KATHOLIEKE
UNIVERSITEIT
LEUVEN

DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

RESEARCH REPORT 0259

MODEL-ROBUST AND MODEL-SENSITIVE DESIGNS

by
P. GOOS
A. KOBILINSKY
T. O'BRIEN
M. VANDEBROEK

D/2002/2376/59

Model-robust and model-sensitive designs

Peter Goos

Katholieke Universiteit Leuven

André Kobilinsky

Institut National de la Recherche Agronomique France

Timothy E. O'Brien

Loyola University Chicago

Martina Vandebroek

Katholieke Universiteit Leuven

Abstract

The main drawback of the optimal design approach is that it assumes the statistical model is known. In this paper, a new approach to reduce the dependency on the assumed model is proposed. The approach takes into account the model uncertainty by incorporating the bias in the design criterion and the ability to test for lack-of-fit. Several new designs are derived in the paper and they are compared to the alternatives available from the literature.

Keywords: precision, bias, lack-of-fit, model-robustness, model-sensitive, model-discrimination, D-optimality, A-optimality

1 Introduction

The assumption that underlies most research work in optimal experimental design is that the proposed model adequately describes the response of interest. It is unlikely however that the experimenter is completely certain that the model will be correct and this should be reflected

in the experimental design. Instead of searching for the optimal design to estimate the stated model several approaches have been proposed to account for model uncertainty, ranging from model-robust to model-sensitive strategies. For a nice overview, see for example Steinberg and Hunter (1984).

In a model-robust approach, one looks for designs that yield reasonable results for the true model even if the postulated model is different. The pioneering work in this area is from Box and Draper (1959). They assume that the true model is composed of a primary model – the one that will eventually be estimated – plus some potential terms. The design strategy they propose minimizes the integrated mean squared error over the region of interest. This criterion can be decomposed into the sum of a bias component and a variance component. The problem with this and similar criteria is that the optimal design will depend on the parameters of the potential terms. Several authors who have worked on the problem of balancing precision and bias have proposed solutions to overcome this dependency on the parameters. Welch (1983) for instance minimized the average variance and the average bias in the extreme points of the design region for maximal parameter values, whereas Montepiedra and Fedorov (1997) develop a method to find designs that strike a balance between the variance and the bias. DuMouchel and Jones (1994) used a Bayesian approach to obtain designs that are less sensitive to the model assumption. The authors claim that their criterion leads to designs that are more resistant to the bias caused by the potential terms, and at the same time yields precise estimates of the primary terms. Inspired by the papers of Box and Draper (1959) and DuMouchel and Jones (1994), Kobilinsky (1998) developed a design criterion combining bias and variance properties in a more explicit way.

On the other hand, model-sensitive design approaches lead to designs that facilitate the improvement of the model by detecting lack-of-fit. Examples of such approaches can be found in Atkinson (1972), Atkinson and Cox (1974) and Atkinson and Fedorov (1975a and 1975b). These authors searched for designs that were good in detecting lack-of-fit by maximizing the dispersion matrix somehow. Jones and Mitchell (1978) elaborated on this idea by maximizing the minimal or average noncentrality parameter over a region of possible values for the potential parameters. Studden (1982) combined the detection of lack-of-fit with the precise estimation of the primary terms. This combined approach was also used in the book on optimum experimental design of Atkinson and Donev (1992).

Experimenters often have multiple desires with respect to the design as they want to generate a

maximum of information over the region of interest, ensure that the fitted response value at each point of the design region is close to the true response value and they want possible lack-of-fit to be detectable. Combining all these aspects in a design criterion will therefore lead to useful designs. A first attempt to combine bias and lack-of fit aspects is given by DeFeo and Myers (1992) who minimize bias and at the same time maximize the power of the lack-of-fit test of the potential terms. They show that a rotated design has the same bias properties as the initial design and use this result to maximize the power of the lack-of-fit test.

In this paper we develop two new design criteria that take into account both model-robust and model-sensitive aspects, 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. In Section 2 we will introduce the notation and describe some existing approaches. In Section 3 we develop our generalized criteria and in Section 4 we illustrate their use with some theoretical examples. Section 5 is devoted to a practical example and Section 6 contains the conclusion.

2 The model

We assume there exists a relationship between the expected response and the experimental factors $x_1, x_2, \ldots, x_k$. The model that will be fitted is

$$Y = \mathbf{x}_1'\boldsymbol{\beta}_1 + \varepsilon \tag{1}$$

with $\mathbf{x}_1$ a p-dimensional vector of powers and products of the factors and $\boldsymbol{\beta}_1$ the p-dimensional vector of unknown parameters. We further assume that the expected response was possibly misspecified and that the true model is given by

$$Y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon = \mathbf{x}_1'\boldsymbol{\beta}_1 + \mathbf{x}_2'\boldsymbol{\beta}_2 + \varepsilon = \eta(\mathbf{x}) + \varepsilon, \tag{2}$$

with $\mathbf{x}_2$ the q-dimensional vector containing powers and products of the factors not included in the fitted model, $\mathbf{x}' = [\ \mathbf{x}_1'\ \mathbf{x}_2'\ ]$ and $\boldsymbol{\beta}' = [\ \boldsymbol{\beta}_1'\ \boldsymbol{\beta}_2'\ ]$. We will refer to $\mathbf{x}_1'\boldsymbol{\beta}_1$ as the primary terms and to $\mathbf{x}_2'\boldsymbol{\beta}_2$ 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. In the examples of Section 4 and Section 5, we will use the uniform measure 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.

2.1 Model-robust design strategies

Box and Draper (1959) were the first to investigate the effect of model misspecification. They introduced the integrated mean squared error (IMSE) with respect to a measure μ on the design region. If we denote the fitted response value for factor settings $\mathbf{x}_1$ under the primary model (1) by $\hat{y}(\mathbf{x}_1)$, the IMSE can be defined as

$$\begin{aligned} \text{IMSE} &= \mathbf{E}_\mu \mathbf{E}_\varepsilon \left[\eta(\mathbf{x}) - \hat{y}(\mathbf{x}_1)\right]^2, \\ &= \mathbf{E}_\mu \mathbf{E}_\varepsilon \left[\eta(\mathbf{x}) - \mathbf{E}_\varepsilon[\hat{y}(\mathbf{x}_1)]\right]^2 + \mathbf{E}_\mu \mathbf{E}_\varepsilon \left[\mathbf{E}_\varepsilon[\hat{y}(\mathbf{x}_1)] - \hat{y}(\mathbf{x}_1)\right]^2, \end{aligned}$$

which consists of the expected squared bias and the expected prediction variance. If we denote by $\mathbf{X}_1$ the $n \times p$ model matrix for the primary terms and by $\mathbf{X}_2$ the $n \times q$ model matrix for the potential terms, we have that $\hat{y}(\mathbf{x}_1) = \mathbf{x}_1'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}$ and $\mathbf{E}_\varepsilon[\mathbf{Y}] = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$. As a result,

$$\begin{aligned} \text{IMSE} &= \mathbf{E}_\mu\Big[\mathbf{x}_1'\boldsymbol{\beta}_1 + \mathbf{x}_2'\boldsymbol{\beta}_2 - \mathbf{x}_1'\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\left(\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2\right)\Big]^2 + \mathbf{E}_\mu\left[\mathbf{x}_1'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_1\sigma^2\right], \\ &= \mathbf{E}_\mu\Big[\mathbf{x}_2'\boldsymbol{\beta}_2 - \mathbf{x}_1'\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_2\boldsymbol{\beta}_2\Big]^2 + \mathbf{E}_\mu\left[\mathbf{x}_1'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_1\sigma^2\right]. \end{aligned}$$

