KATHOLIEKE
UNIVERSITEIT
LEUVEN

DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

KULdtew-or 9918

RESEARCH REPORT 9918

OPTIMAL DESIGNS FOR VARIANCE
FUNCTION ESTIMATION USING SAMPLE
VARIANCES

by
P. GOOS
L. TACK
M. VANDEBROEK

Optimal designs for variance function estimation using sample variances

P. Goos
L. Tack
M. Vandebroek

Katholieke Universiteit Leuven, Belgium

Abstract

Using sample variances for estimating a variance function is intuitively more appealing than using residuals. Main advantage of sample variances over residuals is that they are robust to misspecification of the mean function. However, due to the replication requirement neither standard response surface designs nor small designs generated by design construction algorithms can be used to estimate the variance by means of sample variances. Based on maximum likelihood and weighted least squares estimation, two alternative approaches for the construction of optimal designs for variance function estimation with sample variances are proposed. A generic exchange algorithm and computational results are presented. Irrespective of the link function between the variance and the linear predictor, the algorithm serves as a useful tool to construct tailor-made designs for variance function estimation by means of sample variances.

Keywords: design construction, $\mathcal{D}$ -optimality, multiplicative heteroscedasticity, replicated observations, sample variance, variance function estimation

1 Introduction

As demonstrated by Taguchi (1989), reducing variability is an integral part of quality improvement. This led to a considerable interest in the modeling and analysis of variances in the field of engineering. Among others, Box and Meyer (1986) and Bergman and Hynén (1997) propose methods for identifying dispersion effects from unreplicated experiments, whereas Nair and Pregibon (1988) describe several methods to analyze dispersion effects from replicated factorial experiments. The use of variance functions has also gained popularity after Vining and Myers (1990) adapted the dual response approach to achieve the goals of Taguchi's philosophy. Other fields where heteroscedasticity is modeled include economics and the medical,

biological and physical sciences. Raab (1981) and Davidian (1990) propose methods for variance function estimation with assay data. Davidian and Carroll (1987) and Carroll and Ruppert (1988) give an excellent overview of estimation methods using sample variances and residuals. Main advantages of sample variances over residuals are that they are intuitively appealing and robust to misspecification of the mean function. The large number of observations needed due to the replication requirement is disadvantageous to their use.

The large number of observations needed stems from the fact that experimenters who prefer sample variances over residuals usually replicate standard response surface designs. For instance, using a three variable rotatable central composite design (CCD) with 6 center runs for estimating the variance function with sample variances would require replicating the 8 factorial and the 6 axial points. This would result in 34 observations. Similarly, replicating a 3^3 factorial would imply 54 observations. Another drawback of replicating standard response surface designs is the lack of flexibility in cases where the design region is restricted, in cases where the number of observations available is not a multiple of the number of observations needed to perform a standard response surface design, in cases where the model is of first order in some variables and of second order in other variables or in cases where the effect of qualitative variables is investigated. The solution to these problems lies in design construction algorithms.

The purpose of this paper is twofold. Firstly, we provide the experimenter with two alternative design criteria that are appropriate for the construction of experimental designs for variance function estimation using sample variances. Secondly, we present an exchange algorithm for constructing tailor-made optimal designs satisfying the replication requirement. The design of tailor-made experiments for variance function estimation has received attention only by Atkinson and Cook (1995) and by Vining and Schaub (1996), who developed designs for estimating mean and variance functions simultaneously. The design criterion they propose is not suited for our purpose because it takes into account the mean function, it allows unreplicated observations and assumes the use of residuals for estimating the variance function. Similarly, using the exchange algorithm of Fedorov (1972), the DETMAX algorithm of Mitchell (1974) and the BLKL algorithm of Atkinson and Donev (1989, 1992) does not guarantee replicated observations.

Throughout this paper, we will use the assumption that sample variances follow a χ^2 -distribution. This is a direct consequence of the common assumption of normally distributed observations. Two alternative design criteria are derived based on this assumption. The first criterion is derived from maximum likelihood estimation and the second is derived from weighted least squares regression. We assume

multiplicative heteroscedasticity, i.e. the variance function is taken as

$$\sigma^2 = \exp\{\mathbf{g}'(\mathbf{z})\boldsymbol{\gamma}\} \quad (1)$$

where $\boldsymbol{\gamma}$ is the vector of p unknown variance function parameters or dispersion parameters and $\mathbf{g}(\mathbf{z})$ is the polynomial expansion of the covariates. The first element of $\mathbf{g}(\mathbf{z})$ is equal to one indicating the presence of an intercept. Assumption (1) guarantees positive predictions $\hat{\sigma}^2$ and leads to simple expressions. It is also used by Aitkin (1987), Atkinson and Cook (1995), Box and Meyer (1986), Huele (1998), Myers and Montgomery (1995), Nair and Pregibon (1988), Sohn and Park (1998), Vining and Schaub (1996) and Wolfinger and Tobias (1998). Other functions than the exponential are sometimes used as a link between the variance σ^2 and the linear predictor $\mathbf{g}'(\mathbf{z})\boldsymbol{\gamma}$. For instance, Vining and Myers (1990), Del Castillo and Montgomery (1993), Lin and Tu (1995), and Vining and Schaub (1996) model the standard deviation. The identity and reciprocal link function also arise in practice. The design construction algorithm remains a useful tool to construct designs for estimating the variance function parameters in these cases.

In the next section, we will give an overview of maximum likelihood and weighted least squares estimation and we will derive the variance-covariance matrix of the parameter estimators for both methods. These matrices will be used to derive appropriate design criteria in Section 3. We present a design construction algorithm in Section 4 and discuss our computational results in Section 5.

