KATHOLIEKE
UNIVERSITEIT
LEUVEN

DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

RESEARCH REPORT 0319

TWO-STAGE DESIGNS ROBUST TO MODEL UNCERTAINTY

by

A. RUGGOO
M. VANDEBROEK

D/2003/2376/19

Two-stage designs robust to model uncertainty

Arvind Ruggoo
Martina Vandebroek

Department of Applied Economic Sciences, Katholieke Universiteit Leuven

Abstract

D-optimal designs are known to depend quite critically on the particular model that is assumed. These designs tend to concentrate all the experimental runs on a small number of design points and are ideally suited for estimating the coefficients of the assumed model, but they provide little or no ability for model checking. To address this problem we use the notion of empirical models that have both important and potential terms. We propose within the Bayesian paradigm, a two-stage design strategy for planning experiments in the face of model uncertainty. In the first stage, the experimenter's prime interest is to highlight the uncertainties in the specification of the model in order to refine or modify the model(s) initially entertained. A design criterion is used that accounts for precision of the important terms but also facilitates the improvement of the proposed model(s) by detecting lack of fit. Data from the first stage provide model information enabling the second stage design to be chosen efficiently with reduced model uncertainty. The design in the second stage is obtained using a weighted criterion with weights being posterior model probabilities computed from first stage data. The criterion in the second stage also takes into account precise estimation of important terms as in the first stage but now attempts to minimize bias with respect to potential terms. Results from simulations show that the proposed two-stage strategy performs well. The combined first and second stage design has good properties with respect to precision of important terms, lack of fit and also excellent bias properties with respect to a true assumed model in various simulation studies.

Keywords: Bayesian two-stage procedures, GD-optimality, prior probabilities, posterior probabilities, model-robustness, model-sensitive, bias, lack of fit

1 Introduction

D-optimal designs are known to depend on the assumed model and spend all of the resources in precise estimation of the parameters of the assumed model. They make no explicit provisions for reducing the bias error in case departures from the assumed model occur and also for allowing 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 D-optimal approach 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.

In a model-robust approach, designs are sought that will yield reasonable results for the proposed model even though it is known to be inexact. Box and Draper (1959) were the first authors to consider this problem in depth. They assumed that the true model comprises some primary (important) terms that will eventually be fitted and some potential (questionable) terms. They argue that a more appropriate criterion for comparing experimental designs is the integrated mean squared error (IMSE) over a region of interest. The IMSE can be decomposed as the sum of a bias component and a variance component. The problem with this and similar criterion is that the experimental design will depend on the parameters of the potential terms.

DuMouchel and Jones (1994) (DMJ) illustrate a very practical use of Bayesian methods for design selection that preserves the flexibility of the D-optimal approach, whilst being less sensitive to the model assumptions. Based on the work of DMJ and Box and Draper (1959), Kobilinsky (1998) developed a design criterion combining bias and variance properties. In addition Neff (1996) and Ruggoo and Vandebroek (2002) demonstrate the advantage of the procedure of DMJ and illustrate a two-stage approach in which prior information is updated at the end of the first stage. The two-stage design approach they develop makes it possible to efficiently design experiments when initial knowledge of the model is poor. This is accomplished by using a Bayesian D-optimality criterion in the first stage and the second stage design is then generated from an optimality procedure which incorporates the improved model knowledge from the first stage.

In a model-sensitive design strategy, one looks for designs that facilitate the improvement of the model by detecting lack of fit. Such approaches, also referred to as model discrimination procedures, are elaborated by Atkinson and Donev (1992). The crucial idea of such designs centers around determination of lack of fit by maximizing the dispersion matrix somehow.

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 errors works counter to a strategy designed for minimum variance. It is also seen that the minimum variance approaches like the D-optimal designs results 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 a design.

In this respect, DeFeo and Myers (1992) propose 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 new criterion developed by Goos, Kobilinsky, O'Brien and Vandebroek (2002), henceforth referred to as GKOV. 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 and thereby increasing the knowledge on the true model. They term their new design criterion, the Generalized D-optimal (GD) criterion and show that the new criterion perform well with respect to bias and detection of lack of fit.

The overall objective of the present paper is to develop two-stage designs which incorporate both model-robust and model-sensitive aspects in the design criterion. More specifically, we propose to extend the Bayesian two-stage strategy developed by Neff (1996) and also studied by Ruggoo and Vandebroek (2002) and incorporate the GD criterion of GKOV in both stages. The incorporation of the GD criterion in a two-stage procedure is intuitively very appealing: 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 that accounts for precision of the primary terms but also facilitates the improvement of the proposed models by detecting lack of fit. At the end of the first stage, the experimenter will thus have more information on the true model. The criterion in the second stage should also take into account precise estimation of important terms but most importantly, provide protection against bias induced by an incorrect primary model specification.

The paper will be organized as follows. In Section 2, the design criterion introduced by GKOV will be reviewed. In Section 3, we set the notation to be used in the paper and our two-stage approach is developed including the prior and posterior formulations on the model space. The approach of Neff (1996) to develop the Bayesian two-stage D-D optimal design is briefly reviewed in Section 4 and we show that her approach is a special case of our more general two-stage strategy. In Section 5, we propose an alternative approach for generating the first stage design that explicitly utilizes prior probabilities as weights in the optimality criterion. We discuss the analysis strategy of the two-stage approaches in Section 6, followed by an illustration of the two-stage procedures in Section 7. Our procedures are then evaluated relative to classical unique stage approaches and the two-stage designs of Neff (1996) in Section 8, and we end with a conclusion in Section 9. We show that our two-stage approaches produce designs with significantly smaller bias errors compared to standard designs used in the literature. They also improve the coverage over the factor space and still have very good variance properties of the assumed model.

2 The GD criterion

In this section we review the GD criterion developed by GKOV. We use a slightly different notation and we shall extend their approach in a two-stage design strategy in Sections 3 and 5.

Let us assume that the linear model that will be fitted by the experimenter is of the form

$$y = \mathbf{x}'_{pri}\boldsymbol{\beta}_{pri} + \varepsilon \tag{1}$$

with $\mathbf{x}_{pri}$ being a p-dimensional vector of powers and products of the factors and $\boldsymbol{\beta}_{pri}$ 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

$$y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon = \mathbf{x}'_{pri}\boldsymbol{\beta}_{pri} + \mathbf{x}'_{pot}\boldsymbol{\beta}_{pot} + \varepsilon = \eta(\mathbf{x}) + \varepsilon, \tag{2}$$

where $\mathbf{x}_{pot}$ is the q-dimensional vector containing powers and products of the factors not included in the fitted model and $\boldsymbol{\beta}_{pot}$ is the q-dimensional vector associated with the potential terms. We shall refer to $\mathbf{x}'_{pri}\boldsymbol{\beta}_{pri}$ as the primary terms and to $\mathbf{x}'_{pot}\boldsymbol{\beta}_{pot}$ as the potential terms. To simplify the notation, we will assume that the model has been reparametrized in terms of the orthonormal polynomials with respect to a measure μ 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.

