Three-Level Main-Effects Designs Exploiting Prior Information About Model Uncertainty

Pi-Wen Tsai
Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, ROC.
pwtsai@stat.sinica.edu.tw

Steven G. Gilmour
School of Mathematical Sciences, Queen Mary, University of London,
Mile End Road, London E1 4NS, UK.
s.g.gilmour@qmul.ac.uk

Roger Mead
School of Applied Statistics, The University of Reading,
PO Box 240, Earley Gate, Reading RG6 6FN, UK.
r.mead@reading.ac.uk

SUMMARY

In this paper, we consider experiments for which experimenters have some prior information on the probability of effects being non-negligible in advance of collecting their data. We suggest using this prior information as a tool to help experimenters to decide the appropriate run-sizes that experimenters should use. In addition, we introduce a criterion for selecting orthogonal, or nearly orthogonal, main effects designs, with robustness to interactions as a secondary consideration under particular priors. We will show that this criterion exploiting prior information about model uncertainty can lead to more appropriate designs reflecting experimenters' prior belief on the importance of each effect.

Some key words: prior information; factor screening; orthogonal array; projection; Bayesian D-optimal designs.

1 Introduction

In industrial experiments, when several factors are of interest, screening designs are widely used to find the few "active" factors which have major effects. Traditionally, design and analysis of screening experiments have been restricted to main effects only by assuming all interactions are small and therefore negligible. However, it may be that, when modelling the data, a few interactions are included in the model while some factors may be dropped completely. Motivated by Hamada and Wu (1992), Tsai, Gilmour and Mead (2000) used a two-stage analysis that considers interactions as well as main effects for the analysis of screening experiments with complex aliasing patterns. First, the main effects model is fitted to identify "active" factors. Secondly, a stepwise regression procedure is used for selecting a suitable fitted model that may contain some interactions as well as main effects, but which obeys functional marginality (McCullagh and Nelder, 1989). It has been emphasised that, for this two-stage analysis, it is important that designs used for screening experiments can project onto good lower dimensional designs for various sets of factors and thus provide reasonably efficient parameter estimates for a range of possible models.

To explore projection properties of screening designs, Tsai *et al.* (2000) introduced a criterion, denoted by $Q(\Gamma^{(k)})$ for a k-factor design, to compare orthogonal, or nearly orthogonal, main effects designs, with robustness to interactions as a secondary consideration. This criterion averages an approximation to A_s-efficiency over lower-dimensional projections of the design. It has been showed that designs with lower $Q(\Gamma^{(k)})$ are more likely to have efficient projections and on average can provide better parameter estimates over a range of models than designs with higher $Q(\Gamma^{(k)})$. For detailed information on the Q-criterion, the reader is referred to Tsai *et al.* (2000).

Here, we extend their method and consider the case when in advance of running an experiment, experimenters have some prior knowledge about the probabilities of effects being non-negligible. We will introduce a general framework for the prior probabilities of effects being non-negligible for the two-stage analysis. An 18-run three-level designs with six factors will be used as an example used to demonstrate how to use these prior knowledge as a tool to help experimenters to decide the appropriate experimental run sizes that they should use.

In addition, to incorporate the prior knowledge into design process, we generalise the $Q(\Gamma^{(k)})$ criterion and propose a new criterion exploiting prior information about model uncertainty for selecting orthogonal, or nearly orthogonal, main effects designs, with robustness to interactions as a secondary consideration. In this paper, we will study the relationship between this criterion and the the generalised minimum aberration criterion. For simplicity, we compare this criterion with the G_2-aberration criterion proposed by Tang and Deng (1999) for designs with two levels. We will demonstrate that the efficiency-type $Q(\Gamma^{(k)})$ criterion is the same as the G_2-aberration criterion (1999) is a good surrogate for model robustness for designs with two levels. Therefore, the $Q_B(\Gamma^{(k)})$ criterion which generalised the Q criterion is a good criterion for selecting designs when experimenters have some prior believe of each factorial effect being non-negligible.

One of the frequently used three-level orthogonal arrays in industrial experiments is the $L_{18}(3^7)$, an 18-run design for seven three-level factors. Wang and Wu (1995) and Cheng and Wu (2001) studied the projection properties of the $L_{18}(3^7)$ orthogonal array when projected onto sets of three and four factors. Cheng and Ye (2004) permuted the levels of the projected designs obtained from the $L_{18}(3^7)$ orthogonal array to generate some additional orthogonal main effects designs. Alternatively, Tsai *et al.* (2000) used a columnwise procedure to generate all possible three-level main effects designs for 3 to 6 factors in 18 runs. It has been shown that, by using their procedure, more designs can be found. The best 18-run designs with six factors that they generated have better projection properties and are more efficient than those obtained by the procedure of Cheng and Ye (2004) and those from the $L_{18}(3^7)$. Details and applications of the Q-criterion were given by Tsai, *et al.* (2000, 2004).

Here, we list some best six-factor main effects designs in 18 runs obtained by Tsai *et al.* (2000) and compare them with those obtained by Cheng and Ye (2004) and those obtained from the $L_{18}(3^7)$ orthogonal array under various priors. We will show that, depending on the prior knowledge, more appropriate designs can be found among the class of good designs. The best designs constructed in their way are very robust to prior specifications, so experimenters can have more confidence in using them.

The paper is organised as follows. In Section 2, we introduce the prior probabilities for the two-stage analysis. In Section 3, we introduce the $Q_B(\Gamma^{(k)})$-criterion to assess designs under model uncertainty for some given prior. A comparison of an 18-run experiment is in Section 4. Finally, in Section 5, we provide some concluding remarks.

