

# A General Criterion for Factorial Designs Under Model Uncertainty

Pi-Wen Tsai

Department of Mathematics  
National Taiwan Normal University  
Taipei 116, Taiwan, R.O.C.  
E-mail: pwtsai@math.ntnu.edu.tw

Steven G. Gilmour

Queen Mary University of London  
School of Mathematical Sciences  
Mile End Road  
London E1 4NS, UK  
E-mail: s.g.gilmour@qmul.ac.uk

## Abstract

Motivated by two industrial experiments in which rather extreme prior knowledge was used to choose the design, we show that the  $Q_B$  criterion, which aims to improve the estimation in as many models as possible by incorporating experimenters' prior knowledge along with an approximation to the  $A_s$  criterion, is more general and has a better statistical interpretation than many standard criteria. The generalization and application of the criterion to different types of designs are presented. The relationships between  $Q_B$  and other criteria for different situations are explored. It is shown that the  $E(s^2)$  criterion is a special case of  $Q_B$  and several aberration-type criteria are limiting cases of our criterion, so that  $Q_B$  provides a bridge between alphabetic optimality and aberration. The two case studies illustrate the potential benefits of the  $Q_B$  criterion. R programs for calculating  $Q_B$  are available online as supplemental materials.

**Keywords:** aberration, generalized minimum aberration, model robust, design optimality criterion, projection efficiency, supersaturated design.

# 1 INTRODUCTION

Many experiments in industrial research and development involve studying the effects of several factors on one or more responses. Multifactor designs of various types form one of the main contributions of statistics to industrial research. The broad class of multifactor designs in common use includes response surface designs, regular and irregular fractional factorials with two or three levels and supersaturated designs. The different types of design are defined largely by the models expected to be useful for analyzing the resulting data. The models might be polynomial models for quantitative factors, or involve factorial effects for qualitative factors. The appropriate polynomial or factorial model might be of first, second or higher order, or of mixed order, e.g. including some, but not all, second order terms.

Another important distinction that can be made is between designs for situations in which the model is known in advance and those for situations of model uncertainty. In practice, the model is almost never known for certain in advance, but “the model is known” should be interpreted as meaning that, when designing the experiment, only one model is considered. Designs for a single model are often based on optimal design criteria, such as  $D$ -optimality or  $A$ -optimality. These criteria are very general in that they can be used whether the factors are qualitative or quantitative, however many levels the factors have and whatever the order of the model. These so-called alphabetic optimality criteria and their applications are well summarized by Atkinson *et al.* (2007). Extensions to situations of model uncertainty have been developed, e.g. Läuter (1974) suggested a weighted average, over several models, of a standard criterion and Zhao *et al.* (2003) gave this a Bayesian interpretation when using the  $A$  criterion. Heredia-Langner *et al.* (2004) developed an algorithm for constructing designs which are optimal averaged across several candidate models, while Ozol-Godfrey *et al.* (2005) studied the prediction variance properties of designs under different models. However, these papers concentrate on small numbers of alternative models, rather than the hundreds of models which can be considered for multifactor designs.

In the case of design under model uncertainty, different criteria are used in different situations. Two-level supersaturated designs are chosen using the  $E(s^2)$  criterion, three-level supersaturated designs using  $\text{Ave}(\chi^2)$ , regular fractional factorials using resolution and aberration, irregular fractions using various generalized aberration criteria and response

surface designs using near-orthogonality. Sometimes described as model-free, these criteria consider the estimation of main effects and interactions or polynomial effects of different orders and so implicitly assume that the model to be fitted will be one of the many possible submodels of the model containing all of these factorial effects.

In this paper we study a general criterion that can be used in any of the above situations for design under model uncertainty. The  $Q_B$  criterion requires the definition of a maximal model of interest (which need not be estimable from the experiment), and we assume that the models fitted will be submodels of the maximal model. The criterion allows different prior weights for each possible model, so that experimenters' belief about the model are incorporated into the design selection process.

In Section 2 we motivate our work through two case studies, which are superficially similar, although very different designs were used. The  $Q_B$  criterion is defined in Section 3 and some notation for the *generalized word count* which represents an overall measure of the partial aliasing of the factorial effects, is introduced. In Sections 4 to 6, the use of the  $Q_B$  criterion in different situations and its relationship with other criteria for these situations are studied. In Section 7, the relationship with the weighted- $A$  criterion for regular fractions is explored. In Section 8 we discuss alternative ways in which the case studies could have been designed. Finally, in Section 9, our overall conclusions are discussed, along with the limitations of the  $Q_B$  criterion.

## 2 CASE STUDIES

The motivation for the work described here comes from discussions with industrial statisticians and the experimenters they support. It is clear that, although the hierarchical ordering of effects in factorial structures is usually sensible, in most experiments any statement that effects up to order  $z$  are of interest, or effects of order higher than  $z$  can be assumed negligible, are at best very crude simplifications of the real prior knowledge. What is needed is not a discrete collection of design-types, described by Bailey (1997) as “butterfly collecting”, but a unified theory of how to choose multifactor designs in the face of model uncertainty. Since most industrial experiments we are directly familiar with are not in the public domain, we will use two published case studies to make our discussions more concrete.

Logothetis (1990) gave a detailed analysis of experimental data from a plasma etching process. The aim of the experiment was to identify the effects of six quantitative factors, the pressure in the reaction chamber, the radio frequency power level, the temperature of the lower electrode on which the wafer is supported, the flow rate of boron trichloride, the flow rate of chlorine and the flow rate of silicon tetrachloride, on three response variables, the etch rate of the aluminum-silicon alloy layer, the photoresist etch rate and the oxide etch rate. Three levels of each factor were used, so that nonlinearity could be assessed, and an 18-run orthogonal main effects design was chosen, based on the assumption that there would be no interactions. Logothetis (1990) stated that the “levels for the factors were chosen by the engineers so that the possibility of existence of strong interaction effects was minimized. This was based on previous evidence.” This is a very strong prior assumption and to base complete faith in the absence of interactions seems unwise. Fearn (1992), Tsai *et al.* (1996) and Cox and Reid (2000) reported further analyses of these data and all of these authors included at least one interaction in their best-fitting models. In this paper we show how the prior belief that interactions were unlikely could have been used, without going to the extreme of excluding the possibility of their existence. All that is required is a prior probability of them being active.

An experiment run at Ford Motor Company Ltd. was described by Grove and Davis (1992). The aim was to learn how to improve the quality of the welding of two halves of a fuel tank. The effects of six quantitative factors, weld time, cool time, weld current, cradle speed, wheel condition and wheel clamping pressure, on two responses, the minimum seam width and the strength of the weld, were to be studied. It was decided that the factors would be run at two levels each and that, in addition to the main effects, the interactions of cool time with cradle speed, cradle speed with wheel condition and weld current with wheel condition should be estimated. A sixteen-run regular fractional factorial design was used. Once again, the assumption that all interactions except those declared are zero seems very extreme. When the data were analysed, the second largest estimated effect on the mean strength and the largest on the strength dispersion were both from contrasts which measured interactions other than those expected. In this experiment, one could also question whether the restriction to two levels of each factor was ideal. In this paper, we show how specifying less extreme prior information allows the experimenters’ knowledge and experience to be used, but the limits of that knowledge to also be recognized.

These two experiments are actually rather similar, in that they have six factors and similar numbers of runs, but the designs used were very different in the two cases. While this was, quite correctly, based on the engineers' prior knowledge, we believe that we ought to be more skeptical about such knowledge, or at least to formally recognize the limits of that knowledge. Experimenters have sometimes expressed to us the view that factorial designs "just tell us what we already know". Asking for what is known before the experiment to be clearly stated in probabilistic terms allows this view to be challenged more convincingly. It also allows designs to be tailored more precisely to finding out what is not known.

### 3 METHODOLOGICAL BACKGROUND

#### 3.1 The $Q_B$ Criterion

To explore the projection efficiency of a design under model uncertainty, Tsai *et al.* (2007) introduced the  $Q_B$  criterion to compare three-level main-effects designs for quantitative factors that allow the consideration of interactions in addition to main effects. The criterion takes into account experimenters' prior probabilities of effects being non-negligible in advance of collecting their data.

Let  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$  denote a maximal model of interest, which need not be estimable from the experiment, where  $\mathbf{y}$  is the  $n \times 1$  vector of responses,  $\mathbf{X}$  is the  $n \times (v + 1)$  design matrix,  $\boldsymbol{\beta}$  is the  $(v + 1) \times 1$  vector of unknown parameters and  $\boldsymbol{\epsilon}$  is the error vector, with  $E(\boldsymbol{\epsilon}) = \mathbf{0}$  and  $V(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$ . We assume that we will end up fitting a submodel of the maximal model, i.e. one containing a subset of the parameters  $\boldsymbol{\beta}$ , but we do not in advance know which one. Here, the marginality principle (McCullagh and Nelder (1989)) is used to define the class of *eligible models* that might be fitted. Marginality means that every term in the model must be accompanied by all terms marginal to it, whether their estimates are large or small, e.g. a two-factor interaction cannot appear without the main effects of both factors involved and a quadratic effect cannot appear without the corresponding linear effect.