If $\mathbf{X}_{pri}$ is the $n \times p$ model matrix for the primary terms and $\mathbf{X}_{pot}$ the $n \times q$ model matrix for the potential terms, then under the assumption of orthonormal polynomials, it can be shown that the expression for the IMSE, suggested by Box and Draper (1959) reduces to

$$\text{IMSE} = \boldsymbol{\beta}'_{pot} \left[\mathbf{A}'\mathbf{A} + \mathbf{I}_q\right] \boldsymbol{\beta}_{pot} + \sigma^2 \text{ trace} \left(\mathbf{X}'_{pri}\mathbf{X}_{pri}\right)^{-1},$$

where $\mathbf{A} = \left(\mathbf{X}'_{pri}\mathbf{X}_{pri}\right)^{-1} \mathbf{X}'_{pri}\mathbf{X}_{pot}$ is the alias matrix which essentially transcribes bias errors to parameter estimates, $\widehat{\boldsymbol{\beta}}_{pri}$. By conceiving a prior distribution of the form $\boldsymbol{\beta}_{pot} \sim N(\mathbf{0}, \tau^2\sigma^2\mathbf{I}_q)$ proposed by DMJ, GKOV establish that

$$\mathbf{E}_{\boldsymbol{\beta}}[\text{IMSE}] \;=\; \tau^2\sigma^2 \text{ trace} \left(\mathbf{A}'\mathbf{A} + \mathbf{I}_q\right) + \sigma^2 \text{ trace} \left(\mathbf{X}'_{pri}\mathbf{X}_{pri}\right)^{-1}.$$

The parameter τ^2 is the common prior variance of the potential terms' coefficients, measured in units of the random error variance σ^2. The approach above aims at finding designs that yield precise estimates of primary terms and ensures protection against the existence of potential terms. The possibility of testing for lack of fit is not made explicit. Atkinson and Donev (1992) consider this problem and combine the D-optimality criterion for the primary model and the D_s-optimality criterion for the potential terms. The D_s-optimality criterion for the potential terms is related to the non-centrality parameter, δ. Using the same prior distribution on the potential terms as before, GKOV show that

$$\mathbf{E}_{\boldsymbol{\beta}}[\delta] \;=\; \tau^2 \text{ trace}[\mathbf{L}],$$

where $\mathbf{L} = \mathbf{X}'_{pot}\mathbf{X}_{pot} - \mathbf{X}'_{pot}\mathbf{X}_{pri}\left(\mathbf{X}'_{pri}\mathbf{X}_{pri}\right)^{-1}\mathbf{X}'_{pri}\mathbf{X}_{pot}$, and is usually referred to as the dispersion matrix.

GKOV 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 α_2 and α_3 to attach more or less importance on the different properties. They propose to find designs that minimize the GD criterion

$$\text{GD: min} \left\{\frac{1}{p}\log\left|\left(\mathbf{X}'_{pri}\mathbf{X}_{pri}\right)^{-1}\right| + \frac{\alpha_2}{q}\log\left|\left(\mathbf{L} + \frac{\mathbf{I}_q}{\tau^2}\right)^{-1}\right| + \frac{\alpha_3}{q}\log\left|\mathbf{A}'\mathbf{A} + \mathbf{I}_q\right|\right\}. \qquad (3)$$

For $\alpha_2 = \alpha_3 = 0$ the GD-optimality criterion produces the D-optimal design for the primary model. For $\alpha_3 = 0$, $\alpha_2 = \frac{q}{p}$ and $\tau^2 = \infty$, we obtain the D-optimal design for the full model. Setting $\alpha_3 = 0, \alpha_2 = \frac{q}{p}$ and for finite values for τ^2, the Bayesian D-optimal designs introduced by DMJ 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. Let us assume that $\mathbf{y}_i|\boldsymbol{\beta} \sim N(\mathbf{X}_i\boldsymbol{\beta}, \sigma^2\mathbf{I})$ for each stage i $(i = 1, 2)$ with n_1 and n_2 observations in the first and second stage respectively so that the total number of observations in the combined stage is $n = n_1 + n_2$. $\mathbf{X}$ is the extended design matrix of dimension $n \times (p + q)$ for the combined stages, so that $\mathbf{X}' = [\ \mathbf{X}_1'\ \mathbf{X}_2'\ ]$. $\mathbf{X}_1 = [\ \mathbf{X}_{pri(1)}\ \mathbf{X}_{pot(1)}\ ]$ is of dimension $n_1 \times (p + q)$ and $\mathbf{X}_2 = [\ \mathbf{X}_{pri(2)}\ \mathbf{X}_{pot(2)}\ ]$ is of dimension $n_2 \times (p + q)$ and represent the first and second stage designs expanded to full model space. $\mathbf{X}_{pri(i)}$ and $\mathbf{X}_{pot(i)}$ correspond to the primary and potential terms respectively for each stage i $(i = 1, 2)$. Finally $\mathbf{X}_{pri}' = [\ \mathbf{X}_{pri(1)}'\ \mathbf{X}_{pri(2)}'\ ]$ is of dimension $n \times p$ and $\mathbf{X}_{pot}' = [\ \mathbf{X}_{pot(1)}'\ \mathbf{X}_{pot(2)}'\ ]$ is of dimension $n \times q$ and are respectively the combined first and second stage design matrices for the primary and potential terms models only.

3.1 Selection and Analysis of first stage design

In the first stage, the experimenter believes that the plausible model comprises primary terms but at the same time would like some knowledge about possible incorrect model specification. In other words he/she 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. The first stage design is thus obtained by finding $\mathbf{X}_1 = [\ \mathbf{X}_{pri(1)}\ \mathbf{X}_{pot(1)}\ ]$ which minimizes the GD criterion of GKOV with a large weight placed on the lack of fit component and setting $\alpha_3 = 0$ in (3). We obtain

$$\text{GD}_1: \ \min \ \left\{ \frac{1}{p} \log \left| \left(\mathbf{X}_{pri(1)}' \mathbf{X}_{pri(1)} \right)^{-1} \right| + \frac{\alpha_2}{q} \log \left| \left(\mathbf{L}_1 + \frac{\mathbf{I}_q}{\tau^2} \right)^{-1} \right| \right\}, \tag{4}$$

where $\mathbf{L}_1 = \mathbf{X}_{pot(1)}'\mathbf{X}_{pot(1)} - \mathbf{X}_{pot(1)}'\mathbf{X}_{pri(1)} \left(\mathbf{X}_{pri(1)}'\mathbf{X}_{pri(1)} \right)^{-1} \mathbf{X}_{pri(1)}'\mathbf{X}_{pot(1)}$.

Before observing the first stage data, the experimenter has a model with $(p+q)$ regressors. The total number of plausible models is $m = 2^q$. Consequently each candidate model M_i contains all primary terms and a subset of q_i $(0 \leq q_i \leq q)$ potential terms. Once the 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. Model knowledge can be updated by scoring each

of the plausible candidate models using posterior probabilities indicating the likelihood that a particular candidate model is actually predicting the response adequately. These resulting scores or posterior probabilities can then be incorporated as weights in a second stage criterion.

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 $\mathbf{y}_1$, the posterior probability of the model M_i given $\mathbf{y}_1$ is

$$p(M_i|\mathbf{y}_1) \propto p(M_i) f(\mathbf{y}_1|M_i), \tag{5}$$

where $p(M_i)$ is the prior probability of model M_i and $f(\mathbf{y}_1|M_i)$ is the predictive density of $\mathbf{y}_1$ given model M_i.

To develop our prior probabilities $p(M_i)$'s, we use the approach based on model building assumptions used for factor screening experiments suggested by Bingham and Chipman (2002). They consider screening experiments when prior information about the significance of some of the effects are known from expert knowledge. They classify the regression effects as a requirement set of effects that should be estimated and a negligible set of effects thought to be less important. In our context, this is akin to our regression effects being classified as the primary and potential effects. For a design with r factors, we consider only first and second order effects of a factor. We separate effects into three groups: linear, quadratic and interaction effects. Under the effect inheritance assumption, an interaction is more likely to be important if one or more of its parent factors are also important. Let $p_{x_ix_j,0} \leq p_{x_ix_j,1} \leq p_{x_ix_j,2}$ denote the conditional probabilities that interaction x_ix_j is active, given 0, 1, 2 of main effects x_i and x_j being active. Let $p_{x_i^2,0} \leq p_{x_i^2,1}$ denote the conditional probabilities that the quadratic effect ${x_i}^2$ is active given that the corresponding linear effect x_i is absent or present in the model.