In this expression, $(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ is the so-called alias matrix. We will denote it by $\mathbf{A}$ in the sequel of the paper. Now, denoting $\mu_{ij} = \mathbf{E}_\mu(\mathbf{x}_i\mathbf{x}_j')$ and using the well-known result that

$$\begin{aligned} \mathbf{E}_\mu\left[\mathbf{x}_1'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_1\right] &= \mathbf{E}_\mu\left[\text{trace}\{\mathbf{x}_1'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_1\}\right] = \mathbf{E}_\mu\left[\text{trace}\{\mathbf{x}_1\mathbf{x}_1'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\}\right], \\ &= \text{trace}\left[\mu_{11}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\right], \end{aligned}$$

we obtain

$$\text{IMSE} = \mathbf{E}_\mu\left[\mathbf{x}_2'\boldsymbol{\beta}_2 - \mathbf{x}_1'\mathbf{A}\boldsymbol{\beta}_2\right]^2 + \sigma^2\ \text{trace}[\mu_{11}(\mathbf{X}_1'\mathbf{X}_1)^{-1}],$$

$$
\begin{aligned}
&= \beta_2' \, \mathbf{E}_\mu \left[(\mathbf{x}_2' - \mathbf{x}_1'\mathbf{A})'(\mathbf{x}_2' - \mathbf{x}_1'\mathbf{A}) \right] \, \beta_2 + \sigma^2 \, \text{trace} \left[\mu_{11} (\mathbf{X}_1'\mathbf{X}_1)^{-1} \right], \\
&= \beta_2' \left[\mathbf{A}'\mu_{11}\mathbf{A} - \mathbf{A}'\mu_{12} - \mu_{21}\mathbf{A} + \mu_{22} \right] \beta_2 + \sigma^2 \, \text{trace} \left[\mu_{11} (\mathbf{X}_1'\mathbf{X}_1)^{-1} \right].
\end{aligned}
$$

As we have assumed orthonormal polynomials, we have that $\mu_{11} = \mathbf{I}_p$, $\mu_{12} = \mathbf{0}_{p \times q}$, $\mu_{21} = 0_{q \times p}$ and $\mu_{22} = \mathbf{I}_q$. As a consequence,

$$
\text{IMSE} = \beta_2' \left[\mathbf{A}'\mathbf{A} + \mathbf{I}_q \right] \beta_2 + \sigma^2 \, \text{trace}(\mathbf{X}_1'\mathbf{X}_1)^{-1}.
$$

From this result, Box and Draper (1959) concluded that bias can be minimized by looking for designs for which that $\mathbf{A} = \mathbf{0}_{p \times q}$. In general however the design that minimizes IMSE will depend on the values of β_2. To cope with this dependence, Kobilinsky (1998) suggested to put a prior distribution on the potential parameters. As it is unlikely that these terms are large, the following distribution was considered to be appropriate:

$$
\beta_2 \sim \mathcal{N}(\mathbf{0}, \tau^2\sigma^2\mathbf{I}_q).
$$

Because $\mathbf{x}_2$ is orthonormalized, it is reasonable to assume that all elements in β_2 have equal variances and that they are uncorrelated with each other. Under this assumption, we obtain that

$$
\begin{aligned}
\mathbf{E}_\beta[\text{IMSE}] &= \mathbf{E}_\beta \left[\beta_2' \left[\mathbf{A}'\mathbf{A} + \mathbf{I}_q \right] \beta_2 + \sigma^2 \, \text{trace}(\mathbf{X}_1'\mathbf{X}_1)^{-1} \right] \\
&= \text{trace}(\mathbf{A}'\mathbf{A}\tau^2\sigma^2\mathbf{I}_q + \tau^2\sigma^2\mathbf{I}_q) + \sigma^2 \, \text{trace}(\mathbf{X}_1'\mathbf{X}_1)^{-1} \\
&= \tau^2\sigma^2 \, \text{trace} \left(\mathbf{A}'\mathbf{A} + \mathbf{I}_q \right) + \sigma^2 \, \text{trace}(\mathbf{X}_1'\mathbf{X}_1)^{-1}.
\end{aligned}
$$

It is clear that $\tau^2 = 0$ indicates that the primary model is the true model. In that case, minimization of the expected IMSE will lead to the minimization of $\text{trace}(\mathbf{X}_1'\mathbf{X}_1)^{-1}$ and thus to an A-optimal design for the primary model (1).

Based on a similar prior distribution of the potential terms, DuMouchel and Jones (1994) proposed a Bayesian D-optimality criterion to find designs that yield precise estimates for the primary terms and give some protection against the existence of the potential terms. As the

posterior covariance matrix of $\hat{\beta}$ is

$$\text{cov}(\hat{\beta}) = \left[\frac{\mathbf{X}'\mathbf{X}}{\sigma^2} + \frac{\mathbf{K}}{\tau^2\sigma^2}\right]^{-1},$$

with $\mathbf{X}' = [\, \mathbf{X}_1' \; \mathbf{X}_2' \,]$ and

$$\mathbf{K} = \begin{pmatrix} \mathbf{0}_{p\times p} & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times p} & \mathbf{I}_q \end{pmatrix},$$

they proposed to maximize the following determinant:

$$\frac{1}{\sigma^2}\left|\mathbf{X}'\mathbf{X} + \frac{\mathbf{K}}{\tau^2}\right|.$$

This criterion has the clear advantage that the information matrix for the full model (2), i.e. $\mathbf{X}'\mathbf{X}$, can be singular without causing problems. Therefore it is possible to use this criterion for design problems in which $p \leq n < p + q$, that is in cases where the number of observations n available is insufficient to estimate the full model.