The  $Q_B$  criterion is to minimize the weighted average of an approximation to the  $A_s$ -criterion,

$$A_s = \sum_{i=1}^v \{(\mathbf{X}'_s \mathbf{X}_s)^{-1}\}_{ii},$$

over all the eligible models, defined by design matrices  $\mathbf{X}_s$  and with weights defined by the probability of each model being the best model. In the common case that effects of the

same type have the same probability of being in the best model, then the prior probability of each model being the best can be easily calculated by obtaining experimenters' prior beliefs about each individual effect being in the best model.

Let  $a_{ij}$ , for  $i, j = 0, \dots, v$ , be the  $(i, j)$ th entry of the information matrix  $\mathbf{A} = \mathbf{X}'\mathbf{X}$ , for the maximal model of interest. Tsai *et al.* (2007) replaced the  $A_s$  criterion by a simple first-order approximation and defined the  $Q_B$  criterion as

$$Q_B = \sum_{i=1}^v \sum_{j=0}^v \frac{1}{a_{ii}} \frac{a_{ij}^2}{a_{ii}a_{jj}} p_{ij}, \quad (1)$$

where  $p_{ij}$  is the sum, over models in which terms  $i$  and  $j$  are both included, of the prior probabilities of a model being the best. Minimizing  $Q_B$  gives low variances for parameter estimates across a range of models, but is inexpensive computationally. When all eligible models have the same prior probability, the reciprocal of the total number of eligible models,  $Q_B$  reduces to the  $Q$  criterion defined earlier by Tsai *et al.* (2000) for three-level main effects designs.

The definition of the  $Q_B$  criterion originated from finding good three-level main-effects designs for quantitative factors. Depending on the prior knowledge of effects being non-negligible, different designs will be chosen among the class of main effects designs to ensure robustness to the two-factor interactions. However,  $Q_B$  can be applied to measure the performance of a design under any given maximal model and so can be used in different design problems under model uncertainty. For example, if the prior knowledge is such that large interactions are expected, then a broader class of designs than main effects plans should be considered.

We note that the  $Q_B$  criterion is based on the  $A_s$  criterion, so that different scales of the contrast functions for the effects in the model will produce different values of  $Q_B$  and different design orderings. In this paper we specify some particular scales for the effects, details of which are given in Appendix A.

The  $Q_B$  criterion has both similarities to and differences from the Bayesian  $D$ , or  $D_B$ , criterion of DuMouchel and Jones (1994). Both use the concept of a maximal model, but whereas  $D_B$  requires the parameters of the model to be split into only two groups, primary and potential terms, on the basis of prior beliefs,  $Q_B$  allows different prior beliefs to be put on each parameter in the model. The aims are also different,  $D_B$  focusing mainly on estimating the primary terms, but without biases caused by the potential terms, while  $Q_B$

tries to allow estimation of many different models.  $Q_B$  will sometimes prefer a Resolution-III design over a Resolution-IV design, whereas  $D_B$  will not, at least with the usual priors. Ultimately the choice between them will depend on how experimenters think about their prior knowledge and how they want the experiment to increase it. We will see that  $Q_B$  is closely related to several other commonly used criteria; no such relationships exist with  $D_B$ . Tsai *et al.* (2007) also noted that  $Q_B$  is less computationally expensive, so that it is sometimes possible to do a complete search when  $D_B$  requires an exchange algorithm.

A clear division has grown up between research using alphabetic optimality and that using aberration-type criteria. The  $Q_B$  criterion was developed from the alphabetic optimality viewpoint, but using a computationally cheap approximation to the  $A_s$  criterion in order to consider many models. By showing how this criterion is closely related to the aberration-type criteria, we provide a bridge between the two strands of research. This criterion possesses the advantages of alphabetic optimality of being statistically meaningful and applicable to very general situations, and the advantage of aberration-type criteria of being able to consider many different candidate models.

### 3.2 Definitions and Notation

Making the effect hierarchy assumption that lower order effects are more important than higher order effects, Fries and Hunter (1980) introduced minimum *aberration* as a refinement of maximum resolution in searching for a regular factorial design. Let  $W_i(d)$  denote the number of  $i$ -factor interactions appearing in the defining contrast subgroup of a regular fractional factorial design  $d$ . Given two regular  $s^{m-l}$  designs, say  $d_1$  and  $d_2$ ,  $d_1$  is said to have lower aberration than  $d_2$  if there exists an integer  $r$ ,  $1 \leq r \leq m$ , such that  $W_r(d_1) < W_r(d_2)$  and  $W_j(d_1) = W_j(d_2)$  for  $j = 1, \dots, r - 1$ . A factorial design is said to have minimum aberration if no other factorial design has lower aberration. See, for example, Fang and Mukerjee (2000); Butler (2003); Jacroux (2004); Cheng and Tang (2005) for recent research on finding minimum aberration regular factorial designs.

Recently, various generalized aberration criteria have been defined for irregular fractional replicates in different situations. For example, Tang and Deng (1999) proposed the  $G_2$ -aberration criterion for comparing irregular two-level designs, Xu and Wu (2001) and, independently, Ma and Fang (2001) suggested an ANOVA-type generalized aberration criterion for asymmetrical factorial designs and Cheng and Ye (2004) defined the  $\beta$ -aberration

criterion for designs with quantitative factors.

In the context of model robustness, particularly for the situation where the models of interest contain all the main effects and a selection of two-factor interactions, Cheng *et al.* (1999) defined *estimation capacity* in the context of regular designs and showed that the minimum aberration criterion gives a good surrogate for model robustness. Li and Nachtsheim (2000) proposed *information capacity* for comparing so-called *model-robust factorial designs*. Cheng *et al.* (2002) showed that the generalized minimum aberration criterion indirectly takes efficiencies into account and tends to provide a good surrogate for model robustness for irregular two-level designs when a few two-factor interactions are of interest.

In this section, we introduce a concept, the *generalized word count*, which can be useful in many different design problems. For an  $n$ -run design with  $m$  factors, each with  $p + 1$  levels, let  $\mathbf{D}^{(z)}$  be the  $n \times m$  matrix in which the columns correspond to the  $z$ th order polynomial effects of individual factors,  $z = 1, \dots, p$ . Let  $\mathbf{v}_S^{(z)}$  be an  $n \times 1$  vector for the componentwise product of the  $c_1$ th,  $\dots$ ,  $c_s$ th columns of  $\mathbf{D}^{(z)}$ , where  $S = \{c_1, \dots, c_s\}$  is a subset of  $\{1, \dots, m\}$ ,  $1 \leq s \leq m$ . For  $S = \emptyset$ ,  $\mathbf{v}_\emptyset^{(z)}$  is defined as a vector of 1s.

Let  $j(R_1, \dots, R_p)$  denote a measure of aliasing (or partial aliasing) between the intercept and a *polynomial effect* with factors  $R_z = \{c_1, \dots, c_{r_z}\}$  appearing as  $z$ th order effects, where  $r_z = |R_z|$ ,  $|\cdot|$  denotes the number of elements in the argument set and  $k = \sum_{z=1}^p r_z$ . This can be computed by the sum of the elements of the componentwise product of  $\mathbf{v}_{R_1}^{(1)}$ ,  $\mathbf{v}_{R_2}^{(2)}$ ,  $\dots$  and  $\mathbf{v}_{R_p}^{(p)}$ .  $j(R_1, \dots, R_p) = 0$  when the effect is orthogonal to the intercept, and  $j(R_1, \dots, R_p) = n$  when the effect is completely aliased with the intercept.

