DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

ONDERZOEKSRAPPORT NR 9745

SEMI-BAYESIAN D-OPTIMAL DESIGNS AND ESTIMATION PROCEDURES FOR MEAN AND VARIANCE FUNCTIONS

by

Martina Vandebroek

Peter Goos

Katholieke Universiteit Leuven

Naamsestraat 69, B-3000 Leuven

ONDERZOEKSRAPPORT NR 9745

SEMI-BAYESIAN D-OPTIMAL DESIGNS AND ESTIMATION PROCEDURES FOR MEAN AND VARIANCE FUNCTIONS

by

Martina Vandebroek

Peter Goos

D/1997/2376/47

Semi-Bayesian D-optimal designs and estimation procedures for mean and variance functions

M. Vandebroek and P. Goos*

Department of Applied Economics, K.U.Leuven, Belgium

(martina.vandebroek@econ.kuleuven.ac.be)

(peter.goos@econ.kuleuven.ac.be)

*Aspirant van het Fonds voor Wetenschappelijk Onderzoek -Vlaanderen

Abstract

Semi-Bayesian D-optimal designs for fitting mean and variance functions are derived for some prior distributions on the variance function parameters. The impact of the mean of the prior and of the uncertainty about this mean is analyzed. Simulation studies are performed to investigate whether the choice of design has a substantial impact on the efficiency of the mean and variance function parameter estimation and whether the D-optimality criterion is appropriate irrespective of the method applied to estimate the variance function parameters.

KEY WORDS: experimental design, D-optimality, heteroscedasticity, variance estimation

1 Introduction

The topic of this paper is an offshoot of two developments in statistical quality control. Firstly, experimental design or design of experiments set itself up as a powerful and complementary quality improvement technique for statistical process control. Although indispensable, statistical process control only allows limited gains when compared to experimental design. The larger flexibility characterizing the design phase allows quality to be built in products and processes from the start, thereby creating opportunities for considerable quality improvements. These insights incited to a substantial literature on optimum experimental designs. Secondly, Taguchi highlighted the necessity to develop experimental strategies to achieve some target values for the expected value of certain characteristics while at the same time minimizing their variance. From these developments emerged the need for designs that are suited for estimating the mean and variance structure simultaneously. Usually, however, the major part of the optimum experimental design theory is concerned with designs optimal for response function estimation under the assumption of homoscedasticity. This can be justified by argueing that one can not use information that is unavailable but nevertheless it remains a rather weird strategy. Recently Mays & Easter (1997) derived D- and I- optimal designs for various hypothetical variance functions. Atkinson & Cook (1995) and Vining & Schaub (1996) described a semi-Bayesian approach in that they use prior information on the variance function in order to determine the optimal design. Atkinson and Cook (1995) derive necessary and sufficient conditions for continuous designs to be semi-Bayesian D-optimal for estimating

mean and variance functions simultaneously, while Vining and Schaub (1996) give the optimality criterion for discrete Bayesian designs.

Although the the optimization issues for experimental design are rarely addressed, much work has been done on describing and analyzing models in which both means and variances are functions of the experimental variables. A careful review of the major statistical techniques used to analyze data with nonconstant variability is given by Carroll and Ruppert (1988). Their work includes an extensive treatment of the different estimation procedures for fitting variance functions. Specific problems with variance function estimation are also addressed by Raab (1981), Davidian and Carroll (1987), and by Davidian (1990).

The purpose of this paper is to give a more thorough analysis of the impact of the prior information on the optimal design and of the impact of the design and the estimation procedure on the efficiency of the mean and variance function estimation. In the next section, we start by deriving the optimality criterion for semi-Bayesian designs. Computational results are given in section 3. Section 4 provides the reader with a brief overview of the estimation techniques for the parameters of both the mean and variance function. Finally, section 5 analyzes whether the designs, computed in section 3, have a substantial impact on the estimation efficiency and whether this influence is similar for all estimation procedures.