The choice of τ^2 is of course an arbitrary one. Kobilinsky (1998) suggests $\tau^2 = 1/q$ so that the global effect of the q potential terms is of the same order of magnitude as the residual error. DuMouchel and Jones (1994) suggest to take $\tau^2 = 1$ so that the effect of any of the potential terms is not larger than the residual standard error. They use a less stringent orthogonalization procedure which only orthogonalizes the potential terms with respect to the primary terms. The primary terms are not orthogonalized relative to each other, nor are the potential terms. The orthonormalization used in this paper leads to simpler mathematical derivations.

The approaches of Box and Draper (1959), DuMouchel and Jones (1994) and Kobilinsky (1998) aim at finding designs that yield precise estimates of the primary terms and ensure that predictions are close to the expected response. They do not explicitly consider the possibility of performing a lack-of-fit test and therefore do not provide information on the appropriateness of the primary model. In the next section we consider some existing approaches to deal with this discrimination problem.

2.2 Model-sensitive design strategies

An approach which takes into account both the experimental effort for determining which model is true and the effort for precise estimation of the parameters is given by Atkinson and Donev (1992). They proposed to combine the D-optimality criterion for the primary model and the D_s-optimality criterion for the potential terms. The resulting criterion is given by

$$\max \left\{ \frac{\alpha}{p} \log |\mathbf{X}_1'\mathbf{X}_1| + \frac{1-\alpha}{q} \log |\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2| \right\},$$

where $\alpha \in [0,1]$ represents the belief in the primary model (1). When $\alpha = 1$, this criterion reduces to the D-optimality criterion for the primary model, whereas for $\alpha = 0$ it becomes the D_s-optimality criterion for the potential model parameters $\boldsymbol{\beta}_2$. When $\alpha = p/(p+q)$, the combined criterion leads to D-optimal designs for the full model (2).

Note that the D_s-optimality criterion for the potential terms is related to the noncentrality parameter

$$\delta = \frac{\boldsymbol{\beta}_2' \left[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\right] \boldsymbol{\beta}_2}{\sigma^2} \tag{3}$$

to test for lack-of-fit in the direction of the potential terms. Therefore, it is likely that the power of the lack-of-fit test will increase with decreasing α. The matrix $\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ is well known in the literature on model-sensitive designs. It is usually referred to as the dispersion matrix. In the sequel of this paper, we will denote it by $\mathbf{L}$.

3 A combined approach

The advantages of the approaches described in the previous section will be combined in a flexible criterion that includes three important aspects: precise estimation of the primary model, minimization of the bias caused by the potential terms and possibility to test for lack-of-fit.

The criterion of Kobilinsky (1998) that was derived in the previous section

$$\min \left\{ \tau^2\sigma^2 \text{ trace } (\mathbf{A}'\mathbf{A} + \mathbf{I}_q) + \sigma^2 \text{ trace } (\mathbf{X}_1'\mathbf{X}_1)^{-1} \right\}$$

takes into account precision and bias but not lack-of-fit. As this criterion was derived by computing the expected IMSE over the prior distribution of potential terms, it is natural to apply the same idea to the lack-of-fit term. As the noncentrality parameter also depends on the values of $\boldsymbol{\beta}_2$, we will maximize the expected noncentrality parameter over the prior distribution. The expected noncentrality parameter can be computed as

$$\begin{aligned}
\mathbf{E}_{\boldsymbol{\beta}}[\delta] &= \mathbf{E}_{\boldsymbol{\beta}}\left[\frac{\boldsymbol{\beta}_2'\{\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\}\boldsymbol{\beta}_2}{\sigma^2}\right], \\
&= \tau^2 \text{ trace}\left[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\right], \\
&= \tau^2 \text{ trace}\left[\mathbf{L}\right].
\end{aligned}$$

To combine the three aspects in one criterion we specify weights α_2 and α_3 to attach more or less importance on the different properties. A possible criterion is then given by

$$\min \left\{\frac{1}{p}\text{trace } (\mathbf{X}_1'\mathbf{X}_1)^{-1} - \frac{\alpha_2}{q}\text{trace } (\mathbf{L}) + \frac{\alpha_3}{q}\text{trace } (\mathbf{A}'\mathbf{A} + \mathbf{I}_q)\right\}.$$

Similarly, the criterion

$$\max \left\{\frac{\alpha}{p}\log|\mathbf{X}_1'\mathbf{X}_1| + \frac{1-\alpha}{q}\log|\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2|\right\}$$

of Atkinson and Donev (1992), which takes into account precision and lack-of-fit, can be augmented with a term that represents the bias. As this criterion deals with determinants, a natural extension is given by

$$\min \left\{\frac{1}{p}\log|(\mathbf{X}_1'\mathbf{X}_1)^{-1}| + \frac{\alpha_2}{q}\log\left|\mathbf{L}^{-1}\right| + \frac{\alpha_3}{q}\log|\mathbf{A}'\mathbf{A} + \mathbf{I}_q|\right\}.$$

Because these criteria do not allow for singular design matrices for the full model, we can use the idea of DuMouchel and Jones to allow for smaller designs and generalize the previous criteria to the following generalized A- and D-optimality criteria:

$$\text{GA: min } \left\{\frac{1}{p}\text{trace}(\mathbf{X}_1'\mathbf{X}_1)^{-1} - \frac{\alpha_2}{q}\text{trace}\left(\mathbf{L} + \frac{\mathbf{I}_q}{\tau^2}\right) + \frac{\alpha_3}{q}\text{trace } (\mathbf{A}'\mathbf{A} + \mathbf{I}_q)\right\},$$

and

$$\text{GD: } \min \left\{ \frac{1}{p} \log |(\mathbf{X}_1'\mathbf{X}_1)^{-1}| + \frac{\alpha_2}{q} \log \left| (\mathbf{L} + \frac{\mathbf{I}_q}{\tau^2})^{-1} \right| + \frac{\alpha_3}{q} \log |\mathbf{A}'\mathbf{A} + \mathbf{I}_q| \right\}.$$

It is easy to see that these criteria generalize those proposed by Atkinson and Donev (1992), DuMouchel and Jones (1994) and Kobilinsky (1998) as well as the ordinary D- and A-optimality criteria. For $\alpha_2 = \alpha_3 = 0$ the GD-optimality criterion produces the D-optimal design for the primary model. We will refer to this design as D_1-optimal in the sequel. For $\alpha_3 = 0$, $\alpha_2 = \frac{q}{p}$ and $\tau^2 = \infty$, we obtain the D-optimal design for the full model, denoted by D_{full}. For $\alpha_3 = 0, \alpha_2 = \frac{q}{p}$ and finite values for τ^2, we find the Bayesian D-optimal designs introduced by DuMouchel and Jones (1994). This is because

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

4 Illustrations

In this section, we will illustrate the use of the GD-optimality criterion in a number of simple experimental situations. The GA-optimality criterion leads to different designs but to similar results.

