Model-sensitive sequential optimal designs

Arvind Ruggooa,∗, Martina Vandebroekb

a Faculty of Agriculture, University of Mauritius, Reduit, Mauritius
b Faculty of Economics and Applied Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium

Received 12 July 2004; received in revised form 9 November 2005; accepted 10 November 2005
Available online 5 December 2005

Abstract

The increasing number of experimenters using computer-generated experimental designs creates an increasing need to have design procedures that are less sensitive to model misspecification. To address this problem, the notion of empirical models that have both important and potential terms is used. A two-stage design strategy for planning experiments in the face of model uncertainty is proposed. The advantage of this procedure resides in the rearrangement of active potential terms at the end of the first stage using marginal posterior probabilities of different candidate models. The two-stage procedure has better estimation efficiency than its one-stage alternatives available from the literature.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Two-stage procedures; GD-optimality; Marginal posterior probabilities; Primary terms; Potential terms; Lack of fit; Bias; Rearrangement

1. Introduction

D-optimality and alphabetic optimality criteria in general are known to depend on the assumed model and spend all of the resources on precise estimation of the parameters of the assumed model. They neither make any explicit provisions for reducing the bias in case departures from the assumed model occur nor allow the fit of higher order terms in case model inadequacy is diagnosed. There have been several attempts to develop algorithms that retain the flexibility of the “optimal” approaches whilst avoiding these common criticisms. Steinberg and Hunter (1984) provide a nice overview of the different approaches proposed to account for model uncertainty, ranging from model-robust to model-sensitive strategies.

Since the form of the final model that will be required is usually unknown prior to conducting the experiment, and in practice the final model often differs considerably from the original one, there is a great practical interest in constructing designs that are very efficient relative to the chosen optimality criterion for a set of potential or candidate models (Heredia-Langner et al., 2004). We refer to such designs as model-robust designs. Box and Draper (1959) were the first authors to consider this issue in depth. They introduced an average mean squared error criterion in which the mean squared error is averaged over a region of interest in the design variables. An overview of a considerable part of
the work on model-robust designs is given in Chang and Notz (1996). According to them, a good model robust design should:

1. Allow one to fit the assumed model.
2. Detect the model inadequacy when the fitted model is a poor approximation to the true model.
3. Allow one to make reasonable efficient inferences concerning the assumed model when the assumed model is adequate.

See also for example the non-exhaustive list: Montepiedra and Federov (1997), Federov et al. (1998), Fang and Wiens (2000, 2003), Borkowski and Valeroso (2001) and Heredia-Langner et al. (2004) for additional relevant references. However, we note that most of the work in this area has considered the class of approximate designs and has not necessarily led to practical designs for the experimenter. The use of the approximate theory by several researchers is natural because the various minimizations involved are tractable as opposed to those in the theory for exact designs. Hence, as argued by Chang and Notz (1996), the practical value of all the work on continuous designs serve mainly in alerting us of the dangers of ignoring the approximate nature of any assumed model and to provide some insight regarding the features a design should possess to be robust against departures from an assumed model whilst allowing good fit of the assumed model. We see for example from all these papers that if the fitted model is a rough approximation to the true model, standard optimal design may yield very inefficient results. The space-filling designs, i.e., those designs that take observations spread uniformly over the design region are often sub-optimal.

The interesting work by DuMouchel and Jones (1994) illustrates a very elegant use of Bayesian methods to obtain designs which are more resistant to the bias caused by an incorrect model. They assume that there are two kinds of model terms, namely primary and potential terms which are high and low priority terms in the model respectively. By conceiving a prior distribution on the model coefficients that takes these primary and potential terms under consideration, they generate a Bayesian D-optimal design by maximizing the posterior information matrix. Similar type of work has been done by Andere-Rendon et al. (1997) for experiments involving mixtures where they show that the performance of their Bayesian designs is superior to standard D-optimal designs by producing smaller bias and improved coverage over the factor space.

In a model-sensitive design strategy, one looks for designs that facilitate the improvement of the model by detecting lack of fit. A lot of work in this area is due to Atkinson (1972), Atkinson and Cox (1974) and Atkinson and Fedorov (1975a, 1975b). The crucial idea of such designs lies around detection of lack of fit by maximizing the dispersion matrix somehow. Jones and Mitchell (1978) have also elaborated on this idea and Studden (1982) combined the detection of lack-of-fit with a precise estimation of the primary terms. This combined approach of Studden (1982) and other model discrimination procedures are described in detail by Atkinson and Donev (1992).

