual control main effects, noise main effects, control×noise interactions and control×control interactions are assumed to be active with possibly different probabilities. These results are mathematically more complex than those for two-stage group screening experiments where the group sizes are equal and individual factorial effects of the same type are assumed to be active with the same probability, these types being control main effects, noise main effects, control×noise interactions and control×control interactions (Lewis and Dean (2001)). This is due to the loss of exchangeability in the factors. When counting the number of effects that need to be estimated in the second stage experiment, it is now necessary to know exactly which grouped factors were declared active at the first stage and, in the case of interaction group screening, the reason why they were declared active.

In Section 2.2, the necessary notation is set up for the investigation of unequal group sizes and unequal probabilities of effects being active. Formulae for the probabilities of grouped effects being active and being declared active are developed in terms of the probabilities of individual effects being active. Criteria are given for

consideration when choosing group sizes and screening strategies. In Section 2.3, an explicit formula is obtained for the expected number of effects requiring estimation in a two-stage experiment under classical group screening (CGS), and an example is given to illustrate the use of the formula. The form of the distribution and the variance of the number of effects requiring estimation is also obtained. In Section 2.4, the corresponding results for interaction group screening (IGS) are derived. All of the theory presented in this chapter apply also to the special case where there are no noise factors under consideration. In Section 2.5, alternative methods for assigning interaction probabilities are given. Conclusions are presented in Section 2.6. The work in this chapter is presented in Vine, Lewis and Dean (2004) together with some of the work in Chapter 3 .

2.2 Grouping of factors and probabilities for grouped factorial effects

The labelling of individual factors is now changed, compared with Chapter 1, in order to make explicit the grouping of the factors. The factors are labelled

$$A_{11}, A_{12}, ..., A_{1g_1}, A_{21}, A_{22}, ..., A_{2g_2}, ..., A_{b1}, A_{b2}, ..., A_{bg_b},$$

and each takes two levels. The individual factors are divided into b groups in such a way that the ith group, represented by the grouped factor B_i $(i = 1, ..., b)$ contains individual factors $A_{i1}, A_{i2}, ..., A_{ig_i}$. When all factors in the ith group are set to their high (low) level then grouped factor B_i is at its high (low) level. Only grouped factors formed only from individual control (noise) factors are considered so that grouped control×noise interactions can be investigated.

Let q_{ik} denote the probability that the main effect of individual factor A_{ik} is active, and let $q_{ik,jl}$ denote the probability that the two-factor interaction between individual factors A_{ik} and A_{jl} is active (as defined in Section 1.4.1). Suppose that F grouped factors labelled $B_1^{(c)}, ..., B_F^{(c)}$ are each formed from individual control factors and the remaining N grouped factors labelled $B_1^{(n)}, ..., B_N^{(n)}$ are formed from individual noise factors. The sizes of the corresponding groups are denoted by $g_1^{(c)}, ..., g_F^{(c)}, g_1^{(n)}, ..., g_N^{(n)}$, respectively. The total number of individual control factors

and individual noise factors are, respectively,

$$n_C = \sum_{i=1}^{F} g_i^{(c)} \text{ and } n_N = \sum_{i=1}^{N} g_i^{(n)}.$$

2.2.1 Probabilities of grouped main effects and interactions being active

We make the following simplifying assumptions:

(a) the main effects and interactions of individual factors are independently active or non-active

(b) any non-active effect is zero

(c) the level of each individual factor that is labelled 'high' produces the higher response.

Grouped main effect probabilities:

We denote the probabilities of the grouped main effects being active by

$$\rho_1^{(c)}, ..., \rho_F^{(c)}, \rho_1^{(n)}, ..., \rho_N^{(n)},$$

where the first F probabilities correspond to main effects for grouped control factors and the remaining N probabilities correspond to main effects for grouped noise factors. Using assumption (a), we find these probabilities by subtracting the product of the probabilities that the main effect of each individual factor within the group is not active from 1, giving

$$\rho_i^{(c)} = 1 - \prod_{A_{ik} \in B_i^{(c)}} (1 - q_{ik}); \qquad k = 1, ..., g_i^{(c)}; 1 \leq i \leq F \tag{2.1}$$

and

$$\rho_j^{(n)} = 1 - \prod_{A_{jk} \in B_j^{(n)}} (1 - q_{jk}); \qquad k = 1, ..., g_j^{(n)}; 1 \leq j \leq N. \tag{2.2}$$

Following Lewis and Dean (2001), let $p_i^{(c)}$ ($p_j^{(n)}$) be the probability that the analysis of the data from the first stage experiment leads to the main effect of the ith grouped control (jth grouped noise) factor being declared active. If it were possible to always detect active main effects, and non-active main effects were always close

to zero, so that correct conclusions were always made, then $p_i^{(c)}$ ($p_j^{(n)}$) would be equal to $\rho_i^{(c)}$ ($\rho_j^{(n)}$) for $1 \leq i \leq F$ ($1 \leq j \leq N$). Otherwise, the value of $p_i^{(c)}$ ($p_j^{(n)}$) is likely to be in the region of $\rho_i^{(c)}$ ($\rho_j^{(n)}$).

Grouped interaction probabilities:

Under interaction group screening we need to calculate the probabilities of grouped interactions being active because, in the first stage experiment, each of the FN grouped control×noise interactions is investigated. Using assumption (a), the probability of the grouped control×noise interaction $B_i^{(c)}B_j^{(n)}$ being active is found by subtracting the probabilities that all of the interactions between the individual factors from each group are inactive from 1, giving

$$\rho_{i,j}^{(cn)} = 1 - \prod_{A_{ik} \in B_i^{(c)}} \prod_{A_{jl} \in B_j^{(n)}} (1 - q_{ik,jl}) \tag{2.3}$$

where $k = 1, ..., g_i^{(c)}$; $l = 1, ..., g_j^{(n)}$; $1 \leq i \leq F$ and $1 \leq j \leq N$.

In the first stage experiment the $\binom{F}{2}$ grouped control×control interactions are also investigated. Hence calculation is also needed of the probability of the grouped control×control interaction $B_i^{(c)}B_k^{(c)}$ being active. By a similar argument to that used for $\rho_{i,j}^{(cn)}$,

$$\rho_{i,k}^{(cc)} = 1 - \prod_{A_{il} \in B_i^{(c)}} \prod_{A_{km} \in B_k^{(c)}} (1 - q_{il,km})$$

where $l = 1, ..., g_i^{(c)}$; $m = 1, ..., g_k^{(c)}$; and $1 \leq i < k \leq F$.

We can also calculate the probability $\rho_{j,l}^{(nn)}$ of the grouped noise×noise interaction $B_j^{(n)}B_l^{(n)}$ being active as

$$\rho_{j,l}^{(nn)} = 1 - \prod_{A_{jk} \in B_j^{(n)}} \prod_{A_{lm} \in B_l^{(n)}} (1 - q_{jk,lm})$$

where $k = 1, ..., g_j^{(n)}$; $m = 1, ..., g_l^{(n)}$; and $1 \leq j < l \leq N$.

Let $p_{i,j}^{(cn)}$ be the probability that the analysis from the first stage experiment leads to the interaction between the ith grouped control factor and jth grouped noise factor being declared active (where $1 \leq i \leq F$ and $1 \leq j \leq N$), with similar definitions for $p_{i,k}^{(cc)}$ and $p_{j,l}^{(nn)}$ (where $1 \leq i < k \leq F; 1 \leq j < l \leq N$). As for the grouped main effect probabilities, if it were possible to always draw the correct

conclusions in the first stage analysis, then $p_{i,j}^{(cn)}$ would be equal to $\rho_{i,j}^{(cn)}$. Similarly, $p_{i,k}^{(cc)}$ and $p_{j,l}^{(nn)}$ would equal $\rho_{i,k}^{(cc)}$ and $\rho_{j,l}^{(nn)}$, respectively. For the work in this thesis we use the ρ_i and $\rho_{i,j}$ values to approximate the p_i and $p_{i,j}$ values.

2.2.2 Criteria and their implementation

The total number of effects requiring estimation in a two-stage group screening experiment depends on the designs used at each stage and the choice of probabilities and grouping at stage 1. The number of effects to be estimated in the first stage experiment is determined by the number of grouped factors of each type. However, the number of effects to be estimated at the second stage depends on the conclusions of the analysis of the first stage experiment. This leads to the total number of effects that require estimation over the two-stage experiment being regarded as a random variable whose probability distribution depends on the groupings used, the choice of screening strategy and the choices of probabilities of different factorial effects being active. This random variable is denoted by S and, in later sections, we use S_{CGS} when the strategy is classical group screening and S_{IGS} when the strategy is interaction group screening.

In order to choose between the strategies and the group sizes for the first stage experiment a variety of criteria can be considered:

1. minimise the expected total number of effects to be estimated in the two-stage experiment

2. minimise the standard deviation of the total number of effects to be estimated, in addition to criterion 1

3. minimise the probability that the two-stage experiment will require more than a target size, r, of observations

4. maximise the probability of detecting the active control main effects and interactions involving a control factor

5. minimise the probability of incorrectly identifying individual factorial effects as active when they are not (Type I error).

Criteria 1 and 3 are concerned with keeping the number of effects requiring estimation small. Criterion 2 is closely related to criterion 1, and ideally we would hope to achieve these two criteria simultaneously. Criteria 4 and 5 are concerned with improving the reliability with which conclusions are drawn from the experiment and can be achieved by increasing the number of observations. Hence, the aims of these criteria conflict with requirements 1 to 3 and so a trade off between them is necessary.

The research I have carried out mainly concerns criteria 1 to 3. However, I have been able to examine criterion 4 through using other outputs from this project for an example (see Chapter 4). I have not considered the fifth criterion in this thesis since it leads to minimising the number of effects requiring estimation in the experiment and therefore, it acts in the same direction as criterion 1. The role of the first three criteria in the selection of group sizes and strategy is illustrated in Chapter 3.

In order to implement criteria 1 to 3, ideally we would like to have the probability distribution and the expected value of the total number of effects that have to be estimated across stages 1 and 2. Finding a general form for this distribution is not possible because it depends on the choice of design for each of the two stages of experiment and, in particular, the aliasing structure of the design chosen. Instead, we develop theoretical results for a lower bound S on the total number of effects to be estimated at stage 1 and at stage 2. The lower bound is calculated under the assumption that the designs used at both stages allow estimation of all factorial effects of interest in the smallest number of runs possible (called minimal plans). It is assumed that interactions between three or more factors are negligible and that noise×noise interactions are not of interest and may be aliased together. A total aliasing amongst the grouped noise×noise interactions would be ideal because it would allow information from the experiment to be concentrated on the factorial effects of interest.

Because in general we cannot formulate the exact number of degrees of freedom that will correspond to the strings of aliased noise×noise interactions, we use the lower bound on the number of degrees of freedom given by Lewis and Dean (2001). If there are N_f noise factors then this bound is $N_f - 1$ and corresponds to the

ideal situation where there is as much aliasing as possible amongst the noise×noise interactions. Such an aliasing may not always be achievable in practice but S is calculated assuming that it is possible. When minimal plans are used (that is, designs which allow estimation of all factorial effects of interest in the smallest number of runs possible) the difference between the expected and actual number of effects to be estimated will be small. When experiments with larger numbers of runs are used, for example, full factorials, there will be a greater discrepancy. For example, a minimal plan for a first stage experiment under interaction group screening with three grouped control factors and two grouped noise factors would require 16 runs (see Section 2.4), or as close to 16 as possible. A full factorial design would use $2^5 = 32$ runs.

In this thesis the criteria 1 to 3 are implemented for the lower bound S which will serve as a surrogate for the experiment size and will be used to guide the choice of grouping and screening strategy (CGS or IGS). For simplicity, S will be referred to as the experiment size. In the remainder of this chapter, for each strategy, a closed form expression is obtained for $E(S)$. In addition, the form of the probability distribution for the number of individual factorial effects that have to be estimated at the second stage experiment is given, for specified group sizes and probabilities of individual effects being active. This expression allows computation of the probability distribution and also the variance of S.

2.3 Classical group screening

We first obtain an expression for $E(S)$ under classical group screening and give a small example of its use.

2.3.1 Number of effects for estimation at the second stage

The results of the first stage experiment are unknown when the experiment is being planned, so it is necessary to consider random variables when counting the number of effects to be estimated at the second stage experiment.

Note that if no grouped control factors are declared active at the first stage, then there are no groups of control factors to be investigated at the second stage. There-

fore, the experiment does not proceed to the second stage, because the purpose of the experiment is to find important individual control main effects, control×control interactions and in particular, important control×noise interactions.

We first count the number of main effects of individual control factors that have to be investigated in the second stage experiment. To do this we use various random indicator functions. Define a random indicator vector

$$\boldsymbol{I}^{(c)} = (\delta_1^{(c)}, \delta_2^{(c)}, ..., \delta_F^{(c)})$$

with kth entry equal to 1 when the corresponding grouped control factor $B_k^{(c)}$ is declared active after the first stage experiment, and equal to 0 otherwise. Similarly, let

$$\boldsymbol{I}^{(n)} = (\delta_1^{(n)}, \delta_2^{(n)}, ..., \delta_N^{(n)})$$

with kth entry equal to 1 when the corresponding grouped noise factor $B_k^{(n)}$ is declared active after the first stage experiment, and equal to 0 otherwise.

We use these indicator functions to form an expression for the number, $S_C^{(c)}$, of individual control main effects that have to be investigated at the second stage, under classical group screening. This is a random variable given by

$$S_C^{(c)} = \sum_{i=1}^{F} g_i^{(c)} \delta_i^{(c)}, \tag{2.4}$$

because the ith group of control factors contains $g_i^{(c)}$ individual factors $(1 \leq i \leq F)$. Here the subscript C means that the strategy used is classical group screening and the superscript c indicates that we are counting main effects of control factors.

In the same way, the number of individual control×control interactions to be investigated at the second stage is a random variable which can be written as

$$S_C^{(cc)} = \binom{S_C^{(c)}}{2}, \tag{2.5}$$

using the notation for a binomial coefficient. The number of individual noise main effects to be investigated at the second stage, denoted by $S_C^{(n)}$, depends on whether $S_C^{(c)}$ is zero or not. In particular, if $S_C^{(c)} = 0$ then the experiment stops after the first stage. If we define a indicator random variable by

$$\eta_C^{(c)} = 1 \text{ when } S_C^{(c)} \geq 1, \text{ and 0 otherwise,}$$

then

$$S_C^{(n)} = \eta_C^{(c)} \sum_{j=1}^{N} g_j^{(n)} \delta_j^{(n)}. \tag{2.6}$$

The number of sets of aliased individual noise×noise interactions that have to be included in the model for the second stage observations, given that at least one grouped control main effect is declared active after the first stage, depends on the number of individual noise main effects $S_C^{(n)}$ to be examined. We use the lower bound $S_C^{(n)} - 1$ of Lewis and Dean (2001) for this number of interactions. To formulate the number $S_C^{(nn)}$ of individual noise×noise interactions included at the second stage we again have to take account of whether or not the experiment stops at the first stage because no control factors are taken forward. This leads to

$$S_C^{(nn)} = S_C^{(n)} - \eta_C^{(n)}, \tag{2.7}$$

where $\eta_C^{(n)}$ is another indicator random variable defined by

$$\eta_C^{(n)} = 1 \text{ when } S_C^{(n)} \geq 1 \text{ and } 0 \text{ otherwise.}$$

The number of individual control×noise interactions that need to be studied at the second stage is formulated easily because this is simply given by the product of the numbers of individual control and noise factors brought forward to the second stage. This number is

$$S_C^{(cn)} = S_C^{(c)} S_C^{(n)}. \tag{2.8}$$

We sum (2.4) to (2.8) and $\eta_C^{(c)}$ to find a lower bound, $U_{CGS}^{(2)}$, on the number of effects which must be estimated at the second stage experiment under classical group screening, i.e.

$$\begin{aligned} U_{CGS}^{(2)} &= S_C^{(c)} + S_C^{(n)} + S_C^{(cc)} + S_C^{(cn)} + S_C^{(nn)} + \eta_C^{(c)} \\ &= S_C^{(c)} + S_C^{(n)} + S_C^{(c)} \frac{(S_C^{(c)} - 1)}{2} + S_C^{(c)} S_C^{(n)} + S_C^{(n)} - \eta_C^{(n)} + \eta_C^{(c)} \\ &= (S_C^{(c)} + 1)(\frac{1}{2} S_C^{(c)} + S_C^{(n)}) + S_C^{(n)} - \eta_C^{(n)} + \eta_C^{(c)}. \end{aligned} \tag{2.9}$$

Inclusion of $\eta_C^{(c)}$ is necessary to add an effect for the mean at the second stage, if the number of individual control factors brought forward to the second stage is non-zero.

2.3.2 Number of effects for estimation over both stages

At the first stage experiment only grouped main effects are estimated together with the mean under classical group screening. Hence the number of effects examined at stage 1 is

$$U_{CGS}^{(1)} = 1 + F + N. \tag{2.10}$$

Notice that this is a fixed quantity, i.e. not a random variable. We now use (2.9) and (2.10) to find an expression for the expected total number of effects to be examined in terms of the sizes of the grouped factors and the probabilities that the grouped main effects will be declared active, assuming maximum aliasing of the individual noise×noise interactions at the second stage:

$$\begin{aligned}
E(S_{CGS}) &= U_{CGS}^{(1)} + E(U_{CGS}^{(2)}) \\
&= 1 + F + N + \frac{1}{2}E(S_C^{(c)}) + 2E(S_C^{(n)}) + \frac{1}{2}E([S_C^{(c)}]^2) + E(S_C^{(c)}S_C^{(n)}) \\
&\quad + E(\eta_C^{(c)}) - E(\eta_C^{(n)}) \\
&= 2 + F + N + \frac{1}{2}\Sigma_{i=1}^{F} g_i^{(c)} p_i^{(c)} + 2[\Sigma_{j=1}^{N} g_j^{(n)} p_j^{(n)}][1 - \Pi_{i=1}^{F}(1 - p_i^{(c)})] \\
&\quad + \frac{1}{2}\Sigma_{i=1}^{F}[g_i^{(c)}]^2 p_i^{(c)}(1 - p_i^{(c)}) + \frac{1}{2}[\Sigma_{i=1}^{F} g_i^{(c)} p_i^{(c)}]^2 \\
&\quad + [\Sigma_{i=1}^{F} g_i^{(c)} p_i^{(c)}][\Sigma_{j=1}^{N} g_j^{(n)} p_j^{(n)}] - \Pi_{i=1}^{F}(1 - p_i^{(c)}) \\
&\quad - [1 - \Pi_{i=1}^{F}(1 - p_i^{(c)})][1 - \Pi_{j=1}^{N}(1 - p_j^{(n)})].
\end{aligned} \tag{2.11}$$

We now illustrate this result with a small example.

Example 2.3.1 Consider an experiment with 7 individual control factors and 2 individual noise factors. Two groupings of these factors are considered.

(i) $F = 2$ and $N = 1$ with $g_1^{(c)} = 4$, $g_2^{(c)} = 3$ and $g_1^{(n)} = 2$. Suppose that the probabilities of the individual control factors' main effects being active are

$$q_{11}^{(c)} = 0.2,\ q_{12}^{(c)} = 0.3,\ q_{13}^{(c)} = 0.4,\ q_{14}^{(c)} = 0.5,$$

$$q_{21}^{(c)} = 0.6,\ q_{22}^{(c)} = 0.7,\ q_{23}^{(c)} = 0.8,$$

and the probabilities of the individual noise factors' main effects being active are

$$q_{11}^{(n)} = 0.3,\ q_{12}^{(n)} = 0.3.$$

Then the probabilities of the grouped main effects being declared active are calculated from Section 2.2.1 with p for ρ as

$$\begin{aligned} p_1^{(c)} &= 1-(1-0.2)(1-0.3)(1-0.4)(1-0.5) = 0.832, \\ p_2^{(c)} &= 1-(1-0.6)(1-0.7)(1-0.8) = 0.976, \\ p_1^{(n)} &= 1-(1-0.3)(1-0.3) = 0.51. \end{aligned}$$

Then (2.11) gives $E(S_{CGS}) = 36.82$.

(ii) $F = 3$ and $N = 1$ with $g_1^{(c)} = 3$, $g_2^{(c)} = 2$, $g_3^{(c)} = 2$ and $g_1^{(n)} = 2$. Suppose that the probabilities of the individual control factors' main effects being active are as in the first grouping but with different labelling due to the change of grouping:

$$q_{11}^{(c)} = 0.2,\ q_{12}^{(c)} = 0.3,\ q_{13}^{(c)} = 0.4,$$

$$q_{21}^{(c)} = 0.5,\ q_{22}^{(c)} = 0.6,$$

$$q_{31}^{(c)} = 0.7,\ q_{32}^{(c)} = 0.8,$$

$$q_{11}^{(n)} = 0.3,\ q_{12}^{(n)} = 0.3.$$

The probabilities of the grouped main effects being declared active are calculated as before:

$$\begin{aligned} p_1^{(c)} &= 1-(1-0.2)(1-0.3)(1-0.4) = 0.664, \\ p_2^{(c)} &= 1-(1-0.5)(1-0.6) = 0.8, \\ p_3^{(c)} &= 1-(1-0.7)(1-0.8) = 0.94, \\ p_1^{(n)} &= 1-(1-0.3)^2 = 0.51. \end{aligned}$$

Then (2.11) gives $E(S_{CGS}) = 32.25$, a smaller value than for the previous grouping.

The issues of optimal choice of group sizes and sensitivity of $E(S_{CGS})$ to the choice of values for the probabilities of factorial effects being active under classical group screening, will be discussed in Chapter 3.

2.3.3 Probability distribution of S_{CGS} under classical group screening

The explicit formula (2.11) for the value of $E(S_{CGS})$ is useful to implement criterion 1, but this is not an adequate summary of the distribution of S_{CGS} as it does not give information on the spread of the distribution of S_{CGS} and the probability of exceeding a target size. However, it has not been possible to find an explicit closed form for the distribution of S_{CGS}. This is due to the complexity of formulating the very large numbers of possible scenarios which would cause S_{CGS} to take each particular value s. However, we can obtain an expression for the probability distribution via the random index vectors.

For any given experiment under classical group screening, the total number of effects $U^{(2)}_{CGS}$ to be estimated at the second stage is determined by which grouped control and noise factors are declared active at the first stage, and hence is determined by the realisations of the random index vectors $\boldsymbol{I}^{(c)}$ and $\boldsymbol{I}^{(n)}$. The probability, under classical group screening, that the total number of effects to be estimated S_{CGS} is equal to s, again assuming maximum aliasing, can be expressed as

$$P(S_{CGS} = s) = \sum_{R_{CGS}} P(\boldsymbol{I}^{(c)}_{t_1}, \boldsymbol{I}^{(n)}_{t_2})$$

where, using equation (2.10),

$$R_{CGS} = \{(\boldsymbol{I}^{(c)}_{t_1}, \boldsymbol{I}^{(n)}_{t_2}); U^{(2)}_{CGS} = s - (1 + F + N)\} \tag{2.12}$$

and where

$$1 + F + N \leq s \leq 1 + F + N + 2^{n_C + n_N}. \tag{2.13}$$

Under the simplifying assumptions (a) - (c) of Section 2.2.1 the joint probability function of $\boldsymbol{I}^{(c)}$ and $\boldsymbol{I}^{(n)}$ is

$$P(\boldsymbol{I}^{(c)} = \boldsymbol{I}^{(c)}_{t_1}, \boldsymbol{I}^{(n)} = \boldsymbol{I}^{(n)}_{t_2}) = \prod_{i=1}^{F} (p_i^{(c)})^{\delta^{(c)}_{i:t_1}} (1 - p_i^{(c)})^{(1 - \delta^{(c)}_{i:t_1})} \prod_{j=1}^{n} (p_j^{(n)})^{\delta^{(n)}_{j:t_2}} (1 - p_j^{(n)})^{1 - \delta^{(n)}_{j:t_2}} \tag{2.14}$$

where $1 \leq t_1 \leq 2^F$, $1 \leq t_2 \leq 2^N$ and $\delta^{(c)}_{i:t_1}$ and $\delta^{(n)}_{j:t_2}$ are realisations of the random variables $\delta^{(c)}_i$ and $\delta^{(n)}_j$ respectively.

2.3.4 The variance of S_{CGS}

From Section 2.3.3, the variance of S_{CGS} can be expressed as

$$Var(S_{CGS}) = \sum_{s:\, R_{CGS}} [s^2 P(\boldsymbol{I}_{t_1}^{(c)}, \boldsymbol{I}_{t_2}^{(n)}) - [sP(\boldsymbol{I}_{t_1}^{(c)}, \boldsymbol{I}_{t_2}^{(n)})]^2] \tag{2.15}$$

where R_{CGS} and the range of s are defined in (2.12) and (2.13), respectively.

The corresponding results for interaction group screening are derived in Section 2.4.

2.4 Interaction group screening

In this section an expression is obtained for $E(S)$ under interaction group screening, and the form of the probability distribution and variance of S is given. The use of these expressions is then illustrated by a small example.

2.4.1 Number of effects for estimation at the second stage

The same notation is used for random indicator vectors and random variables as in Section 2.3. Two new random indicator vectors are needed because grouped interactions are also examined at the first stage. Define

$$\boldsymbol{I}^{(cc)} = (\delta_{1,2}^{(cc)}, ..., \delta_{F-1,F}^{(cc)}), \tag{2.16}$$

of length $c_F = F(F-1)/2$, where $\delta_{i,k}^{(cc)}$ is equal to 1 if the interaction between grouped control factors $B_i^{(c)}$ and $B_k^{(c)}$ is declared active at stage 1 and 0 otherwise. Also define

$$\boldsymbol{I}^{(cn)} = (\delta_{1,1}^{(cn)}, ..., \delta_{F,N}^{(cn)}), \tag{2.17}$$

of length FN, where $(\delta_{i,j}^{(cn)})$ is equal to 1 if the interaction between grouped control factor $B_i^{(c)}$ and grouped noise factor $B_j^{(n)}$ is declared active at stage 1 and 0 otherwise.

We also need an expression for the number of individual control×noise interactions to be examined at the second stage under interaction group screening. This is a random variable, given by

$$S_I^{(cn)} = \sum_{i=1}^{F} \sum_{j=1}^{N} g_i^{(c)} g_j^{(n)} \delta_{i,j}^{(cn)}, \tag{2.18}$$

where the subscript I means that the strategy is interaction group screening.

To count the number of individual noise main effects to be examined at the second stage, we recall that, under interaction group screening, a grouped noise factor is only brought forward to the second stage if it is involved in at least one grouped control×noise interaction that is declared active at the first stage. The number of individual noise factors brought forward is therefore a random variable and is denoted by $S_I^{(n)}$ and can be written

$$S_I^{(n)} = \sum_{j=1}^{N} g_j^{(n)} \gamma_j^{(n)} \tag{2.19}$$

where $\gamma_j^{(n)} = 1$ when at least one grouped control×noise interaction is declared active, i.e. when $\sum_{i=1}^{F} \delta_{i,j}^{(cn)} \geq 1$, and zero otherwise.

In considering the individual control×control interactions that are examined at the second stage, we must count two types:

1. those involving individual factors within a grouped control factor
2. those involving individual factors that are in different grouped control factors, i.e. between grouped control factors.

Case 1

Define a new indicator random variable $\gamma_i^{(c)}$ which equals 1 if the ith grouped control factor is taken forward to the second stage; that is,

$$\sum_{j=1}^{N} \delta_{i,j}^{(cn)} + \sum_{k=1,\, k\neq i}^{F} \delta_{i,k}^{(cc)} + \delta_i^{(c)} \geq 1\,,$$

and zero otherwise. If random variable $S_I^{(cc)w}$ is the number of individual control×control interactions within the same grouped control factor, then

$$S_I^{(cc)w} = \sum_{i=1}^{F} g_i^{(c)} \frac{g_i^{(c)} - 1}{2} \gamma_i^{(c)}. \tag{2.20}$$

Case 2

The number of interactions between individual control factors from different groups involved in an active grouped control×control interaction is a random variable which can be expressed as

$$S_I^{(cc)b} = \sum_{i=1}^{F-1} \sum_{k=i+1}^{F} g_i^{(c)} g_k^{(c)} \delta_{i,k}^{(cc)}. \tag{2.21}$$

The indicator function $\gamma_i^{(c)}$ can also be used to obtain a simple expression for the number of individual control main effects to be examined at the second stage:

$$S_I^{(c)} = \sum_{i=1}^{F} g_i^{(c)} \gamma_i^{(c)}. \tag{2.22}$$

The number of contrasts allowed for the individual aliased noise×noise interactions at the second stage corresponds to maximum aliasing, as in Section 2.3, and is given by $S_I^{(nn)} = S_I^{(n)} - \eta_I^{(n)}$, where $\eta_I^{(n)} = 1$ when $S_I^{(n)} \geq 1$ and 0 otherwise (as in equation (2.7)).

We sum (2.18) to (2.22) to give the total number of effects to be estimated at the second stage under interaction group screening:

$$\begin{aligned} U_{IGS}^{(2)} &= S_I^{(c)} + S_I^{(n)} + S_I^{(cc)b} + S_I^{(cc)w} + S_I^{(cn)} + \eta_I^{(c)} + S_I^{(n)} - \eta_I^{(n)} \\ &= S_I^{(c)} + 2S_I^{(n)} + S_I^{(cn)} + S_I^{(cc)b} + S_I^{(cc)w} + \eta_I^{(c)} - \eta_I^{(n)}. \end{aligned} \tag{2.23}$$

Inclusion of $\eta_I^{(c)}$ is necessary because the mean effect has to be estimated at the second stage, when the number of individual control factors brought forward to the second stage is non-zero.

2.4.2 Number of effects for estimation over both stages

At the first stage of the experiment, the number of effects examined (together with the mean) is

$$\begin{aligned} U_{IGS}^{(1)} &= 1 + F + N + c_F + FN + (N - \zeta) \\ &= 2N + c_{F+1} + FN + (1 - \zeta), \end{aligned} \tag{2.24}$$

where $\zeta = 0$ if $N = 0$ and 1 otherwise. This inclusion of ζ is necessary for the situation where there may be no noise factors in the experiment, in which case it would be unnecessary to include $N - 1$ effects for the lower bound on the number of effects corresponding to aliased grouped noise×noise interactions.

The expected total number of effects to be examined under interaction group screening can be calculated in terms of the probabilities that the grouped control main effects, grouped control×control and control×noise interactions will be

declared active, using (2.23) as follows.

$$\begin{aligned}
E(S_{IGS}) &= U_{IGS}^{(1)} + E(U_{IGS}^{(2)}) \\
&= 2N + c_{F+1} + FN + E(S_I^{(cn)}) + 2E(S_I^{(n)}) + E(S_I^{(cc)b}) \\
&\quad + [E(S_I^{(cc)w}) + E(S_I^{(c)})] + E(\eta_I^{(c)}) - E(\eta_I^{(n)}) \\
&= 2N + c_{F+1} + FN + \Sigma_{i=1}^{F}\Sigma_{j=1}^{N} g_i^{(c)} g_j^{(n)} p_{i,j}^{(cn)} \\
&\quad + 2[\Sigma_{j=1}^{N} g_j^{(n)}][1 - \Pi_{i=1}^{F}(1 - p_{i,j}^{(cn)})] + \Sigma_{i=1}^{F-1}\Sigma_{k=i+1}^{F} g_i^{(c)} g_k^{(c)} p_{i,k}^{(cc)} \\
&\quad + \Sigma_{i=1}^{F}[g_i^{(c)}(g_i^{(c)} + 1)/2] \\
&\qquad \left[1 - \Pi_{j=1}^{N}(1 - p_{i,j}^{(cn)}) \Pi_{k=1,\, k\neq i}^{F}(1 - p_{i,k}^{(cc)})(1 - p_i^{(c)}) \right] \\
&\quad - \Pi_{i=1}^{F}\Pi_{j=1}^{N}\Pi_{k=1,\, k\neq i}^{F} \left[(1 - p_{i,j}^{(cn)})(1 - p_{i,k}^{(cc)})(1 - p_i^{(c)})\right] \\
&\quad + \Pi_{i=1}^{F}\Pi_{j=1}^{N}(1 - p_{i,j}^{(cn)}). \qquad (2.25)
\end{aligned}$$

Example 2.4.1 Under interaction group screening, consider the two different groupings of Example 2.3.1 with the same probabilities of the individual control factors' main effects being active. Suppose also that the probabilities of every individual control×control and control×noise interaction being active have the same value of 0.2.