2 Prior probabilities

In this paper, we consider experiments for which experimenters have some prior information on the probabilities of effects being non-negligible in advance of collecting their data. The idea of using prior probability of particular factors being non-negligible for the analysis of designed experiments with complex aliasing was first introduced by Box and Meyer (1993). They proposed a model-discrimination criterion, based on posterior probabilities, for designing follow-up experiments that allow maximum discrimination among the plausible models. Chipman, Hamada and Wu (1997) suggested a more elaborate Bayesian analysis approach for analysing data from designs with complex aliasing. They used the hierarchical variable structure to reduce the model space, and implemented variable selection using the stochastic search variable selection (SSVS) methods, which sample from the posterior probability distribution over the model space. Recently, Beattie, Fong and Lin (2002) proposed a two-stage Bayesian model selection strategy for supersaturated designs, which can be summarized as a combination of SSVS (the first stage) to find the active factors and intrinsic Bayes factors (the second stage) to test whether their effects are significant.

Here, we do not use fully Bayesian methods. Instead, we propose a method to exploit experimenters' prior beliefs for selecting orthogonal, or nearly orthogonal, main effects designs, with robustness to interactions as a secondary consideration. The two-stage analysis is used to analyse the data for screening experiments. First, the main effects model is fitted to identify "active" factors. Secondly, a stepwise regression procedure is used for selecting a suitable fitted model that may contain some interactions as well as main effects, but which obeys functional marginality (McCullagh and Nelder, 1989). Functional marginality means that every term in the model must be accompanied by all terms marginal to it, whether these are large or small. We draw a distinction between functional marginality and the strong heredity principle, defined by Chipman (1996) to be "the belief that for an interaction to be active . . . both corresponding main effects must also be active." Marginality defines the class of models that might be fitted, whereas strong heredity describes a set of prior beliefs about which effects might be large. Unfortunately, these two concepts have become blurred in the literature, but the distinction is crucial to the work presented here. We believe models that break functional marginality should not be formulated or used unless some good practical reason can be given for so doing. The reader who is interested in this issue can refer to Nelder (1997, 1998).

For a three-level design with k factors, we consider only first- and second-order effects when the k-factor second-order model is the maximal model of interest. Assume that there is a "true model" that explains the relationship between the response and the factors. Assume further that experimenters have some prior knowledge about the probabilities of each effect being in the true model in advance of running an experiment. Let π_1 denote experimenter's prior belief that *"a linear term x_i is in the true model"*, which is defined as

$$\Pr(\delta(x_i) = 1) = \pi_1,$$

where $\delta(x_i)$ indicates whether the linear effect of factor i is in the true model.

Under the functional marginality rule, a quadratic effect term may be in the model only if the linear effect of the same factor is already in the model. Therefore, we define π_2 as the prior probability that a quadratic effect is in the true model *given that the linear effect of the same* *factor is in the model*, so that

$$\Pr(\delta(x_i^2) = 1 \mid \delta(x_i)) = \begin{cases} \pi_2 & \text{if } \delta(x_i) = 1; \\ 0 & \text{if } \delta(x_i) = 0. \end{cases}$$

Similarly, we define π_3 as the prior probability that an interaction is in the true model *given that the linear effects of both the factors involved are in the model*, so that

$$\Pr(\delta(x_i x_j) = 1 \mid \delta(x_i), \delta(x_j)) = \begin{cases} \pi_3 & \text{if } \delta(x_i) = \delta(x_j) = 1; \\ 0 & \text{otherwise.} \end{cases}$$

These definitions imply that effects of the same type have the same probability of being in the true model, and that the event that a linear effect is in the true model is independent of the event that any other factor's linear effect is in the true model. Also, for a given set of linear effects being in the model, the inclusion of second order effects in the model is independent of each other.

Note that the definition of π_1 does not denote the probability that the linear effect of factor i is large, but rather the probability that term x_i, *or any effect to which it is marginal*, is large. This is different from the prior probabilities defined by Chipman (1996). Detail discussion on the relationship between the definition above and strong heredity and weak heredity principles defined by Chipman (1996) can be found in Appendix A.

Given prior believe of each factorial effect being in the model, one can explicitly calculate the prior probability of each model being the true model. Let $\Pr(\mathcal{M}_s)$ denote such prior probability of model M_s being the true model under a specific prior, which is computed as the joint probability of each individual effect being in the true model.

$$\Pr(\mathcal{M}_s) = \pi_1^a \, (1 - \pi_1)^{k-a} \, \pi_2^b \, (1 - \pi_2)^{a-b} \, \pi_3^c \, (1 - \pi_3)^{\frac{a(a-1)}{2} - c}, \tag{1}$$

where a, b and c are the numbers of linear, quadratic and interaction effects that are in the model, respectively.

We suggest use this information as a reference to help experimenters to decide the appropriate run-sizes that they should use. First, we compute the probabilities of each of the models being the true model. Second, we add up the probabilities for models with a specific number of parameters. This sum represents the probability that a model with this number of parameters is the true model. Finally, a histogram of these added probabilities over various numbers of parameters is plotted. Experimenters can use this histogram plot as a reference to decide if the experimental run-size they are going to use is appropriate for their prior beliefs.