We further define  $b_k(i_1, \dots, i_p)$  to be the overall aliasing between the intercept and all the sets of  $k$ -factor effects with  $i_z$  factors appearing as  $z$ th order effects. This is the sum of squared  $j(R_1, \dots, R_p)$  divided by  $n^2$ , where the sum is taken over all mutually disjoint  $R$ 's with  $|R_z| = i_z$ , and  $\sum_{z=1}^p i_z = k$ . This is formulated as

$$\begin{aligned} b_k(i_1, \dots, i_p) &= \sum_{\mathcal{R}} \{j(R_1, \dots, R_p)\}^2 / n^2 \\ &= \sum_{\mathcal{R}} \left\{ \sum_{g=1}^n \left( \prod_{\mathcal{H}} \mathbf{d}_{zh_1}^{(1)} \dots \mathbf{d}_{zh_p}^{(p)} \right) \right\}^2 / n^2, \end{aligned} \quad (2)$$

where  $\mathcal{R} = \{R_1, \dots, R_p : |R_1| = i_1, \dots, |R_p| = i_p\}$ ,  $\mathcal{H} = \{h_1, \dots, h_p : h_1 \in R_1, \dots, h_p \in R_p\}$ ,  $\mathbf{d}_{zh_s}^{(z)}$  is the  $(z, h_s)$ th column of  $\mathbf{D}^{(z)}$ ,  $1 \leq k \leq m$  and  $\sum_{z=1}^p i_z = k$ . This is called

the *generalized word count* for words referring to the effects of  $k$  factors where  $i_z$  factors appear as  $z$ th order effects. It measures the partial aliasing of all the effects with  $k$  factors having the same number of factors at a particular order. If all these effects are orthogonal to the intercept, then the generalized word count for words referring to these effects is 0. For designs with two-level factors, i.e.  $p = 1$ , the definition of  $j(R)$  is the same as the  $J$ -characteristics defined by Tang and Deng (1999). Therefore the generalized word count,  $b_k(k)$ , defined above is the same as the  $B_k$  defined in their paper for irregular designs in that context. It is a measure of the partial aliasing between the intercept and the interactions with  $k$  factors. For regular factorial designs, the generalized word count is the same as the usual word count. For three-level designs in a second order polynomial model,  $b_k(i_1, i_2)$  is a measure of the partial aliasing between the intercept and the interactions with  $k$  factors where  $i_1$  factors appear as linear effects and  $i_2$  factors appear as quadratic effects, where  $k = i_1 + i_2$ .

## 4 FIRST ORDER MODELS

First order maximal models are used for screening experiments when the goal is to identify a few important main effects from a large number of potential factors under investigation. The designs might be nearly saturated, saturated or supersaturated, i.e. there might be a few extra degrees of freedom after fitting main effects, just enough to fit all main effects or not enough to fit all main effects. As noted in Appendix A, qualitative factors are easily dealt with by giving equal weights in the  $Q_B$  criterion to all main effects, i.e. linear main effects, quadratic main effects and so on.

### 4.1 Two-Level Factors

For an  $n$ -run design with  $m$  factors, each with two levels, let the first-order model be the maximal model of interest, so that  $v = m$ . The  $Q_B$  criterion in (1) selects a two-level design that minimizes

$$\sum_{i=1}^m p_{i0} \frac{a_{i0}^2}{n^2} + \sum_{i=1}^m \sum_{\substack{j=1 \\ j \neq i}}^m p_{ij} \frac{a_{ij}^2}{n^2},$$

where  $p_{ij}$  is the sum of the probabilities of a model being the best model, the sum being taken over models containing both terms  $i$  and  $j$ , and is determined by the given prior for

each effect being in the model. In the case that each factor has the same probability of being included in the best model, each model containing a particular number of factors has the same prior probability of being the best model. Then  $p_{i_1 0}$ , the sum of probabilities of a model being the best model when the sum is taken over models containing the  $i_1$ th factor, is equal to  $p_{i_2 0}$ , the sum for models containing the  $i_2$ th factor, for  $i_1 \neq i_2$  and  $i_1, i_2 = 1, \dots, m$ ; and  $p_{i_1 i_2}$ , the sum of probabilities of a model being the best when the sum is taken over models containing both the  $i_1$ th and  $i_2$ th factors is equal to  $p_{i_3 i_4}$  the sum for models containing the  $i_3$ th and  $i_4$ th factors,  $i_1 \neq i_2 \neq i_3 \neq i_4$ . We let  $p_{i0} = \xi_1$ ,  $\forall i$  for the sum of probabilities of models containing the main effect of the  $i$ th factor and  $p_{ij} = \xi_2$ ,  $\forall i, j$ ,  $i \neq j$  for those containing the main effects of the  $i$ th and  $j$ th factors.

We note that for two level factors,  $a_{ii} = n$  for all  $i$ . By writing out the elements  $a_{ij}$  and carefully collecting them, we note that the sum of  $a_{i0}^2/n^2$ , for all  $i$ , is the sum of the aliasing between a factor's main effect and the intercept. This is equal to  $b_1(1)$ , the generalized word count for words referring to the effects of one factor as defined in (2) with  $p = 1$  and  $k = 1$ . Similarly, the sum of  $a_{ij}^2/n^2$ , for all  $i \neq j$ , is the sum of the aliasing between a pair of factors' main effects, which is equal to  $2b_2(2)$ , the generalized word count for the effects of two factors.

Then the  $Q_B$  criterion is equal to

$$\xi_1 b_1(1) + 2\xi_2 b_2(2),$$

which is a weighted average of  $b_1(1)$  and  $b_2(2)$ , with weights  $\xi_1$  and  $\xi_2$  depending on the sum of prior probabilities of the relevant models being the best model. It is shown in the following example that different best designs may be selected taking into account experimenters' prior beliefs and the  $Q_B$  criterion provides an adequate framework to accommodate such prior beliefs.

We note that when it is further assumed that each submodel of the first-order model is equally likely to be the best model, then the  $Q_B$  criterion reduces to the  $Q$  criterion which minimizes

$$\delta_1 b_1(1) + 2\delta_2 b_2(2),$$

where  $\delta_1 = \xi_1 \delta_0$ ,  $\delta_2 = \xi_2 \delta_0$  and  $\delta_0$  is the total number of eligible models obeying functional marginality. In most cases ( $n \geq 4$ ),  $\delta_1 = 2\delta_2$ , and the  $Q$  criterion selects a design that minimizes  $b_1(1) + b_2(2)$ .

For a *level-balanced* design in which the two levels appear the same number of times,  $b_1(1) = 0$ , i.e. the main effect of each factor is orthogonal to the intercept. In this case, the  $Q_B$  criterion selects a design that minimizes  $b_2(2)$ . This is proportional to the  $E(s^2)$  criterion, defined by Booth and Cox (1962) in the context of supersaturated designs, where  $E(s^2) = \sum_{i < j} a_{ij}^2 / \binom{m}{2}$ . This criterion has become the standard method of choosing two-level supersaturated designs.

The  $Q_B$  criterion allows more flexibility than this. For example, if a subset of the factors can be identified which is thought more likely to have non-negligible effects than the others, more prior weight can be put on their estimation. This provides an alternative method of obtaining designs for such situations to that of Yamada and Lin (1997) who chose  $E(s^2)$ -optimal designs subject to all  $|a_{ij}|$  involving the more important factors being below a particular bound.

The  $Q_B$  criterion is also closely related to the  $p$ -efficient designs of Lin (1993) who studied the projection efficiencies of first-order saturated designs. To obtain  $p$ -efficient designs, he minimized  $E(s^2)$  among the class of near-level-balanced and near-orthogonal designs. Under the near-level-balance property,  $b_1(1) = 0$  for even  $n$  and  $b_1(1) = m/n^2$  for odd  $n$ . Therefore, the approach of Lin is equivalent to first minimizing  $b_1(1)$ , secondly minimizing  $\max |a_{ij}|$  and thirdly minimizing  $b_2(2)$ . An alternative way to choose a  $p$ -efficient design would be to use  $Q_B$  with  $\xi_1 \gg \xi_2$ . This would have the disadvantage of not guaranteeing near-orthogonality, but would be more flexible by allowing different priors to lead to different designs.

**Example** Here we give a simple example of five factors in six runs in which each factor has two levels. First we restrict ourselves to level-balanced designs and obtain all possible designs. Up to isomorphism of the generalized word count pattern, there are nine different first-order saturated designs. With these designs, every main effect is orthogonal to the intercept, but the main effects of a pair of factors are neither completely aliased together nor estimated independently. The  $Q$ -optimal level-balanced design, which is also the  $p$ -efficient design obtained by Lin (1993), has  $b_1(1) = 0$  and  $b_2(2) = 1.11$ .

Alternatively, consider the 3/16 fractional replicate of John (1971), shown in Table 1, obtained by deleting two points from the regular  $2^{5-2}$  design with defining relation  $I = ACD = ABCE = BDE$ . In this design, the main effects of factors  $A$  and  $E$  are

partially aliased with the intercept and  $b_1(1) = 0.22$ . For effects of pairs of factors, each of the main effects of  $A$  and  $E$  is orthogonal to each of the main effects of  $B$ ,  $C$  and  $D$ , the remaining pairs,  $\{(A,E), (B,C), (B,D), (C,D)\}$ , are partially aliased together and  $b_2(2) = 0.44$ . In terms of the  $Q$  criterion, this design is better than the best level-balanced design. That is, by sacrificing some orthogonality for factors' main effects, we can gain more design efficiency when models with different numbers of factors' main effects are of interest. If  $\xi_1 \geq 6.11\xi_2$ , then Lin's design is optimal with respect to  $Q_B$ . This would imply that there is a strong prior expectation that a model with a small number of main effects will be fitted which, of course, is the situation Lin's designs were intended for. At the other extreme, with  $p_{ij} = 1, \forall i, j$ , only the model with all main effects is considered and John's design can be shown to be  $Q_B$ -optimal and  $A$ -optimal. Thus, whereas Lin contrasted his designs with  $D$ - and  $A$ -optimal designs, the use of  $Q_B$  shows that there is a smooth transition from one to the other as the prior probability of factors being active changes.