(i) Grouping with $F = 2$, $N = 1$, $g_1^{(c)} = 4$, $g_2^{(c)} = 3$ and $g_1^{(n)} = 2$. The probabilities of the grouped control main effects being declared active remain as in Example 2.3.1 as

$$p_1^{(c)} = 0.832,\ p_2^{(c)} = 0.976.$$

We do not need to consider the probability of the grouped noise main effect being declared active as, under interaction group screening, groups of noise factors are only taken forward to the second stage when they are in grouped control×noise interactions that are declared active.

The probabilities of the grouped control×noise interactions being declared active are, from Section 2.2.1,

$$\begin{aligned}
p_{11}^{(cn)} &= 1 - (1 - 0.2)^8 = 0.83, \\
p_{21}^{(cn)} &= 1 - (1 - 0.2)^6 = 0.74.
\end{aligned}$$

The probability of the grouped control×control interaction between group $B_1^{(c)}$ and $B_2^{(c)}$ being declared active is

$$p_{12}^{(cc)} = 1 - (1 - 0.2)^{12} = 0.93.$$

Then (2.25) gives $E(S_{IGS}) = 49.11$.

(ii) For the grouping with $F = 3$, $N = 1$, $g_1^{(c)} = 3$, $g_2^{(c)} = 2$, $g_3^{(c)} = 2$ and $g_1^{(n)} = 2$. The probabilities of the grouped main effects being declared active remain as in Example 2.3.1:

$$p_1^{(c)} = 0.664,\ p_2^{(c)} = 0.8,\ p_3^{(c)} = 0.94,\ p_1^{(n)} = 0.51.$$

The probabilities of the grouped control×control interactions being declared active are

$$\begin{aligned} p_{12}^{(cc)} &= p_{13}^{(cc)} = 1 - (1 - 0.2)^6 = 0.74, \\ p_{23}^{(cc)} &= 1 - (1 - 0.2)^4 = 0.59, \end{aligned}$$

and the probabilities of the grouped control×noise interactions being declared active are

$$\begin{aligned} p_{11}^{(cn)} &= 1 - (1 - 0.2)^6 = 0.74, \\ p_{21}^{(cn)} &= p_{31}^{(cn)} = 1 - (1 - 0.2)^4 = 0.59. \end{aligned}$$

Then (2.25) gives $E(S_{IGS}) = 47.16$ which is a smaller value than for the previous grouping. For both groupings, the value of $E(S)$ is larger under interaction group screening than under classical group screening.

2.4.3 Probability distribution of S_{IGS} under interaction group screening

As for classical group screening it has not been possible to find a closed form for the distribution of S_{IGS}, but a formulation for computation can be obtained.

For any given experiment under interaction group screening, the total number of effects $S_{IGS}^{(2)}$ to be estimated at the second stage is determined by which grouped control and noise factors are declared active at the first stage and why they were

declared active, and hence is determined by the realisations of the random index vectors $\boldsymbol{I}^{(c)}$, $\boldsymbol{I}^{(cc)}$ and $\boldsymbol{I}^{(cn)}$. The maximum value for s, the number of effects examined across the two stages, corresponds to the situation where every grouped factor is declared active at the first stage and brought forward to the second stage. Hence

$$s \leq 2^{n_C+n_N} + 2N + c_{F+1} + FN + (1-\zeta)$$

where $\zeta = 0$ if $N = 0$ and 1 otherwise, for the same reason as in (2.24). The minimum value for s corresponds to the situation where no grouped factors are declared active after the first stage experiment. Hence

$$s \geq 2N + c_{F+1} + FN + (1-\zeta).$$

Hence, the probability under interaction group screening that the total size S_{IGS} of the experiment is equal to s, again assuming maximum aliasing, can be expressed as

$$P(S_{IGS} = s) = \sum_{R_{IGS}} P(\boldsymbol{I}_{t_1}^{(c)}, \boldsymbol{I}_{t_2}^{(cc)}, \boldsymbol{I}_{t_3}^{(cn)}) \tag{2.26}$$

where, using (2.24),

$$R_{IGS} = \{(\boldsymbol{I}_{t_1}^{(c)}, \boldsymbol{I}_{t_2}^{(cc)}, \boldsymbol{I}_{t_3}^{(cn)});\ U_{IGS}^{(2)} = s - (2N + c_{F+1} + FN) + (1-\zeta)\}, \tag{2.27}$$

and where

$$2N + c_{F+1} + FN + (1-\zeta) \leq s \leq 2^{n_C+n_N} + 2N + c_{F+1} + FN + (1-\zeta). \tag{2.28}$$

Under the simplifying assumptions that the grouped effects are independently active or non-active, that any non-active grouped effect is zero, and that the designated high level of each grouped factor produces the higher response, the joint probability function of $\boldsymbol{I}^{(c)}$, $\boldsymbol{I}^{(cc)}$ and $\boldsymbol{I}^{(cn)}$ is

$$\begin{aligned} P(\boldsymbol{I}^{(c)} &= \boldsymbol{I}_{t_1}^{(c)}, \boldsymbol{I}^{(cc)} = \boldsymbol{I}_{t_2}^{(cc)}, \boldsymbol{I}^{(cn)} = \boldsymbol{I}_{t_3}^{(cn)}) \\ &= \Big(\prod_{i=1}^{F} (p_i^{(c)})^{\delta_{i:t_1}^{(c)}} (1-p_i^{(c)})^{1-\delta_{i:t_1}^{(c)}}\Big)\Big(\prod_{i=1}^{F-1}\prod_{k=i+1}^{F} (p_{i,k}^{(cc)})^{\delta_{i,k:t_2}^{(cc)}} (1-p_{i,k}^{(cc)})^{1-\delta_{i,k:t_2}^{(cc)}}\Big) \\ &\qquad \Big(\prod_{i=1}^{F}\prod_{j=1}^{N} (p_{i,j}^{(cn)})^{\delta_{i,j:t_3}^{(cn)}} (1-p_{i,j}^{(cn)})^{1-\delta_{i,j:t_3}^{(cn)}}\Big) \end{aligned} \tag{2.29}$$

where

$$1 \leq t_1 \leq 2^F,\ 1 \leq t_2 \leq 2^{c_F},\ 1 \leq t_3 \leq 2^{FN}$$

and $\delta^{(c)}_{i:t_1}$, $\delta^{(cc)}_{i,j:t_2}$ and $\delta^{(cn)}_{i,j:t_3}$ are realisations of the random variables $\delta^{(c)}_i$, $\delta^{(cn)}_{i,j}$ and $\delta^{(cc)}_{i,k}$ respectively.

2.4.4 The variance of S_{IGS}

The formulation of (2.29) enables the variance of S_{IGS} to be calculated as

$$\begin{aligned} Var(S_{IGS}) &= \sum_{s:\, R_{IGS}} [s^2 P(\boldsymbol{I}^{(c)} = \boldsymbol{I}^{(c)}_{t_1}, \boldsymbol{I}^{(cc)} = \boldsymbol{I}^{(cc)}_{t_2}, \boldsymbol{I}^{(cn)} = \boldsymbol{I}^{(cn)}_{t_3}) \\ &\quad -[sP(\boldsymbol{I}^{(c)} = \boldsymbol{I}^{(c)}_{t_1}, \boldsymbol{I}^{(cc)} = \boldsymbol{I}^{(cc)}_{t_2}, \boldsymbol{I}^{(cn)} = \boldsymbol{I}^{(cn)}_{t_3})]^2] \end{aligned} \tag{2.30}$$

where R_{IGS} and the limits on s are as given in (2.27) and (2.28), respectively.

2.5 Further methods of assigning interaction probabilities

So far the theory of Section 2.4 requires specification of probabilities of individual interactions being active which have to be elicited from subject specialists, typically engineers. When eliciting information about the importance of factorial effects, subject specialists are usually more able to give opinions in the case of main effects than for interactions. An approach for tackling this uncertainty is to make the effect heredity assumption of Hamada and Wu (1992). This assumes that when a two-factor interaction is significant, at least one of the main effects of the factors involved is also significant. Chipman (1996) also used this assumption when assigning probabilities to given interactions, and called it the *weak heredity principle.* Chipman also allowed for the possibility of an active interaction occurring between two factors with non-active main effects. He assigned a very small probability to such interactions and called this the *relaxed weak heredity principle.* He also defined the *strong heredity principle* to correspond to the prior belief that for an interaction to be active, both of the main effects of the factors involved must also be active. We have focussed on the *relaxed weak heredity principle* as the risk of missing important interactions is reduced with this type of heredity.

This section incorporates the relaxed weak heredity principle into the general formulation of Section 2.4. Following Chipman (1996), we assume that

(i) the probabilities of active main effects of the factors in the experiment are mutually independent

(ii) that conditional on the status of the main effects of the factors, the probabilities of active interactions are mutually independent

(iii) the probability of an interaction being active depends only on the status of the main effects of the factors involved in that interaction.

These assumptions lead to the following assignment of the conditional probability that the interaction between individual control factors A_{ik} and A_{jl} is active.

$$P(q_{ik,jl}^{(cc)} = 1 | q_{ik}^{(c)} = s, q_{jl}^{(c)} = t) \;=\; w_{st}^{(cc)}, \qquad s,t = 0,1\,,$$

where the superscript (cc) indicates that the interaction is between two control factors, and the superscript (c) indicates that the individual main effect probabilities are for control factors. The unconditional probability is, therefore,

$$\begin{aligned} P(q_{ik,jl}^{(cc)} = 1) \;&=\; \sum_{s=0}^{1}\sum_{t=0}^{1} P(q_{ik,jl}^{(cc)} = 1 | q_{ik}^{(c)} = s, q_{jl}^{(c)} = t)\, P(q_{ik}^{(c)} = s) P(q_{jl}^{(c)} = t) \\ &=\; w_{00}^{(cc)}(1 - q_{ik}^{(c)})(1 - q_{jl}^{(c)}) \;+\; w_{01}^{(cc)}(1 - q_{ik}^{(c)})(q_{jl}^{(c)}) \\ &\quad +\; w_{10}^{(cc)}(q_{ik}^{(c)})(1 - q_{jl}^{(c)}) \;+\; w_{11}^{(cc)} q_{ik}^{(c)} q_{jl}^{(c)}\,. \end{aligned} \tag{2.31}$$

The conditional probabilities $w_{00}^{(cc)}, w_{01}^{(cc)}, w_{10}^{(cc)}, w_{11}^{(cc)}$ and the conditional probabilities $w_{00}^{(cn)}, w_{01}^{(cn)}, w_{10}^{(cn)}, w_{11}^{(cn)}$ need to be specified and Chipman (1996) used the values 0.01, 0.25, 0.25. 0.5 respectively in his example using relaxed weak heredity.

A similar formulation for $P(q_{ik,jl}^{(cn)} = 1)$, the unconditional probability for an individual control×noise interaction being active can be obtained. We have found that, for the factor screening situation, the conditional probabilities $w_{01}^{(cc)} = w_{10}^{(cc)} =$ 0.25 can be too large, as is illustrated in the following example.

Example 2.5.1 Suppose that there are 26 individual control factors in five groups of sizes 4, 5, 5, 6 and 6 and with probabilities of main effects being active in each group of (0.4, 0.4, 0.4, 0.4), (0.15, 0.15, 0.15, 0.15, 0.15), (0.1, 0.1, 0.125, 0.15, 0.15),

(0.05, 0.05, 0.05, 0.05, 0.05, 0.05), and (0.1, 0.1, 0.1, 0.1, 0.1, 0.1). Suppose there are also 11 individual noise factors in two groups of sizes 6 and 5 with corresponding main effects probabilities of (0.1, 0.1, 0.125, 0.15, 0.15, 0.15) and (0.45, 0.4, 0.35, 0.35, 0.25).

Then, using Chipman's values for $w_{00}^{(cc)}$, $w_{01}^{(cc)}$, $w_{10}^{(cc)}$ and $w_{11}^{(cc)}$ (and similarly for $w_{ij}^{(cn)}$) we can calculate from equation (2.31) the probabilities for individual control×noise interactions being active between the factors in the most likely control group and the least likely noise group. These probabilities are 0.1304, 0.1304, 0.1365, 0.1426, 0.1426, 0.1426, each occurring 4 times. The probabilities for individual control×noise interactions being active between the factors in the least likely control group and the most likely noise group are found similarly and have values 0.1302, 0.1182, 0.1062, 0.1062, 0.0821, each occurring 6 times.

The probabilities for individual control×control interactions being active between the factors in the most likely control group and the least likely control group consist of 0.1182 occurring 24 times. The probabilities for individual control×control interactions being active between the factors in the most likely control group and the next most likely control group consist of 0.1426 occurring 24 times.

The probabilities for the individual control×noise interactions being active between factors in the least likely control group and the least likely noise group consist of 0.0461, 0.0461, 0.0521, 0.0581, 0.0581, 0.0581, each occurring 36 times. The probabilities for the individual control×noise interactions being active between factors in the most likely control group and the most likely noise group consist of 0.2157, 0.2036, 0.1914, 0.1914, 0.167, each occurring 4 times. The probabilities for the individual control×control interactions being active between factors in the least likely control group and the next least likely control group consist of 0.0461 occurring 36 times.

Some of the interaction probabilities given above seem to be quite high; in particular, the first three sets of probabilities have values mainly around 10-14%, even though the main effect of one of the factors involved in each interaction is unlikely to be active.

Using Chipman's full values 0.01, 0.25, 0.25 and 0.5 leads to an expectation of 29 or 30 individual control×noise interactions out of 285 being active and 26 or

27 individual control×control interactions out of 325 being active. This choice of probabilities for individual effects, when there are large numbers of factors, leads to a large number of interactions being considered as active.

In the following we investigate the effect of the choices for probabilities of active effects.

Example 2.5.2 Based on the main effects probabilities and groupings of Example 2.5.1, we have carried out a systematic study of the effect of the w_{ij} on the expected total number and standard deviation of effects to be estimated and, for illustration, the probabilities of exceeding 200 and 350 runs. Starting from $w_{00}^{(cc)} = w_{00}^{(cn)} = 0.01$, $w_{01}^{(cc)} = w_{01}^{(cn)} = w_{10}^{(cc)} = w_{10}^{(cn)} = 0.25$ and $w_{11}^{(cc)} = w_{11}^{(cn)} = 0.5$ we divided the conditional probabilities by integer b, for $b = 1, \ldots, 100$. A selection of the results is shown in Table 2.1. We see that a reduction in the w_{ij}'s by a factor of 5 has the effect of more than halving $E(S_{IGS})$ and reducing $P(S_{IGS} > 350)$ from 1.0 to 0.483. The expected numbers of active individual interactions are reduced by a factor of 5, giving much more reasonable numbers of expected active individual interactions. A further halving of the w_{ij}'s again has a large effect but, thereafter, the reduction is much smaller. The values for the expected numbers of active effects decrease from about 29 and 28 out of 286 control×noise and 325 control×control interactions respectively, to around 0.3 in both cases.

Bingham and Chipman (2002) gave another method of choosing the prior probabilities of interactions being active by assuming that they are proportional to the probability of a main effect being active, with the proportionality constant depending on which 'parent' main effects are active.

We have found that care needs to be taken in assigning the values of the conditional probabilities so that the expected number of active interactions is consistent with the principle of effect sparsity.

2.6 Conclusions

The results of this chapter can be used to explore the effect on the total number of effects that needs to be estimated in two-stage group screening of different choices

Table 2.1: Expected numbers of active individual control×noise (c×n) and control×control (c×c) interactions out of 286 and 325 possible interactions respectively, $E(S_{IGS})$, $s.d.(S_{IGS})$ and $P(S_{IGS} > 350)$ for the experiment of Example 2.5.1 and selected values of b.

b	$E(c \times n)$	$E(c \times c)$	$E(S)$	$s.d.(S)$	$P(S > 200)$	$P(S > 350)$
1	29.26	27.64	637.2	36.67	1.0	1.0
5	5.85	5.23	346.7	64.89	0.99	0.483
10	2.93	2.76	240.0	63.29	0.73	0.040
20	1.46	1.38	166.8	56.50	0.28	0.001
30	0.98	0.92	137.7	50.94	0.12	0.000
40	0.73	0.69	122.0	46.78	0.06	0.000
50	0.59	0.55	112.2	43.60	0.03	0.000
75	0.39	0.37	98.7	38.18	0.01	0.000
100	0.30	0.28	91.7	34.75	0.01	0.000

of strategy and group sizes. Two screening strategies have been investigated: classical group screening and interaction group screening for unequal group sizes and unequal probabilities of individual main effects and interactions being active. Explicit formulae for $E(S_{CGS})$ and $E(S_{IGS})$ have been obtained and the forms of the distributions and variance of S_{CGS} and S_{IGS} have been presented. Small examples have been used to illustrate the results.

The following chapter describes software which incorporates the theory presented in this chapter and enables investigations of experiments with large numbers of factors. The issues of how to choose group sizes and the sensitivity of $E(S)$ to the choice of values for the probabilities of factorial effects being active under both classical and interaction group screening will also be discussed in Chapter 3.

Chapter 3

Comparative Studies

3.1 Introduction

In this chapter a variety of grouping and screening strategies is investigated, where screening strategy refers to a choice between classical group screening (CGS) and interaction group screening (IGS). Specifically, the following questions are considered:

- how does the grouping of factors affect the distribution, expectation and standard deviation of the total number of effects S requiring estimation in a two-stage experiment?
- how do changes to the probabilities of individual factorial effects being active change the probability distribution, expectation and standard deviation of S?

These questions are investigated using grouping software that I have written in C to implement the ideas of Chapter 2 through the calculation of the probability distribution, expected value and standard deviation of S and the total number of effects requiring estimation in a two-stage experiment, under classical group screening and interaction group screening. This software allows unequal group sizes and unequal probabilities of individual factorial effects being active and gives a range of assessments of the grouping strategies under CGS and IGS. In these two ways it is an advance on the software described by Lewis and Dean (2001). This software also applies to the special case where there are no noise factors under consideration. The software also calculates $E(S_{CGS})$, $s.d.(S_{CGS})$, $P(S_{CGS} > u)$,

$E(S_{IGS})$, $s.d.(S_{IGS})$ and $P(S_{IGS} > u)$ where u is a target experiment size. Details of the input required by the software are given in Appendix A.

In Section 3.2 the software is used to investigate screening strategies for a large example with a specified grouping. In Section 3.3 an investigation is presented which examines how changing the numbers of groups and group sizes affects the distribution of S. In Section 3.4, it is shown how the sensitivity of the expected size to changes in the probabilities of the individual factorial effects being active can be investigated under CGS and IGS using the software and fractional factorial designs. Conclusions are presented in Section 3.5.

The following notation will be used to specify a grouping of factors, that is, the sizes of each group and the particular factors held in each group. Suppose that an ordering of the individual control and noise factor labels is specified as

$$A_1^{(c)}, ..., A_{n_C}^{(c)}; A_1^{(n)}, ..., A_{n_N}^{(n)},$$

where n_C (n_N) is the total number of individual control (noise) factors. The expression

$$(a_1, ..., a_F; b_1, ..., b_N)$$

denotes the grouping of adjacent control factors into F (N) consecutive groups of control (noise) factors with sizes a_i for $i = 1, ..., n_1$ (b_j for $j = 1, ..., n_2$), where the grouping is made starting from the leftmost control (noise) factor. This notation will be used for the studies of different groupings in Section 3.3.

Example 3.1.1 For the ordering of 7 individual control factors and 4 individual noise factors given by

$$A_2^{(c)}, A_4^{(c)}, A_6^{(c)}, A_1^{(c)}, A_3^{(c)}, A_7^{(c)}, A_5^{(c)}; A_3^{(n)}, A_1^{(n)}, A_2^{(n)}, A_4^{(n)}$$

the expression $(2, 2, 3; 2, 2)$ denotes the grouping

$$\{A_2^{(c)}, A_4^{(c)}\}, \{A_6^{(c)}, A_1^{(c)}\}, \{A_3^{(c)}, A_7^{(c)}, A_5^{(c)}\}; \{A_3^{(n)}, A_1^{(n)}\}, \{A_2^{(n)}, A_4^{(n)}\}.$$

My software has been incorporated into a system called GISEL (grouping in screening with elicitation) developed on the larger EPSRC project (GR/R72693/01) (see Dupplaw et al. (2004)). This web-based system has a more user friendly

interface than my software and elicits information about the individual effects from users (also see Section 4.1). Users are able to create a list of individual control and noise factors and to submit their opinion on how important each factor is likely to be for a specified response and also the anticipated direction of influence of the factor on the response. The user is also able to enter probabilities of the various individual factorial effects being active, and can specify a particular grouping. The software described above is then implemented and the output is displayed in both a numerical and graphical form. The graphical output allows comparison of the probability distribution of the size of the two-stage experiment, and the probability that the number of effects requiring estimation exceeds a certain target size, under different groupings and choices of interaction or classical group screening.

GISEL also incorporates simulation software to assess the risk of failing to detect important interactions. This software, which I did not write, is based on Dean and Lewis (2002), extended to accommodate the situation where some of the control and/or noise main effects are believed very likely to be active. These are then grouped together in groups of size specified by the user and the remaining factors, whose main effects are believed less likely to be active, are partitioned into groups of size specified by the user. The factors whose main effects are believed very likely to be active are assumed to have main effects whose directions are known. The user can also specify how many directions of the main effects are known. As described in Section 1.4.1 the software is able to investigate how the choice of grouping and screening strategy affects the proportions of active individual main effects and active two-factor interactions involving a control factor that are undetected in the simulated screening experiment. There are far too many possible combinations of groupings, strategy and choices of probabilities to make the simulation of an experiment for all possible groupings computationally feasible. Hence my software is first used to identify a small number of promising options. The simulation software is then run to identify the risk of missing important effects before the final choice is made.

3.2 A large example

The following large example illustrates how the software and ideas of Chapter 2 can be used to investigate screening strategies with unequal group sizes for large numbers of factors with unequal probabilities of different individual main effects and two-factor interactions being active.

Example 3.2.1 Consider an experiment with 26 individual control factors in five groups of sizes 4, 5, 5, 6 and 6. Suppose that the probabilities of main effects being active in each group are (0.4, 0.4, 0.4, 0.4), (0.15, 0.15, 0.15, 0.15, 0.15), (0.1, 0.1, 0.125, 0.15, 0.15), (0.05, 0.05, 0.05, 0.05, 0.05, 0.05), and (0.1, 0.1, 0.1, 0.1, 0.1, 0.1) respectively. Suppose that there are also 11 individual noise factors in two groups of sizes 5 and 6 with corresponding main effects probabilities of (0.45, 0.4, 0.35, 0.35, 0.25) and (0.1, 0.1, 0.125, 0.15, 0.15, 0.15). Suppose that the probabilities of individual control×noise interactions being active between the factors in the first (most likely) group of control factors and the most likely group of noise factors are all 0.2 and all other control×noise interaction probabilities are 0.005. The probabilities of control×control interactions being active between individual factors in the first control group and the others are all 0.1. The probabilities of all other individual control×control interactions being active are assumed to be 0.005.

The expectation and standard deviation of S for the two-stage experiment and the probability of the number of effects requiring estimation exceeding targets of 200, 250, 300, 350, 400 and 450 runs, under classical and interaction group screening, are shown in the first two rows of Table 3.1. Classical group screening (CGS) has a smaller expected value of S than interaction group screening (IGS). However, the probabilities that S exceeds the targets of 350, 400 and 450 runs are larger for CGS than for IGS. This is explained by the increased standard deviation for CGS which can be seen from the probability distributions for the size of the two-stage experiment under CGS and IGS, shown in Figure 3.1. The explanation of this result is that investigation of grouped interactions at the first stage under IGS may lead to several groups of factors being dropped at this stage. Fewer individual interactions would need to be examined at the second stage. For example, if a grouped noise factor is judged to have an active main effect but is not involved in a

control×noise interaction, found to be active, then all the individual noise factors in the group will not be investigated further. For this example, this property of IGS shows strongly because of the low probabilities of individual interactions being active. By contrast, under CGS, all the interactions involving the individual noise factors in a noise group declared active would be examined at the second stage.

Figure 3.2 shows the probability of the number of effects requiring estimation in a two-stage experiment exceeding a target size u under CGS and IGS. For this example, CGS gives a lower probability of exceeding a target u when $u < 342$. Beyond this target value, CGS gives a higher probability of exceeding the target. This is because there are values of s greater than 342 with much higher probabilities of occurring under CGS than under IGS. In particular, under CGS, the probability of every grouped factor being declared active at the first stage (the maximum value S can take), is higher than for IGS as can be seen from Figure 3.1.

Table 3.1: The expectation and standard deviation of S for the experiment of Example 3.2.1, and the probabilities of S exceeding 200, 250, 300, 350, 400 and 450.

	$P(S > u)$						$E(S)$	$s.d.(S)$
u	200	250	300	350	400	450		
CGS	0.52	0.38	0.33	0.15	0.12	0.12	234.16	132.55
IGS	0.97	0.80	0.40	0.10	0.01	0.00	287.61	48.03
IGS(i)	0.99	0.93	0.57	0.20	0.03	0.00	310.11	46.93
IGS(ii)	0.98	0.82	0.42	0.11	0.01	0.00	289.89	47.11
IGS(iii)	1.00	0.95	0.59	0.22	0.03	0.00	312.39	45.99

The effect of changing the probabilities of individual interactions being active can be investigated and is now illustrated in this example.

Control×noise: The effect of increasing the probability of individual control×noise interactions being active between the individual factors in the second grouped control factor and the most likely grouped noise factor from 0.005 to 0.25 is shown in Table 3.1, as IGS(i). The expected size increases by 22.5, but the standard deviation decreases by 1.1.

Figure 3.1: The probability distributions for the number of effects requiring estimation in the two-stage experiment under (a) CGS and (b) IGS.

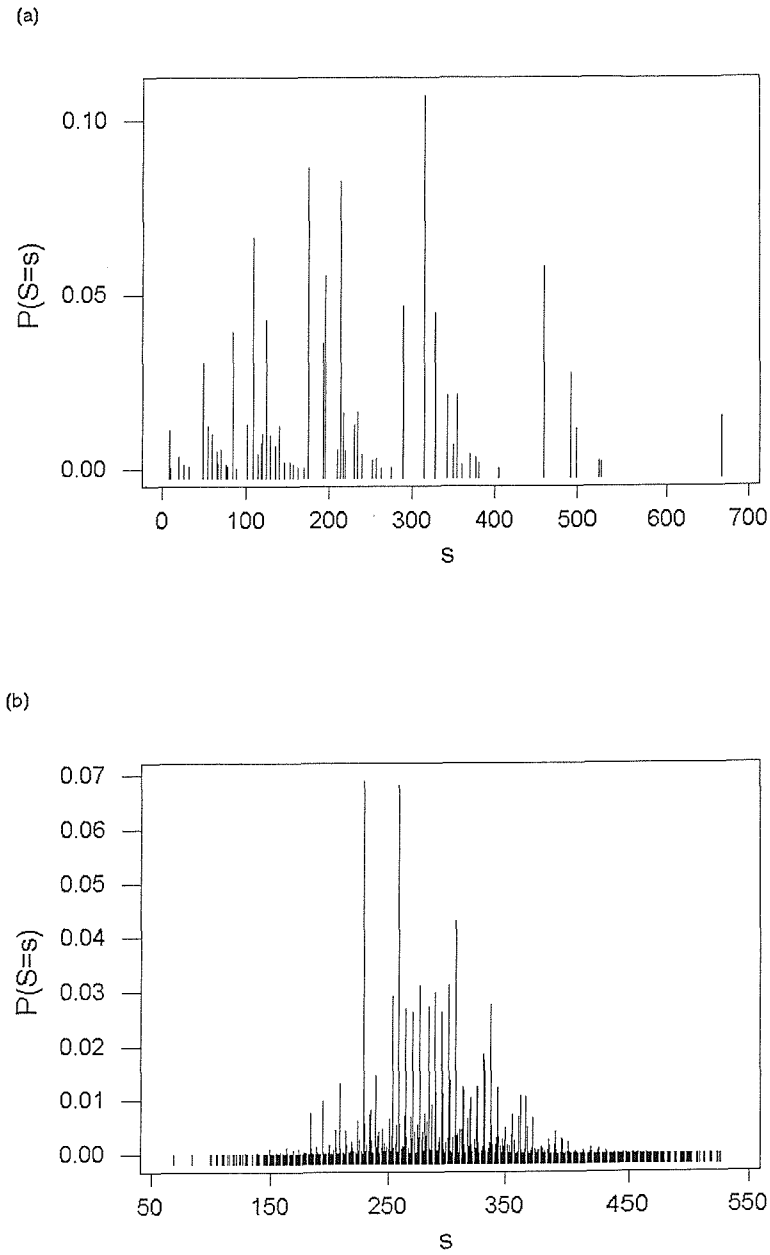
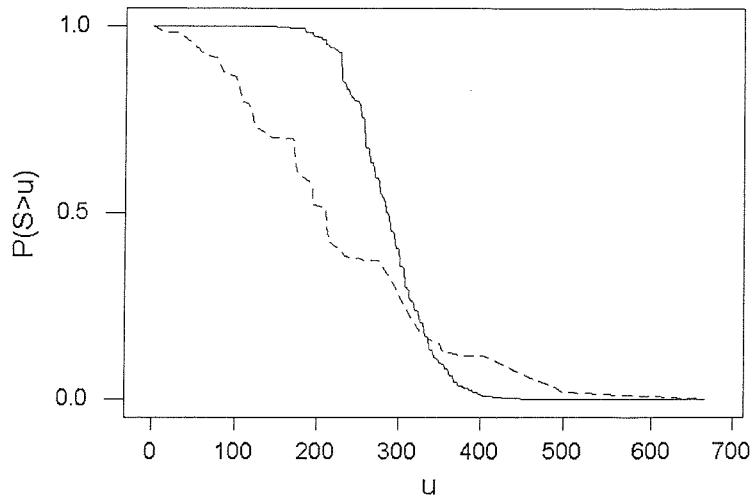


Figure 3.2: The probability of exceeding a target size u under CGS (– – –) and IGS (_____).



Control×control: The effect of increasing the probability of individual control×control interactions being active between the factors in the most likely grouped control factor and the fourth grouped control factor from 0.1 to 0.25 is shown in Table 3.1, as IGS(ii). The expected size increases by 2.28 and the standard deviation by 0.92.