2 Estimation methods

Our alternative design construction algorithms are based on two different methods for variance function estimation with residuals. Both methods use the following distributional assumption:

$$\frac{(n_i - 1)s_i^2}{\sigma_i^2} \sim \chi_{n_i - 1}^2, \quad (2)$$

with n_i the number of observations on which the sample variance s_i^2 is based. In the sequel of this section, we present in succession maximum likelihood estimation and weighted least squares regression. The number of observations, the number of different design points and the number of replications of each design point i will be denoted by n , m and n_i respectively. Note that $\sum_{i=1}^m n_i = n$ and assume that $n_i \geq 2$ for all design points. The variance and the sample variance in the i th design point are given by σ_i^2 and s_i^2 respectively.

2.1 Maximum likelihood estimation

Given the distributional assumption (2), the loglikelihood function, up to constant terms, of one observation s_i^2 can be written as

$$l_i = -\frac{n_i - 1}{2} \mathbf{g}'(\mathbf{z}_i) \boldsymbol{\gamma} - \frac{(n_i - 1) s_i^2}{2 \exp\{\mathbf{g}'(\mathbf{z}_i) \boldsymbol{\gamma}\}}, \quad (3)$$

such that the ML estimator $\hat{\boldsymbol{\gamma}}_{\text{ML}}$ for $\boldsymbol{\gamma}$ for an experiment with n observations in m different design points is obtained by solving the score equations

$$\sum_{i=1}^m (n_i - 1) \mathbf{g}(\mathbf{z}_i) \left[\frac{s_i^2}{\exp\{\mathbf{g}'(\mathbf{z}_i) \boldsymbol{\gamma}\}} - 1 \right] = \mathbf{0}. \quad (4)$$

This set of equations is nonlinear in the unknown parameter $\boldsymbol{\gamma}$ and can be obtained by using $T(s_i) = s_i^2$ in the general set of equations defined by Davidian and Carroll (1987). The likelihood equations in Nair and Pregibon (1988) can be obtained as a special case of (4). The information matrix is given by

$$\begin{aligned} \mathbf{M}_{\text{ML}} &= \sum_{i=1}^m \frac{n_i - 1}{2} \mathbf{g}(\mathbf{z}_i) \mathbf{g}'(\mathbf{z}_i), \\ &= \mathbf{Z}' \mathbf{W}_{\text{ML}} \mathbf{Z}, \end{aligned} \quad (5)$$

where $\mathbf{Z}$ is the $m \times p$ design matrix and $\mathbf{W}_{\text{ML}}$ is the $m \times m$ diagonal matrix with diagonal elements $w_i^{\text{ML}} = (n_i - 1)/2$. The asymptotic variance-covariance matrix of $\hat{\boldsymbol{\gamma}}_{\text{ML}}$ is the inverse of the information matrix.

2.2 Weighted least squares

From model (1), it is just a small step to regression of $\ln s_i^2$ on $\mathbf{g}'(\mathbf{z}_i) \boldsymbol{\gamma}$. The resulting model is given by

$$\ln s_i^2 = \mathbf{g}'(\mathbf{z}_i) \boldsymbol{\gamma} + \nu_i. \quad (6)$$

Ordinary least squares (OLS) regression has a number of deficiencies due to the facts that $\ln s_i^2$ is a biased estimate of $\ln \sigma_i^2$ and that its variance depends on the number of replicates n_i used. In Appendix A is shown that

$$E(\ln s_i^2) = \ln \sigma_i^2 + \Psi\left(\frac{n_i - 1}{2}\right) - \ln\left(\frac{n_i - 1}{2}\right), \quad (7)$$

$$\text{var}(\ln s_i^2) = \Psi'\left(\frac{n_i - 1}{2}\right). \quad (8)$$

The functions Ψ and Ψ' are known in the literature as the digamma and trigamma function respectively. Their values can be computed by any mathematical package

and are tabulated in Abramowitz and Stegun (1970). Correcting the $\ln s_i^2$ for the bias and using weighted least squares, the parameter estimate can be written as

$$\hat{\gamma}_{\text{WLS}} = (\mathbf{Z}'\mathbf{W}_{\text{WLS}}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{W}_{\text{WLS}}\mathbf{c}, \quad (9)$$

where $\mathbf{c}$ contains the corrected logarithms of the m sample variances

$$c_i = \ln s_i^2 - \Psi\left(\frac{n_i - 1}{2}\right) + \ln\left(\frac{n_i - 1}{2}\right), \quad (10)$$

$\mathbf{Z}$ is the $m \times p$ design matrix and $\mathbf{W}_{\text{WLS}}$ is the $m \times m$ diagonal matrix with diagonal elements

$$w_i^{\text{WLS}} = \left\{ \Psi'\left(\frac{n_i - 1}{2}\right) \right\}^{-1}.$$

This approach was suggested by Raab (1981). The information matrix, which is the inverse variance-covariance matrix of $\hat{\gamma}_{\text{WLS}}$, can be written as

$$\begin{aligned} \mathbf{M}_{\text{WLS}} &= \sum_{i=1}^m w_i^{\text{WLS}} \mathbf{g}(\mathbf{z}_i) \mathbf{g}'(\mathbf{z}_i), \\ &= \mathbf{Z}'\mathbf{W}_{\text{WLS}}\mathbf{Z}. \end{aligned} \quad (11)$$

$\mathbf{M}_{\text{ML}}$ as well as $\mathbf{M}_{\text{WLS}}$ will be used to derive a suitable design criterion for the problem under consideration. Both possess the attractive property that they can be obtained as the weighted sum of the outer products of the design points. The weights w_i^{ML} as well as w_i^{WLS} are increasing functions of the number of replicates. As both matrices have the same structure, we will use the shorthand notation w_i in the sequel of this paper for statements which are valid for both w_i^{ML} and w_i^{WLS} . For this reason, the design criterion derived in the next section as well as the design construction algorithm outlined in Section 4 represent two alternative approaches. In Section 5, we will investigate the differences between both approaches.