In our model formulation, primary terms are always present in all $m = 2^q$ plausible models, so that we can assign a probability one for their occurrence in all m models. For effects in the potential set, we set the prior probability of a significant main effect to $\phi = 0.2$ and for the interactions and quadratic effects to be

$$p_{x_ix_j,s} = \begin{cases} 0.01\phi & \text{if} \quad s = 0 \\ 0.5\phi & \text{if} \quad s = 1 \\ \phi & \text{if} \quad s = 2 \end{cases} \qquad \text{and} \qquad p_{x_i^2,s} = \begin{cases} 0.01\phi & \text{if} \quad s = 0 \\ \phi & \text{if} \quad s = 1 \end{cases} \tag{6}$$

as suggested by Bingham and Chipman (2002). Given these prior probabilities, the prior prob-

ability of a particular model, $p(M_i)$, $(i = 1, 2, \ldots, m)$ can be computed as the product of the probability of each individual effect being in the model. These definitions above also imply that the event that a linear effect in a model is independent of the event that any other factor's linear effect is in the model. For a given set of linear effects being in the model, the inclusion of second order effects in the model are independent of each other (Chipman, 1996). The prior probabilities, $p(M_i)$'s can then be scaled so that they sum to one.

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 - that is an arbitrary prior mean and prior variance tending to infinity. On the other hand, potential terms are unlikely to have huge effects and the assumption $\boldsymbol{\beta}_{pot} \sim N(\mathbf{0}, \tau^2\sigma^2\mathbf{I}_q)$ proposed by DMJ and also used by GKOV is appropriate. Following our orthonormalization procedure, which ensures that the effects are well separable and independent, the joint prior distribution assigned to $\boldsymbol{\beta}_{pri}$ and $\boldsymbol{\beta}_{pot}$ is $N(\mathbf{0}, \sigma^2\tau^2\mathbf{K}^{-1})$ 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. Since we have assumed a normal linear model, the probability density of $\mathbf{y}_1$ given M_i and $\boldsymbol{\beta}_i$ is given by

$$f(\mathbf{y}_1|M_i, \boldsymbol{\beta}_i) \propto \sigma^{-n_1}\exp\left[-(\mathbf{y}_1 - \mathbf{X}_i\boldsymbol{\beta}_i)'(\mathbf{y}_1 - \mathbf{X}_i\boldsymbol{\beta}_i)/2\sigma^2\right].$$

The resulting posterior probability for model M_i given $\mathbf{y_1}$ can then be obtained along the lines shown in Box and Meyer (1993) and (5) becomes

$$p(M_i|\mathbf{y}_1) = C\, p(M_i)\, \tau^{-q_i} \left|\mathbf{X}_i'\mathbf{X}_i + \frac{\mathbf{K}_i}{\tau^2}\right|^{-1/2} \left(\mathbf{S}(\widehat{\boldsymbol{\beta}}_i) + \frac{1}{\tau^2}\,\widehat{\boldsymbol{\beta}}_i'\,\mathbf{K}_i\,\widehat{\boldsymbol{\beta}}_i\right)^{-(n_1-1)/2}, \tag{7}$$

where $\mathbf{X}_i$ is the first stage design in model M_i space and

$$\mathbf{K}_i = \begin{bmatrix} \mathbf{0}_{p\times p} & \mathbf{0}_{p\times q_i} \\ \mathbf{0}_{q_i\times p} & \mathbf{I}_{q_i\times q_i} \end{bmatrix},$$

$$\widehat{\boldsymbol{\beta}}_i = \left(\mathbf{X}_i'\mathbf{X}_i + \frac{\mathbf{K}_i}{\tau^2}\right)^{-1}\mathbf{X}_i'\mathbf{y}_1 = E(\boldsymbol{\beta}_i|\mathbf{y}_1), \text{ assuming model } M_i,$$

$$\mathbf{S}(\widehat{\boldsymbol{\beta}}_i) = (\mathbf{y}_1 - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_i)'(\mathbf{y}_1 - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_i) = \text{Residual Sum of Squares for model } M_i$$

and finally C is the normalization constant that forces all probabilities to sum to one.

With our choice of the parameters α_2 and α_3 in the first stage, we expect the first stage design, $\mathbf{X}_1$ to have the power to diagnose any model inadequacy and reflect knowledge on the true

model. The Box and Meyer probabilities in (7) will also capture model importance and the true model will 'enjoy' the highest posterior probability.

3.2 Selection of second stage design

We shall now incorporate all the improved model information from the first stage in selecting the second stage design. Since $\boldsymbol{\beta}$ contains all $(p+q)$ parameters of the full model, we can extend the approach of Neff (1996) so that a Bayesian second stage GD optimal design for the full model is found by choosing $\mathbf{X}_2 = [\, \mathbf{X}_{pri(2)} \; \mathbf{X}_{pot(2)} \,]$ so as to minimize

$$\begin{aligned} \mathrm{GD}_2 : \min \; \Bigg[\; & \frac{1}{p} \log \left| \left(\mathbf{X}'_{pri(1)} \mathbf{X}_{pri(1)} + \mathbf{X}'_{pri(2)} \mathbf{X}_{pri(2)} \right)^{-1} \right| \\ & + \frac{\alpha_2}{q} \log \left| \left(\mathbf{L}_2 + \frac{\mathbf{I}_q}{\tau^2} \right)^{-1} \right| + \frac{\alpha_3}{q} \log \left| \mathbf{A}'_2 \mathbf{A}_2 + \mathbf{I}_q \right| \Bigg], \end{aligned} \tag{8}$$

where

$$\begin{aligned} \mathbf{L}_2 \; = \; \Bigg[& \left(\mathbf{X}'_{pot(1)} \mathbf{X}_{pot(1)} + \mathbf{X}'_{pot(2)} \mathbf{X}_{pot(2)} \right) - \left(\mathbf{X}'_{pot(1)} \mathbf{X}_{pri(1)} + \mathbf{X}'_{pot(2)} \mathbf{X}_{pri(2)} \right) \times \\ & \left(\mathbf{X}'_{pri(1)} \mathbf{X}_{pri(1)} + \mathbf{X}'_{pri(2)} \mathbf{X}_{pri(2)} \right)^{-1} \left(\mathbf{X}'_{pri(1)} \mathbf{X}_{pot(1)} + \mathbf{X}'_{pri(2)} \mathbf{X}_{pot(2)} \right) \Bigg], \end{aligned}$$

and

$$\mathbf{A}_2 = \left(\mathbf{X}'_{pri(1)} \mathbf{X}_{pri(1)} + \mathbf{X}'_{pri(2)} \mathbf{X}_{pri(2)} \right)^{-1} \left(\mathbf{X}'_{pri(1)} \mathbf{X}_{pot(1)} + \mathbf{X}'_{pri(2)} \mathbf{X}_{pot(2)} \right).$$

However, the full model is only one of the candidate models and in most cases not the most appropriate. Based on our judicious choice of α_2 and α_3 in the first stage, we expect the design at the end of the first stage to be able to discriminate between the different plausible models and hence reflect the most likely terms in the true model. The experimenter will still be interested in precise estimation of primary terms in the second stage but would now like to minimize bias caused by any of the potential terms which may now be present in the true unknown model. The design criterion in a second stage needs to give a high weight to the bias component and also ensure precision of the primary terms as these are the terms that the experimenter will eventually fit. An obvious choice would be to set $\alpha_2 = 0$ and use a large value of α_3 in (8) for generating the second stage design.