Control×noise and control×control: The effect of making both the above changes to the probabilities for interactions is shown in Table 3.1, as IGS(iii). The expected size increases by 24.78 (an increase of $\sim 9\%$) and the standard deviation decreases by 2.04 (a reduction of $\sim 0.24\%$).

This example has illustrated how the distribution of S can be investigated for a problem with a large number of factors, unequal group sizes and unequal probabilities. In the following section, investigations are presented of how the choice of the number of groups and the number of factors within each group affect the distribution of S.

3.3 Investigations of groupings

3.3.1 Study 1

In this study, a total of 12 individual factors (6 control and 6 noise) and a variety of different values for the probabilities of main effects being active are considered. The main effects probabilities for the control factors, $A_1^{(c)}$, $A_2^{(c)}$, $A_3^{(c)}$, $A_4^{(c)}$, $A_5^{(c)}$ and $A_6^{(c)}$ are 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8 respectively. The main effects probabilities for the noise factors, $A_1^{(n)}$, $A_2^{(n)}$, $A_3^{(n)}$, $A_4^{(n)}$, $A_5^{(n)}$ and $A_6^{(n)}$ are 0, 0.2, 0.4, 0.6, 0.8 and 1.0 respectively. For interaction group screening, the weak heredity principle (see Section 2.5) is used to calculate the individual interaction probabilities. This takes into account the probabilities of the corresponding main effects probabilities of the factors involved. The values of the conditional probabilities $w_{00}^{(cc)}, w_{01}^{(cc)}, w_{10}^{(cc)}, w_{11}^{(cc)}$ and conditional probabilities $w_{00}^{(cn)}, w_{01}^{(cn)}, w_{10}^{(cn)}, w_{11}^{(cn)}$ are 0.005, 0.125, 0.125, and 0.25 respectively, which are half the values used by Chipman (1996) (see Table 2.1).

In the study a range of different groupings of the factors was considered for up to five groups of control factors and three groups of noise factors. A consequence of using different choices of factors within each group is that different probabilities of individual main effects being active are combined, through equation (2.1) and equation (2.2) of Section 2.2.1, to produce the probability of each grouped main effect being active. In order to gain some understanding of how this might affect the distribution of S, groupings were imposed on the individual factors in two ways.

I - Grouping together factors with similar main effect probabilities:
The individual control factors are ordered in increasing size of main effect probabilities and the individual noise factors are ordered in the same way to give

$$A_1^{(c)}, A_2^{(c)}, A_3^{(c)}, A_4^{(c)}, A_5^{(c)}, A_6^{(c)}; A_1^{(n)}, A_2^{(n)}, A_3^{(n)}, A_4^{(n)}, A_5^{(n)}, A_6^{(n)}.$$

When groupings are imposed on the individual factors in this order (see Section 3.1) then factors within a group have similar main effects probabilities.

II - Grouping together factors with dissimilar main effect probabilities:
The grouping is imposed on the individual factors when held in the order

$$A_1^{(c)}, A_6^{(c)}, A_2^{(c)}, A_5^{(c)}, A_3^{(c)}, A_4^{(c)}; A_1^{(n)}, A_6^{(n)}, A_2^{(n)}, A_5^{(n)}, A_3^{(n)}, A_4^{(n)}$$

so that the factors within a group have more dissimilar main effects probabilities, than in I.

The distinct groupings for the control factors under I are listed in the Appendix in Table B.1. Table B.2 lists the 16 distinct noise groupings that are possible for $N = 1$ up to $N = 3$ under I. Similarly, the groupings for the control and noise factors under II are listed in B.3 and B.4 respectively. In the study every choice of control factor grouping was used with every choice of noise factor grouping.

Interaction Group Screening

Table 3.2 gives the maximum and minimum values of $E(S_{IGS})$, $s.d.(S_{IGS})$ and $P(S_{IGS} > 65)$ found in Study 1, together with the groupings which produced those values.

Table 3.2: Maximum and minimum values for $E(S_{IGS})$, $s.d.(S_{IGS})$ and $P(S_{IGS} > 65)$ for Study 1 and interaction group screening according to similar probabilities (I) and dissimilar probabilities (II), together with the corresponding grouping.

	I	II
max $E(S_{IGS})$	72.98 (6; 5,1)	72.73 (6; 6)
min $E(S_{IGS})$	60.02 (2,2,2; 2,2,2)	60.90 (2,2,2; 2,2,2)
max $s.d.(S_{IGS})$	10.93 (6; 3,2,1)	10.60 (6; 2,2,2)
min $s.d.(S_{IGS})$	3.55 (6; 6)	3.55 (6; 6)
max $P(S_{IGS} > 65)$	0.99 (6; 6)	0.99 (6; 6)
min $P(S_{IGS} > 65)$	0.3 (2,2,2; 2,4)	0.35 (2,2,2; 2,4) (2,2,2;4,2)

From Table 3.2, it can be seen that, for groupings based on similar probabilities in this study, the formation of groups that are as equally sized and as small as possible (excluding groups of size one), that is (2,2,2; 2,2,2), gives the best results under criterion 1 (see Section 2.2.2) of minimising the expected number of effects to be estimated ($E(S_{IGS}) = 60.02$). This is also one of the best groupings in terms of minimising the probability of exceeding a target of 65 runs ($P(S_{IGS} > 65) = 0.34$). However, under the criterion of minimising the standard deviation of S_{IGS}, this is not the best strategy giving the value of $s.d.(S_{IGS}) = 8.79$. As shown in the table, the best overall strategy for minimising the standard deviation is to group all the control factors together in a single group of size six and all the noise factors in a single group. There are then only two groups to consider at the first stage of the experiment and much less scope for variation in the number of grouped factors to be brought forward to the second stage.

For groupings based on similar probabilities, putting all the control factors in one group is nearly always the worst strategy, across all the different groupings of the noise factors, in terms of minimising the expected size. This is because the one grouped control factor is very likely to be brought forward to the second stage and, when this occurs, no control factors will have been ruled out. More observations at the second stage will then be needed.

The results from the study allow a more detailed examination of how the grouping of factors affects the distribution of S_{IGS}. Tables C.1 to C.2 in Appendix C show how the different noise factor groupings affect the maximum and minimum values of $E(S_{IGS})$, $s.d.(S_{IGS})$, and $P(S_{IGS} > 65)$ under orderings I and II for IGS. An interesting comparison involves the grouping (1,1,2,2; 2,2,2), under Method I (similar probabilities). A comparison of the results of this grouping (which are not shown) with those of grouping (2,2,2; 2,2,2) shows that the only benefit of splitting one grouped control factor of size 2 into two groups of size 1 is that the standard deviation of S is decreased from 8.79 to 7.94. However, the expected size is increased from 60.02 to 62.55 and the probability of exceeding 65 runs is increased from 0.34 to 0.38. Figure 3.3 shows the distributions for (2,2,2; 2,2,2) and (1,1,2,2; 2,2,2). It is clear from the distributions that there are fewer possible experiment sizes for grouping (2,2,2; 2,2,2) than for (1,1,2,2; 2,2,2), but there is a wider range

of sizes. Hence a larger value for the standard deviation is obtained.

Grouping together factors with higher main effects probabilities generally seems to be a good approach for control and noise factors in the example. For example, using the control grouping (1,1,4) as opposed to (4,1,1) reduces the expected size by approximately one run and consistently reduces the standard deviation of S, and hence the probability of exceeding a target size, across all of the noise groupings.

For groupings based on dissimilar probabilities, the results are very similar to those for groupings with similar probabilities, with mainly slightly increased values for $E(S_{IGS})$ and $P(S_{IGS} > 65)$ and sometimes decreased values of the standard deviation. However, the overall minimal values arise from grouping together factors with similar main effect probabilities of being active, that is, ordering I. An increase in values for expected size was anticipated under ordering II, as the more likely factors will bring with them the less likely factors with which they are grouped to the second stage. The increase is only slight here and a possible explanation is that the probabilities for the factors within each group are not greatly dissimilar.

Classical Group Screening

The above study was repeated for classical group screening and a summary of the results are given in Table 3.3. and $P(S_{CGS} > 65)$ for 12 factors (6 control, 6 noise) with $F \leq 5$, and $N \leq 3$ for groupings based on similar probabilities and dissimilar probabilities as was done for interaction group screening. Tables C.3 to C.4 in Appendix C show how the different noise factor groupings affect the maximum and minimum values of $E(S_{CGS})$, $s.d.(S_{CGS})$, and $P(S_{CGS} > 65)$ under orderings I and II for CGS.

The control factor grouping (1,1,1,1,2) based on ordering I, similar probabilities, consistently gave the smallest $E(S_{CGS})$ value but a reasonably large value for $s.d.(S_{CGS})$ across all the noise groupings. Grouping the control factors together in one group of size six consistently gave the largest $E(S_{CGS})$ value and, in most cases, the minimum value for the standard deviation. Hence this is also the worst strategy in terms of reducing the probability of exceeding 65 runs. This is because the group of six control factors is extremely likely to be brought forward to the second stage and, when this occurs, all individual control$\times$control interactions will

Figure 3.3: The probability distributions for the size of the two-stage experiment under grouping (2,2,2; 2,2,2) and (1,1,2,2; 2,2,2) based on similar probabilities, for interaction group screening.

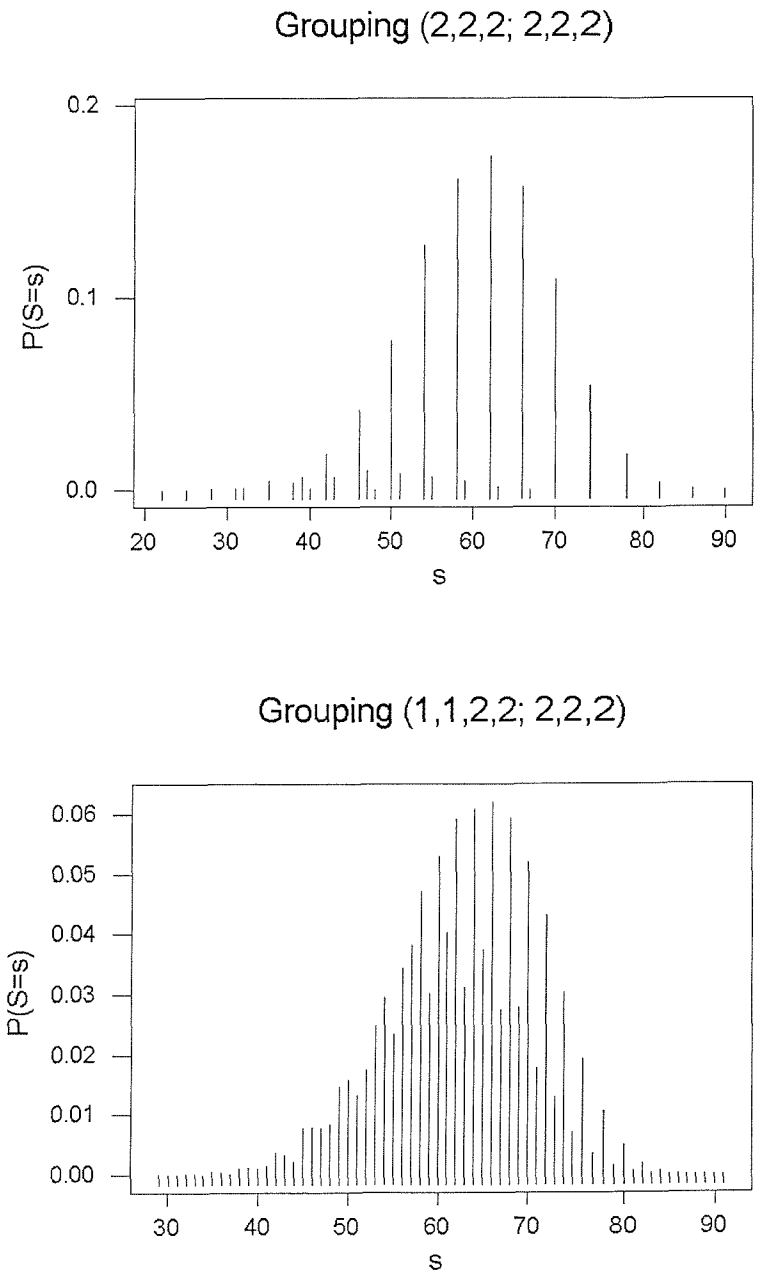


Table 3.3: Maximum and minimum values for $E(S_{CGS})$, $s.d.(S_{CGS})$ and $P(S_{CGS} > 65)$ for Study 1 and classical group screening according to similar probabilities (I) and dissimilar probabilities (II), together with the corresponding grouping.

	I	II
max $E(S_{CGS})$	71.65 (6; 6)	71.65 (6; 6)
min $E(S_{CGS})$	39.74 (1,1,1,1,2; 2,1,3)	44.97 (1,1,1,2,1; 1,3,2)
max $s.d.(S_{CGS})$	4.27 (3,2,1; 4,2)	3.94 (2,2,2; 2,2,2)
min $s.d.(S_{CGS})$	4.32 (6; 1,5)	4.32 (6; 1,5)
max $P(S_{CGS} > 65)$	0.99 (6; 6)	0.99 (6; 6)
min $P(S_{CGS} > 65)$	0.00 (6; 1,5)	0.00 (6; 1,5)

need to be examined. In the case where all the noise factors are also in one group (i.e. grouping (6; 6)), *all* individual interactions are extremely likely to be examined at the second stage. In the study, this grouping gave the overall maximum value for $E(S_{CGS})$ and $P(S_{CGS} > 65)$.

Under ordering I, grouping the more likely control factors together rather than separating them into groups of size one is nearly always better under all three criteria of minimising $E(S_{CGS})$, $s.d.(S_{CGS})$ and $P(S_{CGS} > 65)$. For example, the grouping (1,1,1,3; 2,2,2) gave the values $E(S_{CGS}) = 43.02$, $s.d.(S_{CGS}) = 10.88$ and $P(S_{CGS} > 65) = 0.01$, whereas the grouping (3,1,1,1; 2,2,2) gave the values $E(S_{CGS}) = 46.60$, $s.d.(S_{CGS}) = 14.36$ and $P(S_{CGS} > 65) = 0.04$. This is also true for the noise groupings.

For groupings based on dissimilar probabilities, the results are very similar to those for groupings with similar probabilities, with mainly slightly increased values for $E(S_{CGS})$ and $P(S_{CGS} > 65)$ of about 1 and 0.02 respectively, and sometimes decreased values of standard deviation. However, the overall minimal values arise from groupings of factors having similar probabilities that their main effects are active.

Comparison of classical group screening and interaction group screening

In this study, the minimum value of the expected size occurs under CGS. However, the maximum standard deviation is also produced by CGS and the minimum standard deviation by IGS. This is because, under CGS, the number of observations needed at the second stage is much more variable than under IGS due to the fact that no interactions are ruled out at the first stage.

Under CGS the expected size is always less than the corresponding expected size under IGS as is the probability of exceeding 65 runs. This apparent advantage of CGS fails to reflect the risk of failing to miss important effects or the risk of declaring unimportant factors to be important. This drawback to CGS can be shown by a simulation study.

Under interaction group screening, the best grouping in terms of minimising $E(S)$ was (2,2,2;2,2,2) for orderings I and II, indicating that a good strategy is to keep group sizes as equal and as small as possible (without groups of size 1). Under

classical group screening, the same grouping gave values for $E(S)$ of 46.99 and 58.78 for orderings I and II respectively. These values were not near the corresponding minimal values indicating that this is not a good grouping strategy for classical group screening. The results from this study indicate using as many groups as possible (including groups of size 1) is a good strategy for minimising $E(S_{CGS})$.

The following example provides a counterexample to the hypothesis that using group sizes as small as possible minimises the expected value of S for interaction group screening.

Example 3.3.1 Consider an experiment with 6 individual control factors and 6 individual noise factors. Let the main effects of the individual factors have probability 0.2 of being declared active, and the individual control×noise and control×control interactions have corresponding probabilities of 0.05 of being active. Table 3.4 shows, for this example, what happens when the group sizes and numbers of groups are varied under interaction group screening. The last two rows show that equal sized groups (all of size 2) is slightly less efficient, in terms of minimising expected size, than having the control factors in groups of size 2 and the noise factors in groups of size 3. It also gives a slightly larger probability of exceeding a target of 40 but smaller probabilities of exceeding targets of 50 and 60.

For this example group sizes need to be as equal as possible but not necessarily as small as possible in order to minimise the expected size of the experiment. Also, keeping group sizes as small as possible gives minimal values for standard deviation in this example. However, the differences in values of $E(S_{IGS})$ and $s.d.(S_{IGS})$ for these two groupings are small.

From the studies in this section, it is clear that general statements about the effect of groupings for factors and strategy cannot be made. It is necessary to investigate the impact on the distribution of S of various choices using software to perform the necessary calculations before reaching a decision.

Table 3.4: The expected size, standard deviation and probabilities of exceeding 40, 50 and 60 effects requiring estimation for different groupings of control and noise factors for IGS.

6 control		6 noise		$E(S_{IGS})$	$s.d.(S_{IGS})$	$P(S_{IGS} > u)$		
groups	sizes	groups	sizes			40	50	60
4	2,2,1,1	4	2,2,1,1	50.91	7.79	0.92	0.53	0.11
3	2,2,2	1	6	46.97	15.12	0.75	0.48	0.13
3	3,2,1	3	3,2,1	45.99	11.17	0.69	0.38	0.10
3	2,2,2	4	2,2,1,1	45.68	8.78	0.67	0.34	0.04
2	3,3	2	3,3	45.49	14.70	0.61	0.37	0.23
3	3,2,1	3	2,2,2	44.55	10.37	0.63	0.29	0.06
2	3,3	3	2,2,2	44.00	12.23	0.63	0.31	0.10
3	2,2,2	3	2,2,2	42.51	9.61	0.58	0.18	0.03
3	2,2,2	2	3,3	42.09	11.75	0.56	0.26	0.06

3.3.2 Study 2 - Practical special case

This concerns the practical case, discussed in Chapter 4, of unequal probabilities arising when the main effects of one group of control factors are thought very likely to be active and the main effects of the other control factors are believed to be less likely to be important, or there is little prior knowledge about their importance.

Example 3.3.2 The strategy used in this example is interaction group screening. Consider an experiment with 15 individual control factors and 4 individual noise factors. Suppose that the main effects of 7 of the individual control factors are thought very likely to be active and are assigned probability 1.0. The main effects of the remaining 8 individual control factors are assigned probability 0.2 of being active. Suppose also that there is little information about the 4 individual noise factors, and so their main effects are assigned probability 0.3 of being active. Individual control×noise and control×control interactions have probabilities 0.07 and 0.05 respectively. These values are chosen to be slightly smaller than those that would be obtained using effect heredity with conditional probabilities as in

Example 2.5.1.

The impact of changing groupings for these individual factors within the three sets of 7, 8 and 4 factors under interaction group screening, can be seen from Table 3.5 for $E(S_{IGS})$, $s.d.(S_{IGS})$ and three targets 120, 150 and 180. The first column shows all the possible groupings (excluding groups of size one) of the 7 very-likely-to-be-active control factors; that is, one group of 7, two groups of sizes 2 and 5, two groups of sizes 3 and 4, or three groups of sizes 2, 2 and 3. Column 2 shows possible groupings of the 8 remaining control factors in one, two, three or four groups, as listed. The four noise factors are grouped into one group of 4 or two groups of 2, shown in column 3.

Table 3.5: Investigation of different groupings for Example 3.3.2 under interaction group screening.

7 v likely indiv con		8 indiv con		4 indiv noise		$E(S_{IGS})$	$s.d.(S_{IGS})$	$P(S_{IGS} > u)$		
gps	sizes	gps	sizes	gps	sizes			120	150	180
1	7	1	8	1	4	184.90	19.08	0.99	0.93	0.74
1	7	1	8	2	2,2	174.75	20.31	0.98	0.89	0.55
1	7	2	2,6	1	4	168.03	21.95	0.96	0.80	0.29
1	7	2	2,6	2	2,2	158.41	21.97	0.94	0.70	0.16
1	7	2	3,5	1	4	165.33	23.52	0.95	0.75	0.29
1	7	2	3,5	2	2,2	155.32	23.13	0.92	0.61	0.13
1	7	2	4,4	1	4	164.59	24.10	0.96	0.78	0.36
1	7	2	4,4	2	2,2	154.43	23.56	0.91	0.60	0.12
1	7	3	2,2,4	1	4	151.39	22.01	0.90	0.56	0.07
1	7	3	2,2,4	2	2,2	143.16	21.46	0.85	0.38	0.03
1	7	3	2,3,3	1	4	149.40	22.26	0.90	0.51	0.07
1	7	3	2,3,3	2	2,2	141.14	21.61	0.83	0.35	0.02
1	7	4	2,2,2,2	1	4	140.34	20.08	0.84	0.31	0.01
1	7	4	2,2,2,2	2	2,2	134.04	19.51	0.76	0.20	0.01
2	2,5	1	8	1	4	168.40	21.44	0.97	0.81	0.38
2	2,5	1	8	2	2,2	159.22	21.71	0.94	0.68	0.13
2	2,5	2	2,6	1	4	151.08	22.02	0.90	0.53	0.07
2	2,5	2	2,6	2	2,2	142.42	21.19	0.84	0.37	0.02
2	2,5	2	3,5	1	4	147.53	22.70	0.88	0.48	0.05
2	2,5	2	3,5	2	2,2	138.49	21.45	0.79	0.29	0.01
2	2,5	2	4,4	1	4	146.48	22.96	0.87	0.44	0.05
2	2,5	2	4,4	2	2,2	137.28	21.57	0.78	0.27	0.01
2	2,5	3	2,2,4	1	4	135.26	20.51	0.77	0.23	0.01
2	2,5	3	2,2,4	2	2,2	127.99	18.97	0.67	0.12	0.00
2	2,5	3	2,3,3	1	4	133.09	20.42	0.74	0.20	0.01

continues overleaf

Table 3.5: continued

7 v likely indiv con		8 indiv con		4 indiv noise		$E(S_{IGS})$	$s.d.(S_{IGS})$	$P(S_{IGS} > u)$		
gps	sizes	gps	sizes	gps	sizes			120	150	180
2	2,5	3	2,3,3	2	2,2	125.79	18.76	0.62	0.09	0.00
2	2,5	4	2,2,2,2	1	4	126.19	18.13	0.64	0.09	0.00
2	2,5	4	2,2,2,2	2	2,2	120.85	16.42	0.52	0.04	0.00
2	3,4	1	8	1	4	166.01	22.60	0.96	0.79	0.32
2	3,4	1	8	2	2,2	156.64	22.60	0.93	0.63	0.15
2	3,4	2	2,6	1	4	148.01	22.42	0.88	0.47	0.06
2	3,4	2	2,6	2	2,2	139.17	21.34	0.80	0.30	0.02
2	3,4	2	3,5	1	4	144.17	22.79	0.85	0.41	0.04
2	3,4	2	3,5	2	2,2	134.94	21.28	0.75	0.24	0.01
2	3,4	2	4,4	1	4	143.01	22.93	0.80	0.37	0.03
2	3,4	2	4,4	2	2,2	133.63	21.27	0.73	0.21	0.01
2	3,4	3	2,2,4	1	4	131.89	20.40	0.72	0.18	0.00
2	3,4	3	2,2,4	2	2,2	124.45	18.56	0.60	0.08	0.00
2	3,4	3	2,3,3	1	4	129.66	20.22	0.68	0.15	0.00
2	3,4	3	2,3,3	2	2,2	122.18	18.23	0.54	0.06	0.00
2	3,4	4	2,2,2,2	1	4	122.93	17.92	0.57	0.06	0.00
2	3,4	4	2,2,2,2	2	2,2	117.41	15.84	0.43	0.02	0.00
3	2,2,3	1	8	1	4	153.59	20.80	0.93	0.59	0.07
3	2,2,3	1	8	2	2,2	146.15	20.86	0.88	0.43	0.04
3	2,2,3	2	2,6	1	4	137.30	20.02	0.80	0.26	0.01
3	2,2,3	2	2,6	2	2,2	130.39	18.86	0.71	0.15	0.00
3	2,2,3	2	3,5	1	4	133.49	20.20	0.75	0.20	0.01
3	2,2,3	2	3,5	2	2,2	126.19	18.55	0.63	0.09	0.01
3	2,2,3	2	4,4	1	4	132.33	20.28	0.73	0.18	0.00
3	2,2,3	2	4,4	2	2,2	124.89	18.44	0.61	0.08	0.00
3	2,2,3	3	2,2,4	1	4	123.37	17.80	0.58	0.06	0.00
3	2,2,3	3	2,2,4	2	2,2	117.85	15.72	0.44	0.02	0.00
3	2,2,3	3	2,3,3	1	4	121.24	17.59	0.52	0.04	0.00
3	2,2,3	3	2,3,3	2	2,2	115.69	15.34	0.38	0.01	0.00
3	2,2,3	4	2,2,2,2	1	4	116.56	15.40	0.41	0.01	0.00
3	2,2,3	4	2,2,2,2	2	2,2	112.97	13.00	0.27	0.00	0.00

Table 3.5 shows that, for this example, equal sized groups tend to give rise to smaller values of $E(S_{IGS})$ and smaller corresponding standard deviations. The same pattern has occurred in other examples including those in Section 3.3.1. In this particular study, the smallest group sizes are the best under criteria 1 to 3, but this is not necessarily true for other examples, as was demonstrated in Example 3.3.1.

One way of reducing the expected number of effects to be estimated is to keep fixed, during the experiment, the levels of the factors that are very likely to be active since their effects are assumed to be already known. When this is done for the grouping in the last line of Table 3.5, $E(S_{IGS})$ drops to 49 under interaction group screening. This would require fewer resources but there would be no information about the interactions between the likely active factors and the other factors from the experiment.

Suppose in interaction group screening that several individual control factors are believed to have main effects that are very likely to be active, and the same is not true for the other factors. This raises the question of whether it is better to put the very-likely-to-be-active factors in a single group or to spread them across several groups. The latter strategy could involve forming a group of control factors that contain both very likely and very unlikely-to-be-active factors.

From equation (2.1) of Section 2.2.1, the approximation of the probability of the main effect of a grouped control factor $B_i^{(c)}$ being declared active is

$$p_i^{(c)} = 1 - \prod_{A_{ik} \in B_i^{(c)}} (1 - q_{ik})$$

where q_{ik} is the probability that the main effect of individual factor A_{ik} is active, with a similar definition for grouped noise factors. If $q_{ik} < 1, \forall k$, then $p_i^{(c)} < 1$. However, if $q_{ik} = 1$ for any k, then $p_i^{(c)} = 1$. Hence, any group containing an individual control factor whose main effect has probability 1 of being active will be brought forward to the second stage with probability 1, and *all* the individual main effects, and interactions involving factors within that group, will then have to be estimated in the second stage experiment. This may result in an increase in the experiment size, even when only one highly active factor is put in a group with less likely-to-be-active factors. This suggests that grouping together the very-likely-to-be-active factors is to be preferred.

3.3.3 Different numbers of groups and group sizes for classical group screening

Example 3.3.3 Table 3.6 gives the results for the experiment in Example 3.3.2 when a CGS strategy is adopted, for a variety of group sizes and numbers of groups. It is clear, by comparison with the corresponding IGS results in Table 3.5 that the CGS strategy produces smaller values of expected size. However, the disadvantage of this apparent gain is that a greater number of substantial factorial effects are missed, as can be investigated via simulation.

Notice that the most economical experiment (starred) in terms of minimising the value of $E(S_{CGS})$ is that for which all the control factors believed very likely to be active are put in a single group. This is because the very-likely-to-be-active grouped control factors will be brought forward to the second stage with probability 1. Thus increasing the number of very-likely-to-be-active grouped control factors only increases the number of effects requiring estimation at the first stage. This is unlike Example 3.3.2 (IGS strategy) where increasing the number of very likely grouped control factors decreased the value of $E(S_{IGS})$. This is because grouped interactions could be eliminated at the first stage, reducing the number of individual interactions requiring estimation at the second stage. Having more grouped factors at the first stage increases the number of grouped interactions that can be examined at the first stage.

Table 3.6: Investigation of different groupings for Example 3.3.3 under classical group screening.

7 v likely indiv con		8 indiv con		4 indiv noise		$E(S_{CGS})$	$s.d.(S_{CGS})$	$P(S_{CGS} > u)$		
gps	sizes	gps	sizes	gps	sizes			120	150	180
1	7	1	8	1	4	156.40	51.07	0.83	0.63	0.63
1	7	1	8	2	2,2	141.75	46.16	0.83	0.63	0.22
1	7	2	2,6	1	4	132.03	47.37	0.62	0.56	0.20
1	7	2	2,6	2	2,2	118.89	42.89	0.62	0.32	0.07
1	7	2	3,5	1	4	126.98	47.59	0.59	0.25	0.25
1	7	2	3,5	2	2,2	114.16	43.21	0.42	0.25	0.09
1	7	2	4,4	1	4	125.40	47.81	0.72	0.26	0.26
1	7	2	4,4	2	2,2	112.68	43.46	0.47	0.26	0.09
1	7	3	2,2,4	1	4	111.75	40.43	0.51	0.26	0.06

continues overleaf

Table 3.6: continued

7 v likely indiv con		8 indiv con		4 indiv noise		$E(S_{CGS})$	$s.d.(S_{CGS})$	$P(S_{CGS} > u)$		
gps	sizes	gps	sizes	gps	sizes			120	150	180
1	7	3	2,2,4	2	2,2	99.95	36.45	0.36	0.13	0.02
1	7	3	2,3,3	1	4	109.41	39.71	0.34	0.18	0.07
1	7	3	2,3,3	2	2,2	97.77	35.80	0.25	0.10	0.02
1	7	4	2,2,2,2	1	4	98.94	33.20	0.35	0.10	0.01
* 1	7	4	2,2,2,2	2	2,2	88.07	29.63	0.19	0.04	0.00
2	2,5	1	8	1	4	157.40	51.07	0.83	0.63	0.63
2	2,5	1	8	2	2,2	142.75	46.16	0.83	0.63	0.22
2	2,5	2	2,6	1	4	133.03	47.37	0.62	0.56	0.20
2	2,5	2	2,6	2	2,2	119.89	42.89	0.62	0.32	0.07
2	2,5	2	3,5	1	4	127.98	47.59	0.59	0.25	0.25
2	2,5	2	3,5	2	2,2	115.16	43.21	0.42	0.25	0.09
2	2,5	2	4,4	1	4	126.40	47.81	0.72	0.26	0.26
2	2,5	2	4,4	2	2,2	113.68	43.46	0.47	0.26	0.09
2	2,5	3	2,2,4	1	4	112.75	40.43	0.51	0.26	0.06
2	2,5	3	2,2,4	2	2,2	100.95	36.45	0.36	0.13	0.02
2	2,5	3	2,3,3	1	4	110.41	39.71	0.34	0.18	0.07
2	2,5	3	2,3,3	2	2,2	98.77	35.80	0.25	0.10	0.02
2	2,5	4	2,2,2,2	1	4	99.94	33.20	0.35	0.10	0.01
2	2,5	4	2,2,2,2	2	2,2	89.07	29.63	0.19	0.04	0.00
2	3,4	1	8	1	4	157.40	51.07	0.83	0.63	0.63
2	3,4	1	8	2	2,2	142.75	46.16	0.83	0.63	0.22
2	3,4	2	2,6	1	4	133.03	47.37	0.62	0.56	0.20
2	3,4	2	2,6	2	2,2	119.89	42.89	0.62	0.32	0.07
2	3,4	2	3,5	1	4	127.98	47.59	0.59	0.25	0.25
2	3,4	2	3,5	2	2,2	115.16	43.21	0.42	0.25	0.09
2	3,4	2	4,4	1	4	126.40	47.81	0.72	0.26	0.26
2	3,4	2	4,4	2	2,2	113.68	43.46	0.47	0.26	0.09
2	3,4	3	2,2,4	1	4	112.75	40.43	0.51	0.26	0.06
2	3,4	3	2,2,4	2	2,2	100.95	36.45	0.36	0.13	0.02
2	3,4	3	2,3,3	1	4	110.41	39.71	0.34	0.18	0.07
2	3,4	3	2,3,3	2	2,2	98.77	35.80	0.25	0.10	0.02
2	3,4	4	2,2,2,2	1	4	99.94	33.20	0.35	0.10	0.01
2	3,4	4	2,2,2,2	2	2,2	89.07	29.63	0.19	0.04	0.00
3	2,2,3	1	8	1	4	158.40	51.07	0.83	0.63	0.63
3	2,2,3	1	8	2	2,2	143.75	46.16	0.83	0.63	0.22
3	2,2,3	2	2,6	1	4	134.03	47.37	0.62	0.56	0.20
3	2,2,3	2	2,6	2	2,2	120.89	42.89	0.62	0.32	0.07
3	2,2,3	2	3,5	1	4	128.98	47.59	0.59	0.25	0.25
3	2,2,3	2	3,5	2	2,2	116.16	43.21	0.42	0.25	0.09
3	2,2,3	2	4,4	1	4	127.40	47.81	0.72	0.26	0.26
3	2,2,3	2	4,4	2	2,2	114.68	43.46	0.47	0.26	0.09

continues overleaf

Table 3.6: continued

7 v likely indiv con		8 indiv con		4 indiv noise		$E(S_{CGS})$	$s.d.(S_{CGS})$	$P(S_{CGS} > u)$		
gps	sizes	gps	sizes	gps	sizes			120	150	180
3	2,2,3	3	2,2,4	1	4	113.75	40.43	0.51	0.26	0.06
3	2,2,3	3	2,2,4	2	2,2	101.95	36.45	0.36	0.13	0.02
3	2,2,3	3	2,3,3	1	4	111.41	39.71	0.34	0.18	0.07
3	2,2,3	3	2,3,3	2	2,2	99.77	35.80	0.25	0.10	0.02
3	2,2,3	4	2,2,2,2	1	4	100.94	33.20	0.35	0.10	0.01
3	2,2,3	4	2,2,2,2	2	2,2	90.07	29.63	0.19	0.04	0.00

3.4 Sensitivity of $E(S)$ to specified probabilities

In practice the probabilities of main effects being active are assigned following input from engineers with knowledge or experience of the product. This input is obtained from a web-based questionnaire and meetings. There is therefore uncertainty in the values that should be assigned to the probabilities. The study described in this section aims to investigate, for a 16 factor example, the sensitivity of the expected number of effects requiring estimation ($E(S)$) to the specified probabilities of individual effects being active. Both classical group screening and interaction group screening are considered.