## 4.2 Three-Level Qualitative Factors

For a three-level factor, two comparisons are made for the main effect. With a qualitative factor, these effects are either both included in or both excluded from the model. The first-order maximal model for three-level qualitative factors includes main effects of the factors with a total of  $2m$  effects.

The relationship between  $Q_B$  and the generalized word counts for three-level qualitative factors is given in Appendix B.1. By writing out the form of elements  $a_{ij}$ , it is shown that  $Q_B$  selects designs which minimize

$$\xi_1\{2b_1(1, 0) + b_1(0, 1)\} + 2\xi_2\{b_2(2, 0) + b_2(1, 1) + b_2(0, 2)\}. \quad (3)$$

For a level-balanced three-level design, the  $Q_B$ -criterion selects a design that minimizes

$$b_2(2, 0) + b_2(1, 1) + b_2(0, 2). \quad (4)$$

This criterion is similar to the  $\text{Ave}(\chi^2)$  criterion defined by Yamada and Lin (1999) in the context of three-level supersaturated designs, where  $\text{Ave}(\chi^2)$  is a measure of the dependency between every pair of columns of a design. This criterion minimizes the sum of squared  $n_{ij}$ , where  $n_{ij}$  is the number of occurrences for each of the nine combinations of levels of a pair of factors. For any design with all  $n_{ij}$  equal,  $\text{Ave}(\chi^2)$  is minimized, as is  $Q_B$ , since in this

case  $b_2(2, 0) = b_2(1, 1) = b_2(0, 2) = 0$ . However, departures from this ideal are penalized by the two criteria in different ways. We believe that the direct statistical interpretation of  $Q_B$  makes it preferable and again  $Q_B$  with different weights could be used if some factors are thought more likely to be active than others.

### 4.3 Other Cases

The cases of all factors having the same number of levels, usually being two or three, are by far the most commonly studied and also arise frequently in practice. However, it is not uncommon in practice to have some factors at two levels and others at three levels, or to have some factors having more than three levels. When working with the traditional criteria, it is necessary to define a new criterion for each new class of problems. For example, an extension of the  $\text{Ave}(\chi^2)$  criterion, known as  $E(f(\text{NOD}))$ , has been used for mixed-level supersaturated designs - see Yamada and Matsui (2002); Yamada and Lin (2002); Fang *et al.* (2003).

In contrast,  $Q_B$  can be applied to any problem without modification, apart from a careful choice of the prior probabilities and it has the same, statistically meaningful, interpretation in every case. A particular choice of prior probabilities will allow it to closely mimic any sensible criterion that could be suggested for any particular class of problems.

## 5 SECOND ORDER MODELS FOR QUALITATIVE FACTORS

### 5.1 Two-Level Factors

The second-order model for a two-level design has  $m$  main effects and  $\binom{m}{2}$  two-factor interactions, with a total of  $v = m + \binom{m}{2}$  effects. By writing out the elements  $a_{ij}$  and carefully collecting and counting them we find that the  $Q_B$  criterion defined in (1) selects a design that minimizes

$$\{\xi_{10} + 2(m-1)\xi_{21}\}b_1(1) + \{2\xi_{20} + \xi_{21} + 2(m-2)\xi_{32}\}b_2(2) + 6\xi_{31}b_3(3) + 6\xi_{42}b_4(4), \quad (5)$$

where  $\xi_{ij}$  denotes the sum, over models in which at least main effects of  $i$  factors and  $j$  two-factor interactions of these  $i$  factors are included, of the prior probabilities of a model

being the best model. Details of this relationship between  $Q_B$  and the generalized word counts are given in Appendix B.2.

In the case of orthogonal main effects designs, i.e.  $b_1(1) = b_2(2) = 0$ , the  $Q_B$ -criterion minimizes

$$\xi_{31}b_3(3) + \xi_{42}b_4(4),$$

which is a linear combination of the generalized word counts for words of effects involving sets of three and four factors respectively. This can be used to select the best second-order design among the class of orthogonal main effects designs when the estimation of two-factor interactions is of interest. In addition, one can use it as a refinement of the  $E(s^2)$ -criterion among those designs whose  $E(s^2)$  are the same in the context of saturated or supersaturated designs. Liu and Dean (2004) suggested modifying the  $E(s^2)$  criterion for supersaturated designs including a few interactions, but did not pursue the idea further. The  $Q_B$  criterion with appropriate priors provides such a modification.

An increasingly popular criterion for choosing two-level designs is the  $G_2$  generalized aberration criterion defined by Tang and Deng (1999). The  $G_2$ -aberration criterion sequentially minimizes  $b_2(2)$ ,  $b_3(3)$ ,  $b_4(4)$  and so on. This criterion is consistent with the belief that aliasing among lower-order effects is less desirable, so it concentrates first on minimizing aliasing between pairs of main effects, then on minimizing aliasing between main effects and two-factor interactions, then on minimizing aliasing between pairs of two-factor interactions and so on. The  $Q_B$  criterion on the other hand aims to improve the estimation in as many models as possible by jointly minimizing aliasing of these three types with different weights on each depending on the number of models in which some particular effects are included and prior information of each effect being in the model. The  $G_2$  criterion is a limiting form of  $Q_B$ . If  $2\xi_{20} + \xi_{21} + 2(m-2)\xi_{32} \gg 6\xi_{31} \gg 6\xi_{42}$ , then the  $Q_B$  criterion converges to the  $G_2$  criterion.

## 5.2 Three-Level Factors

For a three-level qualitative factor the maximal model of interest is the model with all  $m$  factors' main effects and all  $\binom{m}{2}$  two-factor interactions. Xu and Wu (2001) defined an ANOVA-type generalized minimum aberration (GMA) criterion for designs with qualitative factors. Effects with the same number of factors are regarded as equally important. Their

criterion functions can be written as

$$A_k = \sum_{\mathcal{K}} b_k(i_1, \dots, i_p), \quad (6)$$

where  $\mathcal{K} = \{i_1, \dots, i_p : i_1 + \dots + i_p = k\}$ . Their criterion is to sequentially minimize  $A_3, A_4, \dots$  among the class of orthogonal main effects designs. This criterion is the same as the minimization of the generalized word count for words of effects with  $k$  factors sequentially for  $k = 3, \dots, m$ .

For orthogonal main effects designs, where all main effects are orthogonal to each other, the  $Q_B$ -criterion selects a design that minimizes

$$\begin{aligned} & 2\xi_3\{b_3(3, 0) + b_3(2, 1) + b_3(1, 2) + b_3(0, 3)\} + \\ & \xi_4\{b_4(4, 0) + b_4(3, 1) + b_4(2, 2) + b_4(1, 3) + b_4(0, 4)\}, \end{aligned} \quad (7)$$

where  $\xi_s$  is the sum of prior probabilities for models in which at least effects of particular sets of  $s$  factors are included. This is a linear combination of the generalized word counts for words of effects involving sets of 3 and 4 factors respectively.

The GMA criterion concentrates first on minimizing aliasing between effects involving three factors and then on minimizing aliasing between effects involving four factors and so on. The  $Q_B$  criterion on the other hand jointly minimizes such aliasing with different weights depending on the model space and the prior information of effects being in the model. The GMA and  $Q_B$  criteria will select the same optimal designs when  $2\xi_3 \gg \xi_4$ . Thus, GMA is a limiting form of  $Q_B$  as  $2\xi_3/\xi_4 \rightarrow \infty$ .

### 5.3 Other Cases

As is the case with first order models, the  $Q_B$  criterion can be used with second order models when different factors have different numbers of levels or when there are factors with more than three levels. On the other hand a new generalized aberration criterion would have to be defined for each of these cases - see, for example, Joseph, Ai and Wu (2009). Since most sensible aberration criteria would correspond to limiting cases of  $Q_B$ , we recommend that  $Q_B$  be used instead. Again, the interpretation of  $Q_B$  is the same in every case.