3 Design criterion

In this paper, we will use the $\mathcal{D}$ -optimality criterion, which minimizes the determinant of the variance-covariance matrix of the estimators (or maximizes the determinant of its inverse) thereby minimizing the volume of the confidence ellipsoid on the unknown parameters. Unlike other criteria, the $\mathcal{D}$ -optimality criterion has the advantage of computational simplicity thanks to the existence of powerful update formulae.

The information matrices obtained by maximum likelihood estimation and weighted least squares are given by (5) and (11). Both can be written as

$$\begin{aligned} \mathbf{M} &= \sum_{i=1}^m w_i \mathbf{g}(\mathbf{z}_i) \mathbf{g}'(\mathbf{z}_i), \\ &= \mathbf{Z}' \mathbf{W} \mathbf{Z}. \end{aligned} \quad (12)$$

However, only part of this information matrix is of interest to the experimenter. As a matter of fact, in many cases the aim of estimating the variance function is to minimize the variance. This implies that the intercept is unimportant for the purpose of the experiment and should not be incorporated in the design criterion. For notational purposes, assume that the first column of $\mathbf{Z}$ corresponds to the intercept and let $\mathbf{Z} = (\mathbf{1} \ \mathbf{X})$. The variance-covariance matrix of all parameters except the intercept is then given by $(\mathbf{X}' \mathbf{W}^* \mathbf{X})^{-1}$ with $\mathbf{W}^* = \mathbf{W} - \mathbf{W} \mathbf{1} (\mathbf{1}' \mathbf{W} \mathbf{1})^{-1} \mathbf{1}' \mathbf{W}$. The appropriate $\mathcal{D}$ -criterion value is given by

$$\begin{aligned} \mathcal{D} &= |\mathbf{X}' \mathbf{W}^* \mathbf{X}|, \\ &= (\mathbf{1}' \mathbf{W} \mathbf{1})^{-1} |\mathbf{Z}' \mathbf{W} \mathbf{Z}|, \\ &= \frac{|\mathbf{Z}' \mathbf{W} \mathbf{Z}|}{\sum_{i=1}^m w_i}. \end{aligned} \quad (13)$$

The derivation of these results is given in Appendix B. In the next section, we will present a design construction algorithm to compute the design matrix $\mathbf{Z}$ that maximizes (13). If, for some reason, estimation of the intercept matters, the appropriate design criterion value would be the determinant of $\mathbf{M}$. In that case, the update formulae in the algorithm have to be slightly modified.

4 Design construction algorithm

The purpose of the algorithm is to generate n -point designs which maximize the $\mathcal{D}$ -criterion value given in (13). The fact that replication is required when using sample variances considerably complicates the design construction algorithm. The effect of adding and deleting observations differs in a fundamental way from the construction of $\mathcal{D}$ -optimal designs intended for mean function estimation. Moreover the exchange policy has to be adapted in order to avoid ending up with suboptimal designs. Despite these complications, the computational effort remains limited thanks to the use of powerful update formulae for the design criterion and the information matrix $\mathbf{Z}' \mathbf{W} \mathbf{Z}$. In this section, we will further explain these considerations and their consequences. We will also give a brief overview of the design construction algorithm we have implemented in Fortran 77.

4.1 Description

In contrast with experimental designs for mean function estimation (e.g. Fedorov 1972, Mitchell 1974, Atkinson and Donev 1992) and with the design criterion de-

veloped in Vining and Schaub (1996), there is no one to one relation between the set of observations and the rows of the design matrix. This is because sample variances are used instead of residuals. Here, all n_i replicates of a given design point i correspond to only one row of the design matrix $\mathbf{Z}$. Hence, addition and/or deletion of one or more observations to the experiment does not necessarily change $\mathbf{Z}$. This will only be the case when observations in a new design point are introduced or when a design point is removed from the experiment. On the contrary, the diagonal matrix $\mathbf{W}$ will be affected by any change. Addition or deletion of design points will change the dimension of $\mathbf{W}$, whereas addition or deletion of one or more replicates of a design point i only alters the corresponding diagonal element w_i of $\mathbf{W}$.

Due to the replication requirement, exchanging observations one by one is not effective and typically results in a local optimum in the neighbourhood of the starting design. However, it is fit for fine tuning once the algorithm approaches the overall optimum. In order to evaluate more substantial changes, two other exchange strategies were considered. Firstly, the effect of deleting two observations in a design point and adding one observation in two other not necessarily distinct design points will be evaluated. Secondly, we also compute the effect of adding two observations in a new design point and deleting one in two other design points. For each preliminary solution, these three strategies are evaluated and the best exchange found is executed. This step is repeated until no further improvement of the design can be realized. Like most exchange algorithms, a number of starting designs should be generated in order to avoid being stuck in a local optimum. The design points are chosen from a predefined grid of candidate points. Main advantage thereof is that disparate design problems are reduced to a common optimization structure. Design problems with restricted or non-regular design regions, e.g. mixture experiments, or design problems with both quantitative and qualitative variables can therefore be solved by the algorithm. Of course, the choice of the candidate points must be done with care in order to obtain a good design. Algorithms with a similar structure can be found in Atkinson and Donev (1989), Donev (1997, 1998) and Mitchell (1974).

4.1.1 The starting procedure

The starting design is partially composed in a random fashion, after which missing observations are added by a greedy heuristic. The random start to the search comes from choosing j additions to a d -point starting design, where j and d are themselves randomly chosen integers. Each addition consists of randomly determining a design point, after which two observations in that point are added to the design if it is not yet in the design. Else, only one observation in the point is added. At this stage, the starting design contains $m \leq d$ different design points. It is augmented by adding sequentially two observations in the $d - m$ points which are not yet in the design and give maximum increase to the $\mathcal{D}$ -criterion value. The number of distinct design points then equals d . Finally, the starting design is completed by

adding one observation in the design points which yields a maximum improvement of the objective value until the number of observations n is attained. The singularity when the number of design points m is less than the number of parameters p is overcome by using $\mathbf{M} + r\mathbf{I}$ instead of $\mathbf{M}$ with r a small positive number. Before trying to improve the design, we make sure the corresponding information matrix is not singular.