For both strategies, an experiment with 8 individual control factors and 8 individual noise factors is considered in which the main effects of 4 of the individual control factors and 4 of the individual noise factors are believed very likely to be active. The main effects of the remaining 4 individual control and 4 individual noise factors are though less likely to be active.

3.4.1 Classical group screening

The effects on $E(S_{CGS})$ of the following four factors were investigated in a factorial experiment:

A: probability of an individual very likely control main effect being active

Table 3.7: The levels of factors A, B, C and D used in the sensitivity study under CGS.

Factor	Level		
	0	1	2
A	0.8	0.9	1.0
B	0.2	0.3	0.4
C	0.8	0.9	1.0
D	0.2	0.3	0.4

B: probability of an individual less likely control main effect being active

C: probability of an individual very likely noise main effect being active

D: probability of an individual less likely noise main effect being active.

A three-level full factorial design was used in order to gain a more thorough understanding of the behaviour of these factors than would be obtained from a two-level factorial experiment. The levels of the factors, are shown in Table 3.7. The software described in Section 3.1 was used to calculate the values of $E(S_{CGS})$ for all $3^4 = 81$ combinations of the factor levels.

Figure 3.4 shows the main effect plots for each of the four factors. The value of $E(S_{CGS})$ for each level of A, averaged over all the levels of the other factors, ranges from about 77 at the low value of the probability of a very likely main effect being active, to 79 at the high value. This range of 2 is small relative to the calculated values of $E(S_{CGS})$. This indicates that, for this example, the impact of increasing a very likely main effect probability from 0.8 to 1 is small. The corresponding range for factor C is slightly smaller than that of A. This difference arises because of the number of individual interactions (specifically control×control interactions) that require estimation at the second stage when a grouped control factor is declared active at the first stage. This number is generally greater than the number of individual interactions requiring estimation as a result of a grouped noise factor being declared active.

Factor B, the probability of the main effect of a less likely control factor being active, has the largest impact. The value of $E(S_{CGS})$ for each level of B, averaged over all the levels of the other factors, ranges from about 71 at the low level to 85 at the high level. This range of 14 indicates that $E(S_{CGS})$ is more sensitive to the probability of a less likely main effect being active being changed from 0.2 to 0.4. In other words, the impact on $E(S_{CGS})$ of increasing a main effect probability for an individual control factor from 0.2 to 0.4 is far greater than that of increasing the probability from 0.8 to 1.0. This increase in probability assigned to a main effect being active can be interpreted as a change from practically inactive (probability 0.2) to quite likely to be active (probability 0.4). It can be viewed as a greater change in prior belief than the increase in probability from 0.8 (very likely) to 1.0.

Similar results are found for the impact of factor D compared with that of C. However, the impact of D is less than that of B, again due to the lack of symmetry at the second stage experiment in terms of the greater number of individual effects requiring estimation due to a grouped control factor being declared active, as opposed to a grouped noise factor being declared active.

Figure 3.5 shows the two-factor interactions among the four factors A, B, C and D, and indicates that the impact of changing the probability of an individual main effect being active (for any factor) changes very little as the probabilities assigned to other factors are changed.

3.4.2 Interaction group screening

A similar study was carried out on the same experiment using interaction group screening and varying the following five factors:

A: probability of an individual very likely control main effect being active

B: probability of an individual less likely control main effect being active

C: probability of an individual very likely noise main effect being active

D: probability of an individual less likely noise main effect being active

E: values of the conditional probabilities used in assigning interaction probabilities.

Figure 3.4: Main effects plot of factors A, B, C and D for the CGS sensitivity study.

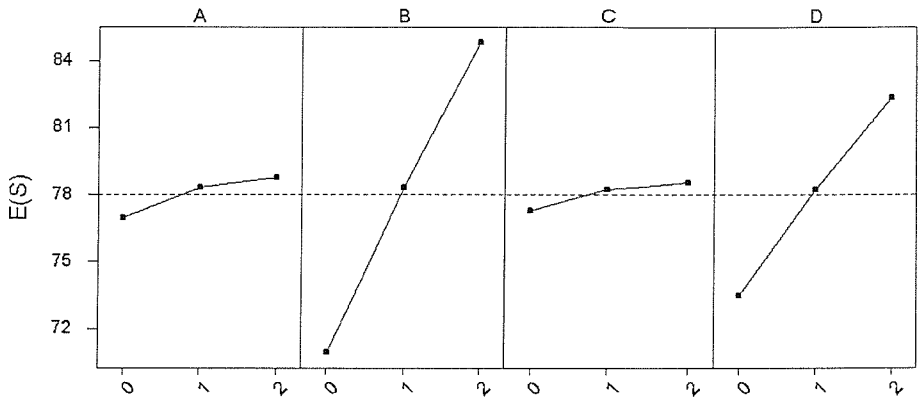


Figure 3.5: Two-factor interactions plot for A, B, C and D for the CGS sensitivity study.

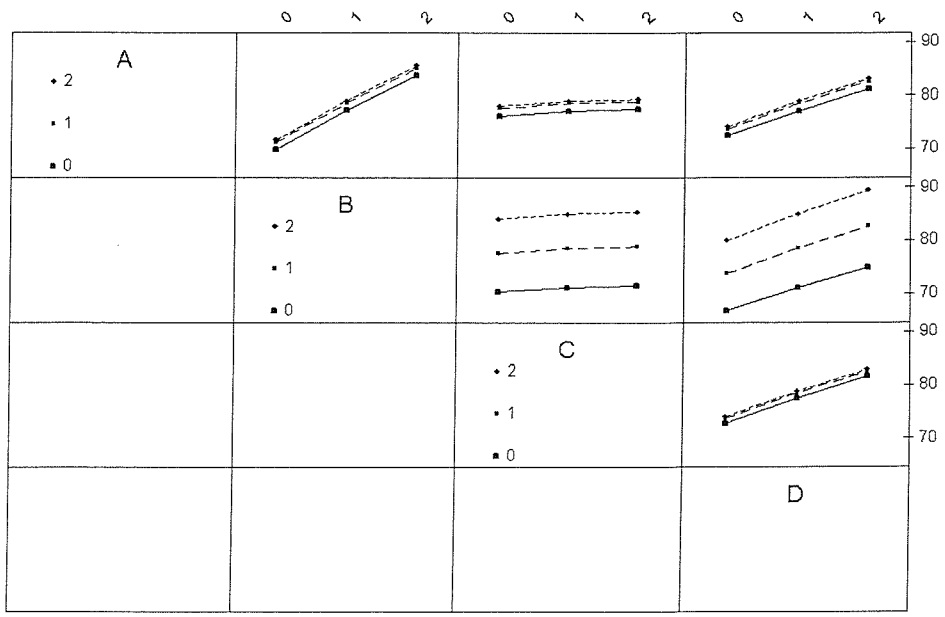


Table 3.8: The levels of factors A, B, C, D and E used in the sensitivity study under IGS.

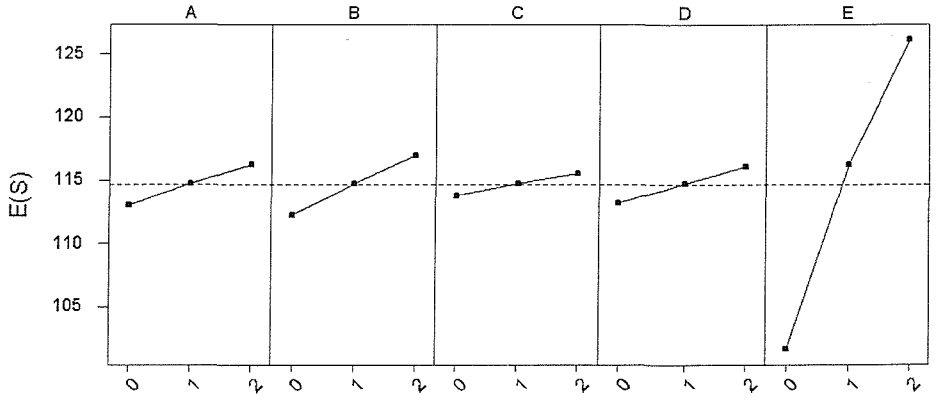
Factor	Level		
	0	1	2
A	0.8	0.9	1.0
B	0.2	0.3	0.4
C	0.8	0.9	1.0
D	0.2	0.3	0.4
E	0.005, 0.125, 0.25	0.0125, 0.1875, 0.375	0.01, 0.25, 0.5

Factors A, B, C and D are the same as those in the CGS study and were given the same levels in this experiment. The high level (2) of factor E was defined as setting the conditional probabilities w_{00}, $w_{10} = w_{10}$ and w_{11} for both control×noise and control×control interactions to the values 0.01, 0.25 and 0.5 respectively, as used by Chipman (1996). The low level (0) of E was half these values. The middle level (1) of E was defined as setting the conditional probabilities to the midpoints between the values at the high and low levels. The levels of the five factors are shown in Table 3.8.

A $\frac{1}{3}$ replicate of a full factorial 3^5 design (with $3^{5-1} = 81$ runs) was used for this investigation with defining contrast $I = ABCDE^2$. The aliasing scheme then allowed clear estimation of all five main effects and all $\binom{5}{2} = 10$ two-factor interactions.

Figure 3.6 shows that there is a slightly larger difference in $E(S_{IGS})$ when A is increased (from 0.8 to 1.0) than when C is increased (from 0.8 to 1.0) than occurred for CGS. An explanation of this difference is that, in the first stage experiment, under CGS, there is no difference between the number of effects requiring estimation due to a grouped control factor and due to a grouped noise factor. This is not the case for IGS due to the need to investigate grouped interactions at the first stage. As was found for CGS, the impact of increasing a very likely main effect probability from 0.8 to 1 is small.

Figure 3.6: Main effects plot of factors A, B, C, D and E for the IGS sensitivity study.

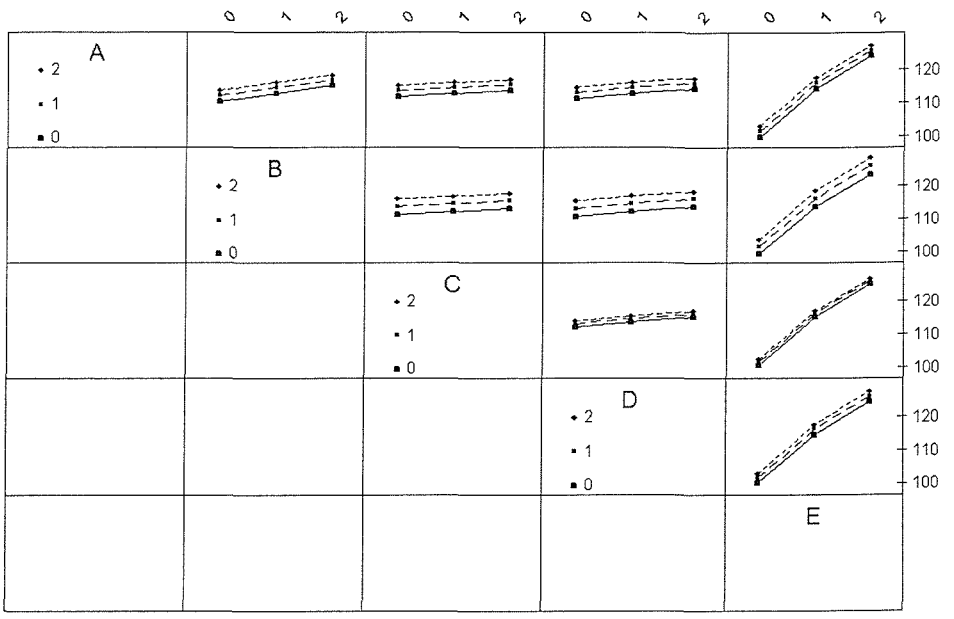


A comparison of Figures 3.4 and 3.6 shows that the value of $E(S_{IGS})$ is much less sensitive than under CGS to the specified probabilities of individual less likely main effects being active. An explanation of this finding is that, under IGS, there are more routes for a grouped factor to be carried forward to the second stage of experimentation than by having its main effect declared active at the first stage.

Figure 3.6 shows that the value of $E(S_{IGS})$ is sensitive to the specification of conditional probabilities used in assigning interaction probabilities. Halving the conditional probabilities has reduced $E(S_{IGS})$ from around 127 to 102, that is a reduction of 25. This sensitivity should be borne in mind when using the software to choose between possible groupings. In particular, when comparing two groupings under IGS, it is important to keep the same value for the conditional probabilities.

Figure 3.7 indicates that, as for CGS, all the two-factor interactions are negligibly small.

Figure 3.7: Two-factor interaction plot for A, B, C, D and E for the IGS sensitivity study.



3.4.3 Investigation of a wider range of probabilities

In order to find out if the above findings were also true for a wider range of values for factors A, B, C and D, the study was repeated for both CGS and IGS, for the example. The factor levels are shown in Table 3.9.

Figure 3.8 shows the main effects plot of factors A, B, C and D for CGS for the increased range of probabilities. The findings are very similar to those of Sections 3.4.1. As expected, for each factor, the value of $E(S_{CGS})$ for each factor level, averaged over the levels of the other factors, has a much wider range.

Figure 3.9 shows the main effects plot of factors A, B, C, D and E for IGS for the increased ranges of probabilities. There is an increased difference between the ranges of values of $E(S_{IGS})$ for A and B (C and D). This indicates that $E(S_{IGS})$ is more sensitive to changes in the main effect probability of a factor believed less likely to be active. Again, the difference in ranges of values of $E(S_{IGS})$ for A and B (C and D) is not as large as for CGS. The value of $E(S_{IGS})$ is still most sensitive to the settings of factor E, the values of the conditional probabilities used when

Table 3.9: The levels of factors A, B, C and D for a wider range of probabilities.

Factor	Level		
	0	1	2
A	0.55	0.775	1.0
B	0.01	0.28	0.55
C	0.55	0.775	1.0
D	0.01	0.28	0.55

assigning the interaction probabilities. Two-factor interactions plots for both CGS and IGS indicated that again, all the two-factor interactions are negligibly small.

Summary of findings for sensitivity study

In this study, $E(S)$ was very sensitive to the specified probabilities of main effects believed less likely to be active under classical group screening. This sensitivity was less under interaction group screening for the larger range of factor levels. For a smaller range of factor levels the sensitivity of $E(S_{IGS})$ to probabilities of very likely and less likely main effects was very similar. Under interaction group screening $E(S)$ was most sensitive to the specified conditional probabilities used in assigning interaction probabilities. All two-factor interactions were negligibly small under both screening strategies.

3.5 Conclusions

This chapter has shown how the ideas of Chapter 2 and the software written to implement them facilitates investigation of different groupings of factors and different assignments of probabilities for individual factorial effects being active, through their impact on the distribution of S, under both classical and interaction group screening. The software and methods also allow the experimenter to try different values for the assigned probabilities of effects being active to study the sensitivity of the results to changes in these probabilities. Such an investigation is an important step in making a decision between different groupings and strategies.

Figure 3.8: Main effects plot of factors A, B, C and D for the wider ranges of probabilities used in the CGS sensitivity study.

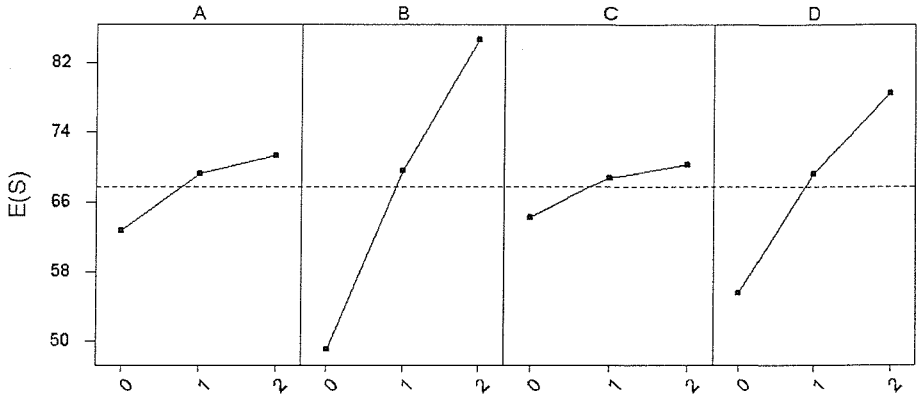
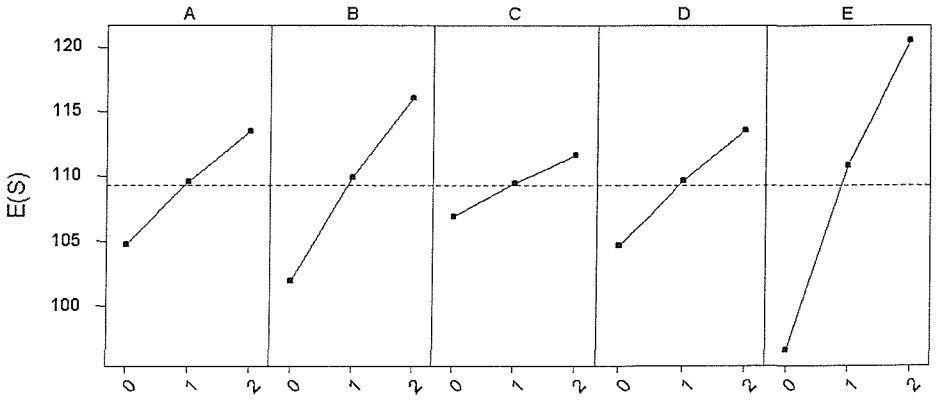


Figure 3.9: Main effects plot of factors A, B, C, D and E for the wider ranges of probabilities used in the IGS sensitivity study.



Example 3.2 demonstrated this software and showed the sensitivity of the results to changes in probabilities for certain individual interactions. It was found that, for this example, increasing the probabilities of individual interactions being active had the effect of increasing the expected experiment size and decreasing the standard deviation of the size of the two-stage experiment.

Examples, such as 3.3.2 and 3.3.3, and Study 1 suggest that grouping together individual factors whose main effects are more likely to be active is a good approach for minimising the expectation and standard deviation of the size of a two-stage experiment and hence minimising the probability of exceeding a target size. Grouping together factors whose main effects have widely differing probabilities of being active has the effect of increasing the expected size of the two-stage experiment. These examples indicate that a good strategy for CGS is to keep all very likely factors together in one group and put all of the remaining factors into as many groups as possible. These examples also indicate that for IGS a good strategy is to use group sizes which are as equal as possible, although empirical evidence, such as Example 3.3.1, suggests that groups sizes need not necessarily be as small as possible in order to minimise the expected size of the two-stage experiment.

The sensitivity study of Section 3.4 indicated that, under classical group screening, the expected number of effects requiring estimation was most sensitive to the probabilities assigned to individual less-likely main effects. Under interaction group screening, there was much less sensitivity to these probabilities. However, the specified values of the conditional probabilities used when assigning interaction probabilities had a large impact on the expected number of effects requiring estimation.

These conclusions give guidance to experimenters in deciding which groupings to consider thus reducing the number of groupings that need to be investigated in practice.

The work in this chapter, and other examples, has found that classical group screening often gives a smaller expectation but a more variable distribution of the size of the two-stage experiment than interaction group screening, and hence not necessarily a smaller probability of exceeding a target size. This apparent advantage of classical group screening does not address the risk of missing important effects or declaring unimportant factors to be important. When a small number of candidate

groupings has been identified, a simulation of a group screening experiment for each candidate grouping can be performed, using an extension of the software of Dean and Lewis (2002) to assess such risks. In examples, it is usually found that classical group screening misses a much larger proportion of important effects than interaction group screening.

Chapter 4

Practical application

4.1 Introduction

The generalised group screening theory described in Chapter 2 has been used in the planning of a screening experiment at Jaguar Cars in a programme whose aim was to investigate, and ultimately to improve, engine cold start optimisation in the company's new generation engine. This new engine has a more sophisticated electronic management system than earlier engines which enables many different factors to be investigated. Also, the development of a new test cell for the engine allowed noise factors to be investigated, such as the temperature surrounding the engine. The aim of the programme is to produce an engine that starts as reliably as possible regardless of, for example, ambient temperature or age of the engine. In order to measure this aspect of engine performance, Jaguar engineers had recently developed a measure of resistance across the spark plugs as the engine is taken through a number of test cycles. Small resistance indicates good starting performance of the engine. Hence the high (low) level of a factor is the level corresponding to a small (large) resistance.

The early stages of planning the experiment concentrated on collecting information from engineers involved in the relevant areas of design and manufacturing to ascertain views, and gather new suggestions, on factors that are likely to affect the engine cold start optimisation process, the likely importance of each factor, and the anticipated direction of influence of each factor on the response. The settings of each factor which would be used as the 'high' and 'low' levels in the experiment

also needed to be defined and agreed. This elicitation is an essential part of the planning process as it allows the number of factors in the experiment to be kept at a manageable level and can help to avoid overlooking important factors. Conventionally, this information is gathered through a local meeting with the engineers, called a 'brainstorming session'. This approach has the disadvantage that experts who are located at other company sites (Ford, Jaguar, Land Rover in the UK and US) are excluded. Also, accurate opinions may sometimes not be obtained from a brainstorming meeting because of the influence of certain members of the group. As part of the EPSRC project, a dynamic web-based questionnaire was developed to overcome these difficulties (see Dupplaw et al. (2004)) and used to allow a wider group of engineers, in the UK and overseas, to contribute their opinions and comments. As described in Section 3.1, this questionnaire forms part of a software system called GISEL (grouping in screening with elicitation) which also incorporates the software described in Chapter 3. The results from the questionnaire were then used in subsequent local discussion meetings before compiling the final list of factors and their settings for experimentation.

The factors chosen for the experiment, together with the results of the grouping investigations under interaction group screening, are given in Section 4.2. This section also describes an initial 'try out' experiment on grouped factors at Jaguar and the lessons learned from this experiment. Results of investigations of different groupings are described in Section 4.3 and the plan is given for the first stage of the full two-stage experiment. In Section 4.4, the analysis of the data from the first stage experiment is described and, in Section 4.5, the plan for the second stage experiment is given. The data from the second stage experiment are presented in Section 4.6, together with the details of their analysis. A discussion of the results is given in Section 4.7.

4.2 Initial investigations

The initial investigations described in this section consist of, first, identifying the factors for inclusion in an experiment and, secondly, a trial grouped experiment which was performed on the currently used Jaguar engine. From the web-based

Table 4.1: Initial list of factors for inclusion in the engine cold start optimisation experiment where * indicates factors whose levels are hard to change.

Control factors		Noise factors
Very likely	Less likely	
*Plug type**	*Plug gap**	*Injector tip leakage**
*Fuel type**	*Spark time during crank*	*Injector spray angle*
Air to fuel ratio (AFR)	*Spark time during run-up*	*& direction*
Injection timing	*Engine off-timer*	*Humidity/temperature*
	Idle flare	*Variable valve timing*
	Higher idle speed	

questionnaire, a list of 40 possible factors for investigation was obtained, together with opinions from 15 experts. From this list, 10 control and 4 noise factors were identified for investigation, based on the engineers' views on their likely importance and how feasible it was to vary them in a controlled way. These factors are listed in Table 4.1 which indicates the four control factors that were believed very likely to have active main effects. For four of the factors, changing from one level to another was quite a slow process; for example, for one factor it involved changing the spark plugs. These hard-to-change factors are identified in the table. In the initial planning discussions, the engineers thought that it would take so long to change these factors that only four runs could be made in a session (morning or afternoon).

The strategy of classical group screening was not considered a sound choice for the experiment at Jaguar. Although the expected number of observations needed in a two-stage experiment is usually smaller under classical group screening, the number of observations tends to be more variable, as was discussed in Chapter 3. Also, importantly, investigations using the simulation software of Dean and Lewis (2002), extended to allow unequal group sizes and probabilities of active effects, indicated that the proportion of missed active control×noise interactions is larger under classical group screening. The engineers were particularly interested in the

investigation of two-factor interactions, as previous experiments in which only main effects had been considered, had often produced smaller improvements in engine performance than had been anticipated.

Probabilities for the main effects and interactions being active were assigned as follows. The main effects for each of the control factors believed very likely to be active were assigned probability 1. The main effects of the control factors believed less likely to be active were each assigned probability 0.167, as this corresponded to an expectation of 1 out of 6 being active. The noise main effects were each assigned a probability 0.5 of being active, which corresponded to 2 out of 4 being active. The individual control×noise and control×control interaction probabilities were calculated using the conditional probabilities $w_{00} = 0.005, w_{01} = w_{10} = 0.125$ and $w_{11} = 0.25$ in formula (2.31) of Section 2.5.

An investigation of different groupings with interaction group screening was carried out using these probabilities and selected results are shown in Table 4.2. In this investigation, the very likely and less likely control factors were kept in separate groups, because any factor that is grouped with a very likely control factor is automatically brought forward to the second stage of experimentation with probability 1. (This is because the probability of a main effect being active is 1, see Section 3.3.2.) An exception to this approach was made in order to investigate groupings that included all the three hard-to-change control factors in a single group. These groupings are the final four listed in Table 4.2, where the group of hard-to-change factors is denoted by 3*.

Although the final four groupings in Table 4.2 did not give the smallest values for $E(S_{IGS})$ and $s.d.(S_{IGS})$, it was decided to use one of these groupings so that all the factors whose levels were hard to change would be changed together in the experiment. This allowed the experiment to be carried out in a reasonable amount of time. The last grouping listed in Table 4.2 gave the smallest value for $E(S_{IGS})$ out of the final four and one of the smallest values for $s.d.(S_{IGS})$. Hence it was chosen for the first stage experiment. The details of this grouping are given in Table 4.3.

The experiment plan chosen was a $\frac{1}{4}$ replicate regular fractional factorial for five grouped control factors (labelled 1-5) and two grouped noise factors (labelled 6 and

Table 4.2: Results of an investigation of different groupings for the initial list of factors in Table 4.1 and interaction group screening, for the engine cold start optimisation experiment, where * indicates groups of factors whose levels are hard to change.

Very likely control	Less likely control	Noise	$E(S_{IGS})$	$s.d.(S_{IGS})$	$P(S_{IGS} > u)$		
Grouping					80	90	100
2*,1,1	5,1*	2,1,1*	96.43	9.88	0.94	0.76	0.36
2*,1,1	5,1*	3,1*	96.14	10.58	0.91	0.72	0.38
2,2*	5,1*	3,1*	95.40	11.05	0.90	0.69	0.35
2,2*	5,1*	2,1,1*	94.40	10.40	0.90	0.66	0.30
2*,1,1	3,2,1*	2,1,1*	92.11	7.11	0.91	0.58	0.15
2,2*	2,1,2,1*	2,1,1*	90.75	7.22	0.91	0.57	0.13
2*,1,1	3,2,1*	3,1*	89.96	9.51	0.85	0.49	0.13
2,2*	3,2,1*	2,1,1*	88.57	9.34	0.81	0.43	0.09
2,2*	2,1,2,1*	3,1*	88.44	9.00	0.82	0.42	0.08
2,2*	3,2,1*	3,1*	87.94	10.19	0.78	0.42	0.10
2 3*	2,1,1,1	2,1,1*	93.66	7.77	0.93	0.65	0.21
2 3*	2,1,1,1	3,1*	91.40	9.08	0.88	0.56	0.16
2 3*	3,2	3,1*	89.39	11.17	0.79	0.48	0.16
2 3*	2,2,1	2,1,1*	89.11	9.33	0.82	0.46	0.11
2 3*	3,2	2,1,1*	88.63	10.45	0.79	0.45	0.12
2 3*	2,2,1	3,1*	88.35	9.97	0.79	0.43	0.11

Table 4.3: Details of the chosen grouping of the factors in Table 4.1.

Grouped control factors	
Group 1	*Plug type* * & *Fuel type* * & *Plug gap**
Group 2	*Air to fuel ratio (AFR) & Injection timing*
Group 3	*Spark time during crank & Spark time during run-up*
Group 4	*Engine-off timer & Higher idle speed*
Group 5	*Idle flare*
Grouped noise factors	
Group 6	*Injector tip leakage**
Group 7	*Variable valve timing & Injector spray angle and direction & Humidity/temperature*

7) with defining contrast

$$I = 1234 = 12567 = 34567.$$

A regular fraction was preferred to a smaller fraction as the estimated factorial effects are independent and hence the hypothesis tests of the sizes of the effects are also independently made. In this particular fraction, in order to keep the fraction small, three pairs of grouped control×control interactions, namely, Group 1×Group 2 and Group 3×Group 4, Group 1×Group 3 and Group 2×Group 4, and Group 1×Group 4 and Group 2 ×Group 3, are aliased together. All the main effects and the remaining two-factor interactions can be estimated.