Example

Consider an 18-run experiment for six factors in which each factor has three levels. When the experimenters' prior beliefs are, for example, $\pi_1 = 1$, $\pi_2 = 1$ and $\pi_3 = 0.5$, we calculate the prior probability of each model being the true model. We then add up the probabilities for models with the same number of parameters. Figure 1 gives the histogram plot of the added probabilities of models with different numbers of parameters, from 1 to 28, being the true model when $\pi_1 = 1$, $\pi_2 = 1$ and $\pi_3 = 0.5$. The value beside the dash line in the plot is the cumulative probability for the models with the number of parameters not greater than a given value being the true model. For example, Figure 1 shows that there is just a 15% chance that a model with no more than 18 parameters will be the true model for this prior. Clearly, an 18-point experiment is not suitable to accommodate such prior. It is better to increase the experimental run-size or to allow for follow-up experiments. By contrast, if the experimenters' prior beliefs are $\pi_1 = 0.8$, $\pi_2 = 0.7$ and $\pi_3 = 0.3$, Figure 2 shows that there is a 97.6% chance that a model with no more than 18 parameters will be the true model. It can be concluded that an 18-point design is appropriate for this prior.

It is clearly seen that this histogram plot can be used a useful tool to help experimenters to decide the appropriate run-sizes for their experiments. If experimenters believe that most factors will have non-negligible effects and models with a larger number of parameters are more likely to be the true model, it is better to increase the experimental run-size or to allow for follow-up

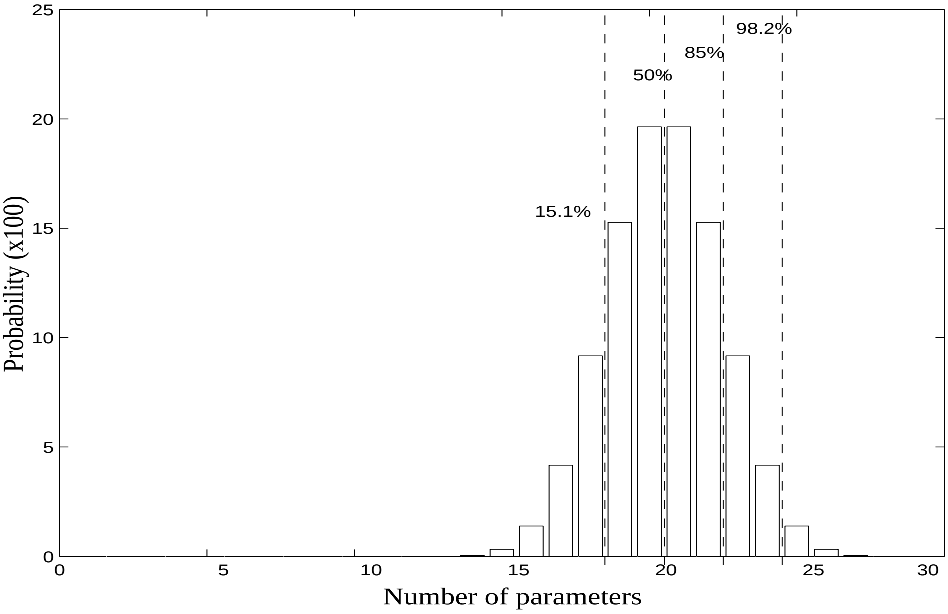

Figure 1: Probabilities for models with various numbers of parameters being the true model for a three-level design for six factors when $\pi_1 = 1$, $\pi_2 = 1$ and $\pi_3 = 0.5$.

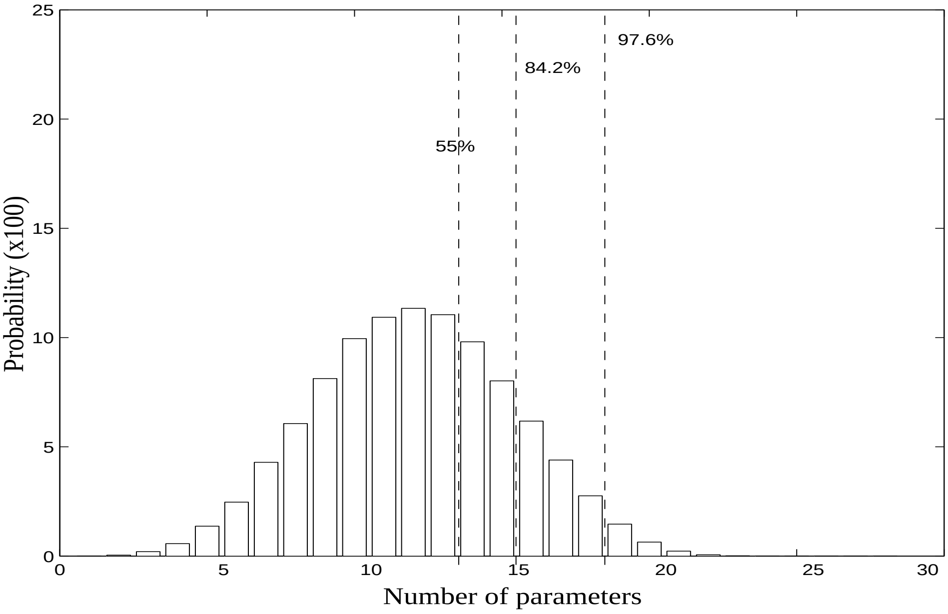

Figure 2: Probabilities for models with various numbers of parameters being the true model for a three-level design for six factors when $\pi_1 = 0.8$, $\pi_2 = 0.7$ and $\pi_3 = 0.3$.

experiments to accommodate their prior belief.