4.1.2 Improving the design

In order to improve the starting design, three exchange strategies are examined. The first strategy evaluates exchanges of one observation in a design point by one observation in another design point. Only design points with at least two observations are considered for addition of one. This is because increasing the number of observations in a point from zero to one does not provide any information. Similarly, we do not consider duplicated design points for deletion of one observation because deleting one of both observations makes the remaining one worthless. In that case, one may as well remove both observations from the design. This situation is covered by the second exchange strategy, which investigates the effect of deleting two observations in a design point and adding one in two other not necessarily distinct design points. As in the first exchange strategy, only design points with at least two observations are considered for addition of one. Only design points with at least four observations are considered for deletion of two. The third strategy considers the exchange of one observation in two not necessarily distinct design points by two observations in a new design point. The best exchange is carried out if it improves the design. This exchange procedure is repeated until no further improvement can be realized.

4.1.3 Update formulae

In order to save computing time, we used several powerful update formulae for the generation of the starting designs and for the evaluation of the exchange strategies of the algorithm. Suppose a new design point a with weight w_a is added to the design $\mathbf{Z}$. Denoting $\mathbf{Z}'\mathbf{W}\mathbf{Z}$ by $\mathbf{M}$ as previously, the $\mathcal{D}$ -criterion value becomes

$$\mathcal{D}' = \frac{|\mathbf{M} + w_a \mathbf{g}(\mathbf{z}_a) \mathbf{g}'(\mathbf{z}_a)|}{\sum_{i=1}^m w_i + w_a}. \quad (14)$$

From matrix algebra, we have that

$$|\mathbf{M} + w_a \mathbf{g}(\mathbf{z}_a) \mathbf{g}'(\mathbf{z}_a)| = \{1 + w_a \mathbf{g}'(\mathbf{z}_a) \mathbf{M}^{-1} \mathbf{g}(\mathbf{z}_a)\} |\mathbf{M}|,$$

which leads to the following update formula for the design criterion value:

$$\mathcal{D}' = \frac{1 + w_a \mathbf{g}'(\mathbf{z}_a) \mathbf{M}^{-1} \mathbf{g}(\mathbf{z}_a)}{1 + w_a / (\sum_{i=1}^m w_i)} \mathcal{D}. \quad (15)$$

In order to update $\mathbf{M}^{-1}$, we used

$$\{\mathbf{M} + w_a \mathbf{g}(\mathbf{z}_a) \mathbf{g}'(\mathbf{z}_a)\}^{-1} = \mathbf{M}^{-1} - w_a \frac{\mathbf{M}^{-1} \mathbf{g}(\mathbf{z}_a) \mathbf{g}'(\mathbf{z}_a) \mathbf{M}^{-1}}{1 + w_a \mathbf{g}'(\mathbf{z}_a) \mathbf{M}^{-1} \mathbf{g}(\mathbf{z}_a)}. \quad (16)$$

Since all components in the right hand sides of (15) and (16) are known, the effect of the design change is readily calculated without computationally intensive matrix inversions and determinant operations. By similar reasoning update formulae can be derived for other design changes involving one particular design point considered in our algorithm. For illustrative purposes, consider addition of one observation in a design point a . In this case, the weight of the point will increase by $\Delta = w'_a - w_a$. The new $\mathcal{D}$ -criterion value can then be written as

$$\mathcal{D}' = \frac{|\mathbf{M} + \Delta \mathbf{g}(\mathbf{z}_a) \mathbf{g}'(\mathbf{z}_a)|}{\sum_{i=1}^m w_i + \Delta}. \quad (17)$$

From here, it is easy to derive the update formulae for $\mathcal{D}$ and $\mathbf{M}^{-1}$. Likewise, expressions for updating $\mathcal{D}$ for changes involving two distinct design points can be derived. Thanks to these update formulae, the effects of design changes can be evaluated at a low computational cost.

4.2 The exchange algorithm

In the outline of the algorithm, we denote the set of g candidate points and the set of m design points by G and D respectively. As previously, we denote the total number of observations by n and the number of observations in each grid point i by n_i . Note that $i \in D$ if and only if $n_i > 0$. We refer to the $\mathcal{D}$ -criterion value of a design D as $\mathcal{D}$. The number of observations in point i in the current optimal design D_o is denoted by n_{oi} and the corresponding criterion value is denoted by $\mathcal{D}_o$. Finally, we denote the number of tries by t and the number of the current try by t_c . The input to the algorithm consists of the parameters n , t , the order of the model, the number of model parameters p , the number of explanatory variables k and the structure of their polynomial expansion. By default, the algorithm computes $G = \{1, 2, \dots, g\}$ as in Atkinson and Donev (1992). The design region is taken as $[-1, +1]^k$ and the grid points are chosen from the $2^k, 3^k, 4^k, \dots$ factorial design depending on whether the model contains linear, quadratic, cubic or higher order terms. Alternatively, the user can specify G if another set of candidate points is desired. Note that construction of a non-singular design requires $n \geq 2p$ and $g \geq p$. At that point, the algorithm proceeds as follows:

1. Set $\mathcal{D}_o = 0$ and $t_c = 1$.
2. Set $m = 0$, $D = \emptyset$, $\mathbf{M} = r\mathbf{I}$ and $\forall i \in G: n_i = 0$.