In order to accommodate the groups of hard-to-change factors (labelled 1 and 6), the 32 runs were arranged so that one grouped control factor and one grouped noise factor were changed every four runs. Each set of four runs thus defines a session or wholeplot. The design is shown in Table 4.4, where the levels of the factors in each treatment combination are written in the factor order 1, 6, 2, 3, 4, 5, 7, that is, so that the groups of factors whose levels are hard to change (the whole-plot factors) are listed first. This ordering shows clearly how the levels of factors 1 and 6 are held fixed throughout each session. All the control×noise interactions except two (16, 47) are estimable within sessions and every control×noise interaction can be

Table 4.4: A plan for the 'try out' grouped experiment. The grouped control factors are labelled 1-5 and the grouped noise factors are labelled 6 and 7. The groups of factors whose levels are hard to change are 1 and 6. The low and high levels of a grouped factor are indicated by 0 and 1 respectively, and the grouped factors are ordered so that the two wholeplot factors are listed first.

Groups	16 23457	16 23457	16 23457	16 23457
Session 1	00 00000	00 10101	00 11010	00 01111
Session 2	01 00001	01 10100	01 11011	01 01110
Session 3	11 10001	11 00100	11 01011	11 11110
Session 4	10 10000	10 00101	10 01010	10 11111
Session 5	01 00010	01 10111	01 11000	01 01101
Session 6	00 00011	00 10110	00 11001	00 01100
Session 7	10 10011	10 00110	10 01001	10 11100
Session 8	11 10010	11 00111	11 01000	11 11101

estimated independently of all main effects and two-factor interactions. Also, the grouped control×control interaction Group 3×Group 5 cannot be estimated within sessions. Two degrees of freedom for error are available for estimating session-to-session error, and four for the within sessions (subplot) error.

This experiment was started at Jaguar Cars but was halted after only four sessions due to commercial pressure on the engine testing facility. However, a number of valuable lessons were learned that informed the planning of the full two-stage group screening experiment (Section 4.3). These were

1. *Engine off-timer*, *Variable valve timing*, *Injector spray angle and direction* and *Humidity* could not be adequately controlled for inclusion in an experiment,

2. the control factor *Fuel type* was too difficult to change and so should no longer be kept as a factor in the experiment, and

3. each run took less time than previously anticipated.

Further, for future experiments, the engineers decided to vary a factor called

Table 4.5: The final list of 10 individual factors for the full two-stage group screening experiment on cold start combustion. * indicates a factor whose levels are hard to change.

Control factors		Noise factors
Very likely	Less likely	
Air to fuel ratio (AFR)	*Spark time during crank*	*Temperature*
Injection timing	*Spark time during run-up*	*Injector tip leakage*
*Plug type**	*Cranking fuel*	
*Plug gap**	*Higher idle speed*	

Cranking fuel instead of the factor *Idle flare* because it was easier to vary and could be set to achieve the same condition as *Idle flare*. The runs achieved in this experiment were very useful as a pilot experiment and guided the planning of the full two-stage experiment.

Several months elapsed before a new generation engine became available and time in the testing cells could be scheduled for the experiment. It was decided in discussions during this period that the control factor *Plug gap* should be promoted from being a factor whose main effect is less likely to be active, to one whose main effect is very likely to be active. It was also decided that the levels of the factor *Injector tip leakage* could be altered indirectly by changing fuel pressure in the engine, with the result that this factor was not regarded as hard-to-change in the full experiment.

4.3 First stage experiment

The knowledge gained from the pilot experiment was used in the planning and running of a full two-stage group screening experiment. The reduced list of 8 control and 2 noise factors identified for investigation are listed in Table 4.5. It was decided that any non-zero factorial effects were of interest, that is, $\Delta = 0$ was used.

The main effects for the control factors believed very likely to be active were again assigned probability 1 of being active and the remaining main effects of control

factors, believed less likely to be active, were assigned probability 0.25. The noise main effects were assigned probability 0.5 of being active, which corresponds to the view that one of them is active. The approach of Chipman (1996) (Section 2.5) was again used in allocating the interaction probabilities using values 0.005, 0.125, 0.125 and 0.25 for the conditional probabilities for both control×noise and control×control interactions.

Table 4.6 shows the results of an investigation of different groupings under interaction group screening using these probabilities. Factors whose levels are hard to change were grouped separately from the other factors. This avoided effects of factors that were not hard to change being estimated at the wholeplot (session) level. Very likely and less likely control factors were also kept in separate groups, as in the pilot study, as any factors grouped with the very likely control factors would be brought forward to the second stage of experimentation with probability 1. In choosing these groupings, account was taken of the engineers' preferences based on their experience at the first stage.

Groupings 1 to 5, listed in Table 4.6, all have similar values for $E(S_{IGS})$ and $s.d.(S_{IGS})$. The simulation software (described in Section 3.1 of Chapter 3) was used to assess the impact of putting the two noise factors in separate groups by comparing grouping 1 and grouping 5. These results are shown in Table 4.7. The proportions of active control and noise main effects and control×control interactions which were not detected in the simulations, were consistently greater for grouping 5 for all the active effect distributions with mean 10. However, for the active effect distributions with means 30 and 50, the proportions missed under groupings 1 and 5 were very similar. Most importantly, the proportions of control×noise interactions missed under grouping 5 were substantially smaller than the proportions missed under grouping 1. Hence, it was decided to keep the noise factors in separate groups and grouping 5 was chosen for the first stage experiment. The details of this grouping are given in Table 4.8.

The chosen grouping has four grouped control factors and two grouped noise factors. A regular 2^{6-1} fraction was chosen as the plan for the first stage experiment, with defining contrast

$$I = 123456$$

Table 4.6: Results of an investigation of different groupings under interaction group screening, for the cold start optimisation experiment where * indicates groups of factors whose levels are hard to change.

	Very likely control	Less likely control	Noise	$E(S_{IGS})$	$s.d.(S_{IGS})$	$P(S_{IGS} > u)$		
	Grouping					40	50	60
1	2,2*	2,2	2	50.79	6.82	0.93	0.56	0.04
2	2,2*	2,1,1	2	52.63	6.01	0.97	0.65	0.10
3	1,1 2*	2,2	2	52.66	6.07	0.97	0.68	0.1
4	2,2*	1,3	2	52.87	6.94	0.95	0.66	0.13
5	2, 2*	2,2	1,1	52.94	6.20	0.97	0.63	0.09
6	2,2*	4	2	54.15	7.59	0.94	0.75	0.19
7	2,2*	4	2	54.73	6.80	0.98	0.73	0.18
8	2,2*	1,3	1,1	54.89	6.27	0.98	0.77	0.19
9	1,1,2*	2,2	2	52.66	6.07	0.97	0.68	0.10
10	1,1,2*	2,2	1,1	55.79	5.495	0.996	0.83	0.2
11	2,2*	2,1,1	1,1	56.04	5.44	1.0	0.84	0.21
12	1,1,2*	2,1,1	1,1	60.49	4.708	1.0	0.98	0.50

Table 4.7: Simulation results for the proportions of active individual control main effects (cme), noise main effects (nme), control×control interactions and control×noise interactions (c×n) that fail to be detected under groupings 1 and 5 of Table 4.6, for interaction group screening.

	Grouping 1				Grouping 5			
Active effect	Proportion missed				Proportion missed			
distribution	cme	nme	c×c	c×n	cme	nme	c×c	c×n
N(10, 4)	0.05	0.13	0.40	0.74	0.08	0.35	0.53	0.58
N(10, 9)	0.05	0.11	0.42	0.74	0.08	0.30	0.52	0.58
N(10, 16)	0.06	0.12	0.42	0.75	0.09	0.26	0.51	0.55
N(10, 25)	0.07	0.10	0.42	0.74	0.10	0.23	0.49	0.53
N(30, 4)	0.00	0.00	0.06	0.56	0.00	0.00	0.05	0.09
N(30, 9)	0.00	0.00	0.06	0.58	0.00	0.00	0.06	0.10
N(30, 16)	0.00	0.00	0.07	0.56	0.00	0.00	0.06	0.07
N(30, 25)	0.00	0.00	0.05	0.56	0.00	0.00	0.04	0.08
N(50, 4)	0.00	0.00	0.05	0.58	0.00	0.00	0.04	0.11
N(50, 9)	0.00	0.00	0.05	0.57	0.00	0.00	0.05	0.09
N(50, 16)	0.00	0.00	0.05	0.56	0.00	0.00	0.05	0.10
N(50, 25)	0.00	0.00	0.05	0.56	0.00	0.00	0.05	0.08

Table 4.8: Details of the chosen grouping for the first stage experiment to investigate engine cold start optimisation.

Grouped control factors	
Group 1	*Plug type* * *& Plug gap**
Group 2	*Air to fuel ratio (AFR) & Injection timing*
Group 3	*Spark time during crank & Spark time during run-up*
Group 4	*Higher idle speed & Cranking fuel*
Grouped noise factors	
Group 5	*Injector tip leakage*
Group 6	*Temperature*

where 1, 2, 3, 4, 5 and 6 are the labels for the grouped factors formed from Groups 1 to 6 respectively. This fraction has 32 runs and, with no blocking, can estimate all main effects and two-factor interactions, when it is assumed that all four-factor interactions are negligible. For the factors whose levels are hard to change, a change every session of eight runs was considered feasible, in light of the pilot study. Thus the fraction had to be arranged in four sessions or wholeplots in such a way that confounding effects of interest was avoided where possible. The experiment plan is shown in Table 4.9, together with the observations. The effects of Group 1 (hard to change), the interaction (56) between the grouped noise factors 5 and 6 and the aliased three-factor interactions $156 = 234$, are confounded with sessions or wholeplots. The remaining five main effects, 14 two-factor interactions and $\frac{1}{2}\binom{6}{3} - 1 = 9$ pairs of aliased three-factor interactions are all estimable within wholeplots. A breakdown of the degrees of freedom is summarised in Table 4.10.

4.4 Analysis of first stage experiment

The results from the first stage experiment are now considered. For clarity, several of the tables of results (Tables 4.11 to 4.15) are located at the end of this chapter. Table 4.11 gives the analysis of variance for the usual split-plot linear model for the resistance measure, where all three-factor interactions and higher are assumed negligible and the estimated factorial effects are given in Table 4.12. It is interesting to note that a very large proportion of the total variation within sessions in the experiment is explained by the main effects and two-factor interactions estimated within the sessions (97.5%). In the wholeplot analysis, there are insufficient degrees of freedom to carry out a formal test, or graphical assessment, of the main effect of factor 1. However, in view of the similar size of the estimated main effect compared with the three-factor interaction, it was decided not to take this factor (Group 1) forward to the second stage. The residual plots resulting from the analysis (not shown) showed no abnormal patterns.

In order to draw conclusions about the 19 main effects and interactions estimated within the sessions, we use a two-sided t-test of the hypothesis that the factorial effect is zero and the Bonferroni method for multiple testing. The critical

Table 4.9: The plan and observations for the Stage I experiment. The grouped control factors are labelled 1, 2, 3, 4 and the grouped noise factors are labelled 5 and 6. The low and high levels of a grouped factor are indicated by 0 and 1 respectively.

Session no.	Test no.	Gpd factor 1 2 3 4 5 6	Response (scaled)	Session no.	Test no.	Gpd factor 1 2 3 4 5 6	Response (scaled)
	1	1 0 1 0 0 0	6.57		17	0 0 1 1 1 1	28.09
	2	1 0 1 0 1 1	2.74		18	0 1 0 1 0 0	-3.86
	3	1 0 0 1 0 0	5.15		19	0 0 0 0 1 1	-5.12
1	4	1 0 0 1 1 1	10.19	3	20	0 0 0 0 0 0	-12.36
	5	1 1 1 1 0 0	8.17		21	0 0 1 1 0 0	-11.36
	6	1 1 0 0 0 0	41.98		22	0 1 0 1 1 1	-2.61
	7	1 1 0 0 1 1	10.27		23	0 1 1 0 0 0	-6.52
	8	1 1 1 1 1 1	9.50		24	0 1 1 0 1 1	-5.69
	9	1 0 0 0 1 0	-11.59		25	0 1 0 0 0 1	-7.43
	10	1 0 0 0 0 1	-6.04		26	0 0 1 0 1 0	-11.23
	11	1 1 1 0 0 1	-5.57		27	0 1 1 1 0 1	-7.25
2	12	1 0 1 1 1 0	-1.75	4	28	0 1 1 1 1 0	-2.73
	13	1 0 1 1 0 1	-3.67		29	0 1 0 0 1 0	-0.82
	14	1 1 0 1 0 1	-2.48		30	0 0 1 0 0 1	-6.16
	15	1 1 0 1 1 0	3.63		31	0 0 0 1 1 0	-1.50
	16	1 1 1 0 1 0	-7.71		32	0 0 0 1 0 1	-2.82

Table 4.10: The breakdown of the degrees of freedom for the experiment plan shown in Table 4.9.

Between sessions - 3 d.f	Within sessions - 28 d.f.
1 main effect (1)	5 main effects (2 - 6)
1 noise×noise (56)	6 control×control (all)
1 3-factor interaction (156=234)	8 control×noise (all)
	9 aliased 3-factor interactions

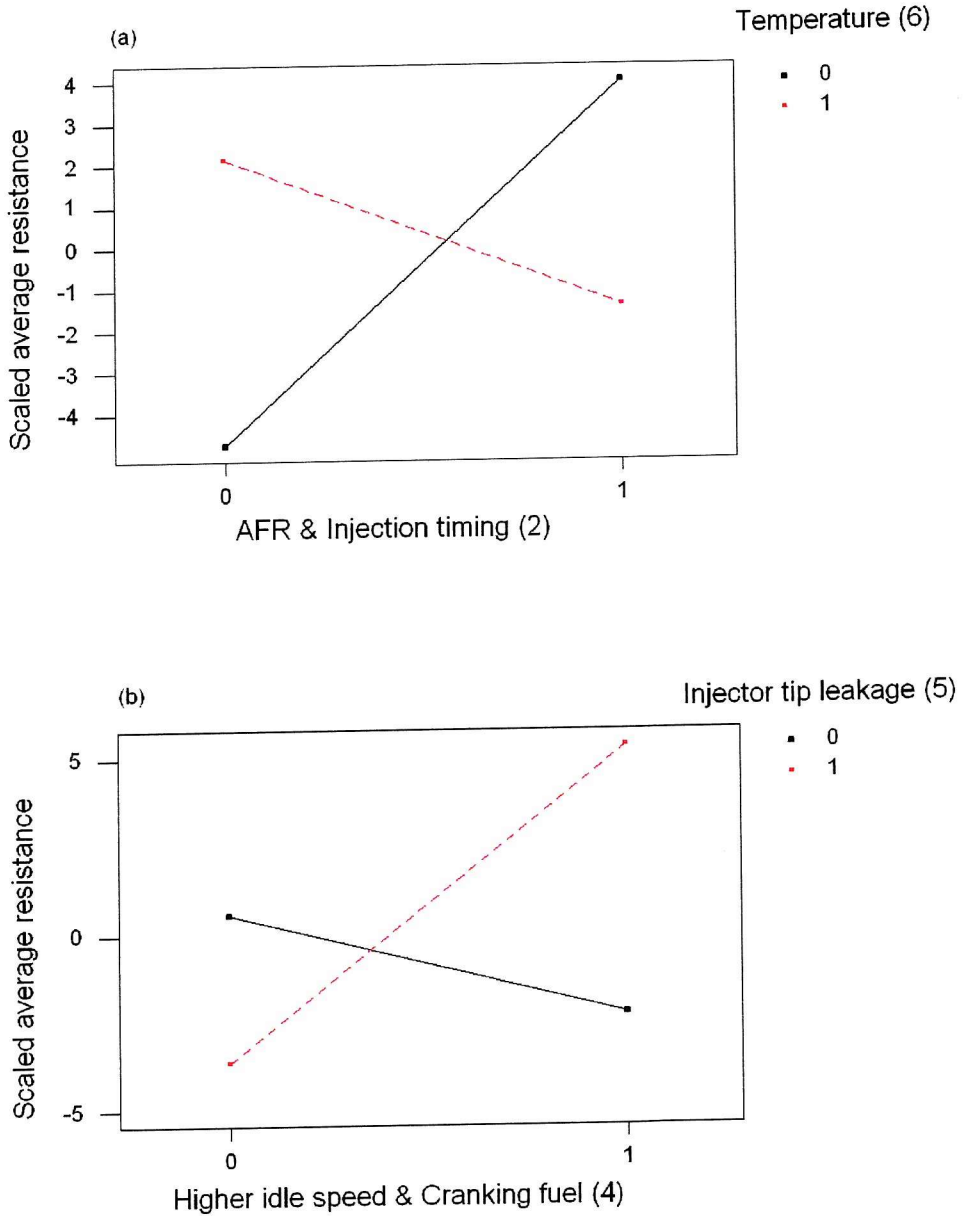
value of an overall significance level of 5% (1%) is then 0.00263 (0.000526). From Table 4.12, there are five grouped factorial effects that are significant at the 1% level. In order of decreasing *t*-test statistic value (on 9 degrees of freedom) these are: 26, 45, 15, 23 and 24. Figure 4.1 shows the interaction plots for each of the two most significant grouped control×noise interactions, 26 and 45. Figure 4.1 (a) indicates that setting both *AFR* and *Injection timing* to their high levels may reduce the variation in resistance as temperature varies. Similarly, Figure 4.1 (b) indicates that setting both *Higher idle speed* and *Cranking fuel* to their low levels may reduce the variation in resistance as the injector tip leakage varies. Plots for the remaining three interactions that are significant at an overall 1% significance level are given in Figures D.1 - D.3 in Appendix D.

In order to keep the second stage experiment to a manageable size, only the two largest significant effects, that is, the above two grouped control×noise interactions, were followed up. Hence the individual factors within Groups 2 and 4 of control factors and Groups 5 and 6 of noise factors were brought forward to the second stage experiment.

4.5 Second stage experiment

A second stage experiment was needed to investigate all main effects of the individual factors within Groups 2, 4, 5 and 6. It was also required to investigate the individual control×control interactions within each of Groups 2 and 4 (which could not be examined at the first stage) and the individual control×noise interactions

Figure 4.1: Interaction plots for the grouped control×noise interactions: (a) Group 2×Group 6 and (b) Group 4×Group 5.



between the factors in Groups 2 and 6, and between the factors in Groups 4 and 5. That is, the following individual factorial effects needed to be estimated:

- the main effects of the control factors *AFR* (A), *Injection timing* (B), *Higher idle speed* (C) and *Cranking fuel* (D)
- the main effects of the noise factors *Injector tip leakage* (E) and *Temperature* (F)
- *AFR*×*Injection timing* (AB)
- *Higher idle speed*×*Cranking fuel* (CD)
- *AFR*×*Temperature* (AF)
- *Injection timing*×*Temperature* (BF)
- *Higher idle speed*×*Injector tip leakage* (CE)
- *Cranking fuel*×*Injector tip leakage* (DE).

The experiment plan chosen to investigate the six factors was again the half replicate (with $2^{6-1} = 32$ runs) with defining contrast

$$I = ABCDEF.$$

This design was used (rather than a smaller fraction) because all main effects and two-factor interactions could be independently estimated. Further, as no hard-to-change factor was included in this study, the experiment could be carried out reasonably swiftly (approximately three days). The plan and observations are shown in Table 4.13.

4.6 Analysis of second stage experiment

A linear model was assumed for the scaled average resistance which included terms for the main effects of factors $A - F$ and for all two-factor interactions. Three-factor interactions and higher order effects were assumed negligible. The random

errors were assumed to be independently and normally distributed with constant variance.

Table 4.14 gives the analysis of variance for this model and shows that the main effects and two-factor interactions account for 85% of the variation. The residual plots (not shown) gave no evidence that the model assumptions were violated. Table 4.15 shows the estimated effects with associated p-values for a t-test of a single hypothesis that the corresponding factorial effect is zero, against the alternative that it is non-zero. The table indicates the six factorial effects which have p-value < 0.15 with their rank ordering. The two noise factors *Injector tip leakage* (E) and *Temperature* (F) have highly significant main effects (p-value < 0.01). The control×control interaction *AFR*×*Higher idle speed* (AC) is the next strongest effect (p-value=0.07), followed by the control×control interactions *Higher idle speed*×*Cranking fuel* (CD) and *AFR*×*Injection timing* (AB), and the control×noise interaction *Cranking fuel*×*Injector tip leakage* (DE) (all with $p < 0.15$). A p-value between 0.10 and 0.15 is usually regarded as weak evidence that a factorial effect is non-zero in an experiment aimed at modelling a response in terms of a small number of factors. In a screening study, however, it is often regarded as sufficient for the effects to be considered for inclusion in the next experiment. From the group screening experiment we conclude that the four control factors, *AFR*, *Injection timing*, *High idle speed*, *Cranking fuel* should be investigated further, together with the noise factor *Injector tip leakage*. Further information about the interactions is given below.

Figure 4.2 shows the main effect of the noise factor *Temperature* and indicates that average resistance is lower when *Temperature* is at its higher level (15^0C) than at its lower level (-5^0C). Note that *Temperature* did not feature in any of the more important interactions at the second stage.

Figure 4.3 shows the control×control interaction *AFR*×*Higher idle speed* and indicates that both factors should be set to their high level in order to achieve low average resistance, and that the high level of *AFR* together with the low level of *Higher idle speed* gives a similarly low resistance.

The control×control interaction *Higher idle speed*×*Cranking fuel* is plotted in Figure 4.4 and this plot indicates that *Higher idle speed* should be set to its low

Figure 4.2: The main effect plot for *Temperature*.

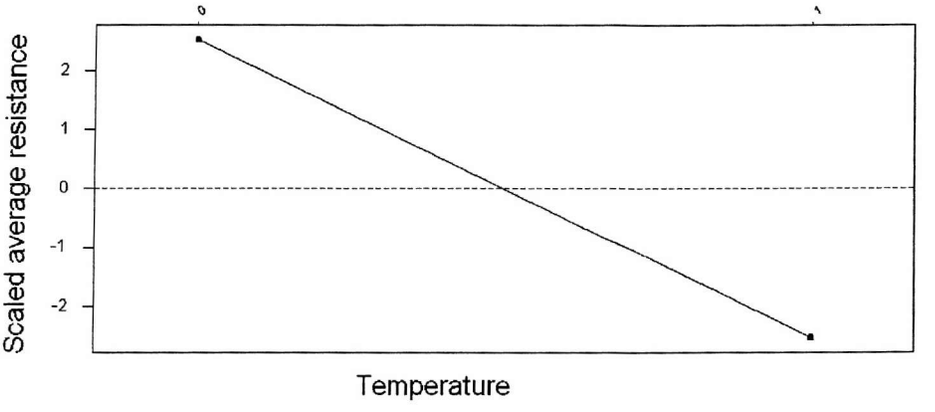
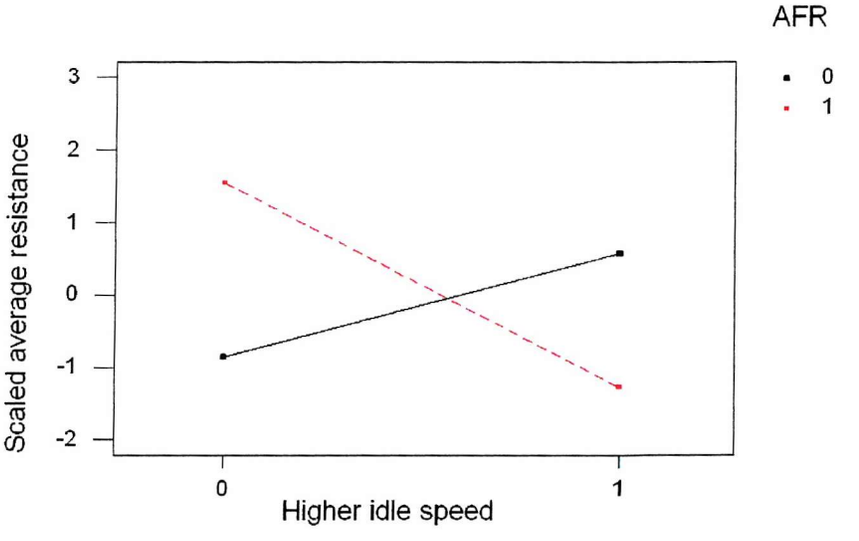


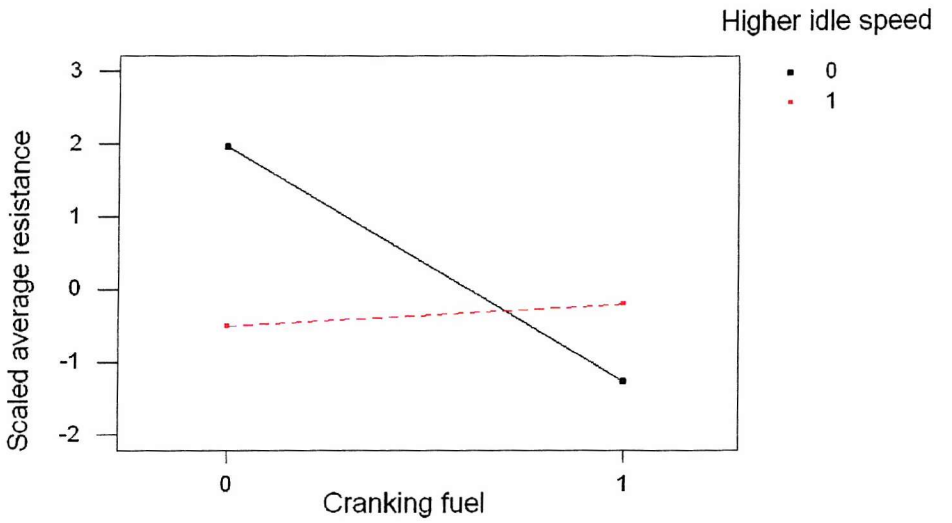
Figure 4.3: The control×control interaction *AFR*×*Higher idle speed*.



level and *Cranking fuel* to its high level in order to reduce the average resistance.

The control×control interaction *AFR*×*Injection timing* is shown in Figure 4.5. The plot indicates that, in order to reduce resistance, the two factors should both

Figure 4.4: The control×control interaction *Higher idle speed*×*Cranking fuel*.



be set to their high levels. This finding is consistent with results from the stage 1 experiment shown in Figure 4.1.

The control×noise interaction *Cranking fuel*×*Injector tip leakage* is illustrated in Figure 4.6. The plot indicates that to reduce variation in the response as injector tip leakage varies (that is, allowing the age or condition of the engine to vary), the control factor *Cranking fuel* should be set to its higher level. We conclude that the screening experiment indicates that *Cranking fuel* may be a useful control factor for achieving low resistance that is insensitive to the age or condition of the engine.

4.7 Discussion

Our experience of designing and analysing this two-stage group screening experiment indicated that it was a practical method of screening. An important aspect of understanding the technique from this experiment is to consider evidence of consistency between the two stages of experiment.

The most significant individual interaction at the second stage, AC, is part of the two-factor interaction between Groups 2 and 4 in the sense that $24 = AC +$

Figure 4.5: The control×control interaction *AFR×Injection timing.*

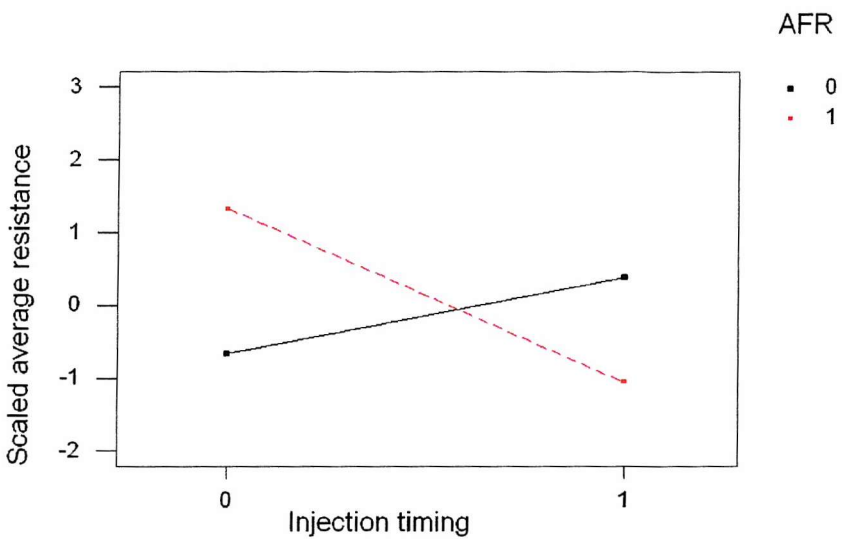
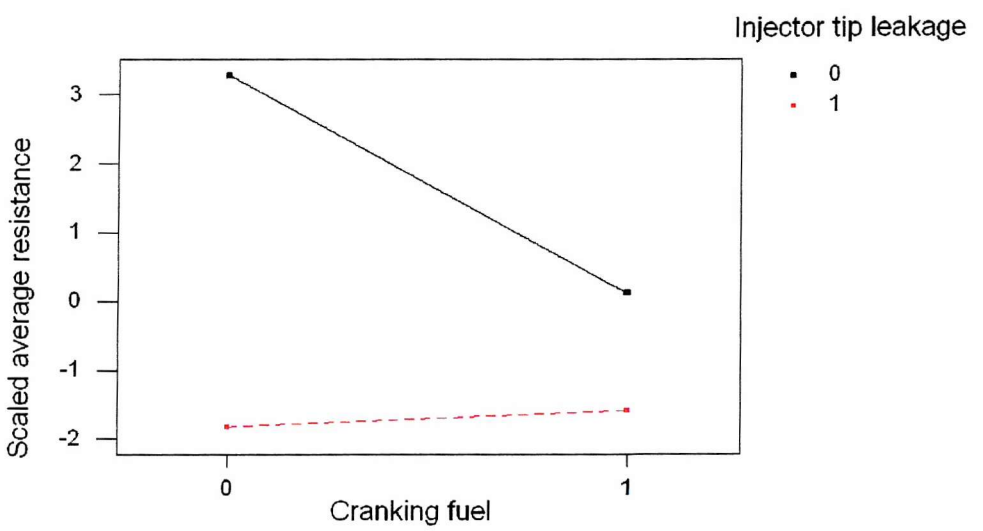


Figure 4.6: The control×noise interaction *Cranking fuel×Injector tip leakage.*



$AD + BC + BD$ (see Lewis and Dean (2001), Theorem 1), i.e., the single contrast 24 in the grouped factors represents four aliased individual factorial effects. The grouped interaction 24 was the third most important effect at stage 1. The next

most significant individual interaction from the stage 2 experiment, CD, was only considered at this second stage as it is an interaction between the two factors within Group 4. Similarly, the individual interaction, AB, is an interaction between the two factors within Group 2. The last of the significant individual interactions, DE, is part of the interaction between grouped factors 4 and 5 ($45 = CE + DE$) which was the second most significant grouped effect. For the grouped factorial effects $26(= AF + BF)$ and $46(= CF + DF)$, none of the individual interactions were important at the second stage. Thus there is some consistency of findings between the first and second stages of the experiment.