Let us consider our subset models $M_0, M_1, \ldots, M_m$ as discussed previously, with each model M_k defined by its parameters $\boldsymbol{\beta}_k$. A Bayesian second stage GD optimal design for model M_k is the

set of design points $\mathbf{X}_2^{(k)} = \left[\mathbf{X}_{pri(2)}^{(k)} \ \mathbf{X}_{pot(2)}^{(k)} \right]$ which minimizes

$$\mathrm{GD}_2^{(k)}: \ \min \left[\frac{1}{p} \log \left| \left(\mathbf{X}_{pri(1)}^{(k)'} \mathbf{X}_{pri(1)}^{(k)} + \mathbf{X}_{pri(2)}^{(k)'} \mathbf{X}_{pri(2)}^{(k)} \right)^{-1} \right| + \frac{\alpha_3}{q} \log \left| \mathbf{A}_2^{(k)'} \mathbf{A}_2^{(k)} + \mathbf{I}_q^{(k)} \right| \right], \quad (9)$$

where $\mathbf{X}_{pri(1)}^{(k)}$, $\mathbf{X}_{pri(2)}^{(k)}$, $\mathbf{A}_2^{(k)}$ and $\mathbf{I}_q^{(k)}$ are the matrices corresponding to $\mathbf{X}_{pri(1)}$, $\mathbf{X}_{pri(2)}$, $\mathbf{A}_2$ and $\mathbf{I}_q$ expanded to model space $\mathbf{M}_k$. Since the Box and Meyer posterior probabilities computed from first stage data in (7) reflect model importance, they can be incorporated as weights to average the GD criterion when the second stage is selected. The objective is to choose the second stage design points so as to minimize $\mathrm{GD}_2^{(k)}$ for each model M_k having a high probability of being the 'best' model. This is achieved by choosing the second stage design points $\mathbf{X}_2$ so as to minimize

$$\sum_{M_k} \mathrm{GD}_2^{(k)} \, p(M_k|\mathbf{y}_1).$$

We shall refer to the two-stage approach developed in this Section as the Bayesian GD-MGD two-stage procedure; the acronym MGD being used to enforce the analogy that all possible models are taken into account in the second stage.

4 Comparison of two-stage procedures

In this section, we give an overview of the development of the Bayesian D-D optimal design for linear models proposed by Neff (1996) and show that our two-stage strategy developed in Sections 3.1 and 3.2, generalize her approach. Neff's (1996) procedure is as follows: The Bayesian D-optimality criterion of DMJ which minimizes $|(\mathbf{X}_1'\mathbf{X}_1 + \frac{\mathbf{K}}{\tau^2})^{-1}|$, is used to select the first stage design. The parameter $\tau = 5$ is recommended in both the first and second stage because of the ability to produce designs which are robust to model misspecification. By letting the second stage prior distribution of $\boldsymbol{\beta}$ be the first stage posterior, a Bayesian D-optimal design for the full model in the second stage is found by choosing $\mathbf{X}_2$ so as to minimize $|(\mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_2'\mathbf{X}_2 + \frac{\mathbf{K}}{\tau^2})^{-1}|$. However as argued in Section 3.2, the full model is only one of the candidate models and in most cases not the most appropriate. Considering the subset models $M_0, M_1, \ldots, M_m$ as discussed previously, the posterior variance of $\boldsymbol{\beta}_i$ is

$$\mathbf{V}_{2(i)} = \sigma^2 \left(\mathbf{X}_{1(i)}'\mathbf{X}_{1(i)} + \mathbf{X}_{2(i)}'\mathbf{X}_{2(i)} + \frac{\mathbf{K}_i}{\tau^2} \right)^{-1},$$

where $\mathbf{X}_{1(i)}$ and $\mathbf{X}_{2(i)}$ are the first stage and second stage design matrices respectively expanded to model space M_i. A Bayesian D-D optimal design for model M_i is the set of design points $\mathbf{X}_{2(i)}$ which minimizes D_i=$|\mathbf{V}_{2(i)}|$. Since the posterior Box and Meyer probabilities computed from

first stage data as described before reflect model importance, they are incorporated as weights as in Section 3.2 so that the second stage design points $\mathbf{X}_2$ are obtained by minimizing

$$\sum_{M_i} D_i \; p(M_i|\mathbf{y}_1).$$

It can be easily established that our criteria in (4) and (8) generalize the first and second stage designs proposed by Neff (1996). We shall for that purpose use the results from Harville (1997), that if $\mathbf{T}$ represent an $r \times r$ matrix and $\mathbf{U}$ is an $r \times t$ matrix, $\mathbf{V}$ an $t \times r$ and $\mathbf{W}$ is an $t \times t$ matrix and if $\mathbf{T}$ is non-singular, then

$$\begin{vmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{vmatrix} = \begin{vmatrix} \mathbf{W} & \mathbf{V} \\ \mathbf{U} & \mathbf{T} \end{vmatrix} = |\mathbf{T}| \left|\mathbf{W} - \mathbf{V}\mathbf{T}^{-1}\mathbf{U}\right|. \tag{10}$$

To obtain the first stage design developed by Neff (1996), we set $\alpha_2 = \frac{q}{p}$ in (4) so that

$$\text{GD}_1^{\text{Neff}}: \;\; \min \left\{ \frac{1}{p} \log \left| \left(\mathbf{X}'_{pri(1)} \mathbf{X}_{pri(1)} \right)^{-1} \right| + \frac{1}{p} \log \left| \left(\mathbf{L}_1 + \frac{\mathbf{I}_q}{\tau^2} \right)^{-1} \right| \right\}.$$

Using (10), we can easily show that

$$\left|\mathbf{X}'_{pri(1)} \mathbf{X}_{pri(1)}\right| \left|\mathbf{L}_1 + \frac{\mathbf{I}_q}{\tau^2}\right| = \left|\mathbf{X}'_1 \mathbf{X}_1 + \frac{\mathbf{K}}{\tau^2}\right|,$$

which yields the first stage design of Neff (1996).

The Bayesian second stage D-optimal design developed by Neff (1996) for the full model is found by choosing $\mathbf{X}_2$ so as to minimize $|(\mathbf{X}'_1\mathbf{X}_1 + \mathbf{X}'_2\mathbf{X}_2 + \frac{\mathbf{K}}{\tau^2})^{-1}|$. For finite values of τ^2 and setting $\alpha_2 = \frac{q}{p}$ and $\alpha_3 = 0$ in (8), we obtain

$$\text{GD}_2^{\text{Neff}}: \;\; \min \left[\frac{1}{p} \log \left| \left(\mathbf{X}'_{pri(1)} \mathbf{X}_{pri(1)} + \mathbf{X}'_{pri(2)} \mathbf{X}_{pri(2)} \right)^{-1} \right| + \frac{1}{p} \log \left| \left(\mathbf{L}_2 + \frac{\mathbf{I}_q}{\tau^2} \right)^{-1} \right| \right].$$

Using (10), we can again show that

$$\left|\mathbf{X}'_{pri(1)} \mathbf{X}_{pri(1)} + \mathbf{X}'_{pri(2)} \mathbf{X}_{pri(2)}\right| \left|\mathbf{L_2} + \frac{\mathbf{I}_q}{\tau^2}\right| = \left|\mathbf{X}'_1 \mathbf{X}_1 + \mathbf{X}'_2 \mathbf{X}_2 + \frac{\mathbf{K}}{\tau^2}\right|. \tag{11}$$

The expression on the right hand side of (11) is identical to the Bayesian second stage D-optimal design criterion of Neff (1996) for the full model.