It would seem desirable to develop a design criterion that would account for both model-robust and model-sensitive aspects of a design. However, as is often the case, a design strategy that for example, involves protection against bias works counter to a strategy designed for minimum variance. In other words, the minimum variance approaches like the D-optimal designs result in pushing the design points to the edge of the region whilst minimum bias designs require adequate placement of design points at a reasonable distance from the design center (see Myers and Montgomery, 2002 for more details). It is imperative that any combined design strategy would result in a compromise design and not specifically suited for any one specific aspect of an experiment.

In this respect, DeFeo and Myers (1992) proposed a new criterion for design robustness that brings together protection against the use of an oversimplified model and detection through lack of fit, for a class of designs called ‘rotated designs’. A more recent development in the area is the criterion developed by Goos et al. (2005). Their criterion accounts for both model-robust and model-sensitive aspects of a design by combining efficiency in estimating the primary terms, protection against bias caused by the potential terms and ability to test for lack of fit, thereby increasing the knowledge about the true model. They term their new design criterion, the Generalized D-optimal (GD) criterion and show that the new criterion performs well with respect to bias and detection of lack of fit.

Sequential procedures have also been used to develop designs in two or more stages that lead to less dependence on model specification. Neff (1996) developed a series of exact Bayesian two-stage designs under model uncertainty for mean estimation models. Lin et al. (2000) extended the approach of Neff (1996) and developed Bayesian two-stage D-D optimal designs for models in experiments involving mixtures. Montepiedra and Yeh (1998, 2002) have proposed a two-stage sequential procedure for approximate designs, where one part of the sample is used for the identification of
the appropriate model, whilst the design in the second stage is chosen for the most efficient inference in the identified model. Biswas and Chaudhuri (2002) have investigated a sequential strategy, where the objective of the design is to select the “correct” model from the family of nested models as well as efficiently estimate the parameters associated with the model. Ruggoo and Vandebroek (2004) have recently reviewed and extended the approach of Neff (1996).

The objective in this paper is to extend the work of Goos et al. (2005) in a two-stage design strategy that will be robust to model uncertainty. In the first stage the true model is unknown and the experimenter’s prime interest is to highlight the uncertainties in the specification of the model in order to refine or modify the model initially entertained. The GD criterion is used since it accounts for precision of the primary terms but also facilitates the improvement of the proposed models by detecting lack of fit. Now, the rationale for the choice of primary terms comes from previous observations of the system, screening experiments conducted or prior information from expert knowledge. Consequently these terms are necessary in modeling the response. Additionally, the potential terms are also possibly important but may not have been noticed during previous observation of the system or may not have been studied thoroughly enough in previous experiments to justify being specified in the primary component. The factor sparsity principle coined by Box and Meyer (1986) implies that only a group of factors drives the system whilst the other factors, though present, will be less important. Consequently, the investigator would be interested to identify any of the active potential terms at the end of the first stage and rearrange the primary and potential components in the second stage. Note that in this scenario, the model fitted at the end of the experiment will be the one that includes all the primary terms after rearrangement, so that the new fitted model will be a much better approximation to the true model than the initial primary model entertained.

For example, suppose that the primary and potential terms are initially specified as follows:

\[
x_{\text{pri}} = \{1, x_1, x_2, x_1^2\}, \quad x_{\text{pot}} = \{x_1 x_2, x_3, x_2^2, x_2 x_3\}.
\]

After conducting the first stage experiment, model information from first stage data may suggest that the new primary and potential terms partition becomes

\[
x_{\text{pri}} = \{1, x_1, x_2, x_1 x_2, x_1^2, x_2^2\}, \quad x_{\text{pot}} = \{x_3, x_2^2\}.
\]

In this case, we note that the process is driven by terms arising from the full quadratic model in \(x_1\) and \(x_2\) with the other terms being less important and included in the potential component. Consequently we can use this updated information in the numerical construction of the second stage design.

To carry out this rearrangement of the potential terms at the end of the first stage, we shall use the approach of Box and Meyer (1993), who propose a Bayesian method to assign an appropriate measure of fit to each model considered (posterior probabilities) that can be accumulated to give rise to marginal posterior probabilities for identifying active effects in a system. The posterior probabilities of Box and Meyer (1993) have also been used in the development of a Bayesian D-D optimal design (see Neff, 1996; Ruggoo and Vandebroek, 2004). The second stage design will then be a D-optimal design for the rearranged primary model.

