

*Design of experiments in the presence of errors
in factor levels*

Donev, A.N.

2003

MIMS EPrint: **2007.163**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: <http://eprints.maths.manchester.ac.uk/>

And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097



Available at

www.ElsevierMathematics.com

POWERED BY SCIENCE @ DIRECT®

Journal of Statistical Planning and
Inference 126 (2004) 569–585

journal of
statistical planning
and inference

www.elsevier.com/locate/jspi

Design of experiments in the presence of errors in factor levels

A.N. Donev*

Department of Probability and Statistics, University of Sheffield, Sheffield S3 7RH, UK

Received 4 February 2003; accepted 3 September 2003

Abstract

This paper is concerned with the statistical properties of experimental designs where the factor levels cannot be set precisely. When the errors in setting the factor levels cannot be measured, design robustness is explored. However, when the actual design could be measured at the end of the investigation, its optimality is of interest. D -optimality could be assessed in different ways. Several measures are compared. Evaluating them is difficult even in simple cases. Therefore, in general, simulations are used to obtain their values. It is shown that if D -optimality is measured by the expected value of the determinant of the information matrix of the experimental design, as has been suggested in the past, on average the designs appear to improve with the variance of the error in setting the factor levels. However, we argue that the criterion of D -optimality should be based on the inverse of the information matrix. In this case it is shown that the experiment could be better or worse than the planned one. It is also recognized that setting the factor levels with error could lead to an increased risk of losing observations, which on its own could reduce considerably the optimality of the experimental designs. Advice on choosing the design region in such a way that such a risk is controlled to an acceptable level is given.

© 2003 Elsevier B.V. All rights reserved.

Keywords: D -optimality; Design region; Errors in variables; Lost observations; Random experiments

1. Introduction

Various experimental designs are used routinely in many statistical investigations. For example two-level factorial or fractional designs are used when a response variable of interest can be explained by a first order polynomial model that may also contain interactions of a number of factors. Response surface designs are used when

* Tel.: +44-114-2223751; fax: +44-144-223759.

E-mail address: a.n.donev@sheffield.ac.uk (A.N. Donev).

the relationship between the response and the factors is more complicated. If the experimental situation is such that a tailor-made design is needed, then computer algorithms can be used to obtain an experimental design that satisfies exactly the needs of the experimenter. In each of these cases the experimental design assigns values of the factors for each observation of the response variable. However, there are situations where setting the levels of some or all of the factors during the experiment is not easy and they are set with some error. A typical case is where the variables are surrogate. For example, in a bioassay the effect of the dose of a drug on the condition of a patient may be studied. It is believed that this effect is dependent on the blood concentration of the drug. However, setting this concentration for a patient is not possible in practice. Instead, the dose is prescribed as per kilogram of body weight of the patient. Consequently, the target blood concentration is achieved only approximately. On the other hand, after the administration of the drug its blood level might be possible to measure precisely.

Even when the experimenter can set the factors directly, this is often done with some errors. In some cases the errors in variables can be measured, i.e. the true values of the variables are known. In others, this cannot be done and only the target values set by the experimental design are known. Common belief is that when the errors in variables are small they can be ignored. However, there is little available practical guidance for how small they should be for this to be justified.

Considerable attention has been given to the analysis of data when there are errors in the variables (see for example Fuller, 1987, and Cheng and Van Ness, 1999). The effect of errors in the factor levels on the statistical properties of the parameters obtained from a two-level factorial and fractional factorial designs was first studied by Box (1963). Draper and Beggs (1970) measure the robustness of experimental designs to errors in the factor levels by the sum of squared differences between the observed and the predicted with the model response values. They assume that the errors are very small and show that zero first and odd second order moments of the designs are desirable. However, they also recognize that finding an analytical proof for any optimality conditions when there are two or more factors is very difficult and recommend searching numerically for a solution. Vuchkov and Boyadjieva (1983) point out that in general the errors in the factor levels introduce heterogeneity in the variance of the response and propose the robustness of the design to such errors to be measured by the maximum element of the covariance matrix for the response or by its trace. Pronzato (1998, 2002) also considers the case when the planned (or target) design in the experiment cannot be achieved due to errors in the factor levels. Unlike the previous authors, he considers the case when the true levels are measured correctly. He gives an expression for the expected value of the determinant of the information matrix for the design and uses it as a D -optimality criterion to show that a design that is optimum according to this criterion may differ from a design that is optimum when there are no errors in the factor levels.

In this paper, we investigate the case when the effect of the errors cannot be ignored. In Section 2 we discuss how the analysis of the data is affected if the actual design is not known and introduce a measure for design robustness that is easier to calculate and interpret than those used previously by Draper and Beggs (1970) and Vuchkov

and Boyadjieva (1983). We show that when the actual design can be recorded exactly and prior information about the errors in setting the factor levels is available, other generalizations of the D -optimality criterion than that used by Pronzato (1998, 2002) are more appropriate. As the criteria of optimality are difficult to evaluate analytically, in Section 3 we illustrate how prior information on the distribution of the errors in the factor levels can be used to obtain the empirical distribution of both the designs of interest and the corresponding criteria of optimality. We also estimate the probability of using a better or worse design than the planned one. Standard designs such as the two-level factorial designs and the three-level factorial designs are used as examples because of their extensive use in practice. Section 4 provides considerations of how the design region should be chosen when errors in setting the factor levels might lead to loss of observations. In this case a compromise is sought between using a large design region and the loss in efficiency of the estimates of the model parameters due to a possible loss of observations. The paper concludes with a discussion of the results.