3 Design criterion

To incorporate experimenters' prior belief into design process, we define a new criterion for selecting orthogonal, or nearly orthogonal, main effects designs, with robustness to interactions as a secondary. This new criterion, denoted by $Q_B(\Gamma^{(k)})$, is defined as an average of an approximation to A_s-efficiency, ignoring the intercept, over all eligible models with each model, weighted by the prior probability of the model.

As defined in Section 2, we use $\{\pi_1, \pi_2, \pi_3\}$ to represent, respectively, experimenters' prior probability that a linear effect is in the true model, the prior probability that a quadratic effect is in the true model given that the linear effect of the same factor is in the model, and the prior probability that an interaction effect is in the true model given that the linear effects of both the factors involved are in the model. The probability that model $\mathcal{M}_s$ is the true model for a given prior can then be computed as the joint probabilities of these effects being in the true model, as defined in equation (1).

For a k-factor three-level design, suppose that there are n_0 possible models under functional marginality rule for the two-stage analysis. Among them, models with more than N parameters are not estimable for an N-point design. Those models whose number of parameters is greater than N is called the *non-eligible models*, and they will not be formulated in the second stage of the model-fitting procedure. In the two-stage analysis, one of eligible model will be the final fitted model. Therefore, we define an adjusted probability, denoted by $\tilde{\Pr}(\mathcal{M}_s)$, to represent the probability that "a particular model is *the best of the eligible models*". The adjusted probability is defined as

$$\tilde{\Pr}(\mathcal{M}_s) = \begin{cases} \Pr(\mathcal{M}_s) & \text{if } |\mathcal{M}_s| < N \\ \Pr(\mathcal{M}_s) + \frac{\gamma}{m} & \text{if } |\mathcal{M}_s| = N \\ 0 & \text{if } |\mathcal{M}_s| > N \end{cases} \tag{2}$$

where $|\mathcal{M}_s|$ is the number of parameters in model $\mathcal{M}_s$, γ is the sum of the prior probabilities for all non-eligible models, and m is the number of models with exactly N parameters. Here, we reallocate the probabilities from non-eligible models between all models one size smaller. Clearly, if a model is more likely to be the best model, the model should be given more weight in the design criterion.

Thus, the $Q_B(\Gamma^{(k)})$ criterion can be computed as an average of an approximation to A_s-efficiency, ignoring the intercept, over the eligible models with each model, weighted by the adjusted probability that the model is the best model, which is

$$Q_B(\Gamma^{(k)}) = \sum_{s=1}^{n_0} \sum_{i=1}^{v} \sum_{j=0}^{v} r_{ij} \; M_s(i,j) \; \tilde{\Pr}(\mathcal{M}_s) \tag{3}$$

where $v = 2k + \binom{k}{2}$ is the maximal number of effects of interest for a k-factor three-level design, and n_0 is the number of eligible models. Symbol $M_s(i,j)$ is an indicator representing whether terms i and j are in the model $\mathcal{M}_s$, which is

$$M_s(i,j) = \begin{cases} 1 & \text{if terms } i \text{ and } j \text{ are both in model } \mathcal{M}_s \\ 0 & \text{otherwise} \end{cases}$$

Symbol r_{ij}, for $i, j = 0, \ldots, v$, is defined as

$$r_{ij} = \frac{1}{a_{ii}} \frac{a_{ij}^2}{a_{ii}a_{jj}},$$

where a_{ij} are the elements in the information matrix for the k-factor second order model whose values are determined by the selection of designs. r_{ij} is a kind of measure for the non-orthogonality among effects for the second-order polynomial model of interest.

Let p_{ij} denote the sum of the adjusted probabilities for models containing both terms i and j, which is

$$p_{ij} = \sum_{s=1}^{n_0} \tilde{\Pr}(\mathcal{M}_s) M_s(i,j).$$

Then $Q_B(\Gamma^{(k)})$ criterion defined in (3) can also be written as

$$Q_B(\Gamma^{(k)}) = \sum_{i=1}^{v} \sum_{j=0}^{v} r_{ij}\; p_{ij},$$

Clearly, the $Q_B(\Gamma^{(k)})$-criterion does not require the inversion of the information matrices for the models of interest. It is simple and computationally inexpensive, and therefore it can quickly provide us with information on the worth of a design over a wide range of models.

This $Q_B(\Gamma^{(k)})$-criterion generalised the $Q(\Gamma^{(k)})$-criterion defined by Tsai *et al.* (2000) which weighs all models equally. Here, we study the relationship between this criterion and the the generalised minimum aberration criterion. For simplicity, we compare this criterion with the G_2-aberration criterion proposed by Tang and Deng (1999) for designs with two levels.

Let $\mathbf{X}_d$ be an $N \times k$ matrix of 1's and -1' for an N run (N even) design for k two-level factors, where each column has $N/2$ 1's and $N/2$ -1's corresponding to a factor's main effect and each row represents a factor-level combination. Consequently, the s-factor interactions oare represented by the componentwise product of s columns of $\mathbf{X}_d$. $s = \{1, \ldots, k\}$. Let B_s be the sum of squares all the componentwise products of s-columns of $\mathbf{X}_d$ divided by N^2. The minimum G_2-aberration criterion proposed by Tang and Deng (1999) is to sequentially minimize B_1, B_2, ... , B_k.