It should be recognised that, in assessing the different groupings and in the simulation of Section 4.3, no account was taken of the blocking required in the first stage experiment. A study of the role of blocking in group screening would be an interesting topic for future research.

Table 4.11: Analysis of variance for the first stage experiment for the grouped factors 1 to 6 listed in Table 4.8. Each factorial effect has 1 degree of freedom; the residual is on 9 degrees of freedom.

Stratum	Source of variation	Sum of squares	Mean Square
Between sessions	1	440.68	440.68
	56	705.38	705.38
	156	372.84	372.84
Within sessions	2	54.48	54.48
	3	26.59	26.59
	4	76.16	76.16
	5	23.33	23.33
	6	4.42	4.42
	12	155.83	155.83
	13	99.78	99.78
	14	88.46	88.46
	15	225.58	225.58
	16	157.11	157.11
	23	219.38	219.38
	24	208.38	208.38
	25	90.18	90.18
	26	304.42	304.42
	34	97.16	97.16
	35	68.12	68.12
	36	132.94	132.94
	45	279.26	279.26
	46	92.77	92.77
	Residual	62.88	6.99
	Total	3986.13	

Table 4.12: Estimated main effects and two-factor interactions and p-values for the response in the first stage experiment run at Jaguar Cars. The p-value shown is for a t-test of the *single* hypothesis that the corresponding factorial effect is zero, against the alternative that it is non-zero.

Stratum	Factorial effect	Estimated effect	p-value
	1	7.42	-
Whole-plot	56	9.39	-
(Session)	156 (Error)	6.83	-
	2	2.61	0.02116
	3	-1.82	0.0832
	4	3.09	0.0092
	5	1.71	0.10072
	6	0.74	0.4469
	12	4.41	0.00116
	13	-3.53	0.00436
	14	-3.33	0.0062
	15	-5.31	0.00030
Sub-plot	16	-4.43	0.0012
	23	-5.24	0.00033
	24	-5.10	0.00040
	25	-3.36	0.0059
	26	-6.17	0.00010
	34	3.48	0.00476
	35	2.92	0.0124
	36	4.08	0.00188
	45	5.91	0.00014
	46	3.41	0.00548

Table 4.13: The plan and scaled response for the second stage experiment.

Run	*AFR*	*Injection timing*	*Higher Cranking speed*	*Idle fuel*	*Injector tip leakage*	*Temp -erature*	Response (scaled)
1	0	0	0	1	0	1	-6.72
2	1	0	0	1	0	0	3.22
3	1	1	0	1	0	1	-1.10
4	1	1	1	1	0	0	2.52
5	0	1	0	1	0	0	4.70
6	0	1	1	1	0	1	-3.81
7	0	0	0	0	0	0	0.99
8	1	0	0	0	0	1	11.76
9	1	1	0	0	0	0	7.03
10	1	1	1	0	0	1	-8.04
11	0	1	0	0	0	1	2.40
12	0	0	1	0	0	1	1.08
13	1	0	1	0	0	0	5.68
14	1	0	1	1	0	1	-1.43
15	0	1	1	0	0	0	5.35
16	0	0	1	1	0	0	3.61
17	0	0	0	1	1	0	-0.73
18	1	0	0	1	1	1	-5.87
19	1	1	0	1	1	0	0.19
20	1	1	1	1	1	1	-5.51
21	0	1	0	1	1	1	-3.80
22	0	1	1	1	1	0	2.20
23	0	0	0	0	1	1	-3.70
24	1	0	0	0	1	0	1.01
25	1	1	0	0	1	1	-3.81
26	1	1	1	0	1	0	0.25
27	0	1	0	0	1	0	0.04
28	0	0	1	0	1	0	1.86
29	1	0	1	0	1	1	-6.19
30	1	0	1	1	1	0	4.03
31	0	1	1	0	1	1	-3.87
32	0	0	1	1	1	1	-1.71

Table 4.14: Analysis of variance for the second stage experiment. Labels A to F represent *AFR*, *Injection timing*, *Higher idle speed*, *Cranking fuel*, *Injector tip leakage* and *Temperature* respectively. Each factorial effect has 1 degree of freedom and the residual has 10 degrees of freedom.

Source of variation	Sum of Squares	Mean Square
A	0.64	0.64
B	3.64	3.64
C	3.96	3.96
D	17.18	17.18
E	92.99	92.99
F	204.63	204.63
AB	23.70	23.70
AC	36.13	36.13
AD	0.26	0.26
AE	11.96	11.96
AF	0.58	0.58
BC	15.37	15.37
BD	7.80	7.80
BE	1.74	1.74
BF	11.29	11.29
CD	24.99	24.99
CE	17.07	17.07
CF	21.68	21.68
DE	23.01	23.01
DF	7.45	7.45
EF	0.28	0.28
Residual	91.10	9.11
Total	617.45	

Table 4.15: Estimated factorial effects and p-values for the response in the second stage experiment, where A to F are labels for *AFR*, *Injection timing*, *Higher idle speed*, *Cranking fuel*, *Injector tip leakage* and *Temperature* respectively. The p-value shown is for a two-sided t-test of the single hypothesis that the corresponding factorial effect is zero. (1)-(6) indicates the six most significant effects ($p < 0.15$).

Factorial Effect	Estimated Effect	p-value
A	0.28	0.7971
B	-0.67	0.5417
C	-0.70	0.5244
D	-1.47	0.1996
E	-3.41	0.0096 (2)
F	-5.06	0.0008 (1)
AB	-1.72	0.1379 (5)
AC	-2.13	0.0744 (3)
AD	-0.18	0.8682
AE	-1.22	0.2786
AF	-0.27	0.8055
BC	-1.39	0.2231
BD	0.99	0.3764
BE	0.47	0.6715
BF	-1.19	0.2916
CD	1.77	0.1287 (4)
CE	1.46	0.2010
CF	-1.65	0.1539
DE	1.70	0.1431 (6)
DF	-0.97	0.3870
EF	-0.19	0.8642

Chapter 5

Supersaturated experiments and a comparison with two-stage group screening

In this chapter supersaturated designs are considered as an alternative to two-stage group screening when interactions as well as main effects are of interest. In the two-stage group screening experiments considered here, a regular fractional factorial design is used at each stage. In order to make a meaningful comparison of these two approaches, the same method of analysis must be applied to each type of experiment. For supersaturated designs, complex aliasing amongst the factorial effects may result in difficulties when a frequentist analysis, for example, a stepwise regression procedure, is applied, as discussed in Section 1.3.1. Chipman, Hamada and Wu (1997) gave a Bayesian approach for the selection of a subset of factors to include in a model and applied it to analysing supersaturated designs. Their method is based on the Stochastic Search Variable Selection (SSVS) procedure of George and McCulloch (1993) and incorporates the hierarchical priors for related predictors of Chipman (1996), see Section 2.5. This chapter applies this method of detecting active effects to group screening, as well as to supersaturated experiments, where 'active' means non-zero in agreement with, for example, Chipman et al. (1997).

Section 5.1 gives a brief introduction to Bayesian model selection and an outline of the SSVS method is given in Section 5.2. In Section 5.3, the results of a small simulation study are presented which compares results from a full facto-

rial experiment analysed using SSVS with those from a regular fractional factorial experiment. The purpose of this study is to investigate the conclusions from an SSVS analysis when the design has total aliasing between pairs of factorial effects. Section 5.4 uses simulated experiments to compare the performance of SSVS with that of all-subsets regression (Section 1.3.1) as a method of analysing data from a supersaturated design. Conclusions on SSVS, based on the findings of Sections 5.3 and 5.4, are given in Section 5.5. In Section 5.6, results are presented from a simulation that compares the two-stage group screening approach with the use of a supersaturated design for a particular experiment. In Section 5.7, a comparison is made of the results from Section 5.6 and some discussion is presented in Section 5.8.

5.1 Bayesian model selection

In a Bayesian approach to model selection, prior uncertainty about which factorial terms should be included in a model is formally incorporated by placing *prior distributions* on the corresponding model parameters. The approach assumes a set of models of interest, M, and uses a parameter γ to represent a model. (γ is fully defined in the next section.) From Bayes theorem, the density function of a particular γ given the observed data in vector $\boldsymbol{y}$, that is, the *posterior distribution* of γ, is

$$f(\gamma|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\gamma)f(\gamma)}{\sum_{\gamma \in M} f(\boldsymbol{y}|\gamma)f(\gamma)} \tag{5.1}$$

where $f(\boldsymbol{y}|\gamma)$ is the likelihood function and $f(\gamma)$ is the prior distribution of γ.

In order to choose a model that best describes the data, ideally we would calculate the posterior probabilities of all possible models, that is, explore the whole model space, and choose the model or models with the largest posterior probability. However, when a large number, p, of factorial effects is of interest, the model space of size 2^p can be very large. This means that an exhaustive exploration is infeasible. In order to overcome this problem, a sample from the posterior distribution $f(\gamma|\boldsymbol{y})$ is generated by using $f(\boldsymbol{y}|\gamma)$ and (5.1). This sample is then used to approximate the posterior distribution. A general class of methods for generating a sequence of samples is Markov Chain Monte Carlo (MCMC) methods and Gibbs sampling is

a particular type of MCMC algorithm which is used in this chapter. The sample of values obtained is called a 'Gibbs sample'. Further details are given in Section 5.2.2.

5.2 Stochastic search variable selection

This section outlines the algorithm of George and McCulloch (1993), which is based on Gibbs sampling, for selecting explanatory variables or predictors to include in a model. The work in this chapter assumes that only main effects and two-factor interactions may be non-negligible and these form the set of all possible predictors. The algorithm identifies models, that is, subsets of predictors which have higher posterior probabilities than other subsets. These are the subsets of predictors that appear most frequently in the Gibbs sample.

The usual linear model assumptions are made in which a set of p predictors labelled $x_1, ..., x_p$ is related to the response $\mathbf{Y}$ as:

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{5.2}$$

where $\mathbf{Y}$ is an $n \times 1$ vector (where n is the number of observations), X is the $n \times p$ model matrix, $\boldsymbol{\beta}$ is a vector of length p containing the p unknown model coefficients, and $\boldsymbol{\epsilon} \sim MN(0, \sigma_*^2 I_{n\times n})$. Note that in equation (5.2) the parameters have been shifted so that the model does not include an intercept term. Also, σ_*^2 will be used to denote the error variance to avoid confusion in later notation.

A vector $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_p)$ of length p is used to indicate the importance of each of the p predictors: $\gamma_i = 0$ means that β_i is close to zero and hence the i-th predictor is not active; $\gamma_i = 1$ indicates that β_i is non-zero and the corresponding predictor is active. Prior distributions are assigned to $\boldsymbol{\beta}$, σ_*^2 and $\boldsymbol{\gamma}$, as described in the following subsection, and then the posterior probabilities (given the data) of the most likely values for $\boldsymbol{\gamma}$ are approximated. These represent the most likely models to describe the data.

5.2.1 Choice of priors

The joint prior distribution of the parameters $\boldsymbol{\beta}$, given a particular value of $\boldsymbol{\gamma}$, is assumed to be

$$\boldsymbol{\beta}|\boldsymbol{\gamma} \sim N_p(0, D_\gamma R D_\gamma) \tag{5.3}$$

where D_γ is a diagonal matrix consisting of elements $a_1\tau_1, ..., a_p\tau_p$, where

$$a_i = \begin{cases} 1 & \text{if } \gamma_i = 0 \\ c_i & \text{if } \gamma_i = 1. \end{cases}$$

Here τ_i and c_i are called tuning constants which are chosen by the user; appropriate choices for these tuning constants are discussed below. The $p \times p$ matrix R in equation (5.3) is specified by the user and determines the prior variance-covariance structure of the β_i. An important special case which has been used successfully in many applications was identified by George and McCulloch (1993) and will be used in this chapter, namely taking $R = I$. Then the prior distribution of β_i, for given γ_i, is a mixture of two normal distributions, one for inactive predictors ($\gamma_i = 0$) and one for active predictors ($\gamma_i = 1$). This distribution can be written

$$\beta_i|\gamma_i \sim (1 - \gamma_i)N(0, \tau_i^2) + \gamma_i N(0, c_i^2\tau_i^2), \tag{5.4}$$

for $i = 1, ..., p$.

In choosing c_i and τ_i, for simplicity, it is usually assumed that $c_i = c$ and $\tau_i = \tau$ for all i. The value of c_i is chosen to reflect the prior belief about the relative size of the parameters of an unimportant and an important predictor, as it is the relative size of the prior standard deviation in each case. Similarly, the value of τ_i represents the size of the parameter of an inactive effect as the larger the value of τ_i, the wider is the spread of $\beta_i|(\gamma_i = 0)$ and so larger values of β_i are more likely. Box and Meyer (1986) and Chipman et al. (1997) used $c = 10$, so that the prior standard deviation of an important effect is ten times larger than that of an unimportant one. For the value of τ_i, Chipman (2003) recommended

$$\tau_i = \frac{1}{3\Delta(x_i)},$$

where $\Delta(x_i)$, for $i = 1, ..., p$, is the difference between the value of the highest and lowest values of predictor x_i. For all the examples in this chapter, $\Delta(x_i)$ has the same value for all values of i, and hence the assumption $\tau_i = \tau$ is appropriate.

The prior for the variance of ϵ in model (5.2) is assumed to be the inverse gamma distribution

$$\sigma_*^2 \sim IG(\nu/2, \nu\lambda/2)$$

where v and λ are again appropriately chosen tuning constants. This is equivalent to $\nu\lambda/\sigma_*^2 \sim \chi_\nu^2$, a chi-squared distribution with ν degrees of freedom. As $Var(\sigma_*^2)$ is not defined for $\nu \leq 4$, Chipman (2003) recommended a value of $\nu = 5$. Chipman et al. (1997) recommended that λ be tuned so that a selected model does not have too many or too few terms and proposed, as a sensible choice, the use of

$$\lambda = \frac{var(\boldsymbol{y})}{25} \tag{5.5}$$

where $var(\boldsymbol{y})$ is the sample variance of the data.

Under the model assumptions described above, the Bayesian analysis now aims to obtain the marginal posterior distribution $f(\boldsymbol{\gamma}|\boldsymbol{y})$. From equation (5.1) this distribution is proportional to $f(\boldsymbol{y}|\boldsymbol{\gamma})f(\boldsymbol{\gamma})$. Use of the data $\boldsymbol{y}$ updates the prior probabilities to give posterior probabilities for each of the 2^p possible values of $\boldsymbol{\gamma}$, that is, the 2^p possible models. The posterior probabilities are then used to provide a ranking of promising models ($\boldsymbol{\gamma}$s) for further investigation.

The choice of prior distribution, $f(\boldsymbol{\gamma})$, for $\boldsymbol{\gamma}$ should incorporate any available prior information about which subsets of predictors are likely to be active. George and McCulloch (1993) used an *independence* prior for $\boldsymbol{\gamma}$, that is,

$$\pi(\boldsymbol{\gamma}) = \prod_{i=1}^{p} p_i^{\gamma_i}(1-p_i)^{1-\gamma_i}, \tag{5.6}$$

where p_i is the probability that $\gamma_i = 1$. However, the use of this prior assumes that the probability of a variable being active does not depend on whether any of the other variables are active.

For the situation where two-factor interactions are considered, there is a more appropriate choice of prior. Chipman (1996) introduced priors for $\boldsymbol{\gamma}$ that incorporate relations between predictors and, in particular, between a two-factor interaction and the main effects of each factor involved in the interaction. This idea was described in Section 2.5 where it was used to calculate individual interaction probabilities in group screening.

5.2.2 Gibbs sampling

The Gibbs sampler is a Markov Chain Monte Carlo technique for generating random samples from a posterior distribution indirectly by using conditional distributions. This approach avoids having to calculate the posterior density. Repeated draws are made from the conditional distribution of a parameter, conditional on the data and the most recently sampled values of the other parameters.

For model (5.2), the Gibbs sampler generates a 'Gibbs sequence'

$$\boldsymbol{\beta}^0, \sigma^0, \boldsymbol{\gamma}^0, ..., \boldsymbol{\beta}^j, \sigma^j, \boldsymbol{\gamma}^j, ...\boldsymbol{\beta}^m, \sigma^m, \boldsymbol{\gamma}^m, \tag{5.7}$$

where m is the number of iterations used in the Gibbs sample, $\sigma^0, \sigma^1, ..., \sigma^m$ are a sample from the successive conditional distributions of σ_*^2 and, similarly, $\boldsymbol{\beta}^0$, $\boldsymbol{\beta}^1$,...,$\boldsymbol{\beta}^m$ are a sample from the successive conditional distributions of $\boldsymbol{\beta}$. Further details are given below. In many cases, the sequence of models

$$\boldsymbol{\gamma}^1, ..., \boldsymbol{\gamma}^m$$

converges rapidly to give, approximately, a sample from the posterior distribution $f(\boldsymbol{\gamma}|\boldsymbol{y})$. As previously discussed, for large numbers of predictors, the number of all possible $\boldsymbol{\gamma}$s (the model space) is very large and then computing the above sequence is much faster than calculating $f(\boldsymbol{\gamma}|\boldsymbol{y})$ for all 2^p possible $\boldsymbol{\gamma}$s.

For models which are not singular (that is, for which $(X'X)^{-1}$ exists), the starting values $\boldsymbol{\beta}^0$ and σ^0 for $\boldsymbol{\beta}$ and σ_*^2 respectively are chosen to be the estimates obtained from model (5.2) for $\boldsymbol{\gamma}^0$, using the method of least squares. Usually, $\boldsymbol{\gamma}^0$ is set to have all the terms active in the full model, i.e. $\boldsymbol{\gamma}^0 = (1, 1, ..., 1)$. Subsequent values of $\boldsymbol{\beta}^j$ $(j = 1, 2, ...)$ are obtained by sampling from the multivariate Normal distribution

$$N_p(\boldsymbol{A}_{\boldsymbol{\gamma}^{j-1}}(\sigma^{j-1})^{-2}X'X\hat{\boldsymbol{\beta}}_{LS}, \boldsymbol{A}_{\boldsymbol{\gamma}^{j-1}}), \tag{5.8}$$

where

$$\boldsymbol{A}_{\boldsymbol{\gamma}^{j-1}} = ((\sigma^{j-1})^{-2}X'X + D_{\boldsymbol{\gamma}^{j-1}}^{-1}R^{-1}D_{\boldsymbol{\gamma}^{j-1}}^{-1})^{-1},$$

as given by George and McCulloch (1993). The distribution (5.8) is the conditional distribution of $\boldsymbol{\beta}^j$ given $\boldsymbol{\gamma}^{j-1}$ and σ^{j-1}. The variance σ^j is obtained by sampling from the updated inverse gamma distribution

$$\sigma^j|\boldsymbol{y}, \boldsymbol{\beta}^j \sim IG(\frac{n+\nu}{2}, \frac{|\boldsymbol{y} - X\boldsymbol{\beta}^j|^2 + \nu\lambda}{2}). \tag{5.9}$$

Each element in the vector $\boldsymbol{\gamma}^j$ is obtained by sampling consecutively from the conditional distribution

$$\gamma_i^j \sim f(\gamma_i^j | \boldsymbol{\beta}^j, \sigma^j, \boldsymbol{\gamma}_{(i)}^j), \tag{5.10}$$

where $\boldsymbol{\gamma}_{(i)}^j = (\gamma_1^j, ..., \gamma_{i-1}^j, \gamma_{i+1}^{j-1}, ..., \gamma_p^{j-1})$. Each distribution (5.10) can then be shown to be Bernoulli with probability

$$P(\gamma_i^j = 1 | \boldsymbol{\beta}^j, \sigma^j, \boldsymbol{\gamma}_{(i)}^j) = \frac{a}{a+b} \tag{5.11}$$

where

$$\begin{aligned} a &= f(\boldsymbol{\beta}^j | \boldsymbol{\gamma}_{(i)}^j, \gamma_i^j = 1) p_i, \\ b &= f(\boldsymbol{\beta}^j | \boldsymbol{\gamma}_{(i)}^j, \gamma_i^j = 0)(1 - p_i) \end{aligned}$$

and $p_i = P(\gamma_i = 1)$. Successive sampling from (5.8), (5.9) and (5.11) gives the Gibbs sequence (5.7). Enough draws must be made so that the Gibbs sequence converges. It is also common practice to discard the first few observations in a Gibbs sequence because the starting point may not be a representative point from the posterior distribution.

If a value of $\boldsymbol{\gamma}$ appears frequently in the Gibbs sample obtained, then this indicates that the corresponding model provides a good description of the data.

5.2.3 Use of SSVS with supersaturated experiments

The SSVS approach can be used with supersaturated designs. However, minor modifications are needed for the starting values $\boldsymbol{\beta}^0$ and σ^0. This is because it is not possible to find the least squares estimates of (5.2) as the information matrix $X'X$ is singular.

In the literature, SSVS is used to select models in several examples including data from supersaturated designs. However, there appear to be no studies available on how reliably the method detects active effects. For this reason, in the next two sections, simulation studies are used to assess SSVS.

For these studies, I have used code written by Professor Hugh Chipman to perform SSVS. (This code is available at www.stats.uwaterloo.ca/~hachipma/code/index.html and is written in C++.) I have incorporated this code into simulation programs which are written in C++.

5.3 Small study to investigate the performance of SSVS

In this study, two small experiments were run, each for three two-level factors A, B, C. The first experiment was a full 2^3 factorial; the second was an experiment using the 2^{3-1} fraction with defining contrast ABC.

The following 'true' model was defined for an observation Y in which a term for the mean is omitted, as recommended by George and McCulloch (1993):

$$Y = 10x_A + 0x_B + 0x_C + 10x_Ax_B + 0x_Ax_C + 0x_Bx_C + \epsilon$$

where $\epsilon \sim N(0, 0.00025)$, the random errors for the observations are assumed independent and x_j $(j = A, B, C)$ denotes the value of factor j. In this model there are 6 predictors and their order in γ is the same as that in the above model. A small value for the variance of the error distribution was chosen in order to gain a clearer understanding of how the aliasing structure in the design affected the ability of SSVS to detect the two active effects. (The investigation was also repeated with $\epsilon \sim N(0, 4)$ and these findings are discussed in Section 5.3.3.)

The design and data for the full factorial experiment are shown in Table 5.1. The SSVS algorithm was run 50,000 times with a burn-in of 100 runs, that is, the first 100 runs were discarded (see Section 5.2.2). The tuning constants $\nu = 5$, $\lambda = 9.14$, $c = 10$ and $\tau = 0.1667$ were used, in accordance with the recommendations described in Section 5.2.1.

Three sets of probabilities $(p_i,\ i = 1, ..., 6)$ were considered for the prior probability that the i-th predictor is active (see equation (5.6)), namely,

(1) $p_i = 0.5,\ i = 1, ..., 6$

(2) $p_i = 0.99,\ i = 1, ..., 6$

(3) $p_1 = p_2 = p_3 = 0.5$, with p_4, p_5, p_6 the respective probabilities of predictors x_Ax_B, x_Ax_C and x_Bx_C being active, calculated using $w_{00} = 0.01$, $w_{01} = w_{10} = 0.1$ and $w_{11} = 0.25$ from formula (2.31) of Section 2.5.

Table 5.1: The design and data for the full factorial experiment, where the responses are given to 4 decimal places.

A	B	C	AB	AC	BC	Response
-1	-1	-1	1	1	1	-0.0078
-1	-1	1	1	-1	-1	-0.0088
-1	1	-1	-1	1	-1	-19.9932
-1	1	1	-1	-1	1	-19.9900
1	-1	-1	-1	-1	1	-0.0031
1	-1	1	-1	1	-1	-0.0054
1	1	-1	1	-1	-1	19.9925
1	1	1	1	1	1	20.0083

5.3.1 Results for the full factorial

Table 5.2 shows the four models that appear most frequently in the Gibbs sample, ranked by posterior probability. The model with the highest posterior probability was found to be the true model $\gamma = 100100$ for sets (1) and (3) of prior probabilities. The term x_A and the interaction $x_A x_B$ occurred in every model visited by the Gibbs sampler, that is, their marginal probabilities were both 1. Here the marginal probability is defined as the number of models visited that included this predictor, divided by the total number of models visited.

When all the main effect and interaction probabilities (p_i) are set to 0.99 (set (2) above), the top model identified is the *full* model with a very high posterior probability of 0.84. This case was included in the study to gauge the influence of the design and the data on the results and shows that the use of high prior probabilities can mask the true situation. Even with very large parameter values and small random error, use of large prior probabilities swamps the information from the experiment. However, the main effect of A and the interaction AB are again included in each of the top four models.

Table 5.2: The four models (γ) with highest posterior probabilities identified by SSVS under three sets of prior probabilities for main effects (m.e.) and interactions (int) being active, and for a full 2^3 experiment. The entries in each γ are ordered according to the terms corresponding to A, B, C, AB, AC, BC.

All m.e. and int. probabilities 0.5		All m.e. and int. probabilities 0.99		All m.e. probs 0.5; $w_{00} = 0.01$ $w_{01} = w_{10} = 0.1$, $w_{11} = 0.25$	
Model	Probability	Model	Probability	Model	Probability
100100	0.44	111111	0.84	100100	0.53
100110	0.10	111101	0.04	110100	0.28
100101	0.10	110111	0.04	101100	0.09
110100	0.10	101111	0.04	111100	0.05
Marginal Probabilities					
A	1.00	A	1.00	A	1.00
B	0.19	B	0.96	B	0.35
C	0.19	C	0.96	C	0.15
AB	1.00	AB	1.00	AB	1.00
AC	0.19	AC	0.96	AC	0.03
BC	0.19	BC	0.96	BC	0.02

5.3.2 Results for the half replicate

The defining relation for the half replicate was $I = ABC$. Hence, the main effect of A was totally aliased with the interaction BC, the main effect of B was totally aliased with the interaction AC and the main effect of C was totally aliased with the interaction AB.

Table 5.3 shows the top ranked models appearing in the Gibbs sample for the sets of prior probabilities (1) - (3), given at the start of Section 5.3, for data generated from the fractional factorial.

For set (1), the top ranked model includes the truly active effects for A and AB, but also for their aliases BC and C respectively. Only these four effects appear in the top four models and their marginal probabilities are approximately equal. This indicates that SSVS has not been able to distinguish between their aliased effects. Note also that the posterior probability for the top model is only 0.09 which is too low for a model to be chosen. The results for the case when all the main effects and interaction probabilities are 0.99 (set (2)), are very similar to those for the full factorial experiment.

For set (3), the marginal probability of x_A is 0.88 but the marginal probability of $x_A x_B$ is only 0.25. Variable x_A appears in the first four models but $x_A x_B$ appears only in the fourth model. In the design, main effect C is totally aliased with the interaction AB which explains the high marginal probability of C. Here the probabilities of main effects being active are higher than those for interactions; SSVS is using the prior probabilities to augment the information from the design to produce the top ranked model.

5.3.3 Results of the investigation with larger error variance used in the data generation

The previous investigation was repeated with $\epsilon \sim N(0, 4)$ used in the data generation. The results are given in Appendix E. The data are shown in Table E.1 and the results for the full 2^3 experiment are summarised in Table E.2.

For prior probability set (1), the ranking of the top four models was exactly the same as in Section 5.3.1 but with a slight decrease in the proportion of times

Table 5.3: The four models (γ) with highest posterior probabilities identified by SSVS under three sets of prior probabilities for main effects and interactions being active, and for a 2^{3-1} experiment. The entries in each model are ordered according to the terms corresponding to A, B, C, AB, AC, BC.

All m.e. and int. probabilities 0.5		All m.e. and int. probabilities 0.99		All m.e. probs 0.5; $p_{00} = 0.01$ $p_{01} = p_{10} = 0.1$, $p_{11} = 0.25$	
Model	Probability	Model	Probability	Model	Probability
101101	0.09	111111	0.95	101000	0.37
101001	0.07	101111	0.01	111000	0.08
100101	0.07	111101	0.01	101001	0.06
101100	0.06	111011	0.01	100100	0.06
Marginal Probabilities					
A	0.71	A	0.99	A	0.88
B	0.30	B	0.99	B	0.33
C	0.70	C	0.99	C	0.88
AB	0.70	AB	0.99	AB	0.25
AC	0.29	AC	0.99	AC	0.11
BC	0.71	BC	0.99	BC	0.27

that the top model occurred in the sample, a slight increase in the proportions of times the next model occurred and a very slight reduction in the proportions of times that the third and fourth models occurred. There was also an increase in the marginal probabilities for x_C and $x_B x_C$. These results indicate a reduction in the, already poor, clarity of findings from SSVS for this investigation.

For prior probability set (2), the top ranked model was again the *full* model which occurred a larger proportion of times than any other model. The fourth ranked model included every term except $x_B x_C$ whereas the fourth ranked model in Section 5.3.1 included every term except x_B. However, the proportions of times the second, third and fourth ranked models occurred in the Gibbs sample were very similar for both investigations.

For prior probability set (3), the ranking of the top four models was exactly the same as in Section 5.3.1 but with a slight decrease in the proportions of times that the top two models occurred and a slight increase in the proportions of times the third and fourth models occurred.

The results for the 2^{3-1} experiment are shown in Table E.3 from which it can be seen that the findings are almost identical to those in Section 5.3.2.

5.4 Comparison of SSVS with all-subsets regression for analysing a supersaturated design

Abraham, Chipman and Vijayan (1999) drew attention to some of the risks in the construction and frequentist analysis of supersaturated designs. They gave the results of a simulation that compared forward stepwise regression with all-subsets regression (described in Chapter 1) using the 14-run supersaturated design of Lin (1993). They concluded that all-subsets regression was a better choice of analysis method than forward stepwise regression, but that this method can be problematic and that caution should be used when using supersaturated designs.

In order to carry out a similar assessment of the SSVS method of analysis, a similar simulation investigation was carried out for the same design of Lin (1993). Only models containing main effects were considered and hence a factor is called

active if its main effect is active. Following Abraham et al. (1999), six true models that involve only subsets of three parameters from β_1, β_2, β_3, β_7, β_{13} were considered. The errors for the observations were generated by independent draws from a $N(0,1)$ distribution. The data were analysed using SSVS and the whole process was repeated 200 times (as was done by Abraham et al. (1999)) and also 1000 times to check the simulation accuracy. The results for 1000 and 200 simulations were almost identical. The probability of a main effect being active was set to 0.5 for all factors. For SSVS, the value of c was set to three different values: 10, 50 and 100. The values of τ, λ and ν were set to 0.167, $\frac{var(y)}{25}$, and 5 respectively (see Section 5.2.1).

Table 5.4 shows results from the SSVS analyses, together with the corresponding results for all-subsets regression from Abraham et al. (1999). SSVS performs markedly better when $c = 10$ and the coefficients of the three active factors are $\beta_1 = 1$, $\beta_2 = 1$ and $\beta_3 = 1$, as compared with all-subsets regression. However, there is not so marked an improvement in any of the other results and, in fact, all-subsets regression performs better for some choices of β_i.

However, an advantage of SSVS is that different probabilities of being active can be assigned to the various factorial terms. A simulation study has also been run with the probabilities of main effects of factors known to be active set to 0.5 and those known to be inactive set to 0.1. The results, shown in Table 5.5, indicate that altering the prior probabilities for SSVS to reflect knowledge of active factors increases (a) the percentage of times that the top model contains all the active factors, and (b) the percentage of times that one of the top five models contains all the active factors.

When $c = 10$, SSVS performs notably better than all-subsets regression in this study but when $c = 50$, SSVS performs only slightly better. This demonstrates the increased flexibility of SSVS, but also highlights the sensitivity of SSVS to the choice of values for the tuning constants.