2. The effect of errors in the settings of the factor levels

We will be concerned with the situations where a response variable of interest can be explained by the linear model

$$\mathbf{Y} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where \mathbf{Y} is a vector of n observations, $\boldsymbol{\beta}$ is a vector of p unknown parameters and \mathbf{F} is the extended design matrix consisting of known functions of m factors x_1, x_2, \dots, x_m . The experimental errors $\boldsymbol{\varepsilon}$ are assumed to be independent, normally distributed and have zero mean and a common variance σ^2 . Customarily, polynomial models are used. We will be interested in first order models with or without interactions, and in models that may also include quadratic terms. Therefore, we derive our results for the model for an observation:

$$Y = \beta_0 + \sum_{i=1}^m \beta_i x_i + \sum_{i=1}^m \beta_{ii} x_i^2 + \sum_{i=1}^{m-1} \sum_{j=i+1}^m \beta_{ij} x_i x_j + \sum_{i=1}^{m-2} \sum_{j=i+1}^{m-1} \sum_{k=j+1}^m \beta_{ijk} x_i x_j x_k + \dots + \varepsilon. \quad (1)$$

Note that the models of interest are easily obtained from Eq. (1) by setting the appropriate parameters to zero. For example, a standard second order polynomial is obtained from it by setting all third and higher order interactions to zero.

In a designed experiment the values of the factors for each of the observations are set to target values (say $x_i = t_i$), which may be chosen using some optimality considerations. When the levels of the factors are set with errors, their values become random variables. For example, if the planned (or target) value of the i th factor in the u th observation is t_{ui} , a value X_{ui} ,

$$X_{ui} = t_{ui} + e_{ui}, \quad (2)$$

rather than t_{ui} generates the data. In Eq. (2) e_{ui} is the error in setting the i th factor in the u th observation. Naturally, the actual extended design matrix, \mathbf{F} , is different to that of the planned experiment, say \mathbf{T} .

In most cases knowledge about the accuracy of the tools used for setting the levels of the factors could be used to define appropriate assumptions regarding the distributions of the errors. We will mainly be concerned with the case where it is possible to assume that the errors are independent and normally distributed with zero mean and variance denoted by σ_i^2 , i.e. $e_i \sim N(0, \sigma_i^2)$, $i = 1, \dots, m$, $E[e_i] = E[e_i e_j] = E[e_i e_j e_k] = 0$, $E[e_i^2] = \sigma_i^2$ and $E[e_i^4] = 3\sigma_i^4$. Hence, $X_i \sim N(t_i, \sigma_i^2)$. More complicated error structures are possible too. We will also review an example given in Pronzato (2002) where σ_i^2 is assumed to be proportional to the value of the i th factor.

The factors can be scaled easily to vary between -1 and 1 and in this case in practice the variances σ_i^2 are considerably smaller than 1 . Substituting Eq. (2) in (1) gives

$$\begin{aligned} Y = & \beta_0 + \sum_{i=1}^m \beta_i(t_i + e_i) + \sum_{i=1}^m \beta_{ii}(t_i + e_i)^2 + \sum_{i=1}^{m-1} \sum_{j=i+1}^m \beta_{ij}(t_i + e_i)(t_j + e_j) \\ & + \sum_{i=1}^{m-2} \sum_{j=i+1}^{m-1} \sum_{k=j+1}^m \beta_{ijk}(t_i + e_i)(t_j + e_j)(t_k + e_k) + \dots + \varepsilon. \end{aligned} \quad (3)$$

The subscript u has been omitted in order to simplify the notation. Partitioning model (3) in order to separate the deterministic and stochastic terms gives

$$\begin{aligned} Y = & \beta_0 + \sum_{i=1}^m \beta_i t_i + \sum_{i=1}^m \beta_{ii} t_i^2 + \sum_{i=1}^{m-1} \sum_{j=i+1}^m \beta_{ij} t_i t_j + \sum_{i=1}^{m-2} \sum_{j=i+1}^{m-1} \sum_{k=j+1}^m \beta_{ijk} t_i t_j t_k + \dots \\ & + \sum_{i=1}^m \left\{ \beta_i + 2\beta_{ii} t_i + \sum_{\substack{j=1 \\ j \neq i}}^m \beta_{ij} t_j + \sum_{\substack{j=1 \\ j \neq i}}^m \sum_{\substack{k=j+1 \\ k \neq i}}^m \beta_{ijk} t_j t_k + \dots \right\} e_i \\ & + \sum_{i=1}^m \beta_{ii} e_i^2 + \sum_{i=1}^{m-1} \sum_{j=i+1}^m \left(\beta_{ij} + \sum_{\substack{k=1 \\ k \neq i; k \neq j}}^m t_k \beta_{ijk} \right) e_i e_j \\ & + \sum_{i=1}^{m-2} \sum_{j=i+1}^{m-1} \sum_{k=j+1}^m \beta_{ijk} e_i e_j e_k + \dots + \varepsilon. \end{aligned}$$

Hence

$$E[Y] = \mathbf{t}^T \boldsymbol{\beta} + \sum_{i=1}^m \beta_{ii} \sigma_i^2, \quad (4)$$

where \mathbf{t}^T is the corresponding row of the extended design matrix \mathbf{T} specified by the experimental design and containing the values of the regressors included in the model: $1, t_1, \dots, t_m, t_1^2, \dots, t_m^2, t_1 t_2, \dots, t_1 t_2 t_3, \dots$ (or a specified subset) evaluated for the observation. Equation (4) reiterates the finding of Box (1963) that in the presence of errors in variables the expectation of a quadratic polynomial model is biased.