For a k-factor two-level design, when the k-factor second-order model is the maximal model of interset, we have all the diagonal entries of the information matrix of the second-order model equal N, i.e., $a_{ii} = N$, for $i = 0, \ldots, k + \binom{k}{2}$. Thus, $r_{ij} = \frac{a_{ij}^2}{N^3}$. Let $F = \binom{k}{2}$, and

$n_0 = n(\delta(0))$ be the number of total models for a k-factor two levels designs, $n_0 = \sum_{i=0}^{k} \binom{k}{i} 2^{\binom{i}{2}}$;

$n_{1,0} = n(\delta(x_i))$ be the total number of models that contain main effect of factor i, $n_{1,0} = \sum_{i=0}^{k-1} \binom{k-1}{i} 2^{\binom{i+1}{2}}$;

$n_{2,0} = n(\delta(x_i), \delta(x_j))$ be the total number of models that contain the main effects of factors i and j, $n_{2,0} = \sum_{i=0}^{k-2} \binom{k-2}{i} 2^{\binom{i+2}{2}}$;

$n_{2,1} = n(\delta(x_i), \delta(x_i x_j))$ be the total number of models that contain main effects of factor i and a two-factor interaction involving the same factor, $n_{2,1} = \sum_{i=0}^{k-2} \binom{k-2}{i} 2^{\binom{i+2}{2}-1}$;

$n_{3,1} = n(\delta(x_i), \delta(x_j x_l))$ be the total number of models that contain the main effect of factor i and a two-factor interaction not involving the same factor, $n_{3,1} = \sum_{i=0}^{k-3} \binom{(k-3)}{i} 2^{\binom{(i+3)}{2}-1}$;

$n_{3,2} = n(\delta(x_i x_j), \delta(x_i x_l))$ be the total number of models that contain two interactions that have one factor in common, $n_{3,2} = \sum_{i=0}^{k-3} \binom{k-3}{i} 2^{\binom{i+3}{2}-2}$;

$n_{4,2} = n(\delta(x_i x_j), \delta(x_l x_m))$ be the total number of models that contain two interactions that have no factor in common, $n_{4,2} = \sum_{i=0}^{k-4} \binom{k-4}{i} 2^{\binom{i+4}{2}-2}$.

Then the $Q(\Gamma^{(k)})$ defined by Tsai *et al.* (2000) can be written as

$$Q(\Gamma^{(k)}) = \frac{1}{n_0 N} \left\{ k n_{1,0} + F n_{2,1} + \left[2n_{2,0} + n_{2,1} + 2(k-2) n_{3,2} \right] B_2 + 6 n_{3,1} B_3 + 6 n_{4,2} B_4 \right\} \tag{4}$$

which is a linear combination of B_2, B_3 and B_4 with decreasing weights. It showns that the G_2-aberration criterion can provide a good surrogate for model robustness.

Consider an alternative model space for a k-factor two-level design with k main effects and f two-factor interactions, where f is a positive integer such that $f \leq F$. Under this model space, we have $n_0 = n_{1,0} = n_{2,0} = \binom{F}{f}$, $n_{2,1} = n_{3,1} = \binom{F-1}{f-1} = \frac{f}{F} n_0$, and $n_{3,2} = n_{4,2} = \binom{F-2}{f-2} = \frac{f(f-1)}{F(F-1)} n_0$. Therefore, the corresponding criterion which averages of an approximation to A_s-efficiency, ignoring the intercept, over the the models with k main effects and f two-factor interactions is as

$$\frac{1}{N} \left\{ k + f + \left[2 + \frac{f}{F} + 2(k-2) \frac{f(f-1)}{F(F-1)} \right] B_2 + 6 \frac{f}{F} B_3 + 6 \frac{f(f-1)}{F(F-1)} B_4 \right\}$$

which is the same as the results in Cheng, Deng and Tang (2002) that the minimum G_2-aberration criterion sequentially minimizing B_1, B_2, . . . , B_k indirectly takes efficiencies into account and provides a kind of overall measure of the partial aliasing and correlation among factorial effects.

The $Q_B(\Gamma^{(k)})$ criterion which generalised the Q criterion with each model weighted by the prior probability of all the models. Therefore, if we partition the the sum of the adjusted probabilities for models containing both terms i and j to be $q_{i,j}$ the sum of the adjusted probabilities for models containing i main terms and j interactions. As done in (4), the $Q_B(\Gamma^{(k)})$ can be written as

$$Q(\Gamma^{(k)}) = \frac{1}{q_0 N} \left\{ kq_{1,0} + Fq_{2,1} + [2q_{2,0} + q_{2,1} + 2(k-2)q_{3,2}]\, B_2 + 6q_{3,1}B_3 + 6q_{4,2}B_4 \right\} \quad (5)$$

which is a linear combination of B_2, B_3 and B_4 where $q_{i,j}$ are determined by the given prior.

There is a conceptual similarity between this Q_B-criterion and the Bayesian D-criterion suggested by DuMouchel and Jones (1994) who proposed a Bayesian modification of D-optimal designs to reduce the dependence of D-optimal designs on the choice of an assumed model for looking at designs that allow the precise estimation of some primary terms while providing detectability for the potential terms. Other useful criteria could be constructed based on similar ideas. The Bayesian D criterion is perhaps more widely applicable than the Q_B criterion, which we recommend only for main effects designs when the robustness to interactions is a secondary consideration. However, this criterion which work with the inverse of a matrix is computationally more expensive and allows us to search over only a few designs, whereas for the specific problem of obtaining main effects designs, the Q_B criterion allows us to consider all possible designs.