The paper will be organized as follows. In Section 2, the design criterion introduced by Goos et al. (2005) will be reviewed. In Section 3, we set the notation to be used in the paper including the prior formulation on the model space. In Section 4, we develop our first stage design and the appropriate analysis of the first stage data. Section 5 contains some simulation studies in connection with the choice of the cut-off point for the identification of active potential terms at the end of the first stage. In Section 6, we obtain our second stage design followed by the evaluation of the combined design with respect to the different one-stage alternatives available from the literature in Section 7. Finally we end up with a small discussion in Section 8.

### 2. The GD criterion

In this section we briefly review the GD criterion developed by Goos et al. (2005). Let us assume that the linear model that will be fitted by the experimenter is of the form

\[
y = X_{\text{pri}}^\prime \beta_{\text{pri}} + \epsilon
\]
with \( \mathbf{x}_{\text{pri}} \) being a \( p \)-dimensional vector of powers and products of the factors and \( \beta_{\text{pri}} \) the \( p \)-dimensional vector of unknown parameters attached to the primary terms. Suppose that the expected response was misspecified so that the true model is actually of the form

\[
y = \mathbf{x}' \beta + \epsilon = \mathbf{x}_{\text{pri}}' \beta_{\text{pri}} + \mathbf{x}_{\text{pot}}' \beta_{\text{pot}} + \epsilon = \eta(\mathbf{x}) + \epsilon,
\]

where \( \mathbf{x}_{\text{pot}} \) is the \( q \)-dimensional vector containing powers and products of the factors not included in the fitted model and \( \beta_{\text{pot}} \) is the \( q \)-dimensional vector associated with the potential terms. We shall refer to \( \mathbf{x}_{\text{pri}}' \beta_{\text{pri}} \) as the primary term and to \( \mathbf{x}_{\text{pot}}' \beta_{\text{pot}} \) as the potential term. To simplify the notation, we will assume that the model has been reparametrized in terms of the orthonormal polynomials with respect to a measure \( \mu \) on the design region. The orthonormalization ensures that the effects are well separable and independent so that a simple prior distribution on the potential terms can be used. Since the primary terms are likely to be active and no particular directions of their effects are assumed, the coefficients of the primary terms are specified to have a diffuse prior distribution. On the other hand, potential terms are unlikely to have huge effects and the assumption \( \beta_{\text{pot}} \sim N(0, \tau^2 \mathbf{I}_q) \) proposed by DuMouchel and Jones (1994) is appropriate. The parameter \( \tau^2 \) is the common prior variance of the potential terms’ coefficients, measured in units of the random error variance \( \sigma^2 \). Following the orthonormalization procedure, the joint prior distribution assigned to \( \beta_{\text{pri}} \) and \( \beta_{\text{pot}} \) is \( N(0, \sigma^2 \tau^2 \mathbf{K}^{-1}) \), where \( \mathbf{K} \) is a \((p + q) \times (p + q)\) diagonal matrix, whose first \( p \) diagonal elements are equal to zero and the remaining \( q \) diagonal elements are equal to one. This results in infinite prior variance for the primary terms, corresponding to the uninformative prior assigned to \( \beta_{\text{pri}} \) and restricted prior on the potential terms as in the DuMouchel and Jones (1994) case.

Let \( \mathbf{X}_{\text{pri}} \) be the \( n \times p \) model matrix for the primary terms and \( \mathbf{X}_{\text{pot}} \) the \( n \times q \) model matrix for the potential terms. Goos et al. (2005) combine the three aspects: precise estimation of the primary model, minimization of the bias caused by potential terms and possibility to test for lack of fit into one criterion. They specify weights \( \alpha_L \) and \( \alpha_B \) to attach more or less importance on the different properties. They propose to find designs that minimize the GD criterion

\[
\text{GD} = \left\{ \frac{1}{p} \log \left| \left( \mathbf{X}_{\text{pri}}' \mathbf{X}_{\text{pri}} \right)^{-1} \right| + \frac{\alpha_L}{q} \log \left| \left( \mathbf{L} + \frac{1}{\tau^2} \mathbf{I}_q \right)^{-1} \right| + \frac{\alpha_B}{q} \log \left| \mathbf{A}' \mathbf{A} + \mathbf{I}_q \right| \right\},
\]

where \( \mathbf{A} = \left( \mathbf{X}_{\text{pri}}' \mathbf{X}_{\text{pri}} \right)^{-1} \mathbf{X}_{\text{pri}}' \mathbf{X}_{\text{pot}} \) is the alias matrix which essentially transcribes bias to the parameter estimates \( \hat{\beta}_{\text{pri}} \) and

\[
\mathbf{L} = \mathbf{X}_{\text{pot}}' \mathbf{X}_{\text{pot}} - \mathbf{X}_{\text{pot}}' \mathbf{X}_{\text{pri}} \left( \mathbf{X}_{\text{pri}}' \mathbf{X}_{\text{pri}} \right)^{-1} \mathbf{X}_{\text{pri}}' \mathbf{X}_{\text{pot}},
\]

and is usually referred to as the dispersion matrix (see Atkinson and Donev, 1992).