If the parameters β are known

$$\begin{aligned} \text{var}(Y) = & \sum_{i=1}^m \left(\beta_i + 2\beta_{ii}t_i + \sum_{\substack{j=1 \\ j \neq i}}^m \beta_{ij}t_j + \sum_{\substack{j=1 \\ j \neq i}}^m \sum_{\substack{k=j+1 \\ k \neq i}}^m \beta_{ijk}t_j t_k + \dots \right)^2 \sigma_i^2 + 2 \sum_{i=1}^m \beta_{ii}^2 \sigma_i^4 \\ & + \sum_{i=1}^{m-1} \sum_{j=i+1}^m \left(\beta_{ij} + \sum_{\substack{k=1 \\ k \neq i; k \neq j}}^m t_k \beta_{ijk} \right)^2 \sigma_i^2 \sigma_j^2 \\ & + \sum_{i=1}^{m-2} \sum_{j=i+1}^{m-1} \sum_{k=j+1}^m \beta_{ijk}^2 \sigma_i^2 \sigma_j^2 \sigma_k^2 + \sigma^2. \end{aligned} \quad (5)$$

As all terms are positive, it is clear that the errors in factor levels lead to increased and heterogeneous variance of the response, hence in loss of efficiency.

In practice there are three situations of interest: (i) the \mathbf{F} matrix is unknown and the variances σ_i^2 of the errors in the variables are unknown; (ii) the \mathbf{F} matrix is unknown but σ_i^2 are known and (iii) the \mathbf{F} matrix is known. We discuss the analysis of the data in these cases in turn and consider measures for the design robustness in the presence of errors in factor levels.

2.1. Unknown \mathbf{F} matrix; unknown σ_i^2

In this case both β and the variance parameters of the model have to be estimated. Standard statistical packages do not provide appropriate options. Fedorov (1974) and Donev (2000) apply iterative procedures where the method of weighted least squares is used to obtain an estimate of β . For example, Donev (2000) calculates the weights using Eq. (5) and updates them at each iteration by finding the maximum likelihood estimates for the variance components, until the search converges. If necessary, β is adjusted for bias. In the first iteration, when no estimates for β are available, the ordinary least squares method is used. A comprehensive treatment of the problem of regression analysis with measurement error is provided by Fuller (1987) and Cheng and Van Ness (1999).

2.2. Unknown \mathbf{F} matrix; known σ_i^2

When σ_i^2 , $i=1, \dots, m$, are known, similar iterative procedures to that described above can be used. Certainly, as the variance components will not have to be estimated a more accurate variance-covariance matrix for the model parameters will be used. An

advantage of this procedure is also that it can easily be programmed in any of the mainstream software packages using the existing modules for data analysis.

Eq. (4) shows that when the model to be fitted is second order polynomial, the \mathbf{F} matrix is not known and the matrix \mathbf{T} is used instead, the intercept may have to be adjusted for bias. The bias does not depend directly on the experimental design but the accuracy of its estimation depends on the accuracy of estimating the parameters corresponding to the quadratic terms (if they are needed in the model). Another negative effect of the errors in variables is the inflated variance of the response which reduces the precision of estimating the model parameters. Hence, we define the following criterion of robustness of an experimental design to errors in the settings of the explanatory variables.

Definition. An experimental design is robust against errors in variables if

$$V = \sum_{u=1}^n \text{var}(Y_u), \quad (6)$$

where $\text{var}(Y_u)$ is calculated using Eq. (5), is smaller than that for any other designs of size n and for the same model.

It seems best to use this as a secondary criterion to choose a design amongst those that are optimum with respect to the main criterion of optimality.

Example 1. Suppose the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 \quad (7)$$

has to be fitted using 4 observations and that the factor x_1 is scaled to vary between -1 and 1 . It can be shown that a design that has one observation at -1 , 0 and 1 , with either of these points being duplicated is D -optimum (i.e. has the largest determinant of the information matrix). We will compare these three possible designs using Eq. (6). The V values

$$V = \sigma_1^2 \sum_{i=1}^4 (\beta_1 + 2\beta_2 T_i)^2 + 4\beta_2^2 \sigma_1^4 + 4\sigma^2$$

of the designs will differ only due to the contribution of the replicated (fourth) observation, i.e. by the value of $(\beta_1 + 2\beta_2 T_4)^2$. Now it can be seen that setting this observation to $T_4 = -\beta_1/2\beta_2$, $\beta_2 \neq 0$, would minimize V . Certainly this value may turn out to be outside of the design region in which case the value -1 or 1 closest to T_4 has to be chosen. Setting T_4 in the middle of the design region (i.e. at zero) would result in a useful compromise when the values of β_1 and β_2 are not known.

Performing similar calculations in order to compare various designs in more complicated cases is not computationally demanding. Similarly to the simple example above, in the absence of prior knowledge about the model parameters, it would be useful to choose a design for which V is relatively small for any feasible values. This also

applies to the case described in Section 2.1 where the variances of the errors in variables are unknown.

2.3. Known \mathbf{F} matrix

There are cases where it is difficult to set precisely the values of the factors as determined by the experimental design, but once they are set their values can be measured exactly. In those cases the \mathbf{F} matrix can be calculated and least squares estimates of $\boldsymbol{\beta}$ can be obtained easily using the standard least squares estimator $\hat{\boldsymbol{\beta}} = \mathbf{G}^{-1}\mathbf{F}^T\mathbf{Y}$, where $\mathbf{G} = \mathbf{F}^T\mathbf{F}$ is the true design information matrix.

Standard criteria of optimality of experimental designs are functions of the information matrix of the design. For example, the D -optimality criterion requires $\sigma^2|\mathbf{G}^{-1}|$, or simply $|\mathbf{G}^{-1}|$, to be minimized because this minimizes the confidence ellipsoid of the model parameters of interest. Often it is more convenient to require maximization of $|\mathbf{G}|$. An equivalent but sometimes easier to tackle criterion requires $\log(|\mathbf{G}^{-1}|) = -\log(|\mathbf{G}|)$ to be minimized.

When there are errors in setting the values of the factors in a designed experiment \mathbf{G} , \mathbf{G}^{-1} , $|\mathbf{G}|$ and $|\mathbf{G}^{-1}|$ are random variables whose distributions are generated through the distributions of the errors in variables. Therefore Pronzato (2002) proposes experimental designs to be compared or constructed using a generalized criterion of D -optimality requiring $E\{|\mathbf{G}|\}$ or $|E\{\mathbf{G}\}|$ to be maximized.