5 An alternative approach for generating the first stage design utilizing prior probabilities

In Section 3, information on the prior model space was not used in the first stage and only the full model was considered when selecting the first stage design. In this section, we present an alternative approach for designing the first stage experiment that explicitly incorporates prior model information and accounts for model uncertainty by considering running the GD optimality criterion over all possible models in the first stage itself. As in Section 3.1, the experimenter has a model with $(p+q)$ regressors and the total number of plausible models is 2^q. Consequently each candidate model M_i contains all primary terms and a subset of q_i $(0 \leq q_i \leq q)$ potential terms. Now as described in Section 3.1, the prior probabilities, $p(M_i)$ reflect model importance of each of the 2^q plausible models and thus can be incorporated as weights to average the GD criterion when a first stage design is selected similar to the approach utilizing posterior probabilities in Section 3.2. Thus the first stage design $\mathbf{X}_1 = [\, \mathbf{X}_{pri(1)}\ \mathbf{X}_{pot(1)}\,]$ can be obtained by minimizing

$$\sum_{M_k} \mathrm{GD}_1^{(k)}\, p(M_k)$$

where

$$\mathrm{GD}_1^{(k)}: \ \min \left[\frac{1}{p} \log \left| \left(\mathbf{X}_{pri(1)}^{(k)\prime}\, \mathbf{X}_{pri(1)}^{(k)} \right)^{-1} \right| + \frac{\alpha_2}{q} \log \left| \left(\mathbf{L}_1^{(k)} + \frac{\mathbf{I}_q^{(k)}}{\tau^2} \right)^{-1} \right| \right], \tag{12}$$

and a large weight placed on the lack of fit component to increase knowledge on the true model. $\mathbf{X}_{pri(1)}^{(k)}$, $\mathbf{L}_1^{(k)}$ and $\mathbf{I}_q^{(k)}$ are the matrices corresponding to $\mathbf{X}_{pri(1)}$, $\mathbf{L}_1$ and $\mathbf{I}_q$ expanded to model space $\mathbf{M}_k$. Once the first stage design $\mathbf{X}_1$ is obtained, the second stage design $\mathbf{X}_2$ can be obtained along the same procedure as in Section 3.2.

The procedure above makes sense as the prior model probabilities $p(M_i)$ are explicitly used in the first stage and the optimality criterion accounts for all possible models. Once data is collected from the first stage, the classical Bayes' Theorem updates these priors to the posteriors, $p(M_i|\mathbf{y}_1)$ for use in the second stage. In essence the two-stage process is summarised as

$$\sum_{M_k} \mathrm{GD}_1^{(k)}\, p(M_k) \quad \Rightarrow \quad \sum_{M_k} \mathrm{GD}_2^{(k)}\, p(M_k|\mathbf{y}_1)$$

We shall refer to this approach as a Bayesian MGD-MGD two-stage procedure; the acronym MGD again enforcing the analogy that the optimality criteria sweeps over all possible models in both stages.

6 Analysis Strategy for the two-stage approaches

In a two-stage design strategy, the first stage is designed with respect to some criteria and then conditional on the information provided by first stage data, the second stage is chosen to create certain desirable conditions in the combined design. Statistical inferences are then based on all the observations as if the experiment had been completed in a single stage. The two-stage procedures described in Sections 3 and 5 are mainly used as devices for generating a combined design with less dependence on the choice of the primary terms. The experimenter will eventually fit the primary model (1) by least squares but with the combined two-stage design, departures of the response estimator $\widehat{y}$ from the true response $\eta(\mathbf{x})$ will be minimized resulting in genuine and less biased predictions. If the experimenter wishes to have more knowledge on the process, the effect of potential terms can be investigated by forward stepwise selection or other regression diagnostics as suggested by DMJ.

7 Illustration of the two-stage procedures

In this section we present a simple example of the two-stage procedures developed in Sections 3 and 5. Consider the two-dimensional problem where the primary model consists of $p = 4$ terms, $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$ and the full model has $q = 2$ extra potential terms: $\beta_{11} x_1^2 + \beta_{22} x_2^2$. The design region we consider is the 5×5 grid on $[-1, +1]^2$. Suppose the experimenter has resources for 16 runs in the experiment. In case model uncertainty is completely ignored, the experimenter can design the experiment using a classical D-optimal design for the primary terms model only. Alternatively if he/she wants to protect against the potential terms, then the design procedure of DMJ which is obtained by minimizing $\left|(\mathbf{X}'\mathbf{X} + \mathbf{K})^{-1}\right|$ can be used.

As an alternative approach to the single stage procedures above, we may also design the experiment using the two-stage approaches developed in Sections 3 and 5. Since the second stage design is dependent on the first stage, response data from the first stage experiment are needed in the computation of the posterior probabilities used as weights in the second stage criterion. We assume that the true model from which data will be simulated is

$$y = 45.0 + 11.5\, x_1 + 12.8\, x_2 + 13.6\, x_1 x_2 - 7.4\, x_1^2 + \varepsilon. \tag{13}$$

Note that the true model comprises all primary terms and one potential term, namely the quadratic effect of x_1. Parameters of the true model reflect more importance on the primary

terms as these are the terms that the experimenter will eventually fit. We assume $\varepsilon \sim N(0,1)$ and the illustration will be for one simulation only. Other simulations showed similar first stage and combined designs. We assume an equal partition in the two stages so that $n_1 = 8$ and $n_2 = 8$. The choices of α_2 and α_3 in the two stages follow the discussions in Section 3. In the first stage, since the true model is unknown, values of $\alpha_2 = 20$, i.e. a large weight on the lack of fit component, and $\alpha_3 = 0$ are appropriate. In the second stage, since bias reduction will now be most important, we use $\alpha_2 = 0$ and $\alpha_3 = 10$. $\tau = 5$ is used in both stages for the GD-MGD approach and $\tau = 1$ in the two-stages of the MGD-MGD procedure. Further justification for these choices of the parameters and distribution of sample sizes in both stages are provided in Section 8.1.

In Table 1, we also present the prior and posterior probabilities on the model space for all possible models computed from (6) and (7) respectively. The primary terms model has highest prior probability as this is the model which the experimenter had certainty on before collecting first stage data. But once first stage data is obtained, as expected, the true model (13) 'enjoys' the highest posterior probability. The values of the different determinants of the GD criterion

Table 1: Prior and Posterior model probabilities

Terms in M_i	$p(M_i)$ using (6), Section 3.1	Scaled $p(M_i)$	$p(M_i\|\mathbf{y}_1)$ (GD-MGD)	$p(M_i\|\mathbf{y}_1)$ (MGD-MGD)
$1\ x_1\ x_2\ x_{12}$	1	0.69444	0	0.00335
$1\ x_1\ x_2\ x_{12}\ x_1^2$ (True Model)	0.2	0.13889	0.93466	0.86550
$1\ x_1\ x_2\ x_{12}\ x_2^2$	0.2	0.13889	0	0.00538
$1\ x_1\ x_2\ x_{12}\ x_1^2\ x_2^2$	0.2×0.2	0.02778	0.06534	0.12577

in (3) will be used as measures of efficiency of the precision, lack of fit and bias components. The measure of precision of the primary terms is given by $\mathrm{D}^*_{\mathrm{X_{pri}}} = \left|\mathbf{X}^{*'}_{pri}\ \mathbf{X}^*_{pri}\right|^{-1/p}$, a measure of the lack of fit component is $\mathrm{D}^*_{\mathrm{lof}} = |\mathbf{L}^*|^{-1/q}$ and $\mathrm{D}^*_{\mathrm{bias}} = \left|\mathbf{A}^{*'}\mathbf{A}^* + \mathbf{I}_q\right|^{1/q}$ represents the degree of bias, where

$$\mathbf{L}^* = \mathbf{X}^{*'}_{pot}\mathbf{X}^*_{pot} - \mathbf{X}^{*'}_{pot}\mathbf{X}^*_{pri}\left(\mathbf{X}^{*'}_{pri}\mathbf{X}^*_{pri}\right)^{-1}\mathbf{X}^{*'}_{pri}\mathbf{X}_{pot} \quad \text{and} \quad \mathbf{A}^* = \left(\mathbf{X}^{*'}_{pri}\mathbf{X}^*_{pri}\right)^{-1}\mathbf{X}^{*'}_{pri}\mathbf{X}^*_{pot}.$$