4 A case study

Consider an experiment described by Logothetis (1990), who gave a detailed analysis of experimental data from a plasma etching process. The aim of the experiment was to identify the effects of six factors, labelled F_1–F_6, on the etch rate (in 10^{-10}m/min) of the aluminum-silicon layer placed on the surface of an integrated circuit. The six factors, each at three levels, were assigned to the columns of an $L_{18}(3^6)$ array given in Table 1. For each of the 18 experimen-

Table 1: An electronics experiment

	Factor						
Run	F_1	F_2	F_3	F_4	F_5	F_6	*Mean*
1	–	–	–	–	–	–	4900
2	–	–	+	+	0	0	5664
3	–	0	0	0	0	0	5777
4	–	0	–	–	+	+	6224
5	–	+	+	+	+	+	9195
6	–	+	0	0	–	–	5569
7	0	–	–	0	0	+	12810
8	0	–	0	+	–	+	5391
9	0	0	0	+	+	–	9591
10	0	0	+	–	0	–	5074
11	0	+	+	–	–	0	5565
12	0	+	–	0	+	0	5203
13	+	–	+	0	+	–	4999
14	+	–	0	–	+	0	8540
15	+	0	–	+	–	0	10750
16	+	0	+	0	–	+	12143
17	+	+	–	+	0	–	6024
18	+	+	0	–	0	+	8806

tal runs, two wafers were tested simultaneously at two levels of over-etch time. Finally, the etch rate was measured at five fixed locations on each wafer. Table 1 gives the mean etch rate across the 10 observations for each of the 18 experimental runs. As analysed by Tsai, Gilmour and Mead (1996), the final model selected for the interpretation of the data is a model which contains linear effects of F_1, F_2, F_5 and F_6 and the linear by linear interaction of F_1 and F_5. Logothetis (1990), Fearn (1992), Tsai *et al.* (1996) and Cox and Reid (2000) discussed other aspects of the analysis of these data.

Table 3 lists some of the best designs for six factors in 18 runs generated by the procedure in Tsai *et al.* (2000). It has been shown that, by using their procedure, more designs are generated. All the designs can all be generated by their procedure are equally efficient for fitting the main effects model and any sub-model of it. Nevertheless, for the two-stage analysis that considers interactions in addition to main effects, these designs are different. The best 18-run designs with six factors that they generated have better projection properties and are more efficient than those obtained by the procedure of Cheng and Ye (2004) and those from the $L_{18}(3^7)$. Design 25 in Table 3 is the preferred design with 6 factors in Cheng and Ye (2004). L18_2 is one of the two six-factor designs that is obtained from $L_{18}(3^7)$. The other is in Table 1 and will be referred to as L18_1 in Table 3. Tsai *et al.* (2004) noted that the designs studied in Cheng and Ye (2004) and obtained from the $L_{18}(3^7)$ always project onto designs for three factors that are formed by putting two regular 3^{3-1} factorials together. Design 11 in Table 3 is the best design generated by Tsai *et al.* (2000) that has this three-factor projection property.

We use $Q_B(\Gamma^{(6)})$ to compare these orthogonal main effects designs for various priors. Table 3 gives the corresponding values of $Q_B(\Gamma^{(6)})$ for the designs given in Table 2. The boldface for $Q_B(\Gamma^{(6)})$ indicates that the design has the lowest $Q_B(\Gamma^{(6)})$ for the particular prior. Clearly, depending on the prior knowledge, slightly more appropriate designs can be found, and the designs chosen by the modified criterion are clearly better than the one actually used and are better than the best design in Cheng and Ye (2004) under several priors. In general, designs with small values of $Q(\Gamma^{(k)})$ tend to have small values of $Q_B(\Gamma^{(k)})$. For picking up the appropriate class of designs, exact specification of prior probabilities is not crucial. The best designs constructed in this way are very robust to prior specifications, so experimenters can have great confidence in using them. Nonetheless, they can use their prior probabilities to select a specific best design for any particular experiment.

5 Conclusion

In this paper, we studied a good way to assess orthogonal main effects designs, with robustness to interactions as a secondary consideration when experimenters have some prior information about the probability of effects being non-negligible in advance of collecting their data. In this paper, we used the $Q_B(\Gamma^{(k)})$-criterion to compare all the 440 three-level main effects designs

Table 2: Some orthogonal main effects designs for six factors in 18 runs

Design 1	Design 2	Design 3	Design 4	Design 5
– – – – – 0	– – – – 0 –	– – – – – 0	– – – – 0 –	– – – – 0 0
– – 0 + + –	– – 0 + – 0	– – 0 + 0 –	– – 0 0 – +	– – 0 0 – –
– 0 – 0 0 –	– 0 – 0 – +	– 0 – 0 + –	– 0 – + + +	– 0 – + + +
– 0 + – + +	– 0 + – + 0	– 0 + – 0 +	– 0 + + – 0	– 0 + + – 0
– + 0 0 – +	– + 0 0 + –	– + 0 0 – +	– + 0 – + 0	– + 0 – + –
– + + + 0 0	– + + + 0 +	– + + + + 0	– + + 0 0 –	– + + 0 0 +
0 – + – 0 –	0 – 0 0 + +	0 – 0 0 0 0	0 – + – + 0	0 – + – + +
0 – + 0 – +	0 – + – – +	0 – + – + –	0 – + + 0 +	0 – + + 0 –
0 0 – + + +	0 0 – + + 0	0 0 – + 0 +	0 0 – 0 + –	0 0 – – – –
0 0 0 0 0 0	0 0 + 0 0 –	0 0 + 0 – 0	0 0 0 0 0 0	0 0 0 0 0 0
0 + – + – –	0 + – + – –	0 + – + – –	0 + – – – +	0 + – 0 + 0
0 + 0 – + 0	0 + 0 – 0 0	0 + 0 – + +	0 + 0 + – –	0 + 0 + – +
+ – – 0 + 0	+ – – 0 0 0	+ – – 0 + +	+ – – 0 – 0	+ – – 0 – +
+ – 0 + 0 +	+ – + + + –	+ – + + – +	+ – 0 + + –	+ – 0 + + 0
+ 0 0 – – –	+ 0 0 – – –	+ 0 0 – – –	+ 0 0 – 0 +	+ 0 0 – 0 +
+ 0 + + – 0	+ 0 0 + 0 +	+ 0 0 + + 0	+ 0 + – – –	+ 0 + 0 + –
+ + – – 0 +	+ + – – + +	+ + – – 0 0	+ + – + 0 0	+ + – + 0 –
+ + + 0 + –	+ + + 0 – 0	+ + + 0 0 –	+ + + 0 + +	+ + + – – 0