Table 5.4: Simulation results for all-subsets regression (Abraham et al. (1999)) and SSVS. The model coefficients are labelled according to the corresponding variable. For SSVS, the results are shown for three values of the tuning constant c. The numbers in parentheses are the percentages of times one of the top five models contained all the active factors.

Actual coefficients					All-subsets regression			SSVS		
					Percentage of times correct factors selected in subsets of size k			Percentage of times top ranked model contains all active factors		
β_1	β_2	β_3	β_7	β_{13}	$k=3$	$k=4$	$k=5$	c=10	c=50	c=100
1	1	1	0	0	52 (77)	50 (77)	51 (70)	79.2 (95.4)	79.9 (95.8)	38.9 (66.4)
0	0.5	0	0.5	0.5	1 (8)	4 (9)	5 (15)	3.8 (13.9)	6.2 (14.2)	3.9 (10.5)
0	1	0	1	1	33 (57)	44 (65)	43 (67)	25 (54.2)	27 (45.6)	22.9 (41.2)
0	5	0	10	20	100 (100)	100 (100)	100 (100)	87.9 (100)	100 (100)	100 (100)
0	14	0	20	20	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
0	20	0	20	20	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)

Table 5.5: Results of the simulation using prior probability 0.5 for active effects and 0.1 for inactive effects, for three settings of 10, 50 and 100 of the tuning constant c. The numbers in parentheses are the percentages of times one of the top 5 models contained all the active factors.

Actual coefficients					Percentages of times the top model contains all active factors		
β_1	β_2	β_3	β_7	β_{13}	$c = 10$	$c = 50$	$c = 100$
1	1	1	0	0	86.7 (99.6)	57.2 (94.3)	42.9 (89.8)
0	0.5	0	0.5	0.5	35.4 (70.2)	18.5 (44.5)	10.9 (32.8)
0	1	0	1	1	78.1 (92.1)	57.1 (82.2)	44.3 (73.1)
0	5	0	10	20	100 (100)	100 (100)	100 (100)
0	14	0	20	20	100 (100)	100 (100)	100 (100)
0	20	0	20	20	100 (100)	100 (100)	100 (100)

5.5 Overall conclusions on Stochastic Search Variable Selection

From the examples studied in this chapter, the overall conclusions on stochastic search variable selection are as follows:

- When there is no aliasing of factorial effects, and the probabilities of effects being active are no larger than 0.5, SSVS selects the correct model about 50% of the time and the second, third and fourth ranked models each contained the true model. The marginal probability of each of the truly active effects is 1. When the probabilities of effects being active are set very high, SSVS finds everything to be important. This shows that for a design with no aliased effects, and a sensible choice of prior probabilities, the conclusions from SSVS are reasonable.

- When there is total aliasing of two factorial effects and the probabilities of effects being active are equal and no larger than 0.5, SSVS will return marginal probabilities for each effect which are roughly equal. This finding indicates

that care should be taken to consider the aliasing structure in a design when interpreting the results from SSVS.

- For experiments where the model space is not large and when there is no prior knowledge available on the importance of factorial effects, all-subsets regression performs no worse than SSVS.

These conclusions are, of course, based only on a small number of examples and further studies would be useful.

5.6 A comparison of a two-stage group screening experiment and a supersaturated experiment

This section describes a further simulation study to compare the ability of a particular two-stage group screening experiment and a particular supersaturated design to detect active factors. The experiment setting used for this study had 8 control factors, labelled

$$A, B, C, D, E, F, G, H,$$

and 6 noise factors, labelled

$$I, J, K, L, M, N.$$

The active factorial effects were chosen to be the main effect of control factor B, the main effect of noise factor I and their interaction BI. All other factorial effects were assumed to be zero. The true model from which the observations for the experiment were generated was

$$Y = 3x_B + 3x_I + 3x_B x_I + \epsilon \tag{5.12}$$

where $\epsilon \sim N(0, 1)$, and the errors for different observations were obtained independently by random draws. The details of the designs used are given in the following sections. Each simulated experiment was then repeated twice, with the values of the coefficients of x_B, x_I and $x_B x_I$ increased, firstly to 10 and secondly to 30.

Each of the resulting six data sets were analysed using SSVS. All probabilities of main effects being active were 0.14 and the values of the conditional probabilities used to calculate the individual interaction probabilities (as described in

Section 2.5) were 0.005, 0.125, 0.125 and 0.25 for both the control×noise and control×control individual interactions. The tuning constants in the SSVS algorithms were: $c = 10$, $\nu = 5$, $\tau = 0.16667$ and $R = I$. The value of λ was calculated for each simulation according to the generated value of the data $\boldsymbol{y}$, using equation (5.5).

5.6.1 Simulation of a supersaturated experiment

For the study, a supersaturated design was required for 14 factors that allowed interactions as well as main effects to be considered. As described in Section 1.3.1, Liu and Dean (2004) gave a method for adding interaction columns to k-circulant main effect supersaturated designs. They also gave a supersaturated design capable of estimating 14 main effects in 8 runs, and also gave a set of 7, 14 or 21 *particular* interaction columns that could be appended to the design matrix which could be used to estimate interactions or further main effects. Here the columns are used to estimate the particular interactions. Each of the three supersaturated designs constructed in this way was used in the study.

A single simulation in the study comprised data generation using model (5.12) and analysis using SSVS with a burn-in of 100 runs and 10,000 iterations. The simulation was performed 10,000 times and the effects appearing in the top two models from the analysis were recorded. These results are shown in the upper part of Table 5.6. The average number of inactive effects which occurred in the top model, averaged over the 10,000 simulated experiments, is in the lower part of Table 5.6.

From the table, it can be seen that the percentage of times that the top model is the true model (indicated by (i) in Table 5.6) becomes smaller as more interaction columns are appended to the design matrix. This is also true for the percentage of times that the top model contains the true model (indicated by (ii)). The percentages of times that one of the top two models contains the true model (indicated by (iii)) are very similar for the designs with 7 and 14 interaction columns, but are smaller for the design with 21 interaction columns. The general trend is that the percentages of correct identification are reduced when more interaction columns are

Table 5.6: The upper part shows the percentage of times that (i) the top model identified was the true model, (ii) the top model contained the active effects and (iii) one of the top two models contained the active effects. Results are shown for simulations with β_i values for active effects of 3, 10 and 30, for three supersaturated designs, each capable of estimating 14 main effects and 7, 14 and 21 interactions respectively. The lower part of this table gives the average number of inactive effects which were included in the top model, averaged over the 10,000 simulations of the experiment.

Coeff. of active effects in simulation	Identification percentages (i), (ii) and (iii) for supersaturated design with 7, 14 and 21 interaction columns								
	7			14			21		
	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
3	66.15	94.89	99.52	26.08	49.55	97.60	20.66	33.26	66.04
10	71.09	100	100	23.63	47.83	100	20.78	34.00	66.23
30	71.14	100	100	23.41	48.55	100	20.92	34.22	67.12

	Average no. of inactive effects in top model		
3	0.3768	1.5128	1.9884
10	0.2891	1.5635	2.1778
30	0.2886	1.5252	2.1145

appended to the supersaturated design.

The table also shows that increasing the values of the active effect coefficients from 3 to 10 nearly always improved the detection of the active effects. However, increasing these coefficients from 10 to 30 made very little further improvement.

The average number of spurious, or non-active, effects in the top models increased with the number of interaction columns in the design. For the design with 7 interaction columns, increasing the values of the active effects from 3 to 10 decreased slightly the average number of effects wrongly declared active. For the designs with 14 and 21 interaction columns, this increase in the active effects coefficients resulted in a slight increase in the average numbers of effects wrongly declared active. A further increase in the coefficient values from 10 to 30 gave very similar results.

These results indicate that, for this particular example, as the number of interaction columns in the design is increased, that is, the aliasing amongst the effects becomes more complicated, the chance of identifying the true model is reduced, even when the values of active effects are large.

The above finding is supported by another investigation in which all $8 \times 6 = 48$ control×noise interaction columns and all $\binom{8}{2} = 28$ control×control interaction columns were appended to the design matrix. For active effect coefficients of size 3, the percentage of times that the top model was the true model was 0. The same was true for the percentage of times that the top model contained the true model. The percentage of times that *one* of the top two models contained the true model was 8.83% and the average number of inactive effects which were declared active was 0.0345. This number is low because the model with no active effects was found to be the highest ranked model in 66.57% of the simulations.

5.6.2 Simulation for the group screening experiment

The simulation of a two-stage group screening experiment is more complicated than simulating a supersaturated experiment. For this reason, an outline of the simulation procedure is given below. The grouping chosen for the investigations is described and the approach used in assigning grouped conditional probabilities is

explained, before the results are presented from applying the simulation method to data generated from model (5.12).

Outline of the simulation

As mentioned in Section 5.2.3, the SSVS code of Professor H. Chipman has been incorporated into both stages of the two-stage group screening simulation.

The user specifies the numbers of individual control and noise factors, the probabilities of their main effects being active and the conditional probabilities to be used in calculating the individual interaction probabilities and the standard deviation, σ_*, to be used for the distribution $N(0, \sigma_*^2)$ for the error terms used in the data generation at both stages. A choice of active effects must also be made, together with their coefficient values. It is assumed that the individual conditional probabilities are the same for both control×noise and control×control interactions. The user can also choose how many of the top ranked models are to be retained at each stage and, in this example, this value was chosen to be two.

The user must also specify the grouping to be used in the first stage experiment, and the model matrix for this grouping in terms of the individual effects. This matrix must have individual effects that are in the same group totally aliased with each other. (The matrix is used in the data generation for the first stage experiment.) The grouped model matrix must also be specified in terms of the grouped factors.

The grouped main effects probabilities are calculated using equations (2.1) and (2.2). The heredity principle is used to assign interaction probabilities at the first stage. Hence appropriate choices for conditional probabilities must be supplied by the user. The approach used in choosing the conditional probabilities for this example is detailed below.

A design that can be used to estimate all grouped main effects and two-factor interactions is used at the first stage. Details of the design and the factor grouping are given in the following subsection. The model for the grouped factors is fitted at the first stage using SSVS. Grouped effects appearing in the top two models are considered for the second stage. More precisely, if a grouped control main effect is declared active, all individual control factors within that group are brought forward to the second stage. If a grouped interaction term is declared active, all the

individual factors involved are brought forward to the second stage. If a grouped noise factor is in the top two models, it will only be declared active if it is involved in an active grouped control×noise interaction.

A regular fraction factorial design, capable of estimating all main effects and two-factor interactions, is selected automatically for the second stage from the table of Russell et al. (2004) according to how many individual control factors and individual noise factors have been brought forward to the second stage. All control×noise and control×control interaction columns are generated by the program from the basic design. This gives a design which is not necessarily a minimal plan. Hence we would expect a higher number of runs for the two-stage experiment using this approach than equation (2.25) would give. Data are generated for the second stage experiment and SSVS used to select a second stage model for individual effects. The models selected by SSVS at the second stage are ranked and the top two are kept. The number of observations required for the second stage experiment is recorded and added to the number of observations used in the first stage experiment.

A single simulation comprises one generation of data for the first stage experiment and their analysis using SSVS (10,000 iterations), together with one second stage data generation and analysis using SSVS (10,000 iterations). The simulation is repeated 10,000 times and the average number of observations taken in the simulated two-stage experiment is calculated.

Choice of grouping and first stage plan

In order to be able to compare the results from the two-stage group screening experiment with those for the supersaturated experiment, it was necessary to consider the same model space, as far as possible. In order to investigate only 7, 14 or 21 specified individual interactions, it was decided to examine only those grouped interactions which contain these particular individual interactions. This meant that a reduced grouped model space was being explored at the first stage. The experiment in which all 76 interactions were investigated was also considered.

In order to find good groupings to use for the simulations, a grouping investigation was performed where the probabilities for individual interactions which were not being investigated were set to zero. Hence three grouping investigations were

Table 5.7: Grouping of control factors A to H and noise factors I to N, for the group screening simulation.

Group	Individual factors
1	A & B
2	C & D
3	E & F
4	G & H
5	I & J & K
6	L & M & N

performed in the same manner as those described in Chapters 3 and 4. A fourth grouping investigation was also performed for the situation where all interactions were to be investigated. Grouping the control factors in four groups of size two and the noise factors in two groups of size three consistently gave minimal or near minimal values of $E(S_{IGS})$ and $s.d.(S_{IGS})$. Hence this grouping was chosen for the simulated experiments. This chosen grouping is shown in Table 5.7.

A regular fractional factorial 2^{6-1} design for 4 control factors and 2 noise factors (as given in Russell et al. (2004)) with defining contrast 123456 was used at the first stage. This design was capable of estimating all grouped main effects and two-factor interactions. However, only the interaction columns of interest were used in the grouped design matrix.

Calculation of grouped conditional probabilities

In order to use the heredity principle (as described in Section 2.5) for the grouped interactions in the first stage experiment, it is necessary to find appropriate choices for the values of the conditional probabilities $w_{ij}^{G(cn)}$ and $w_{ij}^{G(cc)}$, where $i = 0, 1$, $j = 0, 1$ and the superscript G indicates that these are for grouped interactions.

If it is assumed that the relationship between these probabilities is the same as that used for the individual conditional probabilities, then

$$w_{11}^{G(cn)} = 2 \times w_{01}^{G(cn)} = 2 \times w_{10}^{G(cn)} = 50 \times w_{00}^{G(cn)}, \tag{5.13}$$

and

$$w_{11}^{G(cc)} = 2 \times w_{01}^{G(cc)} = 2 \times w_{10}^{G(cc)} = 50 \times w_{00}^{G(cc)}. \tag{5.14}$$

Using the formulae of Chapter 2, we find the probabilities of grouped interactions being active and the probabilities of grouped main effects being active. From the assumed relationships in (5.13) and (5.14), values for $w_{11}^{G(cn)}$ and $w_{11}^{G(cc)}$ can then be found, as follows.

For this example, all the individual main effects probabilities were set to 0.14. Using the conditional probabilities $w_{00}^{(cn)} = 0.005$, $w_{01}^{(cn)} = w_{10}^{(cn)} = 0.125$ and $w_{11}^{(cn)} = 0.25$, and equation (2.31), the probability that an individual control×noise interaction is active is

$$\begin{aligned} & (0.005 \times (1 - 0.14)(1 - 0.14)) + (0.125 \times (1 - 0.14) \times 0.14) \\ & +(0.125 \times 0.14 \times (1 - 0.14)) + (0.25 \times 0.14 \times 0.14) \\ = \; & 0.038698. \end{aligned}$$

Similarly, the probability that an individual control×control interaction is active is also 0.038698. Using equation (2.3), we can then calculate the probability that a grouped control×noise interaction is active to be

$$1 - (1 - 0.038698)^6 = 0.21085,$$

and the probability that a grouped control×control interaction is active to be

$$1 - (1 - 0.038698)^4 = 0.146036.$$

Using equation (2.1), we calculate the probability that a grouped control main effect is active to be

$$1 - (1 - 0.14)^2 = 0.2604.$$

Similarly, using equation (2.2) to calculate the probability that a grouped noise main effect is active, gives

$$1 - (1 - 0.14)^3 = 0.363944.$$

For control×noise interactions we then solve

$$\begin{aligned} 0.21085 &= w_{00}^{G(cn)}(1-0.2604)(1-0.363944) + w_{01}^{G(cn)}(1-0.2604)0.363944 \\ &\quad + w_{10}^{G(cn)}0.2604(1-0.363944) + w_{11}^{G(cn)}0.2604 \times 0.363944 \\ &= \frac{w_{11}^{G(cn)}}{50}(1-0.2604)(1-0.363944) + \frac{w_{11}^{G(cn)}}{2}(1-0.2604)0.363944 \\ &\quad + \frac{w_{11}^{G(cn)}}{2}0.2604(1-0.363944) + w_{11}^{G(cn)}0.2604 \times 0.363944 \end{aligned}$$

to give $w_{11}^{G(cn)} = 0.6557$, $w_{01}^{G(cn)} = w_{10}^{G(cn)} = 0.32785$ and $w_{00}^{G(cn)} = 0.013114$.

Similarly, for control×control interactions it can be shown that $w_{11}^{G(cc)} = 0.01076$, $w_{01}^{G(cc)} = w_{10}^{G(cc)} = 0.2269$ and $w_{11} = 0.5381$.

Results for the simulations of two-stage group screening experiments

The upper part of Table 5.8 shows the percentages of times that (i) the top model identified was the true model, (ii) the top model contained the active effects and (iii) one of the top two models contained the active effects, for two-stage group screening experiments for estimating 14 main effects with 7, 14, 21 or 76 particular individual interactions to be investigated. The average number of inactive effects which were in the top model selected are shown in the lower part of Table 5.8, where the average is over the 10,000 simulated two-stage experiments. Also given is the average number of observations used in the two-stage experiment, again over all simulations.

The percentages of correct identification of the true model were improved as the values of the active effect coefficients were increased from 3 to 10. Increasing these values to 30 had little further effect. The results are very similar across the four different sets of interactions.

5.7 Findings from the simulation study

The simulation study is limited in its scope but some trends can be observed from the examples. The percentages of correct identification of the true model for the two-stage group screening experiment were consistently higher (or equal to) those for the supersaturated design.

Table 5.8: The upper part of this table shows the percentage of times that (i) the top model identified was the true model, (ii) the top model contained the active effects and (iii) one of the top two models contained the active effects. Results are shown for 12 simulations for two-stage group screening experiments with four different first stage designs, each capable of estimating 14 factor m.e.s and 7, 14, 21 and 76 particular interactions respectively and, for each design, active effects with coefficients 3, 10 and 30. The lower part of the table shows the average number of inactive effects included in the top model selected over 10,000 simulations, and the average number of runs required for the two stages of experimentation.

Coefficient of active effects in simulation	Identification percentages (i), (ii) and (iii) for four group screening designs capable of estimating 7, 14, 21 and 76 particular interactions											
	7			14			21			76		
	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
3	94.18	98.14	98.14	88.82	93.42	93.42	80.15	84.12	84.12	79.11	83.23	83.23
10	95.78	99.03	99.03	96.22	100	100	95.67	100	100	95.83	100	100
30	96.77	100	100	96.29	100	100	96.46	100	100	96.27	100	100

	Average no. of inactive effects in the top model (& average number of runs)			
3	0.0403 (48)	0.0378 (47.63)	0.041 (47.45)	0.0417 (47.65)
10	0.033 (48.12)	0.0378 (48)	0.0433 (48)	0.0417 (48)
30	0.0323 (48)	0.0371 (48)	0.0354 (48)	0.0373 (48)

The average numbers of observations required using two-stage group screening, for this example, were consistently larger than the eight required by the supersaturated design. For active effects coefficients of 10 and 30 when all the 14 grouped interactions are considered at the first stage of group screening (the most demanding case), the percentage of times that the top model identified was the true model is approximately 96%. In contrast, when the supersaturated design is used with only 7 interactions (the least demanding case), this percentage drops to 71%. This indicates that group screening performs better on the most demanding case than the supersaturated design on the least demanding case.

The benefits of two-stage group screening, at the cost of using more observations (approximately 40 more), are clear. The average numbers of effects wrongly identified as active were consistently below 0.05 for the group screening simulations, but were as high as 2 for the supersaturated design with 21 interaction columns.

5.8 Discussion

The studies in this chapter have identified that SSVS has the same drawback as the frequentist stepwise regression analysis, namely a failure to detect active effects when the values of active effects are fairly small and complex aliasing exists amongst the predictors. In the example, when two factorial effects were totally aliased together and non-informative priors are used, SSVS was unable to distinguish between the aliased effects. In these circumstances, further knowledge about the factorial effects is required to decide which of an aliased pair of factorial effects should be included.

From a comparison of a particular supersaturated design and a two-stage group screening experiment using simulations, it was found that the correct model was identified more frequently and that fewer effects were wrongly declared active using the group screening approach. However, the two-stage group screening experiments required an average of approximately 48 runs, whereas the supersaturated experiments only required 8 runs. A trade-off must be made between the number of runs an experimenter is able to use, and the level of inaccuracy that can be tolerated in the conclusions from the experiment, when deciding which screening tool

should be used. This trade-off depends on the costs of missing important factors and of including unimportant factors in follow-up experiments. An interesting area for future research would be to perform a similar comparison, but with a smaller imbalance in the number of runs used by each experiment.

An extension to the use of the SSVS method has been proposed by Beattie, Fong and Lin (2002) who presented a two-stage approach for analysing supersaturated designs where SSVS is used at the first stage. The variables in the model that has the highest posterior probability are identified as active, together with any other variables that were not included in the top model but featured, for example, in the second and third models. All the variables identified as important are brought together into a single model. A further analysis is then performed based on the Intrinsic Bayes Factor procedure of Berger and Pericchi (1996a). Suh, Ye and Mendell (2003) used this method in a small simulation study on a problem in genetic linkage analysis. The authors found that the use of Intrinsic Bayes Factor did not completely identify the correct model in their examples. This approach was not considered here, but is a possible area for future work.

Chapter 6

Conclusions and future work

6.1 Conclusions

The work in this thesis has generalised the work of Lewis and Dean (2001) to enable two-stage group screening to be used for experiments where it may be required to have unequal group sizes and unequal probabilities of factorial effects being active. This advance in the theory makes two-stage group screening a much more flexible screening tool for practical use. Software has also been written which implements the generalised theory, thus allowing the investigation of different groupings of factors for an experiment, through their impact on the distribution of the predicted total number of effects requiring estimation in a two-stage group screening experiment. Two group screening strategies, classical and interaction group screening, have been considered in the development of both the theory and software.

The examples presented in Chapter 3 indicated that no 'best way' to group factors can be stated in general, but they do offer some guidance for experimenters. When the strategy used is interaction group screening, it is often better to group together factors with higher main effects probabilities to minimise the expected total number of effects that require estimation in the two-stage experiment. This approach avoids bringing forward factors unnecessarily to the second stage. The examples in Chapter 3 indicated that groupings which give minimal, or near minimal, values of $E(S_{IGS})$ are those for which group sizes are as small and as equal as possible.When the strategy used is classical group screening, it is usually better to keep factors whose main effects are believed very likely to be active together in a

single group, and to spread the remaining factors across as many groups as possible (including groups of size one).

Examples, including those presented in Chapter 3, have indicated that classical group screening generally can require considerably fewer observations than interaction group screening for a two-stage group screening experiment. However, the risk of failing to detect active control×noise interactions is usually, substantially higher for classical group screening. If interactions are not believed to be active in an experiment, classical group screening is the better choice of screening experiment. However, if interactions are believed to be active, or if there is no prior knowledge regarding interactions, interaction group screening is preferable. The increase in the expected number of effects that require estimation under this strategy must be weighted against the increased risk of failing to identify active interactions under classical group screening.

The experiment on cold start optimisation allowed the application of two-stage group screening methods, including the use of the software to explore groupings. It also demonstrated the importance of elicitation of factors.

In Chapter 5, a two-stage group screening strategy and the use of a particular supersaturated design were compared using simulation and a Bayesian variable selection technique, stochastic search variable selection (SSVS). The two-stage group screening experiment was found to correctly identify the correct model more often than the supersaturated experiment. The number of inactive effects declared active was less under two-stage group screening but at the cost of more observations.

In advising an experimenter wishing to investigate a large number of factors in a small number of runs I would, at present, recommend group screening with orthogonal factorial plans at both stages rather than a supersaturated design, provided a sufficient number of runs was available. This is because reliable conclusions may not be available from supersaturated experiments due to the partial aliasing of effects, as shown by the simulations in Chapter 5. For a two-stage group screening design, with regular fractional factorials, the aliasing scheme is known at each stage. Lack of reliability of conclusions from this method may arise from amalgamation and cancellation of effects but simulation evidence (Dean and Lewis (2002)) is that this risk is not very large.

6.2 Future work

The incorporation of information from the first stage experiment into the probabilities used at the second stage is one area in which I would like to direct my future research. Although no information on individual factorial effects would become available as a result of the first stage experiment, the probabilities of factors within a particular group or grouped interaction could all be updated in a uniform way, according to the findings from the first stage.

Another area for future research would be an investigation into the impact of tuning constants and prior information on factorial effects on the results of a Bayesian analysis of data from supersaturated designs and from group screening designs. For example, different choices for R in Chapter 5 (equation (5.2)) could be considered in order to recognise the relationships between the predictors. Also, the appropriateness of the assumption that the tuning constants ν, λ, c and τ are the same for all predictors could be investigated. Other techniques for exploring the model space within a Bayesian framework present further opportunities for research.

An area not addressed in this thesis is the choice of design to be used at the first and second stages of a group screening experiment. In particular, an investigation of choices for first stage designs could be carried out, with a view to choosing those with good projective properties. This would allow data from the first stage experiment to be used as part of the second stage experiment, thus reducing the overall number of observations required.

The idea of using more than two stages in group screening to detect interactions has not been addressed in this thesis and is an interesting area for future research. For example, if three stages of experiment were used, the first experiment would be carried out on grouped factors. Individual factors within those grouped effects declared active from the first stage experiment would be re-grouped for a second stage experiment. At the third stage, the individual factors within the grouped effects declared active at the second stage would be investigated.

Appendix A

Details of the input required for the software described in Chapter 3

A.1 Classical group screening

The software requires the following input from the user.

1. the number of individual control factors
2. the number of individual noise factors
3. the probabilities of each individual control main effect being active (entered in the order they are to be considered for grouping)
4. the probabilities of each individual noise main effect being active (entered in the order they are to be considered for grouping)
5. the number of grouped control factors and the number of grouped noise factors
6. the group sizes for the control factors
7. the group sizes for the noise factors
8. three targets for the number of factorial effects requiring estimation in the two-stage experiment

9. the number of degrees of freedom required for measuring error.

A.2 Interaction group screening

For interaction group screening, the user must specify 1-9, as above. For entering information on interaction probabilities, the user must use one of the following methods:

- enter one probability for all individual control×control interactions being active
- enter the probability of each individual interaction being active
- only enter the main effects probabilities and use the heredity approach of Section 2.5 to calculate the individual interaction probabilities.

Appendix B

Groupings for Section 3.3.1

Table B.1: Groupings for individual control factors, according to their main effects probabilities, for groupings of similar probabilities.

F	Design Grouping
1	(0.3, 0.4, 0.5, 0.6, 0.7, 0.8)
2	(0.3) (0.4, 0.5, 0.6, 0.7, 0.8)
2	(0.3, 0.4, 0.5, 0.6, 0.7) (0.8)
2	(0.3, 0.4) (0.5, 0.6, 0.7, 0.8)
2	(0.3, 0.4, 0.5, 0.6) (0.7, 0.8)
2	(0.3, 0.4, 0.5) (0.6, 0.7, 0.8)
3	(0.3) (0.4) (0.5, 0.6, 0.7, 0.8)
3	(0.3) (0.4, 0.5, 0.6, 0.7) (0.8)
3	(0.3, 0.4, 0.5, 0.6) (0.7) (0.8)
3	(0.3) (0.4, 0.5) (0.6, 0.7, 0.8)
3	(0.3) (0.4, 0.5, 0.6) (0.7, 0.8)
3	(0.3, 0.4) (0.5) (0.6, 0.7, 0.8)
3	(0.3, 0.4) (0.5, 0.6, 0.7) (0.8)
3	(0.3, 0.4, 0.5) (0.6) (0.7, 0.8)
3	(0.3, 0.4, 0.5) (0.6, 0.7) (0.8)
3	(0.3, 0.4) (0.5, 0.6) (0.7, 0.8)
4	(0.3) (0.4) (0.5) (0.6, 0.7, 0.8)
4	(0.3) (0.4) (0.5, 0.6, 0.7) (0.8)
4	(0.3) (0.4, 0.5, 0.6) (0.7) (0.8)
4	(0.3, 0.4, 0.5) (0.6) (0.7) (0.8)
4	(0.3) (0.4) (0.5, 0.6) (0.7, 0.8)
4	(0.3) (0.4, 0.5) (0.6) (0.7, 0.8)
4	(0.3) (0.4, 0.5) (0.6, 0.7) (0.8)
4	(0.3, 0.4) (0.5) (0.6) (0.7, 0.8)
4	(0.3, 0.4) (0.5) (0.6, 0.7) (0.8)
4	(0.3, 0.4) (0.5, 0.6) (0.7) (0.8)
5	(0.3) (0.4) (0.5) (0.6) (0.7, 0.8)
5	(0.3) (0.4) (0.5) (0.6, 0.7) (0.8)

continues overleaf

Table B.1: continued

F	Design Grouping
5	(0.3) (0.4) (0.5, 0.6) (0.7) (0.8)
5	(0.3) (0.4, 0.5) (0.6) (0.7) (0.8)
5	(0.3, 0.4) (0.5) (0.6) (0.7) (0.8)

Table B.2: Groupings for individual noise factors, according to their main effects probabilities for groupings of similar probabilities.

N	Noise Grouping
1	(0.0, 0.2, 0.4, 0.6, 0.8, 1.0)
2	(0.0) (0.2, 0.4, 0.6, 0.8, 1.0)
2	(0.0, 0.2, 0.4, 0.6, 0.8) (1.0)
2	(0.0, 0.2) (0.4, 0.6, 0.8, 1.0)
2	(0.0, 0.2, 0.4, 0.6) (0.8, 1.0)
2	(0.0, 0.2, 0.4) (0.6, 0.8, 1.0)
3	(0.0) (0.2) (0.4, 0.6, 0.8, 1.0)
3	(0.0) (0.2, 0.4, 0.6, 0.8) (1.0)
3	(0.0, 0.2, 0.4, 0.6) (0.8) (1.0)
3	(0.0) (0.2, 0.4) (0.6, 0.8, 1.0)
3	(0.0) (0.2, 0.4, 0.6) (0.8, 1.0)
3	(0.0, 0.2) (0.4) (0.6, 0.8, 1.0)
3	(0.0, 0.2) (0.4, 0.6, 0.8) (1.0)
3	(0.0, 0.2, 0.4) (0.6) (0.8, 1.0)
3	(0.0, 0.2, 0.4) (0.6, 0.8) (1.0)
3	(0.0, 0.2) (0.4, 0.6) (0.8, 1.0)

Table B.3: Groupings for individual control factors, according to their main effects probabilities, for groupings of dissimilar probabilities.