4.1 One explanatory variable

Firstly, assume that the primary model consists of $p = 3$ terms $\beta_0 + \beta_1 x + \beta_2 x^2$ and that there is $q = 1$ potential term $\beta_3 x^3$. As a result, $\boldsymbol{\beta}_1 = [\beta_0\ \beta_1\ \beta_2]'$ and $\boldsymbol{\beta}_2 = [\beta_3]$. Also, assume that $n = 8$ and $\tau^2 = \infty$. By varying the values of α_2 and α_3 we obtain several designs, the extreme ones are displayed in Figure 1. The designs were computed using a grid of 21 equidistant points on $[-1, +1]$. The values of the different determinants in the GD-optimality criterion are given in Figure 1 as well. DX1 represents $|\mathbf{X}_1'\mathbf{X}_1|^{-1/p}$, the measure used for the precision of the primary terms, Dlof=$|\mathbf{L}|^{-1/q}$ provides an idea of the ability to detect lack-of-fit and Dbias=$|\mathbf{A}'\mathbf{A} + \mathbf{I}_q|^{1/q}$ represents the degree of bias. These terms were defined such that the smaller the value obtained, the better the design performs with respect to this criterion. Remark that several designs can

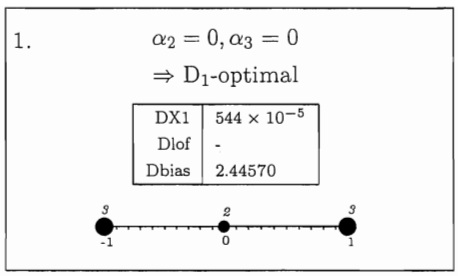
be obtained for large α_2 and α_3. The one presented is one of the symmetric designs we have found.

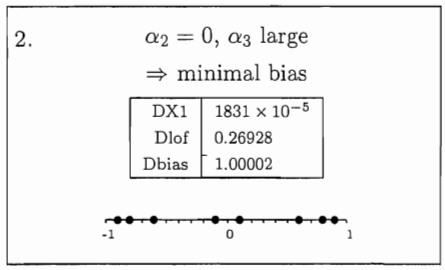
For $\alpha_2 = \alpha_3 = 0$, the D-optimal design for the primary model was obtained. This design is displayed in Panel 1 of Figure 1. When either α_2 or α_3 is increased, different designs are obtained. For example, choosing a large value for α_3 produces the design in Panel 2. This design leads to a small bias. Choosing $\alpha_2 = q/p = 1/3$ and $\alpha_3 = 0$ leads to the D-optimal design for the full model (see Panel 3). The Dlof-value shows that this design allows a good detection of lack-of-fit. Further increasing α_2 allows an even better detection of the lack-of-fit. Choosing large values for both α_2 and α_3 produces a design that is good for detecting lack-of-fit and that leads to a limited amount of bias. Introducing finite values for τ^2 creates no new designs for this example. Probably, this is due to the fact that $n > p + q$.

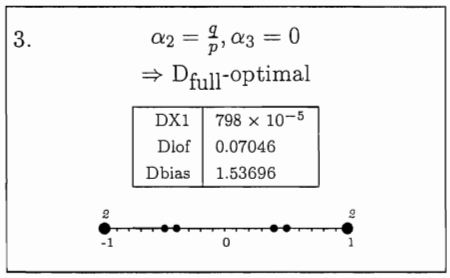
The average squared prediction variance and average squared bias for an arbitrary value of β_3 are given in Table 1. The value chosen is $\beta_3 = 1$. The table also contains the noncentrality parameter for the lack-of-fit test. The table shows that the loss of precision in the estimation of the primary model is compensated by substantial reductions in the bias and by the ability to test for lack-of-fit. Table 1 also shows that choosing positive values for both α_2 and α_3 leads to a design that performs excellently with respect to both bias and detection of lack-of-fit. Using a positive α_2 and setting $\alpha_3 = 0$ provides a design that allows a good detection of the lack-of-fit but it also leads to a substantial reduction in the bias. Using a positive α_3 and setting $\alpha_2 = 0$ leads to a small bias, but the resulting design does not perform that well as to detection of the lack-of-fit. As a result, designs that perform well with respect to lack-of-fit detection also perform reasonably well with respect to the bias, but the opposite is not necessarily true. DuMouchel and Jones (1994) point out that an idea of the significance of the lack-of-fit test can be obtained by assuming that the expectation of the F-statistic

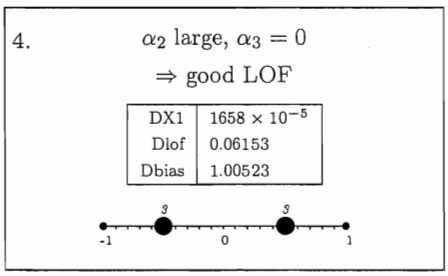
$$F = \frac{\frac{\text{SSE(primary model)}-\text{SSE(full)}}{d_1}}{\frac{\text{SSE(full)}}{n-d_2}},$$

with SSE(M) the sum of squared errors of model M and d_1 and d_2 the degrees of freedom for









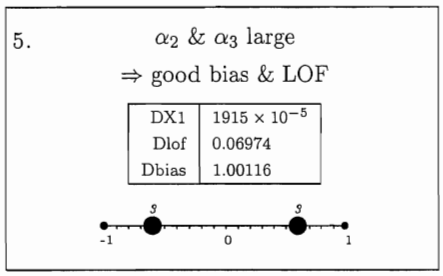


Figure 1: GD-optimal designs for several values of α_2 and α_3, and for $\tau^2 = \infty$.

Table 1: Bias, variance and lack-of-fit measures

design	bias^2	avg var	δ	p-value for lof
1	2.4457	0.134476	-	-
2	1.0000	0.170827	3.77366	0.08203
3	1.5370	0.113033	14.19280	0.01144
4	1.0052	0.156178	16.252114	0.00888
5	1.0012	0.202589	14.337984	0.01122

the test, is equal to

$$F_s \approx \frac{\mathbf{E}\left(\frac{\text{SSE(primary model)}-\text{SSE(full)}}{d_1}\right)}{\mathbf{E}\left(\frac{\text{SSE(full)}}{d_2}\right)} = \frac{\sigma^2 + \delta\sigma^2/d_1}{\sigma^2} = 1 + \frac{\delta}{d_1},$$

where δ is the noncentrality parameter introduced in (3). The number d_1 is equal to q if it is possible to test the full model, whereas $d_2 = n-$ total number of independent parameters in the full model. The p-values obtained using the F_s-statistic are displayed in the last column of Table 1.

4.2 Two dimensions

As another illustration, consider the 2-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 that we considered is the 5×5 grid on $[-1, +1]^2$.