## 6 POLYNOMIAL MODELS

### 6.1 Second Order Model with Three-Level Factors

When the second-order polynomial model is the maximal model of interest, the model has  $m$  linear main effects,  $m$  quadratic main effects and  $\binom{m}{2}$  linear  $\times$  linear two-factor interactions. For level-balanced designs, the  $Q_B$  criterion selects a design which minimizes

$$\begin{aligned} & \{\xi_{201} + 2\xi_{200} + 2\xi_{211} + 2(m-2)\xi_{302}\}b_2(2,0) + 2\xi_{220} b_2(0,2) \\ & + \{2\xi_{210} + \xi_{201}\}b_2(1,1) + 6\xi_{301} b_3(3,0) + \{2\xi_{311} + \xi_{302}\}b_3(2,1) + 6\xi_{402} b_4(4,0), \end{aligned}$$

where  $\xi_{abc}$  denotes the sum, over models which include at least linear effects of  $a$  factors,  $b$  out of the  $a$  factors' quadratic effects and  $c$  linear  $\times$  linear interactions of these  $a$  factors, of the prior probabilities of a model being the best among the eligible models.

As an alternative, consider the  $\beta$ -aberration criterion of Cheng and Ye (2004) in which effects with the same order are regarded as equally important. For designs with three levels, their criterion functions can be written as

$$B_t = \sum_{k=\lceil \frac{t+1}{2} \rceil}^t b_k(2k-t, t-k), \quad (8)$$

where  $\lceil \frac{t+1}{2} \rceil$  denotes the integer part of  $\frac{t+1}{2}$  and  $t = 1, \dots, 2m$ . In general, for designs with  $(p+1)$ -level quantitative factors, the  $\beta$ -aberration criterion sequentially minimizes the sum of the generalized word counts for effects of  $z$ th order where  $z = i_1 + 2i_2 + \dots + pi_p$  for  $z = 1, \dots, pm$ .

For designs with three levels,  $B_1 = b_1(1,0)$ ,  $B_2 = b_1(0,1) + b_2(2,0)$ ,  $B_3 = b_2(1,1) + b_3(3,0)$ ,  $B_4 = b_2(0,2) + b_3(2,1) + b_4(4,0)$  and so on. For level-balanced designs, the  $\beta$ -aberration criterion sequentially minimizes  $B_2$ ,  $B_3$ ,  $B_4$  and so on, while the  $Q_B$  criterion minimizes a linear combination of the individual word counts appearing in  $B_2$ ,  $B_3$  and  $B_4$ . We see that the  $\beta$ -aberration criterion is not a special case of  $Q_B$ , but is a limiting form. If  $\xi_{201} + 2\xi_{200} + 2\xi_{211} + 2(m-2)\xi_{302} = c_2$ ,  $6\xi_{301} = 2\xi_{210} + \xi_{201} = \frac{c_3}{2}$  and  $6\xi_{402} = 2\xi_{311} + \xi_{302} = 2\xi_{220} = \frac{c_4}{3}$ , then  $Q_B = c_2B_2 + c_3B_3 + c_4B_4$ . We see that as  $\frac{c_2}{c_3} \rightarrow \infty$  and  $\frac{c_3}{c_4} \rightarrow \infty$ ,  $Q_B$  converges to the  $\beta$ -aberration criterion.

In general, the  $\beta$ -aberration criterion aims to minimize aliasing, concentrating first on aliasing between linear and linear  $\times$  linear interaction effects, then on aliasing between pairs of linear  $\times$  linear interactions and between quadratic and linear  $\times$  linear interaction effects.

The  $Q_B$  criterion is a combination of these three types of aliasing, with weights for each model being the best model determined by the experimenters' prior belief in each effect being in the best model. Tsai *et al.* (2007) demonstrated that the minimum  $\beta$ -aberration design tends to be  $Q_B$ -optimal if there is more weight on linear effects and the prior information leads to a model of small size; on the other hand, when more weight is placed on the quadratic effects, different designs will be chosen for the given prior.

## 6.2 Mixed-Level Designs

Nguyen (1996) extended the  $E(s^2)$  criterion to factorial designs whose factors have different numbers of levels, although in a rather limited way. He considered designs with one or two three-level factors, which had to be level-balanced and orthogonal to each other and then calculated  $E(s^2)$  by considering the off-diagonal terms relating to the linear and quadratic contrasts of these factors. The criteria illustrated below are much more general than this. Note also that Nguyen did not consider the appropriate scaling of the effects, although this would be a straightforward modification.

Table 2 gives three 12-run mixed-level designs with four two-level factors and one three-level factor. The designs are obtained by adding a three-level factor, say  $X_5$ , to four-factor projections of the Hadamard matrix of order 12. For this design with one three-level factor, let  $b_s(s)$  denote the generalized word count for words involving  $s$  two-level factors,  $\{b_s(s) \oplus b_1(1, 0)\}$  denote the generalized word count for words involving  $s$  two-level factors and the linear effect of the three-level factor and  $\{b_s(s) \oplus b_1(0, 1)\}$  denote the generalized word count for words involving  $s$  two-level factors and the quadratic effect of the three-level factor. The values of the three types of word multiplied by 144 are given in Table 3. With F4D1 and F4D2,  $\{b_1(1) \oplus b_1(1, 0)\} = \{b_1(1) \oplus b_1(0, 1)\} = 0$  and all the two-level factors are orthogonal to the linear and quadratic effects of  $X_5$ ; with F4D3, the two-level factors are orthogonal to the linear effect of  $X_5$  but not to the quadratic effect. When the main effects model for the five factors is taken as the maximal model of interest, the  $Q_B$  criterion selects designs that minimize a linear combination of the generalized word counts of words referring to effects involving two factors, i.e.  $b_2(2)$ ,  $\{b_1(1) \oplus b_1(1, 0)\}$  and  $\{b_1(1) \oplus b_1(0, 1)\}$ . In this case, we find that F4D1 and F4D2 are better than F4D3.

When the second-order maximal model involving only the linear interactions is the maximal model of interest, the  $Q$  criterion uses the following combination of the generalized

word counts of effects of 2nd, 3rd and 4th order to select the best design:

$$450[b_3(3) + \{b_2(2) \oplus b_1(1, 0)\}] + \frac{197}{2}(\{b_2(2) \oplus b_1(0, 1)\} + 179[b_4(4) + \{b_3(3) \oplus b_1(1, 0)\}]),$$

where 450, 197, and 179 are the values of  $\delta_{31}$ ,  $\delta_{32}$  and  $\delta_{42}$  respectively. In this case, design F4D2 has a lower value of  $Q$  than design F4D1.

Another example is a saturated design with nine two-level factors and one three-level factor which is also listed in Table 2 and whose properties are given in Table 3. We note that for this saturated design, the design has two-level factors orthogonal to each other but aliased with the main effects of the three-level factor. These examples show that it is possible to apply the  $Q_B$  criterion to new situations like this and that it keeps its interpretation and remains easy to calculate. On the other hand, new aberration criteria have to be invented for each such new situation.

## 7 REGULAR FRACTIONAL FACTORIAL DESIGNS

In regular designs with two levels, the generalized word counts defined in this paper are the same as the numbers of words of different lengths in the defining relation. The value of  $W_i(d)$  is equal to the generalized word count  $b_i(i)$  for words referring to effects involving  $i$  factors. If  $W_i(d) = 0$ , then the  $i$ -factor interactions are orthogonal to the intercept so  $b_i(i) = 0$ . If  $W_i(d) = 1$ , then one of the  $i$ -factor interactions is completely aliased with the intercept so that  $b_i(i) = 1$ . Thus for regular factorial designs of Resolution-III or higher, for the second-order maximal model, the  $Q_B$  criterion selects designs that minimize

$$\xi_{31} W_3(d) + \xi_{42} W_4(d),$$

where  $\xi_{ab}$  is the sum of prior probabilities for models in which main effects of  $a$  factors and  $b$  two-factor interactions are included. This is a linear combination of the numbers of words of length 3 and 4 in the defining contrast subgroup of the design. The minimum aberration criterion of Fries and Hunter (1980) sequentially minimizes  $W_3(d)$ ,  $W_4(d)$ , etc. Thus, exactly as for the generalized aberration criteria,  $Q_B$  is related to aberration in regular designs.

For regular fractions in which the model is estimable, the approximation to the  $A_s$  criterion used in  $Q_B$  is exact. If the maximal model is estimable then  $Q_B$  represents a true weighted average  $A_s$  criterion over all candidate models. Since the variances of parameter

estimates do not depend on which other parameters are estimated, because of orthogonality, it is a simple weighted- $A$  criterion for the maximal model. For regular fractional replicates, therefore, this establishes an exact link between alphabetic optimality and aberration.

## 8 APPLICATION TO CASE STUDIES

As noted in Section 2 the two case studies were superficially similar, both having six factors and both using similar numbers of runs, but the designs used were quite different. Although this was based on prior knowledge, in both experiments the prior knowledge stated was rather extreme and, in both experiments, the results subsequently obtained from the data analysis strongly indicated that the certainty implied by such extreme prior assumptions was not appropriate. We believe the experiments could have been better designed by expressing the prior beliefs in less extreme form.