3. Generate starting design:

- (i) Randomly choose d subject to $p \leq d \leq \min(n/2, g)$.
- (ii) Randomly choose j subject to $1 \leq j \leq d$.
- (iii) Do j times:
 - a) Randomly choose $i \in G$.
 - b) If $n_i = 0$, then $n_i = n_i + 2$, $m = m + 1$ and $D = D \cup \{i\}$, else $n_i = n_i + 1$.
 - c) Update M .
- (iv) If $n - \sum_{i=1}^m n_i \leq 2(d - m)$ go back to step 2.
- (v) Repeat until $d = m$:
 - a) Add two observations in $i \in G \setminus D$ with largest effect on D .
 - b) Set $n_i = n_i + 2$ and $m = m + 1$. $D = D \cup \{i\}$. Update M .
- (vi) Repeat until $n = \sum_{i=1}^m n_i$:
 - a) Add one observation in $i \in D$ with largest effect on D .
 - b) Set $n_i = n_i + 1$. Update M .

4. Compute $\mathcal{D}$. If $\mathcal{D} = 0$, then go to step 2.

5. Consider exchange strategies.

- (i) Set $\delta = 1$.
- (ii) Find best exchange of Strategy 1: delete one observation in a design point and add one to another.
 - a) $S_1 = \{i : i \in D \wedge n_i \geq 2\}$.
 - b) $S_2 = \{i : i \in D \wedge n_i \geq 3\}$.
 - c) $\forall i \in S_1, j \in S_2, i \neq j$, compute the effect $\delta_j^i = \mathcal{D}'/\mathcal{D}$ of deleting one observation in j and adding one to i . If $\delta_j^i > \delta$, then $\delta = \delta_j^i$ and store i and j .
- (iii) Find best exchange of Strategy 2: delete two observations in a design point and add one to two other not necessarily distinct design points.
 - a) $S_3 = \{i : i \in D \wedge (n_i = 2 \vee n_i \geq 4)\}$.
 - b) $\forall i, j \in S_1, k \in S_3, i \neq k, j \neq k, i \leq j$, compute the effect $\delta_k^{ij} = \mathcal{D}'/\mathcal{D}$ of deleting two observations in k and adding one to i and j each. If $\delta_k^{ij} > \delta$, then $\delta = \delta_k^{ij}$ and store i, j and k .
- (iv) Find best exchange of Strategy 3: delete one observation in two not necessarily distinct design points and add two observations in a new design point.
 - a) $S_4 = \{i : i \in G \setminus D\}$.
 - b) $\forall i, j \in S_2, k \in S_4, i < j$, compute the effect $\delta_{ij}^k = \mathcal{D}'/\mathcal{D}$ of adding two observations in k and deleting one in i and j each. If $\delta_{ij}^k > \delta$, then $\delta = \delta_{ij}^k$ and store i, j and k .

c) $\forall i \in S_3, j \in S_4$, compute the effect $\delta_i^j = \mathcal{D}'/\mathcal{D}$ of adding two observations in j and deleting two in i . If $\delta_i^j > \delta$, then $\delta = \delta_i^j$ and store i and j .

6. If $\delta > 1$, go to step 7, else go to step 8.

7. Carry out the best exchange found. Update $\mathbf{M}$, $\mathcal{D}$ and $n_i, \forall i \in G$. Go to step 5.

8. If $\mathcal{D} \geq \mathcal{D}_o$, then $\mathcal{D}_o = \mathcal{D}$, $D_o = D$, $\forall i \in G$: $n_{oi} = n_i$.

9. If $t_c < t$, then $t_c = t_c + 1$ and go to step 2, else stop.

Sometimes there may be observations which the experimenter wishes to include in the design. This situation is usually referred to as a problem of design augmentation and occurs when further experimentation is required. Design augmentation is therefore used after a screening experiment or when an initial experiment has failed. In this case, the starting and exchange procedures slightly differ. Each starting design contains the observations specified by the experimenter and none of these observations can be removed from the design during the exchange procedure.

Although the algorithm is based on the assumption of multiplicative heteroscedasticity, the experimental designs it produces can be used for the estimation of any other variance model by means of sample variances. However, using another link function than (1) leads to a design criterion which depends on the unknown parameters. The optimal designs for other link functions thus not necessarily coincide with the output of our algorithm. Their construction would require prior information about the unknown parameters in order to compute the weights of the design points.

5 Computational results

In this section, we will illustrate the properties of the optimal designs generated by the algorithm from the previous section. We have computed optimal designs by using both w_i^{ML} and w_i^{WLS} and will denote the corresponding $\mathcal{D}$ -criterion value as $\mathcal{D}_{ML}$ and $\mathcal{D}_{WLS}$ respectively. The benefits of using our design construction algorithm are threefold.

Firstly, our algorithm produces tailor-made designs in cases where no alternatives exist, i.e. for small n . As a matter of fact, when the number of observations available is small, replicating standard response surface designs is impossible and Fedorov's (1972) algorithm, DETMAX (Mitchell 1974) or BLKL (Atkinson and Donev 1989) produce designs which do not satisfy the replication requirement. Secondly, even if alternative design options are available, the algorithm generates superior designs. Thirdly, the design criteria we have developed allow the experimenter to distinguish between several alternative designs he may have. We will demonstrate these benefits in the sequel of this section. We will denote designs

generated by existing algorithms (EA) as $\mathcal{D}_{\text{EA}}$ -optimal. We have chosen the BLKL algorithm to compute them, but Fedorov's algorithm or DETMAX or any other algorithm that maximizes $|\mathbf{Z}'\mathbf{Z}|$ could have been used instead. As was explained in Section 4.1, the design matrix $\mathbf{Z}$ used in these algorithms differs in a fundamental way from the design matrix used in this paper.