Example 2. Pronzato (2002) considers the effect of setting the levels of a factor with errors that are independent and normally distributed with variance $0.03t^2$, where t is the level of the factor specified by the experimental design. The experimental design consists of 9 observations that are evenly divided between 3 design points: 0, ζ and 1, $0 \leq \zeta \leq 1$ and the model that has to be estimated is given by Eq. (7).

For instance, such an experimental design may be used when the amount of the intermediate product of two consecutive chemical reactions can be modelled with a quadratic polynomial model and the factor is the time since the start of the chemical reaction. While it might be difficult to take measurements at the precise times, the times of taking them could easily be recorded accurately.

If there are no errors in setting t and if $\zeta = 0.5000$ the design is D -optimum. Pronzato points out that in the presence of such errors $\zeta \approx 0.5405$ maximizes $E\{|\mathbf{G}|\}$, while $\zeta = 1.0$ maximizes $|E\{\mathbf{G}\}|$.

Note that when $\zeta = 1.0$ the quadratic model cannot be estimated unless there are errors in the settings of the factor. Certainly it is very unlikely that an experimenter would ever consider using an experimental design relying on errors in the factor levels to obtain estimates of the model parameters even if they may turn out to be slightly more precise than when ζ is set at 0.5000.

Apart from that, this small example is instructive. The optimum experimental design for the case where there are errors in the settings of the factors can be different to that when the design can be implemented precisely. On the other hand, it is clear that the criterion of optimality needs to be chosen carefully.

Table 1
Generalized criteria of D -optimality

Criterion	$\Psi\{\mathbf{G}\}$
D_{R1}	$E[\log \mathbf{G}^{-1}]$
D_{R2}	$\log E[\mathbf{G}^{-1}]$
D_{R3}	$\log E[\mathbf{G}^{-1}] $
D_{R4}	$\log\{E[\mathbf{G}]\}^{-1}$
D_{R5}	$\log E[\mathbf{G}] ^{-1}$

While maximizing $|\mathbf{T}^T\mathbf{T}|$ and minimizing $|(\mathbf{T}^T\mathbf{T})^{-1}|$ are equivalent, in general maximizing $E\{|\mathbf{G}|\}$ or $|E\{\mathbf{G}\}|$ is not equivalent to minimizing $E\{|\mathbf{G}^{-1}|\}$. On the other hand it is the latter that ensures minimum expected volume of the confidence ellipsoid of the model parameters and truly implements the D -optimality criterion of optimality.

Clearly, the requirement for D -optimality can be approached in several ways. We adapt generalizations of the criterion requiring minimization of functions of the information matrix or its inverse $\Psi\{\mathbf{G}\}$ as listed in Table 1 (see Chapter 19, [Atkinson and Donev, 1992](#)). We denote them here D_{Ri} , where $i = 1, 2, 3, 4$ or 5 , and the subscript R has been used to indicate that the design is random. A design for which D_{Ri} is minimum is called D_{Ri} -optimum. If no information about the parameters $\boldsymbol{\beta}$ is available, a non-informative prior will be allocated to them.

The criteria of optimality that we use are similar to that of the average, or sometimes referred to as Bayesian criteria, for D -optimality (see p. 101 of [Fedorov and Hackl, 1997](#), for details). However, their interpretation is somehow different. In the case considered in this paper the criteria of optimality refer to properties of a population of designs with errors in variables. In an experiment a single design is used. The probabilities that it would turn out better or worse than the target one are influenced by the choice of experimental design.

A design that is optimum with respect to all criteria in Table 1 may not exist. It can be seen that D_{R1} , D_{R2} and D_{R3} make use of \mathbf{G}^{-1} and therefore require minimization of the expected variance of the model parameters, while D_{R4} and D_{R5} are based on \mathbf{G} and require maximization of the expected information. Note that D_{R5} is easily obtained from the criterion $|E\{\mathbf{G}\}|$ used by Pronzato. However, D_{R2} is the most useful criterion as it gives the expected volume of the confidence region of the estimates of the model parameters.

Evaluating the optimality of experimental designs with respect to the criteria listed in Table 1 is not straightforward even in simple cases. For illustration in the next section we investigate a simple but widely used class of experimental designs, the two-level factorial or fractional factorial designs, for which some progress is possible.

2.4. Two-level factorial designs

A two-level factorial design for m factors consists of all possible combinations of their levels or a fraction of them. The choice of the fraction to be used determines

what statistical models will be estimable from the data obtained in the experiment and what their statistical properties will be. Box et al. (1978, Chapter 12) provide a comprehensive discussion and advice.

If a complete factorial design is used, a first order polynomial model that may include some or all interactions between the factors can be estimated. This model is obtained from Eq. (1) by setting the quadratic terms to zero, i.e.

$$Y = \beta_0 + \sum_{i=1}^m \beta_i t_i + \sum_{i=1}^{m-1} \sum_{j=i+1}^m \beta_{ij} t_i t_j + \sum_{i=1}^{m-2} \sum_{j=i+1}^{m-1} \sum_{k=j+1}^m \beta_{ijk} t_i t_j t_k + \cdots + \varepsilon.$$

If a fractional factorial design is used some or all of the interactions will not be estimable. The information matrix $\mathbf{G} = (\mathbf{T}^T \mathbf{T})$ is diagonal and the estimates of the model parameters are obtained independently from each other. The two-level factorial designs are known to be optimum with respect to many optimality design criteria. For example, they are D -optimum amongst the designs with the same size and for the same model. Customarily the levels of the factors are scaled to -1 and 1 . In this case the information matrix of a complete factorial or fractional factorial design is $\mathbf{G} = n\mathbf{I}_p$ and therefore $D = \log|\mathbf{G}^{-1}| = -p \log n$.