We first note that the design used in the plasma etching experiment was probably not the best among the class of three-level orthogonal main effects designs. Table 4 gives the values of  $B_3$ ,  $B_4$  and  $Q$  for the second-order maximal model for some 18-run three-level main effects designs. Design 220 was the one used, which is based on six columns of the  $L_{18}$  orthogonal array. By broadening the class of designs, we find that the minimum  $\beta$ -aberration design is 25 and the  $Q$ -optimal design is 2 (designs are given in Table 2 of Tsai *et al.* (2007)). Comparing these, we find that a trade-off between the aliasing of linear and interaction effects and that of quadratic and interaction effects occurs. Clearly, the final choice of design will depend on the prior knowledge.

For the fuel tank welding experiment, the data suggest that we should have considered regular two-level designs in 16 runs which do not assume that only three specific interactions will be important, but consider the possibility of any interactions. The minimum aberration design is a resolution IV design with  $(B_3, B_4, B_5) = (0, 3, 0)$ . However, if we do not restrict ourselves to the idea of maximum resolution, some other designs can also be of interest. For example, there is a design with word count pattern  $(B_3, B_4, B_5) = (1, 1, 1)$  (columns 1 2 3 4 8 13 in Table 5). This design is a second-order saturated design in which all degrees of freedom are used to estimate main effects and two-factor interactions, and has 6 clear two-factor interactions which are not aliased with other two-factor interactions.

Here, we consider comparing all of these designs for the two experiments, with less

extreme prior knowledge specified in each case. Two center points are added to the above two-level designs, giving 18-run designs denoted  $Df2.1$  and  $Df2.2$ , respectively. Table 6 gives the values of  $Q_B$  for the  $Q$ -optimal (D2),  $\beta$ -aberration optimal (D25) and these two designs.  $\pi_1$ ,  $\pi_2$  and  $\pi_3$  are, respectively, the prior probability of each linear effect being in the best model, the probability of each quadratic effect being in the best model, given that the corresponding linear effect is, and the probability of each interaction being in the best model, given that the corresponding linear effects are.

We can see that when experimenters are more interested in the linear and interaction effects and less interested in the quadratic effects, then a two-level design might be good to use. In this case if simpler models are of interest then  $Df2.1$  would be preferred since its two-factor interactions are orthogonal to main effects; if models with many interactions are of interest, then  $Df2.2$  might be a better choice since it has more clear two-factor interactions and would allow better parameter estimates for the interactions.

On the other hand, if experimenters want to estimate some factors' quadratic effects as well as linear and interaction effects, then a three-level main effects design might be a better choice. Among the three-level main effects designs, different sets of  $\pi_1$ ,  $\pi_2$  and  $\pi_3$  can lead to different  $Q_B$ -optimal designs. In the case that second-order polynomial models of small size are of interest, D25 is better than the  $Q$ -optimal designs since it has the least aliasing between linear main effects and interactions, but the aliasing between quadratic main effects and interactions is high. In the case of more weight on quadratic effects or for models with more parameters, the  $Q$ -optimal design is preferred.

Thus there are reasonable interpretations of the prior beliefs stated which suggest that the designs used, or similar designs, were suitable. However, there are other interpretations of the prior beliefs which suggest otherwise. This indicates that there are benefits to be gained from stating prior beliefs formally as probabilities.

We also calculate the values of the  $D_B$  criterion of DuMouchel and Jones (1994), for the cases when (1) linear main effects are the primary terms and (2) linear and quadratic main effects are the primary terms. (Note that for six-factor designs with 18 runs there are not enough degree of freedoms to consider linear and two-factor interaction effects as the primary terms.) Table 7 gives the values of  $D_B = \log(\mathbf{X}'\mathbf{X} + \mathbf{K}/\tau^2)$ , where  $\mathbf{K}$  is a diagonal matrix, having diagonal element 0 for each primary term and 1 for each potential term, with  $\tau = 0.25, 1, 4$ , where 1 is the default value suggested by DuMouchel and Jones

(1994),  $\tau = 4$  is when we have a strong faith in the necessity of the potential terms, and  $\tau = 0.25$  is when the potential terms are not really expected to be large. For the case of linear primary terms, *Df2.1* and *Df2.2* are preferred. As we increase  $\tau$ , *Df2.2* which has more clear interactions is preferred, which is coincident with our results in  $Q_B$ . For the case with linear and quadratic primary terms, *D2* and *D25* are preferred since the the linear and quadratic main effect model is not estimable in *Df2.1* and *Df2.2*. In the cases shown *D2* is clearly preferred.

Using  $D_B$  it is not easy to get results which correspond to those in Table 6 with the priors (0.9,0.3,0.2) or (1,0.4,0.4), even though both these cases seem to fit naturally into the description of linear effects as primary and others as potential. It is also impossible to reproduce results like those from the prior (0.7,0.4,0.1), where there are clearly three distinct groups of effects. One advantage of  $Q_B$  is that it separates the prior probability of an effect being in the final model from the size of that effect. For example, a linear effect can be small, but still have a high probability of being in the final model due to effects to which it is marginal being large. So there are at least some respects in which  $Q_B$  is more flexible than  $D_B$ . A possible area for future research would be to develop modifications of  $D_B$  which can match the flexibility of  $Q_B$ .

## 9 SUMMARY AND CONCLUSIONS

In this paper, we have established connections between the  $Q_B$  criterion and other criteria for multifactor designs in different situations. It has been shown that the  $Q_B$  criterion, based on a weighted average of  $A_s$  criteria, has better statistical justification than some popular criteria. This criterion aims to find designs which can provide good parameter estimates in as many models as possible and is recommended for many situations.

Note also that the somewhat elaborate results presented in Sections 4-7 are developed here just to prove the relationships between  $Q_B$  and other criteria. They are not needed to use  $Q_B$  in practice. All that is needed are:

- a maximal model, usually bigger than that we expect eventually to fit;
- a prior probability of each model being the best, usually obtained from prior probabilities of each type of effect being active; and

- a method of searching for an optimal design, which will be different for different classes of designs, but will be the same as for any other optimality criterion for a given class of designs.

The  $Q_B$  criterion concentrates on estimating many models and does not explicitly take account of how to discriminate between these models. However, since all models are sub-models of the maximal model, discrimination between models is indirectly taken account of. To discriminate between two models it is sufficient to be able to fit the model which includes all terms which appear in either model. We have not chosen the weights specifically to reflect this and how to do so might be a fruitful avenue for further exploration. But, clearly, for finding an efficient factorial design under model uncertainty, this criterion is preferable to other related criteria since it has a better statistical justification. We recommend it for practical use.

## ACKNOWLEDGEMENTS

We acknowledge financial support from Queen Mary, University of London, the UK Engineering and Physical Sciences Research Council and the Isaac Newton Institute for Mathematical Sciences. The first author's research was partly supported by a grant from the National Science Council of Taiwan. We also thank Academia Sinica for invited the second author to visit and Ching-Shui Cheng for encouraging us to continue our work on this topic. We thank the referees for thoughtful and constructive comments which led to considerable changes in the way we present this work.

## APPENDIX A: SCALING OF FACTORIAL EFFECTS

For different design situations, we need to clarify the definitions of effects in the maximal model. In general, for a factor with  $p + 1$  levels, there are  $p$  comparisons to be made for the factor's main effect. A set of multiplicative constants for each of the contrast functions is used so that the  $p$  effects of a factor are orthogonal to each other and are estimated with equal variances in the full factorial. For a factor with two levels, the main effect of the

factor is usually represented by a function with coefficients  $(-1, 1)$  which is the comparison of the response at the two different levels. With three levels, there are two comparisons to be made. Often, two orthogonal comparisons with coefficients such as  $\sqrt{3/2}(-1, 0, 1)$  and  $\sqrt{1/2}(1, -2, 1)$  are chosen to define a factor's main effects. When the three levels of the factor are equally spaced quantitative values, these two orthogonal comparisons represent the linear and quadratic effects of the factor. For a factor with four levels, there are three orthogonal effects to be defined. If the levels are equally spaced levels of a quantitative factor, then we use  $\sqrt{1/5}(-3, -1, 1, 3)$ ,  $(-1, 1, 1, -1)$  and  $\sqrt{1/5}(1, -3, 3, -1)$  to represent linear, quadratic and cubic effects of the factor, respectively. In all cases, interaction contrast functions are defined by multiplying the corresponding main effect columns in the full factorial in the usual way. Note that these are normalized contrasts in the full factorial but not, in general, in the designs we are comparing (which each have different normalized contrasts).

For factors with qualitative levels, some alternative definitions of effects might also be useful. For example, for a factor with three levels, one might be interested in all pairwise comparisons among levels. It can be shown that the contrasts for three-level qualitative factors can be derived from the linear and quadratic contrasts, and that the variances for the estimates of the pairwise comparisons are on the same scale, which is  $\frac{3}{2}$  times those for the comparisons with three-level quantitative factors. Thus choosing a design for linear and quadratic effects having equal weight is equivalent to choosing a design for all pairwise comparisons having equal weight.