For $n = 5$ we find only the two designs displayed in Figure 2. The first design is the D-optimal design for the primary model. The second design is obtained as soon as the values of α_2 and/or α_3 become large enough to have some impact. Remark that, as $n < p + q$, τ^2 is given a finite value to obtain a nonsingular dispersion matrix $\mathbf{L}$. These designs, which were also found by DuMouchel and Jones (1994), support the common practice of adding center points to a design in order to carry out a lack-of-fit test.

For $n = 8$ and $\tau^2 = \infty$, we find a lot of different designs, the most important ones are represented

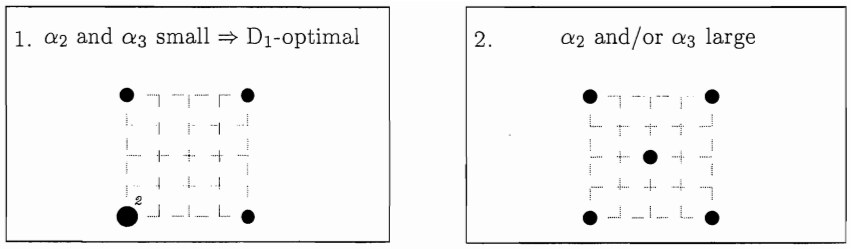


Figure 2: GD-optimal 5-point designs for several values of α_2 and α_3, and for $\tau^2 = 1$.

in Figure 3. Panel 1 shows a duplicated 2^2 factorial design, which is the D_1-optimal 8-point design for the primary model. When α_3 is increased, this design gradually changes into a 2^2 factorial design with 4 center points. When α_2 is increased, then most design points move away from the cornerpoints. This allows the lack-of-fit to be tested and the bias is reduced to some extent. For a good performance on both criteria, it is necessary to choose positive values for both α_2 and α_3. Introducing finite values for τ^2 does not lead to new designs in this example.

4.3 A constrained design region

We reconsider the second example of DuMouchel and Jones (1994) with two constrained variables. In the example, $x_1 + x_2 \leq 1$ so that the set of candidate points only contains 15 points on a triangle. The primary model is the full quadratic model, so $p = 6$, and there are $q = 4$ potential cubic terms, $x_1^3, x_1^2x_2, x_1x_2^2, x_2^3$. The number of observations is equal to 9.

As $n < p + q$, a finite τ-value had to be used. For the same reason, Dlof will not exist for the designs shown. Therefore, Dlof$_\tau$-values that are defined as $\left|\mathbf{L} + \mathbf{I}_q/\tau^2\right|^{-1/q}$ will be reported instead. The results for $\tau = 1$ are displayed in Figure 4. It turns out that the same designs can be found for other values of τ.

From Panel 1 in Figure 4, it can be seen that the D_1-optimal design has minimum support, i.e. the number of distinct design points of the design is equal to the number of parameters in the (primary) model. When α_2 and/or α_3 are increased, the number of distinct design points is increased so that the bias is substantially decreased and the ability to test for lack-of-fit is substantially increased. As in the previous example, it is important to select positive values for α_2 and α_3 for a good performance on both criteria. Note that, when α_2 and/or α_3 are large,

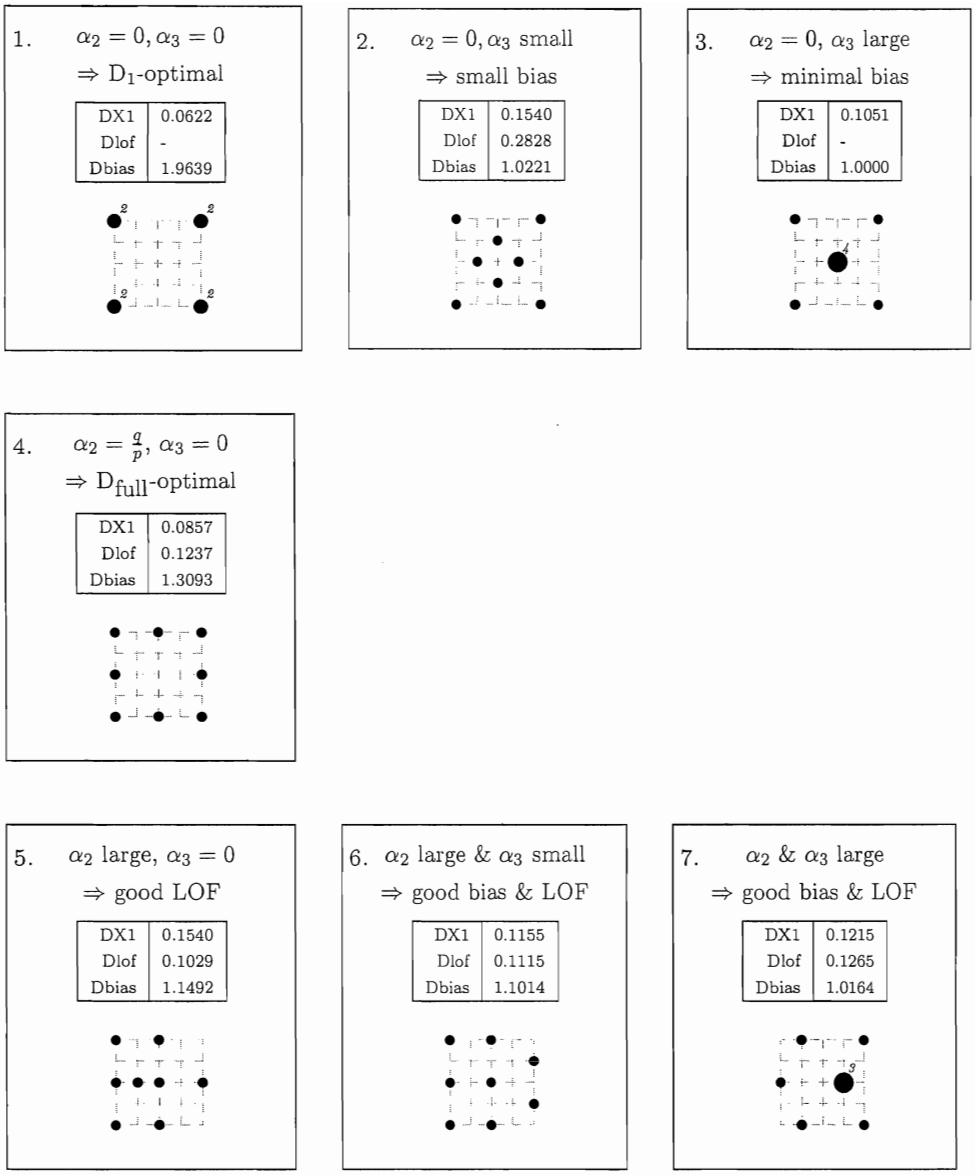


Figure 3: GD-optimal 8-point designs for several values of α_2 and α_3, and $\tau^2 = \infty$.

then the GD-optimal designs contain 9 distinct design points, as can be seen in the Panels 2, 4 and 5 of Figure 4.