2 Semi-Bayesian D-optimal designs

Let Y be the response of interest and let $\mathbf{x}$ and $\mathbf{z}$ denote the $(p \times 1)$ and $(q \times 1)$ vectors of control variables presumed to influence the response and variance function respectively. Denote by $\mathbf{f}(\mathbf{x})$ the $(p_r \times 1)$ vector representing the polynomial expansion of $\mathbf{x}$ for the response model and by $\mathbf{g}(\mathbf{z})$ the $(q_v \times 1)$ vector representing the polynomial expansion of $\mathbf{z}$ for the variance model, with $\mathbf{g}(\mathbf{z})$ containing an intercept. With $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ the $(p_r \times 1)$ and $(q_v \times 1)$ the vectors of unknown parameters, we assume the following heteroscedastic model

$$Y = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + \sqrt{v[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]}\ \epsilon \tag{1}$$

The disturbance term is standardized such that $\mathrm{E}(\epsilon) = 0$ and $\mathrm{VAR}(\epsilon) = 1$ which yields the following mean and variance functions

$$\mathrm{E}(Y) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} \tag{2}$$

$$\mathrm{VAR}(Y) = v[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}] \tag{3}$$

Optimum design theory, dating back to Kiefer and Wolfowitz (1959), suggests that the choice of design should maximize the information on the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. This corresponds to maximizing a function of the information matrix on the unknown parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma})$. The most widely used criterion is D-optimality, which maximizes the determinant of the information matrix. For our model, the per observation information matrix on

$(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is given by

$$\mathbf{I}(x,z) = -\mathrm{E}\begin{pmatrix} \frac{\partial^2 \log \mathbf{L}}{\partial \boldsymbol{\beta}^2} & \frac{\partial^2 \log \mathbf{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}} \\ \frac{\partial^2 \log \mathbf{L}}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}} & \frac{\partial^2 \log \mathbf{L}}{\partial \boldsymbol{\gamma}^2} \end{pmatrix}$$

where L is the likelihood function. Under the assumption of independent standard normal random error terms ϵ, L becomes

$$\mathbf{L}(y,x,z|\boldsymbol{\beta},\boldsymbol{\gamma}) = \frac{1}{\sqrt{v[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]}\sqrt{2\pi}} e^{-\frac{1}{2}\frac{[y-\mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}]^2}{v[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]}}$$

The information matrix on $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ in the point $\mathbf{x}, \mathbf{z}$ is then given by

$$\mathbf{I}(x,z) = \begin{pmatrix} \frac{-\mathbf{f}(\mathbf{x})\mathbf{f}^T(\mathbf{x})}{v[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]} & 0 \\ 0 & \frac{1}{2}\left[\frac{v'[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]}{v[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]}\right]^2 \mathbf{g}(\mathbf{z})\mathbf{g}^T(\mathbf{z}) \end{pmatrix}$$

where v' stands for the first derivative of v.

The total information in the design $\{\mathbf{x}, \mathbf{z}\}_{i=1}^N$ is found by summating the per observation matrices over all design points:

$$\sum_{i=1\ldots N} \mathbf{I}(x_i, z_i) = \begin{pmatrix} \sum_i \frac{-\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)}{v[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]} & 0 \\ 0 & \sum_i \frac{1}{2}\left[\frac{v'[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]}{v[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]}\right]^2 \mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i) \end{pmatrix}$$

A particular design is D-optimal if it maximizes the following determinant over all possible designs

$$\left|\sum_{i=1\ldots N} \mathbf{I}(x_i, z_i)\right| = \left|\sum_i \frac{-\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)}{v[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]}\right| \times \frac{1}{2}\left|\sum_i \left[\frac{v'[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]}{v[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]}\right]^2 \mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i)\right|$$

This expression simplifies a lot if we use the following model for the variance:

$$v[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}] = \exp[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}] \tag{4}$$

This is called *multiplicative heteroscedasticity* and the log-linear form ensures that estimated variances are positive. In this notation, γ_1 or equivalently $\sigma^2 = \exp\gamma_1$, can be seen as a scale parameter. Substituting the exponential function in the total information matrix yields:

$$\sum_{i=1\ldots N} \mathbf{I}(x_i, z_i) = \begin{pmatrix} \sum_i \frac{-\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)}{\exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]} & 0 \\ 0 & \sum_i \frac{1}{2}\mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i) \end{pmatrix}$$

In the homoscedastic case, the information matrix only depends on the scale parameter and consequently the D-optimal design does not depend on the unknown parameters. However, in case of heteroscedasticity, it is clear from the information matrix that the D-optimal design depends on the value of $\boldsymbol{\gamma}$. In order to overcome this dependence, we will adopt a semi-Bayesian approach and take into account any available prior information on the variance function to determine the optimal design.