## APPENDIX B: EXPRESSING $Q_B$ IN TERMS OF GENERALIZED WORD COUNTS

### B1: First order model for three-level qualitative factors

For designs with  $m$  three-level factors, let  $x_L = \sqrt{3/2}x_i$  and  $x_Q = \sqrt{2}(\frac{3}{2}x_i^2 - 1)$  denote the linear and quadratic main effects of a three-level factor with the levels coded as  $-1, 0$  and  $1$ . When all the diagonal elements of the  $\mathbf{X}'\mathbf{X}$  matrix have the same value, we use the generalized word count  $b_k(i_1, i_2)$ , defined in equation (2), to represent the terms appearing in the  $Q_B$ -criterion. Here,  $b_k(i_1, i_2)$  is the overall measure of the partial aliasing of all the factorial effects with  $k(= i_1 + i_2)$  factors where  $i_1$  factors appear in the linear effects and

$i_2(= k - i_1)$  factors appear in the quadratic effects.

1. For  $i$  referring to the intercept and  $j$  referring to a linear effect, the sum of  $a_{ij}^2/n^2$  is equal to  $b_1(1,0)$ , which is an overall measure of the partial aliasing between the intercept and linear effects. With level-balanced designs,  $b_1(1,0) = 0$ , i.e. linear effects are all orthogonal to the intercept.
2. For  $i$  referring to the intercept and  $j$  referring to a quadratic effect, the sum of  $a_{ij}^2/n^2$  is equal to  $b_1(0,1)$ , which is an overall measure of the partial aliasing between the intercept and quadratic effects.
3. For  $i, j$  referring to linear effects of a pair of factors, the sum of  $a_{ij}^2/n^2$  is equal to  $2b_2(2,0)$ , which is twice the overall measure of the partial aliasing between pairs of factors' linear effects.
4. For  $i, j$  referring to quadratic effects of a pair of factors, the sum of  $a_{ij}^2/n^2$  is equal to  $2b_2(0,2)$ , which is twice the overall measure of the partial aliasing between pairs of factors' quadratic effects.
5. For  $i$  referring to a linear effect and  $j$  referring to a quadratic effect of a factor or vice versa, the sum of  $a_{ij}^2/n^2$  is equal to  $b_1(0,1)$  for  $i, j$  referring to effects of the same factor, and is equal to  $2b_2(1,1)$  for  $i, j$  referring to effects of two different factors. We note that for a three-level design, with the coding used in the current paper, the sum of the componentwise product of a factor's linear effect and its quadratic effect, say  $x_L$  and  $x_Q$ , is equal to  $1/\sqrt{2}$  times the sum of the componentwise product of the intercept and the factor's linear effect,  $j(\{x_L x_Q\}) = \frac{1}{\sqrt{2}}j(\{x_L\})$ . Thus, the sum of  $a_{ij}^2/n^2$ s for those  $i, j$  referring to a factor's linear and quadratic effects appearing in  $Q_B$  is equal to  $b_1(0,1)$ .

We note that in  $Q_B$ ,  $p_{ij} = \xi_1$  for  $i$  and  $j$  referring to effects of the same factor, and  $p_{ij} = \xi_2$  for  $i$  and  $j$  referring to effects involving two different factors. The  $Q_B$  criterion in equation (1) selects a design that minimizes

$$\xi_1\{2b_1(1,0) + b_1(0,1)\} + 2\xi_2\{b_2(2,0) + b_2(1,1) + b_2(0,2)\},$$

as in (3).

## B2: Second order model for two-level qualitative factors

For designs with  $m$  two-level factors, the maximal second-order model contains  $m$  main effects and  $\binom{m}{2}$  two-factor interactions. Let  $a_{ij}$  for  $i, j = 0, \dots, v$ , be the  $(i, j)$ th entry of the  $\mathbf{X}'\mathbf{X}$  matrix for the maximal second-order model where  $v = m + \binom{m}{2}$  is the number of effects in the model. For two-level factors,  $a_{ii} = n$  for all  $i$  and we will use the generalized word count  $b_k(k)$  defined in equation (2) to represent the sum of the  $a_{ij}^2/n^2$  terms in the  $Q_B$  criterion in equation (1).

1. For  $i$  referring to the intercept and  $j$  referring to the main effect of a factor, the sum of  $a_{ij}^2/n^2$  is equal to  $b_1(1)$ . With level-balanced designs,  $b_1(1) = 0$ , i.e. main effects are all orthogonal to the intercept.
2. For  $i$  referring to the intercept and  $j$  referring to a two-factor interaction, the sum of  $a_{ij}^2/n^2$  is equal to  $b_2(2)$ , which is an overall measure of the partial aliasing between pairs of factors' main effects.
3. For  $i, j$  referring to main effects of a pair of factors, the sum of  $a_{ij}^2/n^2$  is equal to  $2b_2(2)$ .
4. For  $i, j$  referring to a pair of interactions, the sum of  $a_{ij}^2/n^2$  is equal to  $2(m-2)b_2(2)$  for  $i, j$  referring to interactions with a common factor, and is equal to  $6b_4(4)$  for  $i, j$  referring to interactions with no common factor. We note that with two-level factors, the aliasing between two interactions with a common factor, say  $AB$  and  $AC$ , is equal to the aliasing between the intercept and the  $BC$  interaction. For a given interaction, there are  $2(m-2)$  other interactions with a common factor and  $\binom{m-2}{2}$  interactions with no common factor. Thus, for  $i, j$  referring to interactions, the sum of  $a_{ij}^2/n^2$  is  $2(m-2)b_2(2)$  for those with a common factor and  $6b_4(4)$  for those with four different factors.
5. For  $i$  referring to the main effect of a factor and  $j$  referring to an interaction effect or vice versa, the sum of  $a_{ij}^2/n^2$  is equal to  $2(m-1)b_1(1)$  for  $i, j$  referring to effects of a common factor, and is equal to  $6b_3(3)$  for those with no common factor.

Let  $\xi_{ab}$  denote the sum, over models which include at least main effects of  $a$  factors and  $b$  interaction effects of these  $a$  factors, of the prior probabilities of a model being the best

among the eligible models. We note that the  $p_{ij}$  associated with the  $a_{ij}$  in  $Q_B$  are  $\xi_{10}$  for those in group (1),  $\xi_{21}$  for those in group (2),  $\xi_{20}$  for those in group (3),  $\xi_{32}$  for those in group (4) for effects of three factors,  $\xi_{42}$  for those in group (4) for effects of four factors,  $\xi_{21}$  for those in group (5) with one common factor and  $\xi_{31}$  for those in group (5) with three different factors. Thus, the  $Q_B$  criterion for second-order designs with two-level qualitative factors selects a design that minimizes

$$\{\xi_{10} + 2(m - 1)\xi_{21}\}b_1(1) + \{2\xi_{20} + \xi_{21} + 2(m - 2)\xi_{32}\}b_2(2) + 6\xi_{31}b_3(3) + 6\xi_{42}b_4(4),$$

as in (5). The derivations for the relationships between  $Q_B$  and the generalized word counts for second-order model with three-level qualitative and quantitative factors can be done in similar ways and are not presented here.

## Supplemental Materials

**Qb\_funs.r** A text file containing five R functions and examples of their use.

**gwc\_sup.pdf** A pdf file describing the use of these five R functions.

## References

- Atkinson, A. C., Donev, A. N. and Tobias, R. (2007). *Optimum Experimental Designs, with SAS*. Oxford: Oxford University Press.
- Bailey, R. A. (1998) Statistics and mathematics: the appropriate use of mathematics within statistics (with discussion). *The Statistician*, **47**, 261-271.
- Booth, K. H. V. and Cox, D. R. (1962) Some systematic supersaturated designs. *Technometrics*, **4**, 489-495.
- Butler, N. A. (2003). Some theory for constructing minimum aberration fractional factorial designs. *Biometrika*, **90**, 233-238.
- Cheng, C. S. and Tang, B. (2005). A general theory of minimum aberration and its applications. *Ann. Stat.*, **33**, 944-958.