For small sample sizes, replicating standard response surface designs is clearly not an option. However, well-known design construction algorithms do not produce satisfactory designs either. This is due to the fact that they generate designs in which not all design points are replicated and thus contain observations which do not contribute information to the experiment. Consider the hierarchical variance models in Table 1. For each of them, we have computed $\mathcal{D}_{\text{EA}}$ -optimal designs for $n \geq 2p$ and we have investigated which of these designs satisfy the replication requirement and which do not. The first column of Table 1 shows the explanatory variables included in the model. The intercept is denoted by a "1". The second column gives the minimum number of observations needed to estimate all parameters. Because of the replication requirement, this number is twice the number of parameters p . The last column gives the minimum number of observations for which the $\mathcal{D}_{\text{EA}}$ -optimal designs satisfy the replication requirement. From the discrepancy between the last two columns can be seen that existing algorithms often fail to produce an answer to the design problem at hand. The same conclusions can be drawn for models with more than two explanatory variables.

Explanatory variables	$2p$	n
1, z_1, z_2	6	8
1, $z_1, z_2, z_1 z_2$	8	8
1, z_1, z_2, z_1^2	8	12
1, $z_1, z_2, z_1 z_2, z_1^2$	10	12
1, z_1, z_2, z_1^2, z_2^2	10	18
1, $z_1, z_2, z_1 z_2, z_1^2, z_2^2$	12	22

Table 1: Shortcomings of $\mathcal{D}_{\text{EA}}$ -optimal designs.

Suppose an experiment with 16 observations is to be designed for estimating the linear effects in three variables, i.e. the terms included in the model are 1, z_1, z_2 and z_3 . Three alternative $\mathcal{D}_{\text{EA}}$ -optimal designs were computed and are shown in Table 2. The reader will notice that the second alternative is a full replicate of the 2^3 factorial design and that the third alternative consists of four replicates of the 2^{3-1} fractional factorial design. Unlike alternatives II and III, design I is not an option when sample variances are used because the replication requirement is not satisfied. It turns out that design III is both $\mathcal{D}_{\text{ML}}$ - and $\mathcal{D}_{\text{WLS}}$ -optimal and therefore should be preferred. The $\mathcal{D}_{\text{ML}}$ - and $\mathcal{D}_{\text{WLS}}$ -criterion value of design II amount to 512

Candidate points			Alternative designs		
z_1	z_2	z_3	I	II	III
-	-	-	3	2	4
+	-	-	1	2	0
-	+	-	1	2	0
+	+	-	3	2	4
-	-	+	1	2	0
+	-	+	3	2	4
-	+	+	3	2	4
+	+	+	1	2	0

Table 2: Alternative $\mathcal{D}_{\text{EA}}$ -optimal 16-point designs for a model in 1, z_1 , z_2 and z_3 .

and 4.26 respectively. For design III, we obtain $\mathcal{D}_{\text{ML}}$ - and $\mathcal{D}_{\text{WLS}}$ -criterion values of 1728 and 78.35 respectively. This simple example demonstrates the substantial loss of efficiency that experiments can suffer from an improper design.

In order to illustrate the improvement over $\mathcal{D}_{\text{EA}}$ -optimal designs, consider a full quadratic model in two variables. As can be seen from Table 1, this model has six parameters. The $\mathcal{D}_{\text{EA}}$ -optimal designs for this model do not satisfy the replication requirement unless $n \geq 22$. For $n < 22$, the experimenter could re-allocate the observations by trial and error until he or she finds a non-singular design satisfying the replication requirement. However, it is improbable that this approach results in an optimal solution. For $n \geq 22$, the $\mathcal{D}_{\text{EA}}$ -optimal designs do no longer contain uninformative observations but still are inefficient for practical sample sizes. This is illustrated by Table 3, in which they are compared to $\mathcal{D}_{\text{ML}}$ - and $\mathcal{D}_{\text{WLS}}$ -optimal designs generated by the algorithm in Section 4. The first column of the table displays the number of observations n for which the designs are compared. The second, third and sixth column give the number of different design points m of the $\mathcal{D}_{\text{EA}}$ -, $\mathcal{D}_{\text{ML}}$ - and $\mathcal{D}_{\text{WLS}}$ -optimal designs respectively. When we denote the corresponding optimal designs by $\mathbf{Z}_{\text{EA}}$, $\mathbf{Z}_{\text{ML}}$ and $\mathbf{Z}_{\text{WLS}}$ respectively and by $\mathcal{D}_{\text{ML}}(\mathbf{Z})$ and $\mathcal{D}_{\text{WLS}}(\mathbf{Z})$ the $\mathcal{D}_{\text{ML}}$ - and $\mathcal{D}_{\text{WLS}}$ -criterion value of a given design $\mathbf{Z}$, the relative efficiencies in the columns four, five, seven and eight are computed as

$$\begin{aligned} \Delta_{\text{ML}}^{\text{ML}} &= \mathcal{D}_{\text{ML}}(\mathbf{Z}_{\text{ML}})/\mathcal{D}_{\text{ML}}(\mathbf{Z}_{\text{EA}}), & \Delta_{\text{WLS}}^{\text{ML}} &= \mathcal{D}_{\text{WLS}}(\mathbf{Z}_{\text{ML}})/\mathcal{D}_{\text{WLS}}(\mathbf{Z}_{\text{EA}}), \\ \Delta_{\text{ML}}^{\text{WLS}} &= \mathcal{D}_{\text{ML}}(\mathbf{Z}_{\text{WLS}})/\mathcal{D}_{\text{ML}}(\mathbf{Z}_{\text{EA}}), & \Delta_{\text{WLS}}^{\text{WLS}} &= \mathcal{D}_{\text{WLS}}(\mathbf{Z}_{\text{WLS}})/\mathcal{D}_{\text{WLS}}(\mathbf{Z}_{\text{EA}}). \end{aligned}$$