Theorem 1. *If all factors in a two-level complete or fractional factorial design are set with error $e \sim N(0, \sigma_e^2)$*

$$\log|E[\mathbf{G}]| = p \log n + h \log(1 + \sigma_e^2),$$

where $h = m + \sum_{i=2}^m i k_i$ and k_i is the number of i -factor interactions in the model, $i \in [2, \dots, m]$. The proof is given in Appendix A.

Taking into account that in most practical cases σ_e^2 is small and using Maclaurin's series expansion we obtain

$$\log|E[\mathbf{G}]| \approx p \log n + h \sigma_e^2.$$

Clearly $D_{R5} = -\log|E[\mathbf{G}]|$ decreases with increasing variance of the errors in factor levels. This result can be generalized for the case where the variances of the errors of the factors are different.

Deriving explicit expressions for the remaining criteria of optimality even in simple cases is usually not possible. Similar difficulties in evaluating criteria of optimality analytically have been reported by Draper and Beggs (1970) and Pronzato (2002). On the other hand such results may not be necessary as the criteria of optimality can be estimated relatively easy with the use of simulations. We illustrate this in the following section.

3. Simulation studies

The discussion in Section 2 showed that if the prior distribution of the errors in variables is known it can be taken into account at the stage of designing of the experiment and the analysis of the results. In such cases the distribution of the experimental

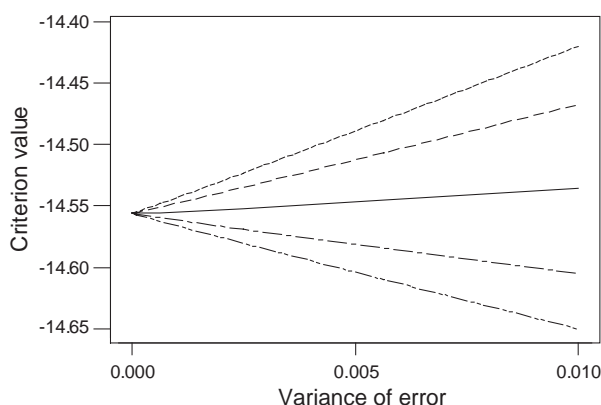


Fig. 1. Values of the criteria of optimality for a two-level complete factorial design for 3 factors and identical variances for the errors for each of them, first order polynomial model with two-factor interactions. From bottom to top: D_{R5} , D_{R4} , D_{R1} , D_{R3} , D_{R2} .

designs that can occur and the corresponding distributions of the criteria of optimality listed in Table 1 can be obtained by simulation. All results in this section are obtained by simulating 10000 designs where we assume that the errors in setting the factors are independent and normally distributed.

Firstly, we investigate the effect of errors of variables on the values of the optimality criteria D_{R1} , D_{R2} , D_{R3} , D_{R4} and D_{R5} for some standard experimental designs. For simplicity we assume that the variances of the errors are identical for all explanatory variables and that they can take values between 0 and 0.01.

Example 3. Fig. 1 shows the values of the optimality criteria D_{R1} , D_{R2} , D_{R3} , D_{R4} and D_{R5} when a two-level complete factorial design in 3 factors is to be used to estimate a first order polynomial model that also includes all three two-factor interactions. Interestingly, the curves for the criteria of optimality are nearly linear with different slopes. Theorem 1 establishes this property only for D_{R5} . As discussed earlier, we consider D_{R2} as the most appropriate criterion to use. The criteria values based on the elements of the information matrix D_{R4} and D_{R5} steadily decrease with the variance of the errors, suggesting an improvement of the experimental designs. However, D_{R1} , D_{R3} and most importantly D_{R2} increase with the variance of the errors, suggesting the opposite. Therefore in this case the evidence provided by D_{R4} and D_{R5} is misleading.

Example 4. We consider the case where the model that has to be estimated is a second order polynomial in two factors. Therefore a three-level complete factorial design is used. Fig. 2 shows the values of the criteria of optimality for different variances of the errors in the factor levels. Similarly to Example 3 each criterion changes with the variance of the error in the factor settings in a nearly linear fashion. The design again worsens with respect to D_{R2} with the variance of the errors in variables. However, a notable difference from the previous example is that D_{R1} decreases slightly with the variance of the errors. Hence, this criterion could also be misleading.

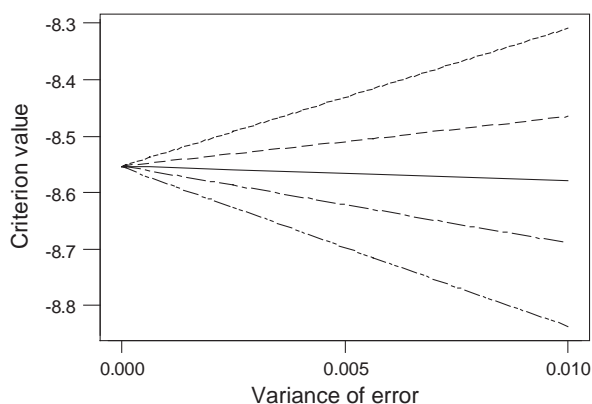


Fig. 2. Values of the criteria of optimality for a three-level complete factorial design for 2 factors and identical variances for the errors for each of them, second order polynomial model. From bottom to top: D_{R5} , D_{R4} , D_{R1} , D_{R3} , D_{R2} .