F	Design Grouping
1	(0.3, 0.8, 0.4, 0.7, 0.5, 0.6)
2	(0.3) (0.8, 0.4, 0.7, 0.5, 0.6)
2	(0.3, 0.8, 0.4, 0.7, 0.5) (0.6)
2	(0.3, 0.8) (0.4, 0.7, 0.5, 0.6)
2	(0.3, 0.8, 0.4, 0.7) (0.5, 0.6)
2	(0.3, 0.8, 0.4) (0.7, 0.5, 0.6)
3	(0.3) (0.8) (0.4, 0.7, 0.5, 0.6)
3	(0.3) (0.8, 0.4, 0.7, 0.5) (0.6)
3	(0.3, 0.8, 0.4, 0.7) (0.5) (0.6)
3	(0.3) (0.8, 0.4) (0.7, 0.5, 0.6)
3	(0.3) (0.8, 0.4, 0.7) (0.5, 0.6)
3	(0.3, 0.8) (0.4) (0.7, 0.5, 0.6)
3	(0.3, 0.8) (0.4, 0.7, 0.5) (0.6)
3	(0.3, 0.8, 0.4) (0.7) (0.5, 0.6)
3	(0.3, 0.8, 0.4) (0.7, 0.5) (0.6)
3	(0.3, 0.8) (0.4, 0.7) (0.5, 0.6)
4	(0.3) (0.8) (0.4) (0.7, 0.5, 0.6)
4	(0.3) (0.8) (0.4, 0.7, 0.5) (0.6)
4	(0.3) (0.8, 0.4, 0.7) (0.5) (0.6)
4	(0.3, 0.8, 0.4) (0.7) (0.5) (0.6)
4	(0.3) (0.8) (0.4, 0.7) (0.5, 0.6)
4	(0.3) (0.8, 0.4) (0.7) (0.5, 0.6)
4	(0.3) (0.8, 0.4) (0.7, 0.5) (0.6)
4	(0.3, 0.8) (0.4) (0.7) (0.5, 0.6)
4	(0.3, 0.8) (0.4) (0.7, 0.5) (0.6)
4	(0.3, 0.8) (0.4, 0.7) (0.5) (0.6)
5	(0.3) (0.8) (0.4) (0.7) (0.5, 0.6)
5	(0.3) (0.8) (0.4) (0.7, 0.5) (0.6)
5	(0.3) (0.8) (0.4, 0.7) (0.5) (0.6)
5	(0.3) (0.8, 0.4) (0.7) (0.5) (0.6)
5	(0.3, 0.8) (0.4) (0.7) (0.5) (0.6)

Table B.4: Groupings for individual noise factors, according to their main effects probabilities for groupings of dissimilar probabilities.

N	Noise Grouping
1	(0.0, 1.0, 0.2, 0.8, 0.4, 0.6
2	(0.0) (1.0, 0.2, 0.8, 0.4, 0.6)
2	(0.0, 1.0, 0.2, 0.8, 0.4) (0.6)
2	(0.0, 1.0) (0.2, 0.8, 0.4, 0.6)
2	(0.0, 1.0, 0.2, 0.8) (0.4, 0.6)
2	(0.0, 1.0, 0.2) (0.8, 0.4, 0.6)
3	(0.0) (1.0) (0.2, 0.8, 0.4, 0.6)
3	(0.0) (1.0, 0.2, 0.8, 0.4) (0.6)
3	(0.0, 1.0, 0.2, 0.8) (0.4) (0.6)
3	(0.0) (1.0, 0.2) (0.8, 0.4, 0.6)
3	(0.0) (1.0, 0.2, 0.8) (0.4, 0.6)
3	(0.0, 1.0) (0.2) (0.8, 0.4, 0.6)
3	(0.0, 1.0) (0.2, 0.8, 0.4) (0.6)
3	(0.0, 1.0, 0.2) (0.8) (0.4, 0.6)
3	(0.0, 1.0, 0.2) (0.8, 0.4) (0.6)
3	(0.0, 1.0) (0.2, 0.8) (0.4, 0.6)

Appendix C

More detailed results for Study 1 of Section(3.3.1)

Table C.1: A comparison of $E(S)$, $s.d.(S)$, and $P(S > 65)$ for 6 individual control factors and 6 individual noise factors with $F \leq 5$, and $N \leq 3$ for groupings based on similar probabilities for interaction group screening.

Noise grouping	max $E(S)$	min $E(S)$	max $s.d.(S)$	min $s.d.(S)$	max $P(S > 65)$	min $P(S > 65)$
(6)	72.73 (6)	65.88 (1,1,2,2)	9.19 (2,2,2)	3.55 (6)	0.99 (6)	0.57 (2,2,2) (1,1,2,2)
(1,5)	70.51 (6)	65.28 (2,2,2)	8.61 (2,2,2)	5.33 (6)	0.99 (6)	0.52 (2,2,2)
(5,1)	72.98 (6)	65.10 (2,2,2)	9.75 (3,3)	6.71 (6)	0.98 (6)	0.5 (2,2,2)
(2,4)	69.87 (6)	62.00 (2,2,2)	9.51 (3,3)	7.40 (1,1,1,1,2)	0.64 (6)	0.3 (2,2,2)
(4,2)	72.53 (6)	62.23 (2,2,2)	10.75 (3,3)	7.89 (1,1,1,1,2)	0.85 (6)	0.33 (2,2,2)
(3,3)	71.26 (6)	61.17 (2,2,2)	11.31 (3,3)	6.02 (1,1,1,1,2)	0.81 (6)	0.31 (2,2,2)
(1,1,4)	70.09 (1,1,1,1,2) (1,1,1,2,1) (1,1,2,1,1)	64.43 (2,2,2)	8.41 (3,3)	6.72 (1,1,1,1,2)	0.64 (6)	0.54 (2,2,2)
(1,4,1)	70.64 (6)	64.26 (2,2,2)	9.34 (3,3)	7.10 (1,1,1,1,2)	0.80 (6)	0.52 (2,2,2)
(4,1,1)	72.15 (6)	63.74 (2,2,2)	10.49 (3,3)	7.62 (1,1,1,1,2)	0.85 (6)	0.49 (2,2,2)
(1,2,3)	68.97 (6)	61.96 (2,2,2)	9.41 (3,3)	6.87 (1,1,1,1,2)	0.72 (6)	0.37 (2,2,2)
(1,3,2)	70.06 (6)	62.00 (2,2,2)	9.93 (3,3)	7.08 (1,1,1,1,2)	0.82 (6)	0.38 (2,2,2)

continues overleaf

Table C.1: continued

Noise grouping	max $E(S)$	min $E(S)$	max $s.d(S)$	min $s.d(S)$	max $P(S > 65)$	min $P(S > 65)$
(2,1,3)	68.84 (6)	61.58 (2,2,2)	9.66 (6)	6.96 (1,1,1,2,1)	0.63 (6)	0.36 (2,2,2)
(2,3,1)	69.77 (6)	61.43 (2,2,2)	10.14 (6)	7.18 (1,1,1,2,1)	0.61 (6) (5,1)	0.36 (2,2,2)
(3,1,2)	70.54 (6)	61.41 (2,2,2)	10.69 (6)	7.42 (1,2,1,1,1)	0.76 (6)	0.37 (2,2,2)
(3,2,1)	70.86 (6)	61.30 (2,2,2)	10.93 (6)	7.50 (1,2,1,1,1) (2,1,1,1,1)	0.73 (6)	0.37 (2,2,2)
(2,2,2)	69.05 (6)	60.02 (2,2,2)	10.81 (6)	7.02 (2,1,1,1,1)	0.48 (6)	0.34 (2,2,2)

Table C.2: A comparison of $E(S)$, $s.d.(S)$, and $P(S > 65)$ for 6 individual control factors and 6 individual noise factors with $F \leq 5$, and $N \leq 3$ for groupings based on dissimilar probabilities for interaction group screening.

Noise grouping	max $E(S)$	min $E(S)$	max $s.d(S)$	min $s.d(S)$	max $P(S > 65)$	min $P(S > 65)$
(6)	72.73 (6)	66.13 (1,2,2,1)	9.07 (2,2,2)	3.55 (6)	0.99 (6)	0.58 (1,2,2,1)
(1,5)	70.51 (6)	65.56 (2,2,2)	8.53 (2,2,2)	5.33 (6)	0.99 (6)	0.53 (2,2,2)
(5,1)	72.29 (6)	65.68 (2,2,2)	9.12 (2,2,2)	6.15 (6)	0.99 (6)	0.53 (2,2,2)
(2,4)	72.15 (6)	62.98 (2,2,2)	10.02 (3,3)	7.69 (1,2,1,1,1)	0.80 (6)	0.35 (2,2,2)
(4,2)	72.10 (6)	62.98 (2,2,2)	10.00 (3,3)	7.68 (1,2,1,1,1)	0.79 (6)	0.35 (2,2,2)
(3,3)	72.27 (6)	62.12 (2,2,2)	10.47 (3,3)	7.72 (1,2,1,1,1)	0.85 (6)	0.41 (1,1,2,2) (1,2,1,2) (1,2,2,1) (2,1,1,2) (2,1,2,1) (2,2,1,1)
(1,1,4)	70.64 (6)	64.50 (2,2,2)	9.28 (3,3)	7.17 (1,2,1,1,1)	0.80 (6)	0.53 (2,2,2)
(1,4,1)	70.13 (2,1,1,1,1)	64.71 (2,2,2)	8.80 (3,3)	6.97 (1,2,1,1,1)	0.73 (6)	0.55 (2,2,2)
(4,1,1)	71.16 (6)	64.75 (2,2,2)	9.34 (3,3)	7.21 (1,2,1,1,1)	0.79 (6)	0.54 (2,2,2)
(1,2,3)	70.05 (6)	62.51 (2,2,2)	9.61 (3,3)	6.97 (1,2,1,1,1)	0.80 (6)	0.4 (2,2,2)
(1,3,2)	69.79 (6)	62.46 (2,2,2)	9.52 (3,3)	6.93 (1,2,1,1,1)	0.78 (6)	0.4 (2,2,2)
(2,1,3)	70.39 (6)	62.52 (2,2,2)	9.75 (3,3)	7.04 (1,2,1,1,1)	0.78 (6)	0.4 (2,2,2)
(2,3,1)	71.07 (6)	62.38 (2,2,2)	10.15 (3,3)	7.23 (1,2,1,1,1)	0.76 (6)	0.4 (2,2,2)
(3,1,2)	71.17 (6)	62.20 (2,2,2)	10.33 (3,3)	7.31 (1,2,1,1,1)	0.74 (6)	0.4 (2,2,2)
(3,2,1)	71.14 (6)	62.32 (2,2,2)	10.24 (3,3)	7.27 (1,2,1,1,1)	0.77 (6)	0.4 (2,2,2)
(2,2,2)	70.42 (6)	60.90 (2,2,2)	10.60 (6)	7.11 (2,1,1,1,1)	0.59 (5,1)	0.39 (2,2,2)

Table C.3: A comparison of $E(S)$, $s.d.(S)$, and $P(S > 65)$ for 6 individual control factors and 6 individual noise factors with $F \leq 5$, and $N \leq 3$ for groupings based on similar probabilities for classical group screening.

Noise grouping	max $E(S)$	min $E(S)$	max $s.d.(S)$	min $s.d.(S)$	max $P(S > 65)$	min $P(S > 65)$
(6)	71.65 (6)	50.21 (1,1,1,1,2)	15.54 (3,2,1)	4.89 (6)	0.99 (6)	0.03 (1,1,1,1,2)
(1,5)	64.69 (6)	45.54 (1,1,1,1,2)	14.07 (3,2,1)	4.32 (6)	0.86 (4,2)	0.00 (6)
(5,1)	71.12 (6)	50.12 (1,1,1,1,2)	16.70 (3,2,1)	9.03 (6)	0.96 (6)	0.03 (1,1,1,1,2)
(2,4)	59.92 (6)	42.14 (1,1,1,1,2)	14.35 (3,2,1)	7.52 (6)	0.20 (6)	0.01 (1,1,1,3)
(4,2)	66.54 (6)	46.86 (1,1,1,1,2)	18.22 (3,2,1)	13.28 (1,5)	0.80 (6)	0.03 (1,1,1,1,2) (1,1,1,2,1)
(3,3)	61.19 (6)	43.05 (1,1,1,1,2)	17.10 (3,2,1)	12.56 (1,5)	0.52 (6)	0.02 (1,1,1,1,2) (1,1,1,2,1)
(1,1,4)	59.32 (6)	42.01 (1,1,1,1,2)	13.20 (3,2,1)	5.01 (6)	0.20 (6)	0.01 (1,1,1,3) (1,1,1,1,2) (1,1,1,2,1) (1,1,2,1,1) (1,2,1,1,1)
(1,4,1)	64.47 (6)	45.67 (1,1,1,1,2)	14.87 (3,2,1)	7.45 (6)	0.96 (6)	0.03 (1,1,1,1,2)
(4,1,1)	65.95 (6)	46.73 (1,1,1,1,2)	18.21 (3,2,1)	13.53 (1,5)	0.80 (6)	0.05 (1,1,1,3)
(1,2,3)	58.05 (6)	41.10 (1,1,1,1,2)	14.49 (3,2,1)	8.82 (6)	0.52 (6)	0.02 (1,1,1,1,2) (1,1,1,2,1)
(1,3,2)	61.11 (6)	43.28 (1,1,1,1,2)	15.64 (3,2,1)	10.24 (6)	0.80 (6)	0.03 (1,1,1,1,2) (1,1,1,2,1)
(2,1,3)	56.14 (6)	39.74 (1,1,1,1,2)	13.98 (3,2,1)	8.32 (6)	0.20 (6)	0.01 (1,1,1,3)
(2,3,1)	59.77 (6)	42.33 (1,1,1,1,2)	14.86 (3,2,1)	9.06 (6)	0.19 (6)	0.01 (1,1,1,3)
(3,1,2)	59.01 (6)	41.78 (1,1,1,1,2)	17.00 (3,2,1)	12.83 (1,1,1,1,2)	0.52 (6)	0.03 (1,1,1,3)
(3,2,1)	60.92 (6)	43.14 (1,1,1,1,2)	17.35 (3,2,1)	13.07 (1,1,4)	0.48 (6)	0.03 (1,1,1,3)
(2,2,2)	57.10 (6)	40.42 (1,1,1,1,2)	14.97 (3,2,1)	10.05 (6)	0.15 (6)	0.01 (1,1,1,3)

continues overleaf

Table C.3: continued

Noise grouping	max $E(S)$	min $E(S)$	max $s.d.(S)$	min $s.d.(S)$	max $P(S > 65)$	min $P(S > 65)$
						(1,1,3,1) (1,1,2,2)

Table C.4: A comparison of $E(S)$, $s.d.(S)$, and $P(S > 65)$ for 6 individual control factors and 6 individual noise factors with $F \leq 5$, and $N \leq 3$ for groupings based on dissimilar probabilities for classical group screening.

Noise grouping	max $E(S)$	min $E(S)$	max $s.d.(S)$	min $s.d.(S)$	max $P(S > 65)$	min $P(S > 65)$
(6)	71.65 (6)	51.54 (1,1,1,2,1)	14.40 (2,2,2)	4.89 (6)	0.99 (6)	0.05 (1,1,1,1,2) (1,1,1,2,1)
(1,5)	64.69 (6)	46.75 (1,1,1,2,1)	12.29 (3,1,2)	4.32 (6)	0.83 (2,4)	0.00 (6)
(5,1)	69.47 (6)	50.22 (1,1,1,2,1)	14.29 (2,2,2)	6.08 (6)	0.86 (3,3)	0.05 (1,1,1,1,2) (1,1,1,2,1)
(2,4)	71.43 (6)	51.65 (1,1,1,2,1)	15.20 (2,2,2)	7.79 (6)	0.96 (6)	0.05 (1,1,1,1,2) (1,1,1,2,1)
(4,2)	68.83 (6)	49.76 (1,1,1,2,1)	15.02 (2,2,2)	8.23 (6)	0.76 (6)	0.04 (1,1,1,1,2) (1,1,1,2,1) (1,1,2,1,1) (1,2,1,1,1)
(3,3)	71.51 (6)	51.71 (1,1,1,2,1)	14.91 (2,2,2)	7.02 (6)	0.95 (6)	0.05 (1,1,1,1,2) (1,1,1,2,1) (1,2,1,1,1)
(1,1,4)	64.47 (6)	46.86 (1,1,1,2,1)	13.95 (2,2,2)	7.45 (6)	0.96 (6)	0.05 (1,1,1,1,2) (1,1,1,2,1) (1,2,1,1,1)
(1,4,1)	62.51 (6)	45.43 (1,1,1,2,1)	12.99 (2,2,2)	5.66 (6)	0.60 (6)	0.03 (1,1,1,1,2) (1,1,1,2,1) (1,2,1,1,1)
(4,1,1)	65.69 (6)	47.75 (1,1,1,2,1)	13.94 (2,2,2)	7.02 (6)	0.76 (6)	0.07 (1,1,1,3) (1,3,1,1)
(1,2,3)	64.55 (6)	46.92 (1,1,1,2,1)	13.64 (2,2,2)	6.64 (6)	0.95 (6)	0.05 (1,1,1,1,2) (1,1,1,2,1) (1,2,1,1,1)
(1,3,2)	61.87 (6)	44.97 (1,1,1,2,1)	13.80 (2,2,2)	7.93 (6)	0.76 (6)	0.04 (1,1,1,2,1) (1,1,2,1,1) (1,2,1,1,1)
(2,1,3)	66.14	48.07	14.18	7.44	0.95	0.08

continues overleaf

Table C.4: continued

Noise grouping	max $E(S)$	min $E(S)$	max $s.d.(S)$	min $s.d.(S)$	max $P(S > 65)$	min $P(S > 65)$
	(6)	(1,1,1,2,1)	(2,2,2)	(6)	(6)	(1,3,1,1)
(2,3,1)	68.18 (6)	49.56 (1,1,1,2,1)	15.25 (2,2,2)	9.23 (6)	0.90 (6)	0.08 (1,1,1,3) (1,3,1,1)
(3,1,2)	68.24 (6)	49.60 (1,1,1,2,1)	15.05 (2,2,2)	8.77 (6)	0.76 (6)	0.07 (1,1,1,3) (1,3,1,1)
(3,2,1)	68.56 (6)	49.83 (1,1,1,2,1)	14.71 (2,2,2)	7.89 (6)	0.88 (6)	0.08 (1,1,1,3) (1,3,1,1)
(2,2,2)	67.28 (6)	48.91 (1,1,1,2,1)	15.52 (2,2,2)	10.02 (6)	0.64 (6)	0.06 (1,1,1,3) (1,3,1,1)

Appendix D

Further interaction plots for the first stage experiment (Chapter 4)

Figure D.1: Interaction plot for the grouped control×noise interaction between Group 1 and Group 5.

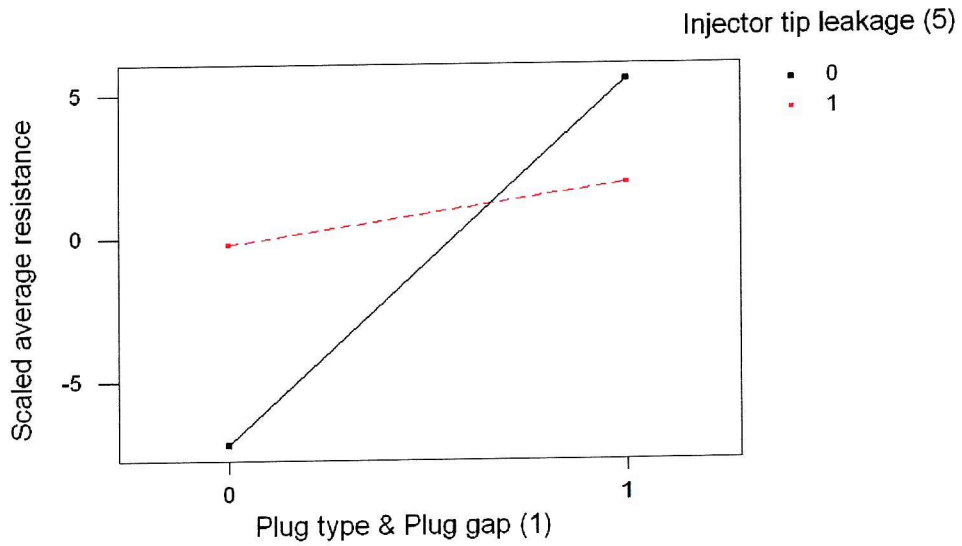


Figure D.2: Interaction plot for the grouped control×control interaction between Group 2 and Group 3.

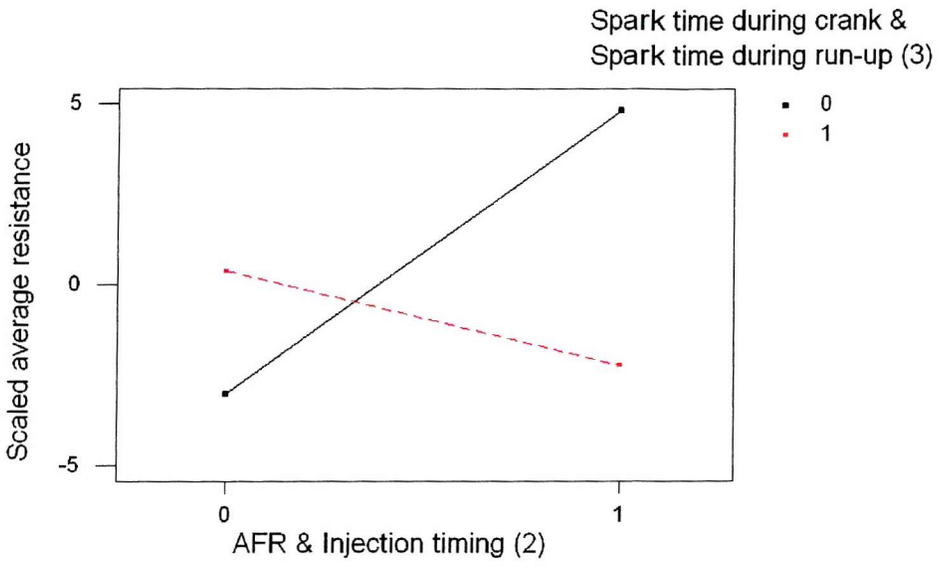
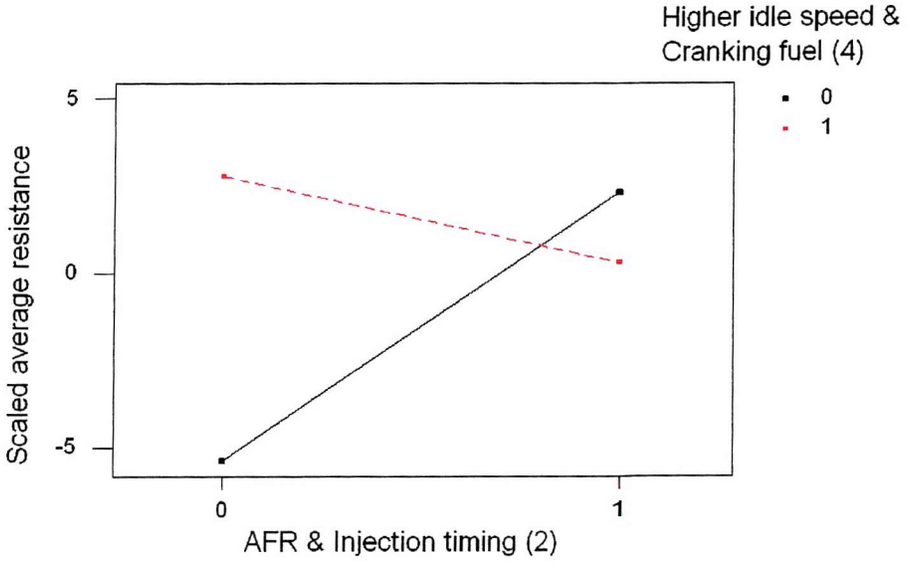


Figure D.3: Interaction plot for the grouped control×control interaction between Group 2 and Group 4.



Appendix E

Results for Section 5.3

Table E.1: The design and data (to 4 decimal places) for the full factorial experiment where the error terms in the data generation were drawn from a $N(0, 4)$ distribution.

A	B	C	AB	AC	BC	Response
-1	-1	-1	1	1	1	-2.4475
-1	-1	1	1	-1	-1	2.6033
-1	1	-1	-1	1	-1	-21.5117
-1	1	1	-1	-1	1	-18.4743
1	-1	-1	-1	-1	1	-2.5308
1	-1	1	-1	1	-1	3.2757
1	1	-1	1	-1	-1	20.3820
1	1	1	1	1	1	21.5441

Table E.2: The four models (γ) with highest posterior probabilities identified by SSVS for the design and data given in Table E.1 under three sets of prior probabilities for main effects and interactions being active, and for a full 2^3 experiment. The entries in each model are ordered according to the terms corresponding to A, B, C, AB, AC, BC.

All m.e. and int. probabilities 0.5		All m.e. and int. probabilities 0.99		All m.e. probs 0.5; $w_{00} = 0.01$ $w_{01} = w_{10} = 0.1$, $w_{11} = 0.25$	
Model	Probability	Model	Probability	Model	Probability
100100	0.31	111111	0.87	100100	0.42
100110	0.20	101111	0.04	110100	0.22
100101	0.09	111101	0.04	101100	0.20
100110	0.07	111110	0.03	111100	0.10
Marginal Probabilities					
A	1.00	A	1.00	A	1.00
B	0.19	B	0.96	B	0.35
C	0.39	C	0.99	C	0.34
AB	1.00	AB	1.00	AB	1.00
AC	0.19	AC	0.96	AC	0.04
BC	0.22	BC	0.97	BC	0.03

Table E.3: The four models (γ) with highest posterior probabilities identified by SSVS under three sets of prior probabilities for main effects and interactions being active, and for the 2^{3-1} experiment where the error terms in the data generation were drawn from a $N(0, 4)$ distribution. The entries in each model are ordered according to the terms corresponding to A, B, C, AB, AC, BC.

All m.e. and int. probabilities 0.5		All m.e. and int. probabilities 0.99		All m.e. probs 0.5; $w_{00} = 0.01$ $w_{01} = w_{10} = 0.1$, $w_{11} = 0.25$	
Model	Probability	Model	Probability	Model	Probability
101101	0.09	011111	0.95	101000	0.37
100101	0.08	101111	0.01	111000	0.08
101001	0.07	111101	0.01	101001	0.06
101100	0.05	100111	0.01	100100	0.05
Marginal Probabilities					
A	0.72	A	0.99	A	0.89
B	0.29	B	0.99	B	0.32
C	0.67	C	0.99	C	0.87
AB	0.68	AB	0.99	AB	0.24
AC	0.29	AC	0.99	AC	0.11
BC	0.73	BC	0.99	BC	0.27

References

Abraham, B., Chipman, H. and Vijayan, K. (1999) Some risks in the construction and analysis of supersaturated designs. *Technometrics*, **41**, 135–141.

Allen, T. and Bernshteyn (2003) Supersaturated designs that maximize the probability of identifying active factors. *Technometrics*, **45**, 90–97.

Ankenman, N. E. and Dean, A. M. (2001) Quality improvement and robustness via design of experiments. In *Handbook of statistics* (eds. R. Khattree and C. R. Rao), vol. 22, chap. 8, 263–317. North-Holland.

Beattie, S. D., Fong, K. H. and Lin, D. K. J. (2002) A two-stage Bayesian model selection strategy for supersaturated designs. *Technometrics*, **44**, 55–63.

Berger, J. O. and Pericchi, L. R. (1996a) The intrinsic Bayes factor for linear models. In *Bayesian Statistics - Proceedings of the 5th Valencia International Meeting Held in Alicante, June 5-9, 1994.*

Bingham, D. and Chipman, H. (2002) Optimal designs for model selection. *Tech. Rep. 388*, University of Michigan.

Booth, K. H. V. and Cox, D. R. (1962) Some systematic supersaturated designs. *Technometrics*, **4**, 489–495.

Box, G. E. P. and Meyer, D. (1986) An analysis for unreplicated fractional factorials. *Technometrics*, **28**, 11–18.

Butler, N. A., Mead, R., Eskridge, K. M. and Gilmour, S. G. (2001) A general method of constructing $E(s^2)$-optimal supersaturated designs. *Journal of the Royal Statistical Society B*, **63**, 621–632.

Cheng, R. (1997) Searching for important factors: sequential bifurcation under uncertainty. In *Proceedings of the 1997 Winter Simulation Conference.*

Cheng, S.-W. and Wu, C. F. J. (2001) Factor screening and response surface exploration. *Statistica Sinica*, **11**, 553–604.

Chipman, H. (1996) Bayesian variable selection with related predictors. *The Canadian Journal of Statistics*, **24**, 17–36.

— (2003) An overview of the use of Bayesian priors in analysis of screening experiments. *Tech. rep.*, University of Waterloo.

Chipman, H., Hamada, M. and Wu, C. F. J. (1997) A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics*, **39**.

Dean, A. M. and Lewis, S. M. (2002) Comparison of group screening strategies for factorial experiments. *Computational Statistics and Data Analysis*, **39**, 287–297.

Deming, W. E. (1986) *Out of the crisis*. Cambridge: Cambridge University Press.

Dorfman, R. (1943) The detection of defective members of large populations. *Annals of Mathematical Statistics*, **14**, 436–440.

Draper, N. R. and Smith, H. (1998) *Applied regression analysis*. New York: Wiley, third edn.

Dupplaw, D., Brunson, D., Vine, A. E., Please, C., Lewis, S., Dean, A., Keane, A. and Tindall, M. J. (2004) Web-based knowledge elicitation system (GISEL) for planning and assessing group screening experiments for product development. *ASME Transactions - Journal of Computing and Information Science in Engineering*. In press.

George, E. I. and McCulloch, R. E. (1993) Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, **88**, 881–889.

Hamada, M. and Wu, C. F. J. (1992) Analysis of designed experiments with complex aliasing. *Journal of Quality Technology*, **24**, 130–137.

John, J. A. and Williams, E. R. (1995) *Cyclic and computer generated designs*. Chapman and Hall: London, second edn.

Kleijnen, J. P. C. (1987) Review of random and group-screening designs. *Communications in Statistics - Theory and Methods*, **16**, 2885–2900.

Lewis, S. M. and Dean, A. M. (2001) Detection of interactions in experiments on large numbers of factors. *Journal of the Royal Statistical Society B*, **63**, 633–672. With discussion.

Lin, D. K. J. (1993) A new class of supersaturated designs. *Technometrics*, **35**, 28–31.

— (1995) Generating systematic supersaturated designs. *Technometrics*, **37**, 213–225.

Liu, Y. and Dean, A. (2004) k-circulant supersaturated designs. *Technometrics*, **46**, 32–43.

Nguyen, N. (1996) An algorithmic approach to constructing supersaturated designs. *Technometrics*, **38**, 69–73.

Russell, K. G., Lewis, S. and Dean, A. M. (2004) Fractional factorial designs for the detection of interactions between design and noise factors. *Journal of Applied Statistics*, **31**, 545–552. In press.

Satterthwaite, F. (1959) Random balance experimentation (with discussion). *Technometrics*, **1**, 111–137.

Shoemaker, A. C., Tsui, K.-L. and Wu, C. F. J. (1991) Economical experimentation methods for robust design. *Technometrics*, **33**, 415–427.

Suh, Y. J., Ye, K. Q. and Mendell, N, R. (2003) A method of evaluating the results of Bayesian model selection: application to linkage analyses of attributes determined by two or more genes. *Human Heredity*, **55**, 147–152.

Taguchi, G. (1986) *Introduction to quality engineering.* Tokyo: Asian Productivity Organization.

Trocine, L. and Malone, L. (2001) An overview of newer, advanced screening methods for the initial phase in an experimental design. In *Proceedings of the 2001 Winter Simulation Conference.*

Vine, A. E., Lewis, S. M. and Dean, A. M. (2004) Two-stage group screening in the presence of noise factors and unequal probabilities of active effects. *Statistica Sinica.* Accepted subject to minor revision.

Watson, G., S. (1961) A study of the group screening method. *Technometrics*, **3**, 371–388.

Williams, K. R. (1968) Designed experiments. *Rubber Age*, **100**, 65–71.

Wu, C. F. J. (1993) Construction of supersaturated designs through partially aliased interactions. *Biometrika*, **80**, 661–669.

Wu, C. F. J. and Hamada, M. (2000) *Experiments: Planning, Analysis and Parameter Design Optimization.* New York: Wiley.