Following Atkinson & Cook (1995) and Vining & Schaub (1996) we will assume a prior distribution for $\boldsymbol{\gamma}$ and maximize the determinant of the expected information matrix. Atkinson and Cook (1995) computed optimal designs assuming a discrete prior distribution for $\boldsymbol{\gamma}$, whereas Vining and Schaub (1996) use a multivariate normal distribution $N(\boldsymbol{\gamma}_0, \rho I)$ as a prior for $\boldsymbol{\gamma}$. Adopting the more realistic assumption of the latter, the determinant of the expected information matrix becomes

$$\left|\sum_i \mathbf{f}(\mathbf{x}_i)\left[\exp[-\boldsymbol{\gamma}_0{}^T\mathbf{g}(\mathbf{z}_i) + \frac{1}{2}\rho\mathbf{g}^T(\mathbf{z}_i)\mathbf{g}(\mathbf{z}_i)]\right]\mathbf{f}^T(\mathbf{x}_i)\right| \times \frac{1}{2}\left|\sum_i \mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i)\right|$$

This expression differs from the result in Vining and Schaub (1996) in that they have omitted the parameter ρ for no obvious reason. This parameter plays a rather important

role in the remainder of this paper because it expresses the degree of uncertainty attached to the prior. The results of Vining and Schaub (1996) can be derived as special cases setting $\rho = 1$.

3 Computational results w.r.t. D-optimal designs

In this section, we derive semi-Bayesian D-optimal designs on the design region $\chi = [-1, 1]^2$. To simplify the representation of our results we will assume $\boldsymbol{x} = \boldsymbol{z}$ which means that the factors influencing the mean and variance functions are the same.

We will compute and compare optimal designs for a number of different settings. Firstly, we distinguish between a *discrete* and an *approximately continuous* design region. In the former case, we use the results of section 2 to find optimum designs for the response and variance function parameters over a 3×3 grid on χ whereas for the latter case, we choose the design points from the 21×21 grid on χ. Secondly, we calculate both *exact* and *approximately continuous* designs consisting of 12 and 54 observations or runs respectively. Finally, four different values for γ_0, the expected value of the variance function parameters in the prior, and three different values for ρ are used.

This extends the results of Vining and Schaub (1996), who dealt with exact designs over a discrete $(3 \times 3 \times 3)$ design region for $\rho = 1$ and four different different values for γ_0 and Mays and Easter (1997) who computed exact designs over a discrete design region for $\rho = 0$ and several variance structures. Atkinson and Cook (1995) computed continuous designs over a continuous design region.

In order to compute the Bayesian D-optimal designs, we adapted the BLKL exchange algorithm described by Atkinson and Donev (1992). This algorithm randomly chooses a few design points, then adds the missing number of points using a greedy heuristic. In each step, the point that leads to the greatest improvement of the optimality criterion is added to the design. Finally, the algorithm tries to improve the design by exchanging one of the K design points, at which the prediction variance is lowest, with one of the L grid points, at which the prediction variance is highest. The parameters K and L are defined by the user.

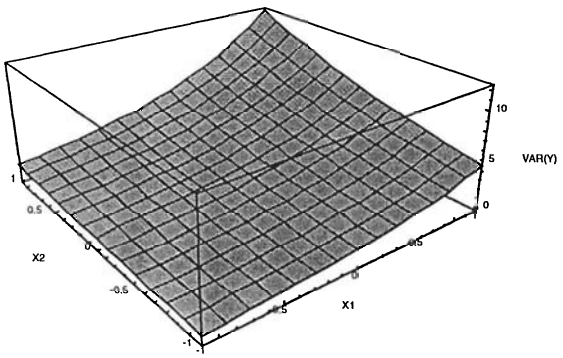

Figure 1: Variance function

The response function we adopted is the full-second order polynomial

$$\mathbf{f}^T(\mathbf{x}) = (1 \quad x_1 \quad x_2 \quad x_1^2 \quad x_2^2 \quad x_1 x_2)^T$$

and the variance function we used to simulate responses is

$$v[\mathbf{g}^T(\mathbf{x})\boldsymbol{\gamma}_0^*] = \exp(1 + x_1 + \frac{1}{2}x_2)$$