5 Practical application

Snee (1981) described a mixture experiment to investigate to what extent the octane of various blends of gasoline depends on the component proportions. In the experiment, five components were investigated: butane, isopentane, reformate, cat cracked and alkylate. The ranges of these five components are given in the following table:

component	range
butane (B)	0-0.15
isopentane (I)	0-0.30
reformate (R)	0-0.35
cat cracked (C)	0-0.60
alkylate (A)	0-0.60

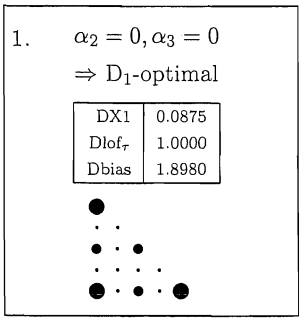
In addition to the traditional mixture constraint $B+I+R+C+A = 1$, the following requirements were defined

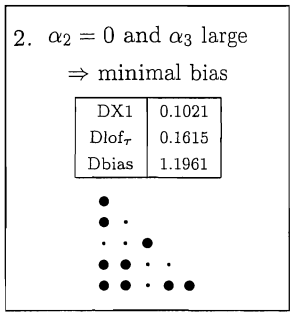
$$\begin{aligned} B + I &\leq 0.30 \\ C + A &\leq 0.70 \\ 97 \leq \quad 101.8B + 99.6I + 112.4R + 94.2C + 99.8A \quad &\leq 101. \end{aligned}$$

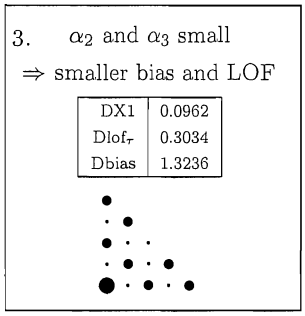
A 25-run D-optimal design for the quadratic Scheffé model augmented by the centroid of the design region was run and it turned out that six of the second order terms could be eliminated. The fitted model was

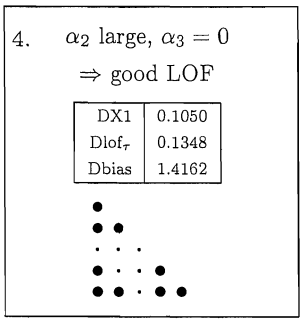
$$155.1B + 97.7I + 108.6R + 95.0C + 101.4A - 44.6BI - 77.0BR - 67.6BC - 60.0BA, \qquad (4)$$

and the residual standard deviation amounted to 0.30.









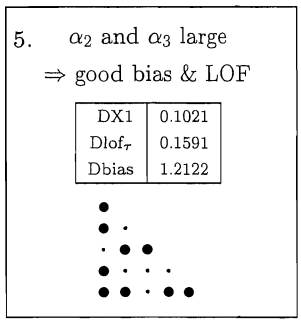


Figure 4: GD-optimal designs for several values of α_2 and α_3, $\tau = 1$.

Table 2: Design 1: D_1-optimal designs for the five-component mixture experiment.

Run	B	I	R	C	A
1	0.0000	0.0000	0.3500	0.6000	0.0500
2	0.0000	0.0000	0.3500	0.6000	0.0500
3	0.0000	0.3000	0.0000	0.1000	0.6000
4	0.0000	0.3000	0.0492	0.6000	0.0508
5	0.0000	0.3000	0.1000	0.0000	0.6000
6	0.0000	0.3000	0.2846	0.4154	0.0000
7	0.1500	0.0336	0.1164	0.1000	0.6000
8	0.1500	0.0336	0.1164	0.1000	0.6000
9	0.1500	0.1273	0.0227	0.6000	0.1000
10	0.1500	0.1273	0.0227	0.6000	0.1000
11	0.1500	0.1500	0.2665	0.4335	0.0000
12	0.1500	0.1500	0.2665	0.4335	0.0000

For illustrative purposes, we will treat this model as the true model in order to compare different designs. We assume that the primary model contains the five linear terms and that the potential terms are the 10 quadratic terms. In order to compare the performances of the GD-optimal designs with those of DuMouchel and Jones (1994), we will consider 12-run designs. With only 12 runs it is impossible to fit all primary and potential terms, so that a finite τ-value has to be used. We will show the results for $\tau = 1$. Other values of τ yield similar designs and results. The set of candidate points we used consists of the extreme vertices, the overall centroid, the constraint plane centroids, the edge centroids and the lattice grid within the constraint region with each proportion being a multiple of 0.05.

We will use the four designs presented by DuMouchel and Jones (1994) in our comparisons. The first design they consider is the D_1-optimal design displayed in Table 2. This design consists of the eight vertices of the design region, four of which are duplicated. The other three designs considered by DuMouchel and Jones (1994) are displayed in Table 3. The first is a design obtained by including two runs at the centroid of the design region and adding 10 design points using the D-optimality criterion. The second design they use in their computations is the design obtained using their Bayesian criterion (see Section 2.1). Finally, they consider a D-optimal design for the true model (4). This approach is called the omniscient approach. These four designs will be compared to three designs obtained by using the GD-optimality criterion: one design obtained with a large α_2-value, one obtained with a large α_3-value, and one obtained using a large α_2 and a large α_3. The design points for these three designs are displayed in Table 4.

Table 3: Designs considered by DuMouchel and Jones (1994).