These efficiencies show how the $\mathcal{D}_{\text{ML}}$ - and $\mathcal{D}_{\text{WLS}}$ -criterion value of $\mathbf{Z}_{\text{EA}}$ relates to the $\mathcal{D}_{\text{ML}}$ - and $\mathcal{D}_{\text{WLS}}$ -criterion value of $\mathbf{Z}_{\text{ML}}$ and $\mathbf{Z}_{\text{WLS}}$. In other words, they show the improvement of our approach over the existing design construction algorithms. For $n \leq 22$, no efficiencies are displayed because the $\mathcal{D}_{\text{EA}}$ -optimal designs do not satisfy the replication requirement.

n	EA	ML			WLS		
	m	m	Δ_{ML}^{ML}	Δ_{WLS}^{ML}	m	Δ_{ML}^{WLS}	Δ_{WLS}^{WLS}
12		6			6		
15		7			6		
18		6			6		
21		7			6		
22	9	7	146%	379%	6	124%	477%
23	9	7	152%	403%	6	131%	517%
24	9	7	159%	426%	6	137%	556%
30	9	7	123%	221%	6	100%	242%
40	9	7	106%	165%	7	106%	165%
45	9	8	101%	119%	7	99%	139%
50	9	9	101%	102%	7	97%	131%
85	9	9	100%	100%	8	93%	100%
100	9	9	100%	100%	9	100%	100%

Table 3: Relative efficiencies of $\mathcal{D}_{ML}$ - and $\mathcal{D}_{WLS}$ -optimal designs to $\mathcal{D}_{EA}$ -optimal designs for a full quadratic model in two variables. Balanced designs are printed in bold.

From Table 3, we see that for n not too large both the $\mathcal{D}_{ML}$ - and $\mathcal{D}_{WLS}$ -optimal designs have a smaller number of design points than the $\mathcal{D}$ -optimal designs. In addition, the $\mathcal{D}_{WLS}$ -optimal designs have a smaller number of design points than the $\mathcal{D}_{ML}$ -optimal designs. It turns out that improvements of more than 20% for the $\mathcal{D}_{ML}$ -criterion value and of more than 100% for the $\mathcal{D}_{WLS}$ -criterion value are no exception for practical sample sizes of $n \leq 40$. As n grows large, the designs all have nine different design points and equal numbers of replications. The gain in efficiency then falls to zero. Similar results can be found for other models in two variables. For simpler models, the convergence of the three designs occurs for smaller n .

For n not too large, the $\mathcal{D}_{ML}$ - and $\mathcal{D}_{WLS}$ -optimal designs are nearly balanced, i.e. all design points nearly have the same number of replicates. When n is a multiple of m , the designs are completely balanced. The balanced designs in Table 3 are printed in bold. This observation is particularly important because balanced designs turn model (6) into a homoscedastic model. The model parameters of interest, i.e. all model parameters except the intercept, can then be estimated by simple ordinary least squares. For small n , the designs are saturated, i.e. $m = p$, in which case ordinary and weighted least squares estimation of model (6) are also equivalent. These conclusions remain valid for models in more than two variables. The gains in efficiency are of the same order of magnitude. In addition, the designs do not converge as fast as in the two variable case even for simple models without interactions and pure quadratic terms. In other words, when more than two explanatory variables are included in the variance model, our algorithm produces designs which are

different from and more efficient than $\mathcal{D}_{\text{EA}}$ -optimal designs for nearly any practical sample size.

In this section, we have outlined the merits of the algorithm in Section 4 by comparing the efficiency of the $\mathcal{D}_{\text{ML}}$ - and $\mathcal{D}_{\text{WLS}}$ -optimal designs with the efficiency of approaches found in practice. We have pointed out that for small n , no alternatives exist. For large n , existing algorithms produce useful but inefficient designs. It also turns out that the more explanatory variables considered, the more useful our algorithm. When n is a multiple of the number of observations required for using a standard response surface design, replicating the design is an option, but it may be a poor one.

6 Conclusion

Although variance function estimation has received considerable attention in the literature, not much work has been done on the design of experiments for that purpose. The few exceptions assume the use of residuals and are intended for estimating mean and variance function simultaneously. Hence, they are not fit for designing experiments for using sample variances to estimate the variance function only. Since replication of standard response surface designs and existing design construction algorithms do not yield satisfactory designs either, experimenters preferring sample variances over residuals will benefit from the algorithm presented in this paper.

We propose two alternative $\mathcal{D}$ -optimality criteria for the evaluation of experimental designs for variance function estimation and minimization by using sample variances. An exchange algorithm is developed to construct the optimal designs according to both criteria. The algorithm takes into account the specific features of the design problem at hand, the most important being the replication requirement. A selection of the computational results shows that substantial benefits can be obtained by using the algorithm. Firstly, there is no alternative to generate efficient tailor-made designs for small sample sizes. Secondly, if alternatives exist, they are outperformed by our algorithms. The algorithm is suited for designing experiments in industrial environments where observations are expensive or screening experiments are conducted, but also for the construction of second order designs, for design augmentation and for the design of experiments with mixture variables and qualitative variables.

Acknowledgement

Peter Goos is supported by a doctoral research grant from the Fund for Scientific Research - Flanders (Belgium).