For \( \alpha_L = \alpha_B = 0 \) the GD-optimality criterion produces the D-optimal design for the primary model. For \( \alpha_B = 0, \alpha_L = q/p \) and \( \tau^2 = \infty \), we obtain the D-optimal design for the full model. Setting \( \alpha_B = 0, \alpha_L = q/p \) and for finite values for \( \tau^2 \), the Bayesian D-optimal designs introduced by DuMouchel and Jones (1994) are obtained.

3. Development of the two-stage approach

In a two-stage strategy, data from the first stage is used to generate parameter information which is subsequently used to select the remaining second stage experimental runs with maximum efficiency. We shall consider the same framework as in Section 2 and set the notation and assumptions to be used in the development of our two-stage designs. We assume that \( y_i | \beta \sim N(\mathbf{X}_i \beta, \sigma^2 \mathbf{I}) \) for each stage \( i = 1, 2 \) with \( n_1 \) and \( n_2 \) observations in the first and second stage respectively so that the total number of observations in the entire experiment \( n = n_1 + n_2 \). \( \mathbf{X} \) is the extended design matrix of dimension \( n \times (p + q) \) for the combined stages, so that \( \mathbf{X}' = \begin{bmatrix} \mathbf{X}_1' & \mathbf{X}_2' \end{bmatrix} \). \( \mathbf{X}_{\text{pri}(i)} \) and \( \mathbf{X}_{\text{pot}(i)} \) correspond to the primary and potential terms respectively for each stage \( i = 1, 2 \). In the sequel of this paper and in the various examples considered, we shall use the uniform measure on the design region because it is assumed that all the points in the design region are equally important.

Before observing the first stage data, the experimenter has specified a set of \((p + q)\) regressors defining the full model. The true relationship between the response and the input variables is believed to contain all primary terms and
a subset of \( q_i \) (\( 0 \leq q_i \leq q \)) potential terms. Consequently the total number of possible models is \( m = 2^q \). Let us consider the subset models \( M_0, M_1, \ldots, M_{m-1} \), with each model \( M_k \) defined by its corresponding parameters \( \beta_k \), where \( M_0 \) corresponds to the primary terms model. Let \( \pi \) be the prior probability that any one of the potential terms is active in the true model. Assuming that any one of the potential terms being active is independent of beliefs about the other terms, the prior probability \( p (M_i) \) of model \( M_i \) is given by

\[
p (M_i) = \pi^i (1 - \pi)^{q - q_i}, \quad i = 0, 1, 2, \ldots, m - 1.
\]

(4)

The value of \( \pi \) should be chosen to represent the proportion of potential terms to be active. Since the experimenter may expect only a few of the potential terms to be really active, Neff (1996), Lin et al. (2000) and Ruggoo and Vandebroek (2004) recommend to use \( \pi = \frac{1}{2} \) in the computations. In general, the procedure is quite robust to the choice of this parameter. However as a last point, it can be noted from (4) that

\[
p (M_0) = (1 - \pi)^q.
\]

(5)

Thus in case the experimenter has a priori knowledge on \( p (M_0) \), then it could also be reasonable to choose \( \pi \) indirectly from (5).

4. Construction of the first stage design

In the first stage, the experimenter believes that the plausible model comprises the primary terms but at the same time would like some knowledge about possible incorrect model specification. In other words s/he would wish to be able to test for lack of fit thereby increasing the knowledge on the true model whilst at the same time ensuring precise estimation of the primary terms. A possible first stage design could be obtained by finding \( X_1' = [X_{pri(1)}' X_{pot(1)}] \) which minimizes the GD criterion of Goos et al. (2005) with a large weight placed on the lack of fit component and setting \( \alpha_B = 0 \) in (3). We thus obtain

\[
GD_1 = \left\{ \frac{1}{p} \log \left( \left| X_{pri(1)}' X_{pri(1)} \right|^{-1} \right| + \frac{\alpha_L}{q} \log \left( \left| L_1 + \frac{I_q}{\tau^2} \right|^{-1} \right) \right\},
\]

(6)

where \( L_1 = X_{pot(1)}' X_{pot(1)} - X_{pot(1)}' X_{pri(1)} \left( X_{pri(1)}' X_{pri(1)} \right)^{-1} X_{pri(1)}' X_{pot(1)} \).