Table 2

Values of criteria of optimality for model (7) for 9 trial designs replicating equally 0, ζ and 1

	Design		
	D1 ($\zeta = 0.47$)	D2 ($\zeta = 0.50$)	D3 ($\zeta = 1.00$)
D_{R1}	-0.8650	-0.8722	-0.7363
D_{R2}	-0.6656	-0.6600	-0.4008
D_{R3}	-0.5468	-0.5420	-0.1257
D_{R4}	-1.0434	-1.0579	-0.9882
D_{R5}	-1.1400	-1.1521	-1.1716

Examples 3 and 4 show that the evidence regarding the optimality of the designs provided by D_{R2} may be contradicted by D_{R1} , D_{R4} and D_{R5} . Therefore the usefulness of the latter is questionable even when their evaluation is considerably easier than that for D_{R2} .

In practice usually one experimental design is used. It could turn out to be better or worse in terms of D -optimality than the planned one. Therefore it would be useful to know the probabilities of these events. When D_{R2} is evaluated by simulation, the generated distribution of the covariance matrices for the estimates of the model parameters β could be used to estimate the probability, say p_B , of using a better design than the target one.

Example 2 (continued). Table 2 lists the values of the criteria listed in Table 1 for various values of ζ (note that small values are desirable!). The design D2 is D -optimum in absence of errors in variables, while design D3 found by Pronzato (2002) with $\zeta = 1$ is best with respect to D_{R5} . However, a value of $\zeta = 0.47$ minimizes D_{R2} and therefore

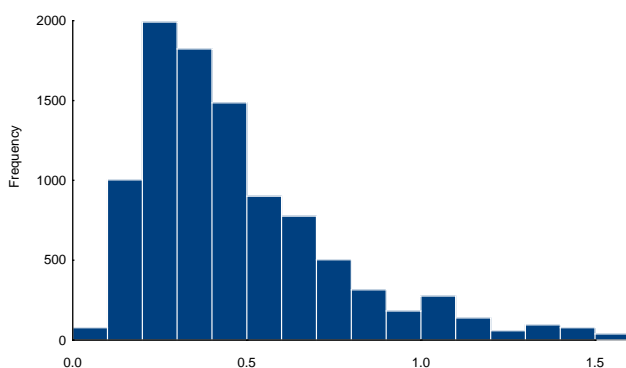


Fig. 3. Frequency distribution of $|\mathbf{G}^{-1}|$ for design D2 in Table 2 for model (7) when the errors in setting x are $e \sim N(0, 0.3x^2)$. $|\mathbf{G}^{-1}| = 0.523$ if there are no errors in the setting of x .

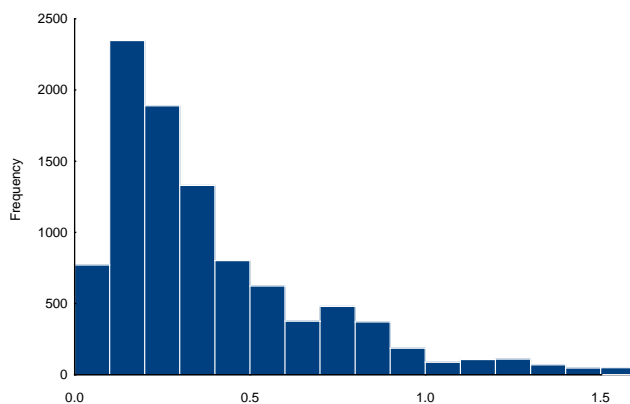


Fig. 4. Frequency distribution of $|\mathbf{G}^{-1}|$ for design D2 in Table 2 for model (7) when the errors in setting x are $e \sim N(0, 0.3)$.

is the D -optimum under the stated assumptions regarding the errors in variables. The probabilities of using a better design than the planned one as a result of the errors in setting the levels of the factor p_B for designs D1, D2 and D3 are 0.7204, 0.7234 and 0.6416, respectively.

By the use of simulations we can also compare the performance of designs under different error assumptions. Figs. 3 and 4 show the frequency distributions of the values for $|\mathbf{G}^{-1}|$ of the design D2 in Table 2 for model (7) when the errors in setting the factor x are $e \sim N(0, 0.3x^2)$ and $e \sim N(0, 0.3)$, respectively. Both distributions are fairly skewed to the right. In both cases the mean of D_{R2} is considerably smaller than $|\mathbf{G}^{-1}|$ for the design without errors. Thus, on average the design improves with respect to D -optimality with the variance of the errors in variables. Despite this there appears

to be a probability of about 0.28 in the first case and 0.23 in the second of getting a worse design than the planned one.

Simulations such as those described in this section can be easily obtained for different models and different assumptions about the distribution of the errors in setting the factors. Sample S-Plus programs are available from the author on request. Clearly, being able to obtain information about the distribution of the random designs from which the one to be used is taken is preferable to dealing only with an expected value of a design criterion.

In all examples that were investigated the inequalities

$$D_{R2} \geq D_{R3} \geq D_{R1} \geq D_{R4} \geq D_{R5}$$

were satisfied. Proving their validity in general is difficult and, we trust, not really necessary.

4. Considerations for the design region

The results of the previous sections show that experimental designs can be improved in terms of D -optimality by the errors in the factor levels in many cases. This is not surprising because in those cases the factors take values smaller than -1 or larger than 1 , so that the design region is effectively increased. On the other hand in practice setting the factor levels precisely can be time consuming and expensive. Therefore, it is important to know when it is worth spending resources on ensuring high precision.

For the statistical analysis of the data to be correct it is important that the actual values of the factors remain within the range of validity of the assumed statistical model. It is a common practice to select the intervals in which the factors vary during the experiment as wide as possible as this leads to gain in D -optimality and in a larger region of validity of the statistical model. Hence, as a result of the errors in setting the factors the actual values may fall outside the valid experimental region. Then, using such measured response values may bring bias to the analysis of the data. There may also be situations where the response may not be possible to measure. On the other hand, the loss of one or more observations would certainly considerably worsen the statistical properties of the estimates of the model parameters. Therefore, in such cases it would be important for the experimenter to be able to control the probability of losing observations by selecting appropriately the design region. We will illustrate this problem by considering the two-level factorial and fractional factorial designs with levels -1 and 1 .