$\mathbf{X}^*_{pri}$ and $\mathbf{X}^*_{pot}$ represent the design points for the primary and potential terms expanded to contain regressors in the true model only. Note that the minimum bias design arises when the alias matrix, $\mathbf{A}^* = \mathbf{0}$ and consequently $\mathrm{D}^*_{\mathrm{bias}} = |\mathbf{I}_q|^{1/q} = 1$, irrespective of the number of potential terms in the true model. Further $\mathrm{D}^*_{\mathrm{X_{pri}}}$, $\mathrm{D}^*_{\mathrm{lof}}$ and $\mathrm{D}^*_{\mathrm{bias}}$ have been defined such that the smaller

the value obtained, the better the design performs with respect to that criterion.

Figure 1, shows the D-optimal design for the primary terms model and that of DMJ for 8 and 16 runs. As expected the D-optimal design spans at the extremes of the design region and have large bias components. The DMJ designs have improved coverage over the factor space, better bias and lack of fit properties than the D-optimal designs. The designs of run size 8 in Panels 1 and 3 are included to enable comparison with the first stage design in the two-stage procedures. Figure 2, shows the first stage and combined stage designs for both the GD-MGD and MGD-MGD approaches developed in Sections 3 and 5. The lack of fit properties in the first stage for both approaches in Panels 1 and 3 are excellent suggesting that the design will have good discriminating properties and diagnose any model inadequacy. For the combined design, both approaches result in drastic reductions in the bias effect compared to those of the D-optimal and DMJ designs in Panels 3 and 4 of Figure 1 and still have good variance properties of the assumed model. The MGD-MGD procedure gives the lowest bias component compared to all the design procedures and good coverage over the factor space in both the first and second stage. This example suggest that the two-stage approaches developed work well and the resulting design has good variance properties, lack of fit and excellent bias properties with respect to potential terms. They also have improved coverage over the factor space.

Using the D-optimal, DMJ, GD-MGD and MGD-MGD designs in Panels 2 and 4 of Figures 1 and 2 respectively, 16 observations were simulated from the true model (13), assuming as before that $\varepsilon \sim N(0,1)$. Since the experimenter will eventually fit the primary terms model only, these 16 runs corresponding to each design were used to fit the primary model (1). We can then for each design obtain the predicted values, $\widehat{y} = \mathbf{x}_{pri}^{'} \widehat{\boldsymbol{\beta}}_{pri}$.

The predicted values for the range of values $-1 \leq x_1 = x_2 \leq +1$ for the four design scenarios are plotted in Figure 3. The actual true model (13) is also included in the Figure. The differences or vertical deviations from the true assumed model represents "model bias" values over the range of x-values. In case of the D-optimal design, the endpoints results in rather small bias compared to the other design scenarios and as expected large bias errors occur near the design center as there is no data available there. In case of the DMJ and the two-stage designs moderate errors are revealed at the design center and the design perimeter with the most reduction in bias occurring with the two-stage procedures. This is so as the two-stage designs have improved coverage over the design region.

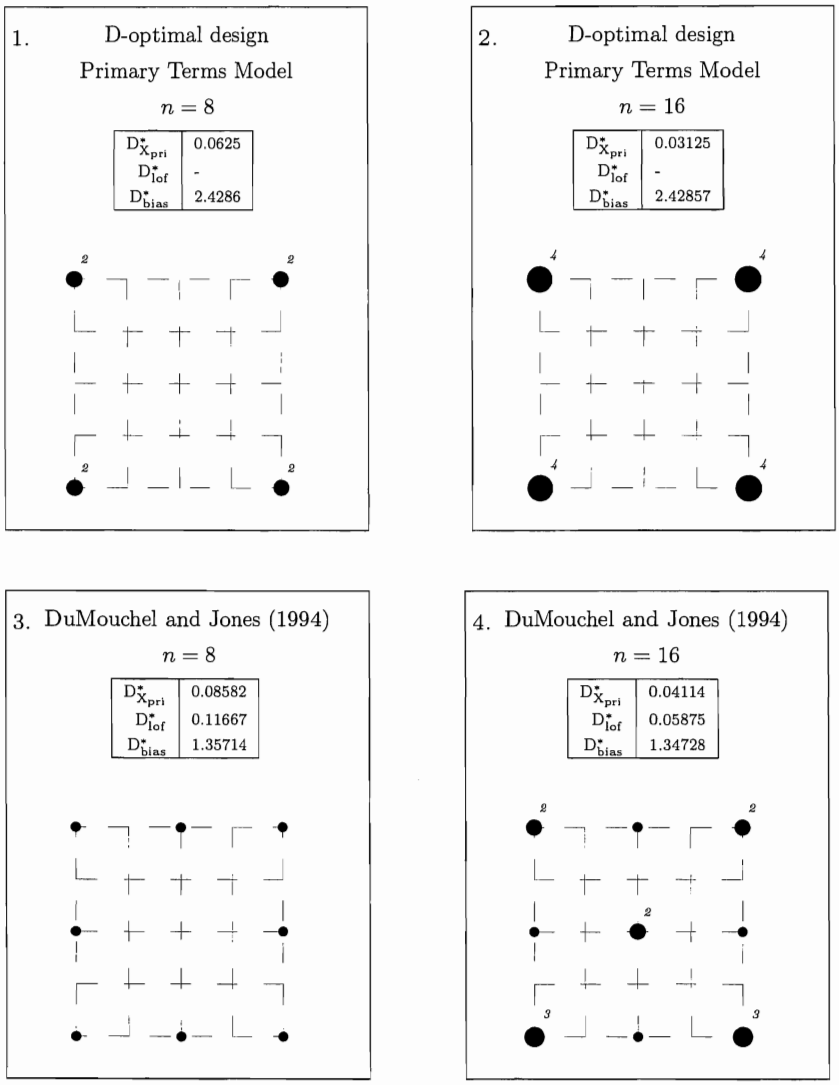
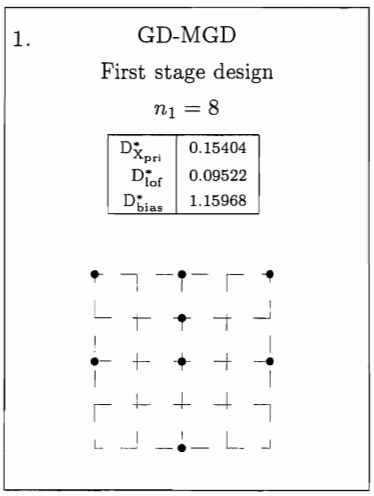
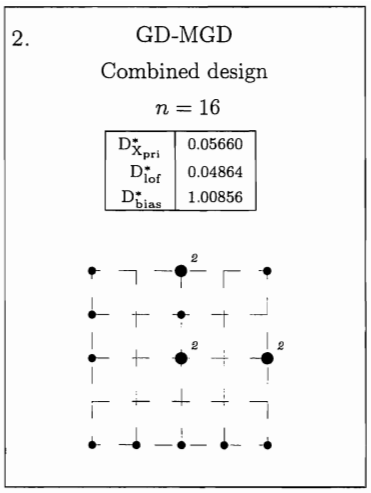
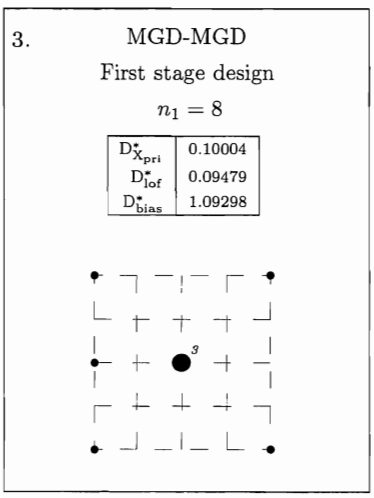