Since the prior probabilities, \( p (M_i) \)’s in (4), reflect a priori model importance for each of the \( m = 2^q \) candidate models, they can be incorporated as weights in the first stage criterion so that our first stage design \( X_1 = [X_{pri(1)} X_{pot(1)}] \) is actually obtained by minimizing

\[
\sum_{M_k} p (M_k) GD_1^{(k)},
\]

(7)

where

\[
GD_1^{(k)} = \left\{ \frac{1}{p} \log \left( \left| X_{pri(1)}^{(k)}' X_{pri(1)}^{(k)} \right|^{-1} \right| + \frac{\alpha_L}{q_k} \log \left( \left| L_1^{(k)} + \frac{I_q^{(k)}}{\tau^2} \right|^{-1} \right) \right\},
\]

and \( X_{pri(1)}^{(k)}, L_1^{(k)} \) and \( I_q^{(k)} \) are the matrices corresponding to \( X_{pri(1)}, L_1 \) and \( I_q \) expanded to model space \( M_k \).

4.1. Analysis of first stage data

Once data from the first stage has been collected, the information from the analysis can be used as prior information to reduce model uncertainty in the next stage. Box and Meyer (1993) propose a general way for calculating the posterior probabilities of different candidate models within the framework of fractionated screening experiments. Given the first stage data \( y_1 \), the posterior probability of model \( M_i \) given \( y_1 \) is

\[
p (M_i | y_1) \propto p (M_i) p (y_1 | M_i),
\]

(8)
where \( p(M_i) \) is the prior probability of model \( M_i \) and \( p(y_1|M_i) \) is the integrated likelihood of \( y_1 \) given model \( M_i \). The resulting posterior probability for model \( M_i \) given \( y_1 \) can then be obtained along the lines shown in Box and Meyer (1993):

\[
p(M_i|y_1) = C \pi_i^{2q_i} (1 - \pi)^{q - q_i} \left| X'_{1(i)}X_{1(i)} + \frac{K_i}{\tau^2} \right|^{-1/2} \times \left( S(\hat{\beta}_i) + \frac{1}{\tau^2} \tilde{K}_i \hat{\beta}_i \right)^{-(n_1 - 1)/2},
\]

(9)

where \( X_{1(i)} \) is the first stage design in model \( M_i \) space and of dimension \( n_1 \times (p + q_i) \),

\[
K_i = \begin{bmatrix} 0_{p \times p} & 0_{p \times q_i} \\ 0_{q_i \times p} & 1_{q_i \times q_i} \end{bmatrix};
\]

\[
\hat{\beta}_i = E(\beta_i|M_i, y_1) = \left( X'_{1(i)}X_{1(i)} + \frac{K_i}{\tau^2} \right)^{-1} X'_{1(i)}y_1,
\]

\[
S(\hat{\beta}_i) = \text{Residual sum of squares for model } M_i,
\]

\[
= (y_1 - X_{1(i)}\hat{\beta}_i)'(y_1 - X_{1(i)}\hat{\beta}_i),
\]

and finally \( C \) is the normalization constant that forces all probabilities to sum to one. In what follows, we show how we can accumulate these posterior probabilities to identify active effects at the end of the first stage.

### 4.2. Rearrangement of terms using marginal posterior probabilities after the first stage

To carry out the rearrangement of terms at the end of the first stage, Box and Meyer (1993) propose to accumulate the posterior probabilities computed in (9), to uncover active effects driving the system. They compute the marginal posterior probability \( P_j \) that effect \( j \) is active as

\[
P_j = \sum_{M_i; \text{effect } j \text{ active}} p(M_i|y_1).
\]

(10)

The probability \( P_j \) is just the sum of the posterior probabilities of all the distinct models in which the effect \( j \) is active. The probabilities \( \{P_j\} \) are thus calculated by direct enumeration over the \( m = 2^q \) possible models \( M_i \). A large value for \( P_j \) would indicate that effect \( j \) was active and a value of \( P_j \) close to zero would indicate that the effect \( j \) was inert (see Box and Meyer, 1993 for more details). We illustrate the steps in the computation of these marginal posterior probabilities, with the following simple example. Suppose \( X^{(pri)} = \{1, x_1, x_2, x_3, x_1^2\} \) and \( X^{(pot)} = \{x_1 x_2, x_2^2, x_3^2\} \). We have \( m = 2^3 = 8 \) possible models. We can use first stage data to compute their respective posterior model probabilities, \( p(M_i|y_1) \)’s as in Table 1. The marginal posterior probabilities for the potential terms are then obtained by direct
The results are displayed in Fig. 1. We have also calculated the percentage of times the marginal posterior probabilities exceed the 0.05 cut-off point over the 1000 simulations and for the different sample sizes. These results are displayed in Fig. 2.