Suppose that the observations can be taken safely between $-(1+a)$ and $(1+a)$, the design region is defined to be between -1 and 1 , and that (for simplicity) $e \sim N(0, \sigma_e^2)$ for each of the factors. Observations for which the factors fall outside the interval $-(1+a)$ and $(1+a)$ will be considered lost. Let $\phi = \Pr(|x_i| < (1+a))$. Hence, the probability that a design point will be lost is $(1 - \phi^m)$. The probability that, say, v of the points of the factorial design will be lost becomes

$$P_v = \binom{n}{v} (1 - \phi^m)^v \phi^{m(n-v)}. \quad (8)$$

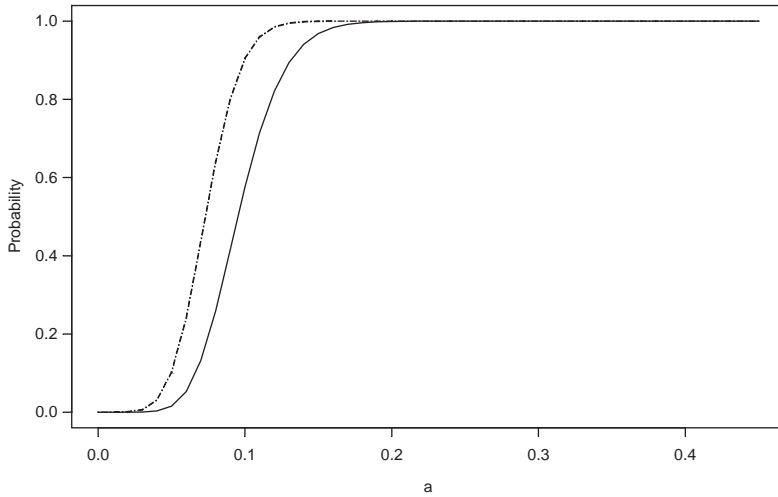


Fig. 5. Probabilities of losing no more than 1 observation: $\sigma_e = 0.05$ —dotted line, $\sigma_e = 0.1$ —solid line.

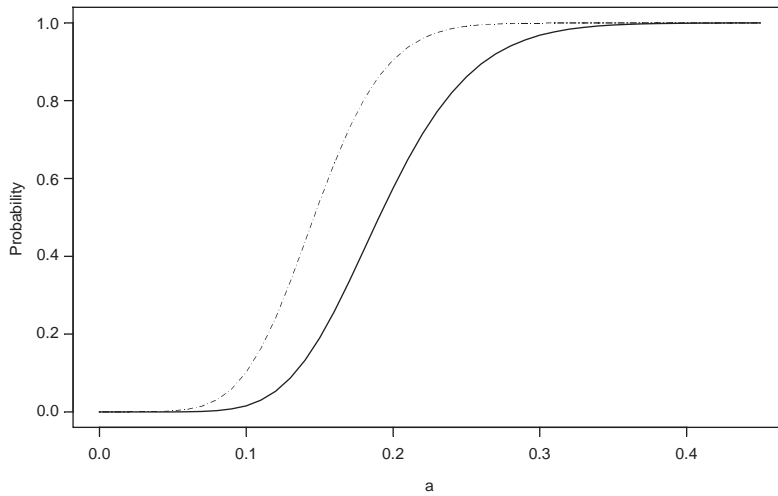


Fig. 6. Probabilities of no lost observations: $\sigma_e = 0.05$ —dotted line, $\sigma_e = 0.1$ —solid line.

This formula can be very helpful when the intervals in which the variables vary during the experiment are chosen.

Example 3 (continued). Fig. 5 shows the probabilities that no more than one observation will be lost for different values of a , for $\sigma_e = 0.05$ and $\sigma_e = 0.1$. Choosing the design region in such a way that $a = 0.09$ and 0.13 , respectively, assures over 99% probability that not more than one observation will be lost. Fig. 6 shows the probabilities that no

observations will be lost, the corresponding values of a are considerably bigger, 0.17 and 0.34, respectively.

On the other hand, in the absence of errors in variables the design region could span between $-(1+a)$ and $(1+a)$. In this case

$$-\log |\mathbf{G}| = -p \log n - 2p \log(1+a) \approx -p(\log n + 2a),$$

and could be considerably smaller than the corresponding value when the design region is placed within the interval $(-1, 1)$ of each of the factors. Hence, loss in D -optimality could come from the attempt to guard against losing observations or due to the loss of observations.

In general, the value a may not be the same for all variables and will depend on the real scale of the variables. However, Eq. (8) can easily be generalized for any particular case.

5. Discussion

Difficulties in setting precisely the levels of the factors in an experimental design affect the results of the investigation. When the errors cannot be measured, the actual design that is used is unknown. Therefore, the planned design is used in the analysis of the data. This introduces heterogeneity in the variance of the experimental error and leads to loss of efficiency in estimating the parameters of the model. In some cases it also brings bias to the prediction. Our results suggest that in such cases, a design that is most robust against errors in variables should be selected from a set of designs that perform well according to the main criterion of optimality. The robustness can be measured by the level of heterogeneity in the variance introduced by the errors in variables.

The problem is different when the actual design is known at the end of the study. Certainly, in this case the analysis of the data is straightforward. We showed that when the distribution of the errors in setting the factors is known it could be possible to obtain a D -optimum design that is different to the optimum design when there are no errors. There are several ways in which D -optimality can be assessed in this case. We consider D_{R2} as the most appropriate criterion to use. It is also worth considering other measures based on the simulated distributions such as median and inter-quartile range for the distribution of the criterion and the probability of using a better design than the planned one. While in some simple cases it might be possible to obtain the analytical form for the chosen criterion of optimality, in general this is not possible. The results of Section 4 show that simulations can be used effectively to generate the distribution of the possible designs and estimate the values of the required criteria of optimality.