Run	B	I	R	C	A
Design 2: D-optimal 10-point design for the primary model augmented with 2 center points.					
1	0.0000	0.0000	0.3500	0.6000	0.0500
2	0.0000	0.0000	0.3500	0.6000	0.0500
3	0.0000	0.3000	0.0000	0.1000	0.6000
4	0.0000	0.3000	0.0490	0.6000	0.0510
5	0.0000	0.3000	0.1000	0.0000	0.6000
6	0.0000	0.3000	0.2850	0.4150	0.0000
7	0.0680	0.1210	0.1750	0.4440	0.1920
8	0.0680	0.1210	0.1750	0.4440	0.1920
9	0.1500	0.0340	0.1160	0.1000	0.6000
10	0.1500	0.1270	0.0230	0.6000	0.1000
11	0.1500	0.1500	0.2660	0.4340	0.0000
12	0.1500	0.1500	0.2660	0.4340	0.0000
Design 3: Design obtained using the criterion of DuMouchel and Jones (1994) ($\tau = 1$, $\alpha_2 = 2$, $\alpha_3 = 0$).					
1	0.0000	0.0000	0.3000	0.4607	0.2393
2	0.0000	0.1258	0.1742	0.6000	0.1000
3	0.0000	0.1578	0.1422	0.1000	0.6000
4	0.0000	0.3000	0.0000	0.4893	0.2107
5	0.0000	0.3000	0.1000	0.0000	0.6000
6	0.0000	0.3000	0.1500	0.5500	0.0000
7	0.0000	0.3000	0.2846	0.4154	0.0000
8	0.1033	0.0000	0.3305	0.5662	0.0000
9	0.1500	0.0000	0.1500	0.1768	0.5232
10	0.1500	0.1273	0.0227	0.6000	0.1000
11	0.1500	0.1500	0.0000	0.1000	0.6000
12	0.1500	0.1500	0.2665	0.4335	0.0000
Design 4: D-optimal design for model (4).					
1	0.0000	0.0000	0.3000	0.4607	0.2393
2	0.0000	0.3000	0.0492	0.6000	0.0508
3	0.0000	0.3000	0.1000	0.0000	0.6000
4	0.0000	0.3000	0.2846	0.4154	0.0000
5	0.0750	0.0000	0.2250	0.6000	0.1000
6	0.0750	0.2250	0.0000	0.1000	0.6000
7	0.0750	0.2250	0.2756	0.4245	0.0000
8	0.1500	0.0000	0.1500	0.1768	0.5232
9	0.1500	0.0000	0.3110	0.5390	0.0000
10	0.1500	0.1500	0.0230	0.6000	0.0770
11	0.1500	0.1500	0.0819	0.0181	0.6000
12	0.1500	0.1500	0.2665	0.4335	0.0000

Table 4: GD-optimal designs for the five-component mixture experiment.

Run	B	I	R	C	A
Design 5: GD-optimal design for $\alpha_2 = 0$ and α_3 large.					
1	0.0000	0.1000	0.2000	0.3000	0.4000
2	0.0000	0.1503	0.1996	0.6000	0.0501
3	0.0000	0.2500	0.1000	0.2000	0.4500
4	0.0000	0.3000	0.1500	0.4000	0.1500
5	0.0283	0.0000	0.3500	0.5833	0.0384
6	0.0500	0.2000	0.2500	0.4000	0.1000
7	0.0500	0.2500	0.0000	0.5000	0.2000
8	0.0750	0.2250	0.0455	0.0545	0.6000
9	0.1000	0.0000	0.2000	0.5000	0.2000
10	0.1500	0.0500	0.1000	0.3500	0.3500
11	0.1500	0.0500	0.2000	0.3000	0.3000
12	0.1500	0.1000	0.1500	0.6000	0.0000
Design 6: GD-optimal design for α_2 large and $\alpha_3 = 0$.					
1	0.0000	0.0000	0.3000	0.4607	0.2393
2	0.0000	0.0796	0.3500	0.5704	0.0000
3	0.0000	0.1258	0.1742	0.6000	0.1000
4	0.0000	0.1578	0.1422	0.1000	0.6000
5	0.0000	0.3000	0.2846	0.4154	0.0000
6	0.0000	0.3000	0.0000	0.4893	0.2107
7	0.0500	0.2500	0.1500	0.1500	0.4000
8	0.0500	0.2500	0.1500	0.5500	0.0000
9	0.1500	0.0000	0.1500	0.1768	0.5232
10	0.1500	0.1273	0.0227	0.6000	0.1000
11	0.1500	0.1500	0.0000	0.1000	0.6000
12	0.1500	0.1500	0.2665	0.4335	0.0000
Design 7: GD-optimal design for α_2 and α_3 large.					
1	0.0000	0.0000	0.3000	0.6000	0.1000
2	0.0000	0.0796	0.3500	0.5704	0.0000
3	0.0000	0.1578	0.1422	0.1000	0.6000
4	0.0000	0.2500	0.0500	0.6000	0.1000
5	0.0000	0.3000	0.0000	0.1500	0.5500
6	0.0000	0.3000	0.1923	0.2077	0.3000
7	0.0500	0.2500	0.1500	0.5500	0.0000
8	0.1000	0.0000	0.2000	0.3000	0.4000
9	0.1500	0.0000	0.2500	0.6000	0.0000
10	0.1500	0.1500	0.0000	0.5482	0.1518
11	0.1500	0.1500	0.0819	0.0181	0.6000
12	0.1500	0.1500	0.2665	0.4335	0.0000

Table 5: Comparison of seven alternative designs for the constrained mixture experiment using the parameter values of the true model (4).

	precision			bias			lack-of-fit		
DESIGN	DX1	average pred.var.	maximum pred.var.	Dbias	average sqd.bias	maximum sqd.bias	Dlof_τ	δ	p-value
$\alpha_2 = 0, \alpha_3 = 0$ $\Rightarrow$ D_1-optimal	0.0441	0.2470	0.4876	1.6784	0.1027	0.7887	0.3076	1.0307	0.3790
$\alpha_2 = 0, \alpha_3 = 0$ with center points	0.0512	0.2825	0.6018	1.5393	0.0687	0.6165	0.2864	3.7601	0.2244
$\alpha_2 = 2, \alpha_3 = 0$ $\Rightarrow$ Bayesian	0.0496	0.2680	0.5614	1.3928	0.0645	0.2583	0.0832	8.4136	0.1896
$\alpha_2 = 0.8, \alpha_3 = 0, \tau = \infty$ $\Rightarrow$ omniscient[1]	0.0497	0.2883	0.6325	1.1049	0.0504	0.3253	0.2658	9.0566	0.1793
$\alpha_2 = 0, \alpha_3$ large $\Rightarrow$ minimal bias	0.0779	0.3918	1.2180	1.0161	0.0403	0.5507	0.1568	2.8128	0.3451
α_2 large, $\alpha_3 = 0$ $\Rightarrow$ best LOF	0.0540	0.3074	0.7173	1.3288	0.0935	0.6346	0.0817	5.4571	0.2527
α_2 large, α_3 large $\Rightarrow$ good bias & LOF	0.0538	0.2853	0.7542	1.2275	0.0511	0.3442	0.0836	11.3362	0.1492

[1] In order to compute this design, only the 4 interactions active in the true model were used.

These seven designs are compared to each other in Table 5. The first column of the table contains the settings for the parameters α_2, α_3 and τ. The next three columns contain detailed information about the precision of the estimation and the prediction. The columns 5 through 7 contain the performances of the design as to bias and the last three columns show the designs' abilities to detect lack-of-fit. A note on the computation of the average squared biases, the maximum squared biases, the noncentrality parameters δ (see Equation (3)) and the corresponding p-values, which all depend on the parameter values of the true model, is given in the Appendix. The coefficients -44.6, -77.0, -67.6 and -60.0 were used for the terms involving BI, BR, BC and BA respectively. Zeroes were used for the coefficients of the other potential second order terms.