In the examples of their paper, Box and Meyer (1993) use the cut-off probability, $P_j = 0.5$, for identification of effects active in the system. They further remark that the absolute magnitudes of the probabilities are less important than their pattern and relative magnitudes. Also probabilities which stand out from the others, even if they are less that 0.5, identify effects worthy of further consideration. In our context, this would suggest that any effect for which $P_{j} \geq 0.5$ would be part of the primary component. In the next section, we shall further examine the choice of this cut-off value.

In general we shall see that the cut-off value of 0.5 is helpful for sorting out potentially active effects whenever the sample size is not too small in the first stage.

### 5. Simulation studies to determine cut-off point

We now conduct a large simulation study to investigate the relationship between sample size, number of potential terms and the cut-off probability point of 0.5. Based on some preliminary evaluations, we further assume that $\tau = 5$ in both stages and $z_L = 10$ (see also Goos et al., 2005 for additional discussions on the choice of $z_L$). We have found that using $z_L > 10$, may provide a design which would enable better detection of lack-of-fit, but the overall first and second stage design resulting may have less desirable properties in terms of precision of the parameters.

**Case I:** Here we consider $p = 5$ primary terms, $x^{(\text{pot})} = \{1, x_1, x_2, x_1 x_2, x_1^2\}$, and $q = 5$ potential terms, $x^{(\text{pot})} = \{x_3, x_1 x_3, x_2 x_3, x_1^2, x_2^2\}$. So here $p + q = 10$. First stage data is simulated from the true model with all primary terms and three of the five potential terms:

$$y = 10.0 + 2.8 x_1 - 3.1 x_2 + 2.0 x_1 x_2 - 2.5 x_1^2 + 1.0 x_3 + 1.0 x_1 x_3 + 1.0 x_2^2 + \varepsilon. \quad (11)$$

The error term $\varepsilon$ is assumed to come from $N(0, 1)$ in all the simulations and we set the coefficients of the potential terms to be 1.0, i.e., coefficients with the minimum impact required to have a significant effect with respect to the noise in the system. We assume the following sample sizes in the first stage $n_1 = 5$–15. We then carry out 1000 simulations and compute the average marginal posterior probabilities for each of the potential terms over the different sample sizes. The standard deviations over the simulations were small indicating that the 1000 simulations were more than sufficient. The results are displayed in Fig. 1. We have also calculated the percentage of times the marginal posterior probabilities exceed the 0.5 cut-off point over the 1000 simulations and for the different sample sizes. These results are displayed in Fig. 2.
Here note that $p + q = 10$, so that we need at least 10 observations in the first stage to be able to estimate the parameters of the full model. The active effects present in the true model and driving the system are all the primary terms and the following potential terms, $x_3$, $x_1x_3$ and $x_2^2$, whilst the inert terms are $x_2x_3$ and $x_3^2$. In general we note that, for the active terms, the average marginal posterior probabilities increase with increasing sample sizes. For sample size less than or equal to seven, the average marginal posterior probabilities for terms $x_3$ and $x_2^2$ are less than the cut-off value of 0.5 and consequently will not be detected as active, but once $n_1 > 7$, we can detect all the active terms. The average marginal probability for the inert terms, in general decreases with increasing sample sizes. We can perceive similar patterns from Fig. 2, where with increasing sample size, the power of using the marginal posterior probabilities in detecting the active effects increases.

**Case II:** In this case we consider $p = 7$ primary terms, $\mathbf{x}^{(\text{pri})} = \{1, x_1, x_2, x_3, x_1x_2, x_1^2, x_2^2\}$ and $q = 6$ potential terms, $\mathbf{x}^{(\text{pot})} = \{x_4, x_1x_3, x_2x_3, x_1x_4, x_3^2, x_4^2\}$. First stage data is simulated from the true model with the seven primary terms and three of the six potential terms:

$$y = 10.0 + 2.8x_1 - 3.2x_2 + 2.0x_3 + 1.9x_1x_2 - 2.2x_1^2 + 3.3x_2^2 - 1.0x_4 + 1.0x_2x_3 + 1.0x_3^2 + \varepsilon. \quad (12)$$

As before, $\varepsilon \sim N(0, 1)$ is assumed in all the simulations and the following sample sizes in the first stage, $n_1 = 8–18$. We carry out 1000 simulations and compute the average marginal posterior probabilities for each of the potential terms over the different sample sizes. The results are shown in Fig. 3.