- Cheng, C. S. Deng, L. W. and Tang, B. (2002). Generalized minimum aberration and design efficiency for non-regular fractional factorial designs. *Statistica Sinica*, **12**, 991–1000.
- Cheng, C. S., Steinberg, D. M. and Sun, D. X. (1999) Minimum aberration and model robustness for two-level fractional factorial designs. *Journal of the Royal Statistical Society, Series B*, **61**, 85–93.
- Cheng, S. W. and Ye, K. Q. (2004). Geometric isomorphism and minimum aberration for factorial designs with quantitative factors. *Annals of Statistics*, **32**, 2168–2185.
- 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.
- Fang, K. T. and Mukerjee, R. (2000). A connection between uniformity and aberration in regular fractions of two-level factorials. *Biometrika*, **87**, 193–198.
- Fang, K. T., Lin, D. K. J. and Liu, M. Q. (2003). Optimal mixed-level supersaturated design. *Metrika*, **58**, 279–291.
- Fearn, T. (1992). Box-Cox transformations and the Taguchi method: an alternative analysis of a Taguchi case study. *Applied Statistics*, **41**, 553–559.
- Fries, A. and Hunter, W. G. (1980). Minimum aberration  $2^{k-p}$  designs. *Technometrics*, **22**, 601–608.
- Grove, D. M. and Davis, T. P. (1992). *Engineering, Quality and Experimental Design*. Harlow: Longman.
- Heredia-Langner, A., Montgomery, D. C., Carlyle, W. M. and Borror, C. M. (2004). Model-robust optimal designs: a genetic algorithm approach. *Journal of Quality Technology*, **36**, 263–279.
- Jacroux, M. (2004). A modified minimum aberration criterion for selecting regular  $2^{(m-k)}$  fractional factorial designs. *J. Statist. Plannng Inf.*, **126**, 325–336.
- John, P. W. M. (1971). *Statistical Design and Analysis of Experiments*. New York: Macmillan.

- Joseph, V. R., Ai, M. Y. and Wu, C. F. J. (2009). Bayesian-inspired minimum aberration two- and four-level designs. *Biometrika*, **96**, 95–106.
- Läuter, E. (1974). Experimental design in a class of models. *Matematische Operationsforschung und Statistik*, **5**, 379–396.
- Li, W. and Nachtsheim, C. J. (2000). Model-robust factorial design. *Technometrics*, **42**, 345–354.
- Lin, D. K. J. (1993). Another look at first-order saturated design: The  $p$  efficient designs. *Technometrics*, **35**, 284–292.
- Liu, Y. and Dean, A. M. (2004).  $k$ -circulant supersaturated designs. *Technometrics*, **46**, 32–43.
- 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.
- Ma, C. X. and Fang, K. T. (2001). A note on generalized aberration in factorial designs. *Metrika*, **53**, 85–93.
- Nguyen, N. K. (1996). A note on the construction of near-orthogonal arrays with mixed levels and economic run size. *Technometrics*, **38**, 279–283.
- Ozol-Godfrey, A., Anderson-Cook, C. M. and Montgomery, D. C. (2005). Fraction of design space plots for examining model robustness. *Journal of Quality Technology*, **37**, 223–235.
- Tang, B. and Deng, L. Y. (1999). Minimum  $G_2$ -aberration for non-regular fractional factorial designs. *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. (2007). Three-level main-effects designs exploiting prior information about model uncertainty. *Journal of Statistical Planning and Inference*, **137**, 619–627.
- Xu, H. and Wu, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *Annals of Statistics*, **29**, 1066–1077.
- Yamada, S. and Lin, D. K. J. (1997). Supersaturated design including an orthogonal base. *Canadian Journal of Statistics*, **25**, 203–213.
- Yamada, S. and Lin, D. K. J. (1999). Three-level supersaturated designs. *Statistics and Probability Letters*, **45**, 31–39.
- Yamada, S. and Lin, D. K. J. (2002). Construction of mixed-level supersaturated design. *Metrika*, **56**, 205–214.
- Yamada, S. and Matsui, T. (2002). Optimality of mixed-level supersaturated designs. *Journal of Statistical Planning and Inference*, **104**, 459–468.
- Zhao, X., Joseph, L, Wolfson, D. B. and Bélisle, P. (2003). A Bayesian  $A$ -optimal and model robust design criterion. *Biometrics*, **59**, 1082–1088.

Table 1: A  $\frac{3}{16}$  fraction of a  $2^5$  factorial design

Run	A	B	C	D	E
1	-	+	-	+	+
2	-	-	+	-	+
3	+	+	-	-	-
4	+	-	+	+	-
5	-	+	+	-	-
6	-	-	-	+	-

Table 2: Some 12-run mixed-level designs

F4D1	F4D2	F4D3	F9D1
-1-1-1 1-1	-1-1 1-1-1	-1-1-1 1-1	-1 1-1-1 1-1 1 1 1-1
-1 1-1-1-1	-1 1-1 1-1	-1 1 1 1-1	1-1 1 1 1-1-1-1 1-1
1-1 1 1-1	1-1-1-1-1	1-1 1 1-1	1 1 1-1-1-1 1-1-1-1
1 1 1-1-1	1 1 1 1-1	1 1-1-1-1	1 1 1 1 1 1 1 1 1-1
-1-1 1-1 0	-1-1-1 1 0	-1-1 1-1 0	-1 1 1 1-1-1-1 1-1 0
-1 1 1 1 0	-1 1-1-1 0	-1 1-1-1 0	-1-1 1-1 1 1 1-1-1 0
1-1-1 1 0	1-1 1 1 0	1-1-1-1 0	1-1-1 1-1 1 1 1-1 0
1 1-1-1 0	1 1 1-1 0	1 1 1-1 0	1 1-1-1-1 1-1-1 1 0
-1-1 1-1 1	-1-1 1-1 1	-1-1 1-1 1	-1-1 1-1-1 1-1 1 1 1
-1 1-1 1 1	-1 1 1 1 1	-1 1-1 1 1	-1-1-1 1-1-1 1-1 1 1
1-1-1-1 1	1-1-1 1 1	1-1-1 1 1	1-1-1-1 1-1-1 1-1 1
1 1 1 1 1	1 1-1-1 1	1 1 1 1 1	-1 1-1 1 1 1-1-1-1 1

Table 3: Generalized word counts for some 12-run mixed-level designs

Design	$s$	$b_s(s)$	$b_s(s) \oplus b_1(1, 0)$	$b_s(s) \oplus b_1(0, 1)$
F4D1	1	0	0	0
	2	0	120	72
	3	64	48	80
	4	16	24	8
F4D2	1	0	0	0
	2	0	48	144
	3	64	96	32
	4	16	0	32
F4D3	1	0	0	72
	2	0	72	0
	3	64	96	56
	4	16	24	32
F9D1	1	0	114	54
	2	0	468	540
	3	1344	996	1164
	4	2016	1476	1260

Table 4: Comparison of criteria for 18-run three-level designs with six factors

	$B_3$	$B_4$	$Q$
1	1.1875	5.0781	1.1657
2	0.9583	5.5312	<b>1.1620</b>
8	1.1458	5.0781	1.1782
25	<b>0.7500</b>	<b>6.9375</b>	1.1836
220	<b>0.7500</b>	8.7188	1.2226

Table 5: A Hadamard matrix of order 16, where the column with all +1 is omitted

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

Table 6:  $Q_B$  for some 18-run three-level designs with six factors under various priors

$\pi_1$	$\pi_2$	$\pi_3$	D2	D25	Df2.1	Df2.2
0.9	0.1	0.5	0.8998	0.9064	<b>0.5500</b>	0.5672
0.9	0.1	0.8	0.9383	0.9604	0.6944	<b>0.6913</b>
0.9	0.2	0.5	0.9220	0.9324	<b>0.6478</b>	0.6653
0.9	0.2	0.8	0.9260	0.9501	0.7389	<b>0.7370</b>
0.9	0.3	0.2	0.6127	<b>0.6126</b>	0.6534	0.6742
0.9	0.5	0.2	<b>0.6850</b>	0.6887	1.0839	1.1047
0.9	0.5	0.5	<b>0.9425</b>	0.9615	1.0472	1.0647
0.8	0.3	0.2	0.5160	<b>0.5159</b>	0.5533	0.5681
0.8	0.3	0.5	<b>0.7776</b>	0.7881	0.6533	0.6661
0.8	0.5	0.5	<b>0.8117</b>	0.8273	0.9298	0.9425
0.7	0.2	0.5	0.6030	0.6083	<b>0.4446</b>	0.4535
0.7	0.2	0.8	0.7456	0.7630	0.5318	<b>0.5261</b>
0.7	0.4	0.1	0.3887	<b>0.3886</b>	0.56409	0.5701
1	0.4	0.4	<b>1.0355</b>	1.0496	1.0465	1.0767
1	1	0.2	<b>0.9713</b>	0.9868	3.2925	3.3195

Table 7:  $D_B$  for some six-level designs with 18 runs with (1) linear primary terms and (2) linear and quadratic primary terms under various  $\tau$

	$\tau$	D2	D25	Df2.1	Df2.2
Case 1	0.25	78.65	78.69	<b>81.25</b>	80.82
	1.00	35.77	35.30	37.73	<b>38.23</b>
	4.00	5.66	2.93	0.87	<b>6.07</b>
Case 2	0.25	<b>68.62</b>	68.54		
	1.00	<b>33.24</b>	32.22		
	4.00	<b>4.22</b>	0.78		