Figure 1: D-optimal and DMJ designs for 8 and 16 runs







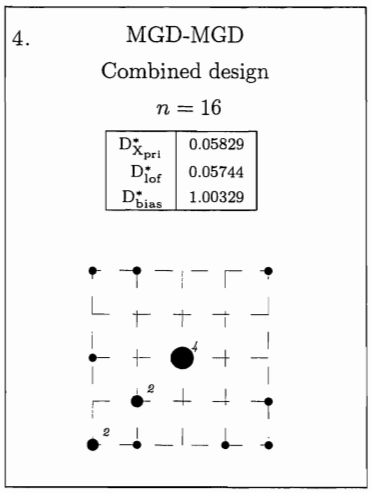


Figure 2: GD-MGD and MGD-MGD designs for the first and combined stage

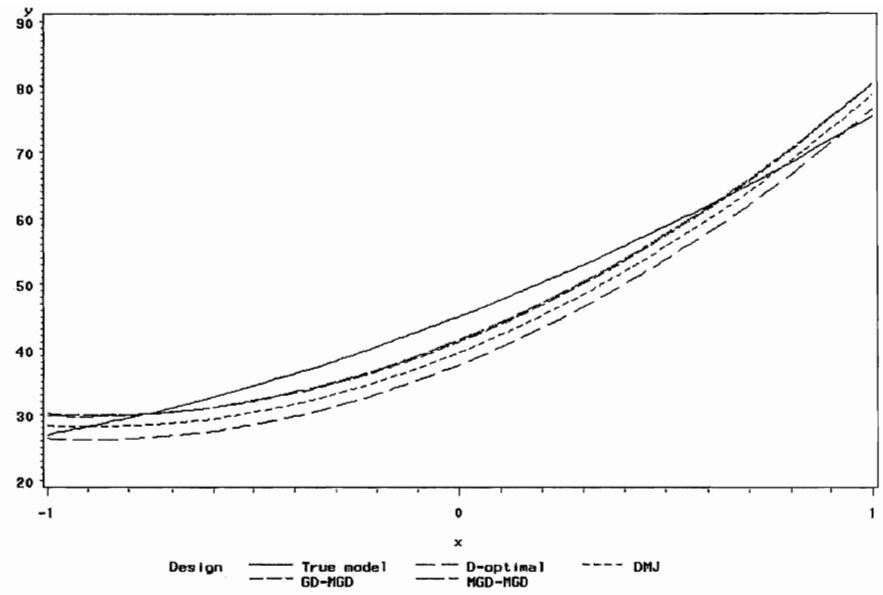


Figure 3: Predicted values under different design scenarios.

8 Further evaluation of the two-stage procedures

The performance of the Bayesian two-stage optimal designs presented in Sections 3 and 5 will be evaluated over several more simulations using additional examples, relative to the classical one-stage designs and the procedure of Neff (1996) presented in Section 4. Since the second stage design of both our procedures and that of Neff (1996) are dependent on first stage data through the Box and Meyer posterior probabilities, we need to evaluate their performance via a simulation approach. The performance of each design will be measured by its efficiency relative to a true assumed model in the simulations. 200 simulations will be performed and each will produce first stage data and consequently the posterior probabilities for use as the measure of fit in selecting the second stage design. The error $\varepsilon \sim N(0, 1)$ is assumed in all the simulations. The unique stage competitors to the Bayesian two-stage optimal design are the traditional D-optimal design for the primary terms model and the Bayesian D-optimal design of DMJ.

As in Section 7, the values of the different determinants of the GD criterion in (3) will be used as measures of efficiency of the precision, lack of fit and bias components of the different designs. The performance of the two-stage procedures are then measured by the average of $D^*_{X_{pri}}$, D^*_{lof} and D^*_{bias} over the different 200 simulations, i.e.

$$AD^*_{X_{pri}} = \frac{\sum_{i=1}^{200} D^*_{X_{pri}}}{200}, \qquad AD^*_{lof} = \frac{\sum_{i=1}^{200} D^*_{lof}}{200}, \qquad AD^*_{bias} = \frac{\sum_{i=1}^{200} D^*_{bias}}{200}.$$

The one-stage traditional non-Bayesian D-optimal design and one-stage Bayesian D-optimal design of DMJ are not data dependent and can thus be evaluated over the n design runs by the single measures $D^*_{X_{pri}}$, D^*_{lof} and D^*_{bias} for the true model.

8.1 Preliminary Evaluations for choice of parameters and sample sizes

Before making recommendations on the choice of the parameters τ, α_2, α_3 and sample sizes in the two-stage procedure, it was necessary to assess the performance of the two-stage designs with several different values of these implicit parameters in a simulation approach. In connection with sample sizes for each stage, we shall follow the recommendations of Neff and Myers (1998), Lin, Myers and Ye (2000) and Ruggoo and Vandebroek (2002) who suggest that efficiency and robustness is gained from a two-stage design of size $n = 2(p + q + 2)$ with half of the design points allocated to each stage of the design. Based on our various simulation studies, it is also recommended to use a value of $\tau = 5$ in the GD-MGD approach and a value of $\tau = 1$ in the MGD-MGD approach in both stages. A good default choice for the weight of the lack of fit component in the first stage is $\alpha_2 = 20$. In the second stage, $\alpha_3 = 10$ gives reliable and good results. Simulation studies undertaken with larger values of α_3 in the second stage do result in some further reduction of the bias component but it takes a toll on the precision component which increases rapidly. As was pointed out earlier, a combined design approach would necessarily be a trade-off between the different components and these values of τ, α_2, α_3 are recommended for their ability to produce satisfactory designs with respect to a combined criterion involving precision, lack of fit and bias properties.

8.2 Evaluation Phase

We consider the following cases for our evaluation purposes. The design region that we consider is the $5 \times 5 \times 5$ grid on $[-1, +1]^3$.

Case I :

Let the primary model under consideration for the numerical evaluation be defined with $p = 5$ terms, $\mathbf{x}^{(\text{pri})} = \{1,\ x_1,\ x_2,\ x_3,\ x_1^2\}$. Suppose that the expected response was possibly misspecified and the full model comprises an additional $q = 3$ potential terms, $\mathbf{x}^{(\text{pot})} = \{x_1x_2,\ x_2^2,\ x_3^2\}$. For simulation purposes, we shall assume that the true model consists of the five primary terms and one of the three potential terms so that

$$y = 42.0 + 11.5\ x_1 + 12.8\ x_2 + 10.5\ x_3 + 14.6\ x_1^2 - 7.4\ x_2^2 + \varepsilon.$$

Case II :

In this case, $p = 5$ primary terms, $\mathbf{x}^{(\text{pri})} = \{1,\ x_1,\ x_2,\ x_3,\ x_1x_2\}$ and the misspecified full model comprises an additional $q = 4$ potential terms, $\mathbf{x}^{(\text{pot})} = \{x_1^2,\ x_1x_3,\ x_2^2,\ x_3^2\}$. First stage data is simulated from the true model with the five primary terms and two of the four potential terms as below

$$y = 42.0 + 11.2\ x_1 + 14.5\ x_2 + 10.6\ x_3 + 12.5\ x_1x_2 + 8.9\ x_1^2 - 9.9\ x_1x_3 + \varepsilon.$$