From the table, it can be seen that taking into account possible misspecification of the model goes at the expense of precision in the estimation of the primary model. This is especially true for the minimal bias design. All designs providing protection against misspecification lead to a substantial reduction in bias. In addition, they allow the experimenter to detect lack-of-fit. It turns out that the inclusion of center points in the designs is however the worst option to decrease the bias and to detect lack-of-fit in this case. The Bayesian design obtained by using the approach of DuMouchel and Jones (1994) is a close competitor to the GD-optimal designs when detecting lack-of-fit is important. It performs considerably worse, however, when the focus is on reducing the potential bias. Surprisingly, the noncentrality parameter of the GD-optimal

Table 6: Comparison of seven alternative designs for the constrained mixture experiment using the parameter values 70, 70, 40 and 40 instead for the model terms involving BI, BR, BC and BA respectively.

	precision			bias			lack- of-fit		
DESIGN	DX1	average pred.var.	maximum pred.var.	Dbias	average sqd.bias	maximum sqd.bias	Dlof_τ	δ	p-value
$\alpha_2 = 0, \alpha_3 = 0$ $\Rightarrow$ D_1-optimal	0.0441	0.2470	0.4876	1.6784	0.0812	0.3183	0.3076	7.3619	0.1311
$\alpha_2 = 0, \alpha_3 = 0$ with center points	0.0512	0.2825	0.6018	1.5393	0.0533	0.2643	0.2864	8.8961	0.1082
$\alpha_2 = 2, \alpha_3 = 0$ $\Rightarrow$ Bayesian	0.0496	0.2680	0.5614	1.3928	0.0655	0.2945	0.0832	6.8270	0.2198
$\alpha_2 = 0.8, \alpha_3 = 0, \tau = \infty$ $\Rightarrow$ omniscient	0.0497	0.2883	0.6325	1.1049	0.0503	0.2434	0.2658	8.7457	0.1841
$\alpha_2 = 0, \alpha_3$ large $\Rightarrow$ minimal bias	0.0779	0.3918	1.2180	1.0161	0.0446	0.2485	0.1568	4.2157	0.2903
α_2 large, $\alpha_3 = 0$ $\Rightarrow$ best LOF	0.0540	0.3074	0.7173	1.3288	0.0585	0.2988	0.0817	8.4291	0.1893
α_2 large, α_3 large $\Rightarrow$ good bias & LOF	0.0538	0.2853	0.7542	1.2275	0.0478	0.2972	0.0836	5.0584	0.2639

design obtained for a large α_2-value and $\alpha_3 = 0$ is small relative to that for other designs. This is due to the parameter values in the true model. Another choice of the coefficients of the potential terms leads to totally different noncentrality parameters. For example, choosing 70, 70, 40 and 40 instead of -44.6, -77.0, -67.6 and -60.0 gives the results displayed in Table 6.

6 Conclusions

In this paper, we have derived a generalization of several existing design criteria in order to take into account possible misspecification of the model when designing an experiment. Traditionally, the optimal design approach assumes that the specified model is known. In most applications, the model is unknown. The design criteria presented take into account the potential bias from the unknown true model as well as the power of a lack-of-fit test. Several simple examples are used to illustrate the properties of the designs produced by the new criteria. A constrained mixture experiment was used to demonstrate the usefulness of the approach. This example showed that the new design criteria used lead to designs that perform well with respect to bias and with respect to the detection of lack-of-fit.

Acknowledgement

The research that led to this paper was carried out while Peter Goos was a Postdoctoral Researcher of the Fund for Scientific Research – Flanders (Belgium) and while Timothy O'Brien was visiting the Faculty of Economic and Applied Economic Sciences of the Katholieke Universiteit Leuven.

Appendix

The coefficients of the primary and potential terms in the practical example are the coefficients in the non-orthonormalized model (4). This has to be taken into account when computing the bias and the noncentrality parameter δ. In order to compute the bias for a particular combination of factor levels, the following formula was used:

$$bias(\mathbf{x}) = \eta(\mathbf{x}) - \mathbf{x}_1' \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1} \mathbf{X}_1'\eta(\mathbf{x}),$$

where $\eta(\mathbf{x})$ represents the response obtained using the non-orthonormalized model (4).

In order to compare the ability to detect lack-of-fit, the noncentrality parameter and the corresponding F-statisic were computed as well. The noncentrality parameter δ is equal to the sum of the squared biases for the runs in the design divided by σ^2:

$$\begin{aligned}\delta &= \frac{1}{\sigma^2}\left[\eta(\mathbf{X}) - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\eta(\mathbf{X})\right]'\left[\eta(\mathbf{X}) - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\eta(\mathbf{X})\right] \\ &= \frac{1}{\sigma^2}\eta(\mathbf{X})\left[\mathbf{I}_n - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\right]\eta(\mathbf{X}),\end{aligned}$$

where $\boldsymbol{\eta}(\mathbf{X})$ represents the vector of responses obtained using the non-orthonormalized model (4) for all design points.

References

Atkinson, A.C. (1972) Planning experiments to detect inadequate regression models, Biometrika, vol 59, 2, pp 275-293.

Atkinson, A.C. and Cox, D.R. (1974) Planning experiments for discriminating between models, Journal of the Royal Statistical Society, B, vol 36, pp 321-348.

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

Atkinson, A.C. and Fedorov, V.V. (1975a) The design of experiments for discriminating between two rival models, Biometrika, vol. 62, pp 57-70.

Atkinson, A.C. and Fedorov, V.V. (1975b) Optimal design: experiments for discriminating between several models, Biometrika, vol. 62, pp 289-303.

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

DeFeo, P. and Myers, R.H. (1992) A new look at experimental design robustness, Biometrika, vol 79, no 2, 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, no 1, pp 37-47.

Jones, E.R. and Mitchell, T.J. (1978) Design criteria for detecting model inadequacy, Biometrika, vol 65, no 3, pp 541-551.

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).

Montepiedra, G. and Fedorov, V. (1997) Minimum bias designs with constraints, Journal of Statistical Planning and Inference, vol 63, pp 97-111.

Snee, R.D. (1981) Developing blending models for gasoline and other mixtures, Technometrics,

vol 23, no 2, pp 119-130.

Steinberg, D. and Hunter, W. (1984) Experimental design: review and comment, Technometrics, vol 26, no 2, pp 71-97.

Studden, W.J. (1982) Some robust-type D-optimal designs in polynomial regression, Journal of the American Statistical Association, vol 77, no 380, pp 916-921.

Welch, W. (1983) A mean squared error criterion for the design of experiments, Biometrika, vol 70, pp 205-213.