Appendices

Appendix A. Mean and variance of $\ln s_i^2$

From (2), we can derive that

$$t_i = \ln\left(\frac{n_i - 1}{2} \frac{s_i^2}{\sigma_i^2}\right)$$

follows a loggamma distribution with parameter $(n_i - 1)/2$ such that its mean and variance amount to $\Psi(\frac{n_i-1}{2})$ and $\Psi'(\frac{n_i-1}{2})$ respectively. Rewriting mean and variance of t_i yields

$$\begin{aligned} E(t_i) &= E(\ln s_i^2) - \ln \sigma_i^2 + \ln\left(\frac{n_i - 1}{2}\right), \\ \text{var}(t_i) &= \text{var}(\ln s_i^2). \end{aligned}$$

Appendix B. Derivation of the design criterion

Substituting $\mathbf{Z} = (\mathbf{1} \quad \mathbf{X})$ in (12) gives

$$\mathbf{M} = \begin{bmatrix} \mathbf{1}'\mathbf{W}\mathbf{1} & \mathbf{1}'\mathbf{W}\mathbf{X} \\ \mathbf{X}'\mathbf{W}\mathbf{1} & \mathbf{X}'\mathbf{W}\mathbf{X} \end{bmatrix}.$$

From matrix algebra, we have that $\mathbf{M}^{-1}$ can be written as

$$\begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{12}' & \mathbf{B}_{22} \end{bmatrix},$$

where

$$\begin{aligned} \mathbf{B}_{22} &= \{\mathbf{X}'\mathbf{W}\mathbf{X} - \mathbf{X}'\mathbf{W}\mathbf{1}(\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}\mathbf{1}'\mathbf{W}\mathbf{X}\}^{-1}, \\ &= [\mathbf{X}'\{\mathbf{W} - \mathbf{W}\mathbf{1}(\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}\mathbf{1}'\mathbf{W}\}\mathbf{X}]^{-1}, \\ &= (\mathbf{X}'\mathbf{W}^*\mathbf{X})^{-1}, \end{aligned}$$

the variance-covariance matrix of the parameters of interest excluding the intercept. Hence, $\mathbf{X}'\mathbf{W}^*\mathbf{X}$ is the matrix of interest. Since

$$\begin{aligned} |\mathbf{Z}'\mathbf{W}\mathbf{Z}| &= |\mathbf{1}'\mathbf{W}\mathbf{1}||\mathbf{X}'\mathbf{W}\mathbf{X} - \mathbf{X}'\mathbf{W}\mathbf{1}(\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}\mathbf{1}'\mathbf{W}\mathbf{X}|, \\ &= (\mathbf{1}'\mathbf{W}\mathbf{1})|\mathbf{X}'\mathbf{W}^*\mathbf{X}|, \end{aligned}$$

the $\mathcal{D}$ -criterion value is given by (13).

References

- Abramowitz, M. and Stegun, I. (1970). *Handbook of Mathematical Functions*, New York: Dover Publications.
- Aitkin, M. (1987). Modeling variance heterogeneity in normal regression using GLIM, *Applied Statistics* **36**: 332–339.
- Atkinson, A. and Cook, R. (1995). D -optimum designs for heteroscedastic linear models, *Journal of the American Statistical Association* **90**: 204–212.
- Atkinson, A. and Donev, A. (1989). The construction of exact D -optimum experimental designs with application to blocking response surface designs, *Biometrika* **76**: 515–526.
- Atkinson, A. and Donev, A. (1992). *Optimum Experimental Designs*, Oxford U.K.: Clarendon Press.
- Bergman, B. and Hynén, A. (1997). Dispersion effects from unreplicated designs in the 2^{k-p} series, *Technometrics* **39**: 191–198.
- Box, G. and Meyer, R. (1986). Dispersion effects from fractional designs, *Technometrics* **28**: 19–27.
- Carroll, R. and Ruppert, D. (1988). *Transformation and Weighting in Regression*, Chapman and Hall.
- Davidian, M. (1990). Estimation of variance functions in assays with possibly unequal replication and nonnormal data, *Biometrika* **77**: 43–54.
- Davidian, M. and Carroll, R. (1987). Variance function estimation, *Journal of the American Statistical Association* **82**: 1079–1091.
- Del Castillo, E. and Montgomery, D. (1993). A nonlinear programming solution to the dual response problem, *Journal of Quality Technology* **25**: 199–204.
- Donev, A. (1997). An algorithm for the construction of crossover trials, *Applied Statistics* **46**: 288–298.
- Donev, A. (1998). Construction of non-standard row-column designs, in R. Payne and P. Green (eds), *COMPSTAT 1998, Proceedings in Computational Statistics*, Harpenden: IACR-Rothamsted, pp. 275–280.
- Fedorov, V. (1972). *Theory of Optimal Experiments*, New York: Academic Press.
- Huele, A. (1998). *Statistical Robust Design*, Ph.d. thesis, University of Amsterdam, The Netherlands.
- Lin, D. and Tu, W. (1995). Dual response surface optimization, *Journal of Quality Technology* **27**: 34–39.

- Mitchell, T. (1974). An algorithm for the construction of D -optimal experimental design, *Technometrics* **16**: 203-210.
- Myers, R. and Montgomery, D. (1995). *Response Surface Methodology: Process and Product Optimization using Designed Experiments*, New York: Wiley.
- Nair, V. and Pregibon, D. (1988). Analyzing dispersion effects from replicated factorial experiments, *Technometrics* **30**: 247-257.
- Raab, G. (1981). Estimation of a variance function, with application to immunoassay, *Applied Statistics* **30**: 32-40.
- Sohn, S. and Park, C. (1998). Random effects linear models for process mean and variance, *Journal of Quality Technology* **30**: 33-39.
- Taguchi, G. (1989). *Introduction to Quality Engineering: Designing Quality into Products and Processes*, Tokyo: Asian Productivity Organization.
- Vining, G. and Myers, R. (1990). Combining Taguchi and response surface philosophies: A dual response approach, *Journal of Quality Technology* **22**: 38-45.
- Vining, G. and Schaub, D. (1996). Experimental design for estimating both mean and variance functions, *Journal of Quality Technology* **28**: 135-147.
- Wolfinger, R. and Tobias, R. (1998). Joint estimation of location, dispersion, and random effects in robust design, *Technometrics* **40**: 62-71.