Case III :

Finally the model we consider is with $(p + q) = 10$ terms comprising 5 primary terms namely, $\mathbf{x}^{(\text{pri})} = \{1,\ x_1,\ x_2,\ x_3,\ x_1^2\}$ and 5 potential terms, $\mathbf{x}^{(\text{pot})} = \{x_1x_2,\ x_1x_3,\ x_2x_3,\ x_2^2,\ x_3^2\}$. For the simulation purposes, the true model has $p = 5$ primary and $q = 3$ potential terms

$$y = 40.0 + 11.5\ x_1 + 12.8\ x_2 + 10.5\ x_3 + 14.6\ x_1^2 + 9.8\ x_1x_2 - 7.4\ x_1x_3 - 8.7\ x_2^2 + \varepsilon.$$

In all the above cases, primary terms are believed to be important and are assigned larger coefficients than potential terms assumed to be unity in the simulated data. The results of the evaluations are shown in Tables 2 to 4. As expected, the D-optimal designs have the most desirable precision characteristics but the worst bias. The designs of DMJ and the two-stage approach of Neff (1996) allow for testing lack of fit and results in some reduction of the bias. In case of our two-stage approaches, the small loss in precision in all cases is compensated by a drastic reduction in the bias component. As argued in Section 7, the minimum possible bias is one, so that our two-stage approaches perform excellently with respect to the bias and outperforms the D-optimal designs and those proposed by DMJ and Neff (1996), whilst still maintaining very good precision of the primary terms. Both the GD-MGD and MGD-MGD

Table 2: Evaluation of the two-stage designs and single stage competitors

Case I $y = 42.0 + 11.5\ x_1 + 12.8\ x_2 + 10.5\ x_3 + 14.6\ x_1^2 - 7.4\ x_2^2 + \varepsilon$.

Two-Stage Approach ($n_1 = n_2 = 10$)	$AD^*_{X_{pri}}$	AD^*_{lof}	AD^*_{bias}
GD-MGD	0.046308	0.046103	1.006845
MGD-MGD	0.046084	0.046428	1.004525
Neff (1996)	0.039017	0.039151	1.113426
One-Stage Approach ($n = 20$)	$D^*_{X_{pri}}$	D^*_{lof}	D^*_{bias}
D-optimal (Primary Terms)	0.034299	-	2.428570
DMJ	0.038914	0.049374	1.279301

Table 3: Evaluation of the two-stage designs and single stage competitors

Case II $y = 42.0 + 11.2\ x_1 + 14.5\ x_2 + 10.6\ x_3 + 12.5\ x_1x_2 + 8.9\ x_1^2 - 9.9\ x_1x_3 + \varepsilon$.

Two-Stage Approach ($n_1 = n_2 = 11$)	$AD^*_{X_{pri}}$	AD^*_{lof}	AD^*_{bias}
GD-MGD	0.037508	0.040165	1.009899
MGD-MGD	0.036782	0.036739	1.008554
Neff (1996)	0.027304	0.028956	1.222622
One-Stage Approach ($n = 22$)	$D^*_{X_{pri}}$	D^*_{lof}	D^*_{bias}
D-optimal (Primary Terms)	0.022887	-	1.581590
DMJ	0.02958	0.031216	1.273629

Table 4: Evaluation of the two-stage designs and single stage competitors

Case III $y = 40.0 + 11.5\ x_1 + 12.8\ x_2 + 10.5\ x_3 + 14.6\ x_1^2 + 9.8\ x_1x_2 - 7.4\ x_1x_3 - 8.7\ x_2^2 + \varepsilon.$

Two-Stage Approach ($n_1 = n_2 = 12$)	$AD^*_{X_{pri}}$	AD^*_{lof}	AD^*_{bias}
GD-MGD	0.035264	0.031280	1.005077
MGD-MGD	0.037010	0.031256	1.006440
Neff (1996)	0.031655	0.021330	1.089927
One-Stage Approach ($n = 24$)	$D^*_{X_{pri}}$	D^*_{lof}	D^*_{bias}
D-optimal (Primary Terms)	0.028421	-	1.344158
DMJ	0.031606	0.023785	1.135410

perform well and can be recommended to generate two-stage designs with reduced dependence on model uncertainty. The MGD-MGD procedure gives the most desirable variance and bias characteristics for the combined design in Cases I and II. Based on our extensive simulations carried out, our preference would be to use the MGD-MGD approach as the procedure intuitively uses both prior and posterior information in the design generation.

9 Conclusion

The increasing number of experimenters turning to computer programs rather than statistical consultants for design assistance, creates an ever increasing need to have D-optimal and similar designs to be less dependent on implicit assumptions and more able to produce designs that are less sensitive to model misspecification. This suggests that a good design should provide protection against the possibility of model inadequacy whilst assuring good estimation of the assumed model. The two-stage procedure developed is flexible as it allows to take care of lack of fit in the first stage of the experimental process. The second stage then allows proper estimation of the proposed model whilst protecting with the greatest sensitivity possible any inadequacies in the model. As Steinberg and Hunter (1984), point out," by designing experiments sequentially, we can in a sense, approximate this happy situation by "peeking" at the answer and modifying

the design accordingly." The two-stage approach suggested within the Bayesian paradigm is powerful and can be easily implemented in a wide range of situations.

Acknowledgements

We thank Dr Peter Goos for comments on an earlier version of the article.

References

Atkinson, A. C. and Donev, A. N. (1992) Optimum Experimental Designs, Clarendon Press: Oxford.

Box, G. E. P. and Draper, N. (1959) A basis for the selection of a response surface design, Journal of the American Statistical Association, vol 54, pp 622-653.

Box, G. E. P. and Meyer, R. D. (1993) Finding the active factors in fractionated screening experiments, Journal of Quality Technology, vol 25, pp 94-105.

Bingham D. R. and Chipman H. A. (2002) Optimal designs for model selection, Technical Report 388, University of Michigan and University of Waterloo.

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

DeFeo, P. and Myers, R. H. (1992) A new look at experimental design robustness, Biometrika, vol 79, pp 375-380.

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

Goos, P., Kobilinsky, A., O'Brien, T. E. and Vandebroek, M. (2002) Model-robust and Model-sensitive designs, Technical Report 0259, Department of Applied Economic Sciences, Katholieke Universiteit Leuven.

Harville, D. A (1997) Matrix algebra from a statistician's perspective, Springer: New York.

Kobilinsky, A. (1998) Robustesse d'un plan d'expériences factoriel vis-à-vis d'un sur-modèle, in Proceedings of the 30th Journées de Statistique, ENSAI, Bruz (France).

Lin, H., Myers, R. H. and Ye, K. (2000) Bayesian two-stage optimal design for mixture models,

Journal of Statistical Computation and Simulation, vol 66, pp 209-231.

Myers, R. H. and Montgomery, D. C. (2002) Response Surface Methodology: Process and Product Improvement with Designed Experiments, John Wiley & Sons: New York.

Neff, A. (1996) Bayesian two-stage designs under model uncertainty, Ph.D. dissertation, Virginia Polytechnic and State University, Virginia.

Neff, A. and Myers, R. H. (1998) Recent developments in response surface methodology and its application in industry, in Statistical Process Control for Quality and Productivity Improvement, edited by S. Park and G. G. Vining. Marcel Dekker. New York.

Ruggoo, A. and Vandebroek, M. (2002) Bayesian sequential D-D optimal model-robust designs, Technical Report 0260, Department of Applied Economic Sciences, Katholieke Universiteit Leuven.

Steinberg, D. and Hunter, W. (1984) Experimental design: review and comment, Technometrics, vol 26, pp 71-97.