Here $p + q = 13$, the active terms from the potential component are $x_4$, $x_2x_3$, $x_3^2$ and the inert terms are $x_1x_3$, $x_1x_4$ and $x_4^2$. When the sample size is small there are not enough resources to estimate all the $2^6 = 64$ potential models.
The Box and Meyer posterior probabilities are then unstable and often fail to identify the true model. Consequently it becomes difficult to identify the active and inert effects. Once \( n_1 > 12 \), we see the average marginal probability for the active terms increases and can easily be detected. As for the inert effects, the average marginal probability rapidly decreases below the 0.5 cut-off point postulated by Box and Meyer (1993).

5.1. Some additional remarks on the simulation studies

The cases above and some others we have investigated indicate that the 0.5 cut-off point for the marginal posterior probability is a good choice for identifying active effects when the sample size in the first stage is not too small and always correctly identifies the active and inert components whenever \( n_1 = p + q \). We have also noticed similar types of patterns for the linear, cross-product and quadratic terms, so that in general we need only one cut-off point for all these types of effects in a model. As we have also seen in Cases I–II, whenever \( n_1 < p + q \), it is not possible to give an indication about the minimum sample size required for correctly identifying the active and inert effects for a specific problem. Our general recommendations would be to carry out rearrangement of terms at the end of the first stage whenever we have \( n_1 \geq p + q \). If \( n_1 < p + q \), a possible solution would be to increase the cut-off point to say 0.8, and consequently we hope to identify the correct active and inert effects at the end of the first stage. We could also wait at the end of the second stage when we have adequate resources to rearrange the terms using the marginal posterior probabilities.

Finally, in connection with the relative sample sizes allocated to each stage, Neff (1996) and Ruggoo and Vandebroek (2004) recommends a 50\% distribution of the sample size in the two stages to produce designs robust to model misspecifications. Lin et al. (2000) also recommend a ratio of 1 : 1 between the two stages to give satisfactory results for the mixture experiments. Additionally, Ruggoo and Vandebroek (2004), based on their simulation studies suggest that if resources permit, try to have a sample size of at least \((p + q + 2)\) in the first stage. In the paper, we are working on the assumption that resources are scarce and we have tried to work with at least \((p + q)\) in the first stage as our results have shown that it may not be worthwhile to adopt our procedure in case \( n_1 < p + q \). In the second stage, we have used sample size, \( n_2 < p + q \), but it is expected that the results for the two-stage procedure to be much better if more resources are available.

6. Selection of the second stage design

Whenever \( n_1 \geq p + q \), the researcher can compute the Box and Meyer posterior probabilities at the end of the first stage and then decide to allocate active potential terms to the primary component if their marginal posterior probabilities exceed 0.5. For e.g., in Case I, suppose that the researcher had 12 runs available in the first stage. If one obtained posterior probabilities similar to the average values in Fig. 1, then at the end of the first stage s/he would have rearranged the terms as follows: \( x^{(pri)} = \{1, x_1, x_2, x_1 x_2, x_1^2, x_3, x_1 x_3, x_2^2\} \) and \( x^{(pot)} = \{x_2 x_3, x_3^2\} \). Suppose that the researcher is interested in obtaining precise estimates of the parameters for the fitted model. At the end of the first stage, the experimenter may then wish to drop the redundant potential terms and augment the first stage design with a D-optimal design for the rearranged primary model in the second stage. We may ignore bias at this stage, because we shall expect the rearranged model to be close to the true model driving the system. In other words, the second stage design will be obtained with reduced model uncertainty. Note that in the procedure we propose, the final model that shall be fitted at the end of the experiment will be the new set of primary terms obtained after rearrangement at the end of the first stage.

We shall refer to this two-stage procedure as the MGD-D two-stage procedure. The acronym, MGD (Model Generalized D), enforcing the analogy that all possible models are taken into account in the first stage criterion.

7. Evaluation of the two-stage designs

We now need to evaluate our MGD-D two-stage procedure and compare it with its competitors which are (i) the one-stage Bayesian D-optimal design developed by DuMouchel and Jones (1994), (ii) the unique stage D-optimal design for the full model and (iii) the GD optimal design of Goos et al. (2005) using \( x_L = 5 \) and \( x_B = 10 \).
All evaluations will be carried out taking into account the regressors present in the true model. Note the Mahalanobis values in brackets are the standard deviations over the 200 simulations.