Design 8	Design 11	Design 25	L18_2
– – – – 0 –	– – – – – –	– – – – – –	– – – – – –
– – 0 + – +	– – + 0 + 0	– – + 0 0 +	– – 0 0 + +
– 0 – 0 – 0	– 0 – + + +	– 0 – + 0 0	– 0 – + 0 +
– 0 + – + +	– 0 0 – 0 0	– 0 0 – + +	– 0 + – + 0
– + 0 0 + –	– + 0 0 – +	– + 0 0 – 0	– + 0 + – 0
– + + + 0 0	– + + + 0 –	– + + + + –	– + + 0 0 –
0 – 0 0 + 0	0 – 0 0 0 –	0 – 0 0 + –	0 – – + + 0
0 – + – – 0	0 – + + – +	0 – + + – 0	0 – + – 0 +
0 0 – + + +	0 0 – 0 – 0	0 0 – 0 – +	0 0 0 0 0 0
0 0 + 0 0 –	0 0 + – + –	0 0 + – 0 –	0 0 + + – –
0 + – + – –	0 + – – 0 +	0 + – – + 0	0 + – 0 – +
0 + 0 – 0 +	0 + 0 + + 0	0 + 0 + 0 +	0 + 0 – + –
+ – – 0 0 +	+ – – + 0 0	+ – – + + +	+ – 0 + 0 –
+ – + + + –	+ – 0 – + +	+ – 0 – 0 0	+ – + 0 – 0
+ 0 0 – – –	+ 0 0 + – –	+ 0 0 + – –	+ 0 – 0 + –
+ 0 0 + 0 0	+ 0 + 0 0 +	+ 0 + 0 + 0	+ 0 0 – – +
+ + – – + 0	+ + – 0 + –	+ + – 0 0 –	+ + – – 0 0
+ + + 0 – +	+ + + – – 0	+ + + – – +	+ + + + + +

Table 3: Comparison of $L_{18}(3^6)$ designs under several priors

$Q_B(\Gamma^{(6)})$	Design									
(π_1, π_2, π_3)	D1	D2	D3	D4	D5	D8	D11	D25	L18_1	L18_2
(1, 1, 0.5)	**272.15**	273.60	272.97	276.06	273.09	276.05	276.76	282.57	293.11	293.12
(1, 1, 0.2)	192.90	191.41	193.06	194.14	192.13	**191.35**	192.63	195.39	201.00	194.24
(0.8, 0.9, 0.2)	130.75	129.79	130.81	131.20	130.50	**129.79**	130.54	131.51	134.90	131.21
(0.8, 0.9, 0.3)	152.72	**151.79**	152.86	153.67	152.33	151.90	152.78	154.71	159.01	154.65
(0.8, 0.8, 0.6)	200.62	200.72	201.04	203.00	**200.31**	201.52	202.28	206.56	212.34	209.26
(0.8, 0.7, 0.3)	139.59	**138.64**	139.72	140.20	139.77	139.05	139.95	140.84	146.13	142.63
(0.8, 0.6, 0.6)	**189.97**	190.13	190.39	192.01	190.34	191.30	192.04	195.16	202.04	200.13
(0.8, 0.5, 0.7)	**199.35**	199.96	199.84	201.69	199.97	201.48	202.09	205.52	212.83	212.12
(0.7, 0.7, 0.4)	128.19	**127.53**	128.33	128.89	128.23	127.86	128.57	129.71	133.92	131.03
(0.6, 0.9, 0.5)	120.04	119.54	120.15	120.87	119.56	**119.53**	120.06	121.87	124.23	121.24
(0.67, 0.8, 0.7)	135.38	135.12	135.57	136.63	**134.87**	135.27	135.80	138.33	141.14	138.30
(0.67, 0.7, 0.5)	132.79	**132.29**	132.95	133.71	132.71	132.63	133.29	134.91	138.89	136.13

with six factors in 18 runs generated by using the procedure by Tsai *et al.* (2000). We showed that, depending on the prior knowledge, more appropriate designs can be found among the class of good designs. The best designs constructed in this way are very robust to prior specifications, so experimenters can have more confidence in using them.