Perhaps the most striking result of this study, related to that of Demidenko and Spiegelman (1997) and Pronzato (2002), is that designs may be improved in terms of D -optimality by allowing for random errors in the factor settings. However, the results of Sections 3 and 4 show that such a claim must be taken cautiously, especially when

the errors in setting the factor levels can lead to loss of observations. When such errors are anticipated, their potential effect should be taken into account in selecting both the design region and the design of the experiment.

Acknowledgements

The author is indebted to the editor and the other referees for their constructive suggestions that led to an improvement of the paper. He is also grateful to Rainer Schwabe and Nick Fieller for useful discussions.

Appendix A

The exact form of the information matrix depends on the model that is fitted. We are interested in fitting model (1) or simpler models. The diagonal elements g_{ii} in this case are as follows (all summations are for $u = 1, \dots, n$):

$$g_{11} = n,$$

$$g_{i+1,i+1} = \sum (T_{ui} + e_{ui})^2,$$

for $i = 1, \dots, m$, the terms corresponding to the second order interactions are

$$\sum (T_{ui} + e_{ui})^2 (T_{uj} + e_{uj})^2,$$

for $i = 1, \dots, m-1$, $j = i+1, \dots, m$. If some or all of the third order interactions are also included, the corresponding elements are

$$\sum (T_{ui} + e_{ui})^2 (T_{uj} + e_{uj})^2 (T_{ul} + e_{ul})^2,$$

for $i = 1, \dots, m-2$, $j = i+1, \dots, m-1$, $l = j+1, \dots, m$, etc.

The off diagonal elements $g_{ij} = g_{ji}$ correspond to first order terms in the model, products of first order terms, products of first order terms and interactions and so on and are as follows: $\sum (T_{ui} + e_{ui})$, for $i = 1, \dots, m$; $\sum (T_{ui} + e_{ui})(T_{uj} + e_{uj})$, for $i = 1, \dots, m-1$, and $j = i+1, \dots, m$; $\sum (T_{ui} + e_{ui})(T_{uj} + e_{uj})(T_{ul} + e_{ul})$, for $i = 1, \dots, m-2$, $j = i+1, \dots, m-1$, $l = j+1, \dots, m$, etc. For any complete or fractional factorial designs $\sum T_{ui} = 0$ for any i , $\sum T_{ui}T_{uj} = 0$ for any $i < j$, etc. As $E[\varepsilon_i] = E[\varepsilon_i\varepsilon_j] = E[\varepsilon_i\varepsilon_j\varepsilon_k] = 0$ for any i, j and k , it can be seen that the expected values of all non-diagonal elements of \mathbf{G} are zero.

All diagonal elements of the information matrix, except g_{11} which is n , are sums of n products of, say q , terms $(T_{ui} + e_{ui})^2$, $i \in [1, \dots, m]$. The value of q is 1 for the terms that correspond to main effects in the model, it is 2 for the terms that correspond to second order interactions, etc. As $E[\sum (T_{ui} + e_{ui})^2] = n(1 + \sigma_e^2)$ for any i , we obtain

$$E[\mathbf{G}] = \text{diag}\{n(1 + \sigma_e^2) \dots n(1 + \sigma_e^2) n(1 + \sigma_e^2)^2 \dots n(1 + \sigma_e^2)^2 n(1 + \sigma_e^2)^3 \dots n(1 + \sigma_e^2)^3 \dots\}$$

and clearly

$$\begin{aligned}\log|E[\mathbf{G}]| &= \log\{n^p(1 + \sigma_e^2)^m(1 + \sigma_e^2)^{2k_2}(1 + \sigma_e^2)^{3k_3} \dots\} \\ &= p \log n + h \log(1 + \sigma_e^2).\end{aligned}$$

If there are no errors in variables $\mathbf{G} = \mathbf{F}^T\mathbf{F} = \mathbf{T}^T\mathbf{T}$ is not random and naturally $\log|E[\mathbf{G}]| = p \log n$.

References

- Atkinson, A.C., Donev, A.N., 1992. Optimum Experimental Design. Clarendon Press, Oxford.
- Box, G.E.P., 1963. The effect of errors in the factor levels and experimental design. *Technometrics* 5, 247–262.
- Box, G.E.P., Hunter, J.S., Hunter, W.G., 1978. Statistics for Experimenters. Wiley, New York.
- Cheng, C.-L., Van Ness, J.W., 1999. Statistical Regression with Measurement Error. Arnold, London.
- Demidenko, E., Spiegelman, D., 1997. A paradox: more measurement error can lead to more efficient estimates. *Commun. Statist.—Theory Methods* 26, 1649–1675.
- Donev, A.N., 2000. Dealing with errors in variables in regression surface exploration. *Commun. Statist.—Theory Methods* 29, 2065–2077.
- Draper, N., Beggs, W., 1970. Errors in the factor levels and experimental design. *Ann. Math. Statist.* 42, 46–58.
- Fedorov, V.V., 1974. Regression problems with controllable variables subject to error. *Biometrika* 61, 49–56.
- Fedorov, V.V., Hackl, P., 1997. Model-Oriented Design of Experiments. Springer, New York.
- Fuller, W.A., 1987. Measurement Error Models. Wiley, New York.
- Pronzato, L., 1998. On a property of the expected value of a determinant. *Statist. Probab. Lett.* 39, 1961–1965.
- Pronzato, L., 2002. Average D-optimum designs for randomly varying experimental conditions. *J. Statist. Plann. Inference* 108, 189–200.
- Vuchkov, I.N., Boyadjieva, L.N., 1983. The robustness of experimental designs against errors in the factor levels. *J. Statist. Comput. Simulation* 17, 31–41.