The experimenter has a difficult choice to make: s/he may be willing to have a minimum variance design but it may turn out to be for the wrong model. It is exactly in this direction that the MGD-D two-stage procedure in this paper was developed. The approach has an added advantage in terms of increased precision of the parameters of the fitted model, namely

\[ \beta = \hat{\beta} + \sum_{i=1}^{m} \frac{\hat{\beta}_i}{a} + \epsilon, \]

where \( \beta \) and \( \hat{\beta} \) are the estimates and true values of the parameters respectively, and \( X \) is the corresponding model matrix.

We consider details for Case I, where \( p + q = 10 \). Suppose there are resources for \( n_1 = 12 \) in the first stage and \( n_2 = 8 \) in the second. The first stage design can be obtained using the approach in Section 4. Once we obtain first stage data, we can compute the Box and Meyer posterior probabilities and consequently compute the marginal posterior probabilities for the potential terms using (10). For e.g., from Fig. 1, for \( n_1 = 12 \), we can identify the active terms, namely \( \{x_1, x_2, x_1x_2, x_1^2, x_3, x_1x_3, x_2^2\} \) and the inert potential terms \( \{x_2x_3, x_1^2\} \). In the second stage, we can then drop the inert potential terms and obtain the D-optimal design for the rearranged primary terms as explained earlier. Once we have the second stage design, we can simulate second stage data to compute the Mahalanobis distances for the combined design.

We have used simulated data from the first and second stage designs and computed the average Mahalanobis distance over 200 simulations under the assumption that \( \epsilon \) is drawn from a \( N(0, 1) \) distribution. We have also computed the Mahalanobis distances over 200 simulated data for the one-stage alternatives. The results are shown in Table 2. Similar computations have been performed for Case II assuming \( n_1 = 14 \) and \( n_2 = 10 \). All the results are displayed in Table 2. Values in brackets are the standard deviations over the 200 simulations.

From Table 2, the two-stage MGD-D procedure, which involves rearranging terms at the end of the first stage gives much better estimation efficiency of the parameters for the true model than its competitors. The one-stage Bayesian D-optimal design of DuMouchel and Jones (1994) also perform well as opposed to the design obtained using the D-optimality approach for the full model. The designs of Goos et al. (2005) are the worst in terms of efficiency of the parameters of the true model. The results suggest that using the marginal posterior probabilities at the end of the first stage is very useful in identifying the most appropriate model activating the system. The MGD-D procedure we have thus considered, provides us with a way to tackle model uncertainty which is inherent in most practical problems.

### Table 2
Comparisons of the different designs using the Mahalanobis distances for Cases I and II

<table>
<thead>
<tr>
<th>Design</th>
<th>Case I</th>
<th>Case II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goos et al. (2005)</td>
<td>0.041 (0.0392)</td>
<td>0.038 (0.0345)</td>
</tr>
<tr>
<td>D-optimal (Full model)</td>
<td>0.029 (0.0287)</td>
<td>0.032 (0.0323)</td>
</tr>
<tr>
<td>DuMouchel and Jones (1994)</td>
<td>0.025 (0.0285)</td>
<td>0.029 (0.0282)</td>
</tr>
<tr>
<td>MGD-D two-stage procedure</td>
<td>0.020 (0.0194)</td>
<td>0.024 (0.0221)</td>
</tr>
</tbody>
</table>

8. Discussion

The increasing number of experimenters using computer programs for design assistance creates an ever increasing need to have designs that are less dependent on implicit assumptions and less sensitive to model misspecifications. The experimenter has a difficult choice to make: s/he may be willing to have a minimum variance design but it may turn out to be for the wrong model. It is exactly in this direction that the MGD-D two-stage procedure in this paper was developed. The approach has an added advantage in terms of increased precision of the parameters of the fitted model.
model as opposed to the standard one-stage alternatives in the literature. Another interesting point to stress here is that as opposed to a lot of work devoted to continuous designs and design measures, this work has given a procedure to construct exact two-stage designs which are required in practice and can easily be implemented by the experimenter.

Acknowledgements

This work was supported by the Fund for Scientific Research Flanders-Belgium. The authors are grateful to the Editor, Associate Editor and the referees for their constructive comments and suggestions.

References


Atkinson, A.C., Fedorov, V.V., 1975a. The design of experiments for discriminating between two rival models. Biometrika 62, 57–70.