In this paper, the $Q_B(\Gamma^{(k)})$-criterion is used as a secondary consideration for orthogonal main effects designs. Clearly, it can be used in exactly the same way for nearly orthogonal designs. However, there is a question about how far we can go towards expecting large interactions. If their prior probability becomes larger than that for quadratic effects, then we should be using a different class of designs from orthogonal main effects plans. If this is the case, we could use the Q_B criterion as a primary, rather than a secondary criterion, and we could use this criterion in an exchange algorithm to construct designs in a way similar to Bayesian D-optimal designs. However, the objectives for such designs will be different from those given in this paper and we have not explored this possibility.

References

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

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

Cheng, C. S., Deng, L. W. and Tang, B. (2002) Generalized minimum aberration and design efficiency for nonregular fractional factorial designs. *Statistica Sinica*. **12**, 991-1000.

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

Cheng, S. W. and Ye, K. Q. (2004) Geometric isomorphism and minimum aberration for factorial designs with quantitative factors. To appear in *The Annals of Statistics*.

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

Chipman, H., Hamada, M. and Wu, C. F. J. (1997) A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics*, **39**, 372-381.

Cox, D. R. and Reid, N. (2000) *The Theory of the Design of Experiments.* London: Chapman & Hall.

DuMouchel, W. and Jones, B. (1994) A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model. *Technometrics*, **36**, 37-47.

Fearn, T. (1992) Box-Cox transformations and the Taguchi method: an alternative analysis of a Taguchi case study. *Applied Statistics*, **41**, 553-559.

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

Logothetis, N. (1990) Box-Cox transformations and the Taguchi method. *Applied Statistics*, **39**, 31-38.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edition. London: Chapman & Hall.

Nelder, J. A. (1997) Letter to the Editors: Functional marginality is important. *Applied Statistics*, **46**. 281-282.

Nelder, J. A. (1998) The selection of terms in response-surface models: how strong is the weak-heredity principle? *American Statistician*, **52**, 315-318.

Tang, B. and Deng, L. Y. (1999) Minimum G_2-aberration for non-regular fractional factorial designs. *The Annals of Statistics*, **27**, 1914-1926.

Tsai, P.-W., Gilmour, S. G. and Mead, R. (1996) Letter to the Editors: An alternative analysis of Logothetis's plasma-etching data. *Applied Statistics*, **45**, 498-503.

Tsai, P.-W., Gilmour, S. G. and Mead, R. (2000) Projective three-level main effects designs robust to model uncertainty. *Biometrika*, **87**, 467-475.

Tsai, P.-W., Gilmour, S. G. and Mead, R. (2004) Some new three-level orthogonal main effects designs robust to model uncertainty. To appear in *Statistica Sinica*.

Wang, J. C. and Wu, C. F. J. (1995) A hidden projection property of Plackett-Burman and related designs. *Statistica Sinica*, **5**, 235-250.

Appendix A

Chipman (1996) defined the prior for a term of a given order being active is conditional on all terms of its lower order such as

$$\Pr\,(x_i \text{ is active}) = p$$

$$\Pr\left(x_i^2 \text{ is active} \mid \delta(x_i)\right) = \begin{cases} p_0 & \text{if } \delta(x_i) = 0 \\ p_1 & \text{if } \delta(x_i) = 1 \end{cases}$$

$$\Pr\,(x_i x_j \text{ is active} \mid \delta(x_i), \delta(x_j)) = \begin{cases} p_{00} & \text{if } \delta(x_i) = \delta(x_j) = 0 \\ p_{01} & \text{if } \delta(x_i) = 0 \text{ and } \delta(x_j) = 1 \\ & \text{or } \delta(x_i) = 1 \text{ and } \delta(x_j) = 0 \\ p_{11} & \text{if } \delta(x_i) = \delta(x_j) = 1 \end{cases}$$

In our definition, π_1 is defined as the probability that a linear effect x_i is in the true model, which is the probability that term x_i, or any effect to which it is marginal, is large. This is the joint probability of the following events such that x_i is active and x_i^2 is active and any $x_i x_j$ interaction, for $j \neq i$ and $j = 1, \ldots, k$, is active, which can be written as

$$\begin{aligned} \pi_1 &= p + (1-p)p_0 + (1-p)(1-p_0)[p_{01}p + p_{00}(1-p)] + \\ &\quad (1-p)(1-p_0)\{1 - [p_{01}p + p_{00}(1-p)]\}[p_{01}p + p_{00}(1-p)] + \\ &\quad \cdots + (1-p)(1-p_0)\{1 - [p_{01}p + p_{00}(1-p)]\}^{k-2}[p_{01}p + p_{00}(1-p)] \\ &= p + (1-p)p_0 + (1-p)(1-p_0)\left\{1 - \{1 - [p_{01}p + p_{00}(1-p)]\}^{k-1}\right\} \end{aligned}$$

π_2 is defined as the probability that a quartic effect is in the true model given that the linear effect of the same factor is in to model, so that

$$\pi_2 = p_1 \tag{6}$$

Similarly, π_3 is defined as the probability that an interaction effect is in the true model given that the linear effects of both the factors involved are in to model, so that

$$\pi_3 = p_{11} \tag{7}$$

Under the strong heredity principle defined by Chipman (1996) that the belief that for an interaction to be active both corresponding main effects must also be active, we have $p_0 = p_{00} = p_{01} = 0$. Therefore,

$$\pi_1 = p. \tag{8}$$

Under the weak heredity principle that the belief that for an interaction to be active one of the corresponding main effects must also be active, we have $p_{00} = 0$. Thus

$$\pi_1 = p + (1-p)p_0 + (1-p)(1-p_0)\left\{1 - [1 - p_{01}p]^{k-1}\right\} \tag{9}$$