This function is depicted in figure 1. This choice implies

$$g^T(x) = (1 \quad x_1 \quad x_2)^T \quad \text{and} \quad \gamma_0^* = (1 \quad 1 \quad \frac{1}{2})^T$$

For each combination of design region (discrete or continuous) and number of observations, the resulting designs are presented in the following scheme:

	$\rho = 0$	$\rho = 1$	$\rho = 10$
$\gamma_0 = (1,0,0)^T$, constant prior			
$\gamma_0 = \gamma_0{}^* = (1,1,\frac{1}{2})^T$, true value			
$\gamma_0 = (1,1,\frac{3}{2})^T$, slightly wrong			
$\gamma_0 = (1,-1,-\frac{1}{2})^T$, completely wrong			

In this table, each combination of γ_0 and ρ represents a different prior distribution for the variance function parameters. In the ideal situation, we know with certainty ($\rho = 0$) the true value of the variance function parameters, that is $\gamma_0 = \gamma_0{}^*$. Since the design resulting from this prior distribution is the ideal design for a given estimation problem, every design should be compared to this one in order to evaluate its performance. Moving to the right in the table means we are increasingly uncertain about the prior distribution on γ_0 and we attach more weight to variance function estimation. Special attention will be given to the top left cell in which we assume a constant variance and do not take into account a possible misspecification ($\rho = 0$). This assumption will lead to the classical designs.

As mentioned in the previous paragraph, to evaluate the performance of a given design,

we compare it to the *ideal* design:

$$\left[\frac{\left|\sum_{considered\ design} \mathbf{f}(\mathbf{x}_i)\left[e^{-\gamma_0{}^{*T}\mathbf{g}(\mathbf{z}_i)}\right]\mathbf{f}^T(\mathbf{x}_i)\right| \times \frac{1}{2}\left|\sum_i \mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i)\right|}{\left|\sum_{optimal\ design} \mathbf{f}(\mathbf{x}_i)\left[e^{-\gamma_0{}^{*T}\mathbf{g}(\mathbf{z}_i)}\right]\mathbf{f}^T(\mathbf{x}_i)\right| \times \frac{1}{2}\left|\sum_i \mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i)\right|}\right]^{\frac{1}{p_r+p_v}}$$

The denominator of this measure computes the determinant of the information matrix corresponding to the ideal design. The numerator computes how the design considered will perform in reality, in casu with $\gamma_0 = \gamma_0{}^*$ and $\rho = 0$. Tables A1 and A2 in the appendix divide this total efficiency into the efficiency for mean function estimation and the efficiency for variance function estimation.

The resulting designs for the four combinations of design region and number of observations are shown in figures 2, 3, 4 and 5. From these figures, it is clear that the impact of the prior in case of a discrete design region and few observations is rather small. The reason is that there is little flexibility in the choice of design points. For this situation, having precise knowledge on the variance function is no better than having only a fairly good idea. Moreover, there is no need to assess the uncertainty about this prior information since the optimal design remains the same for each value of ρ. It is also clear that using a classical design with constant prior is better than using a completely wrong prior distribution (95.15% vs. 67.78%).

In case of many runs on a discrete design region, the results are similar although there is somewhat more flexibility in choosing the number of replications of each design point. The optimal design will therefore depend to a small extent on the degree of certainty attached to the prior.

The real impact of the prior distribution can be derived from the 12- and 54-point designs

generated on a continuous design region. For each prior γ_0 the design points selected shift towards the corner points as ρ increases. This is because increasing ρ attaches more importance to the variance function estimation and since $\mathbf{g}(\mathbf{z})$ is linear in the independent variables, choosing the points near the corners yield maximum information for variance estimation. However, as the response function is quadratic, one needs observations for at least three different levels of the control variables. These design points tend to lie where the variance is expected to be small. As before, for small ρ and under the assumption of variance homogeneity, classical designs are obtained.

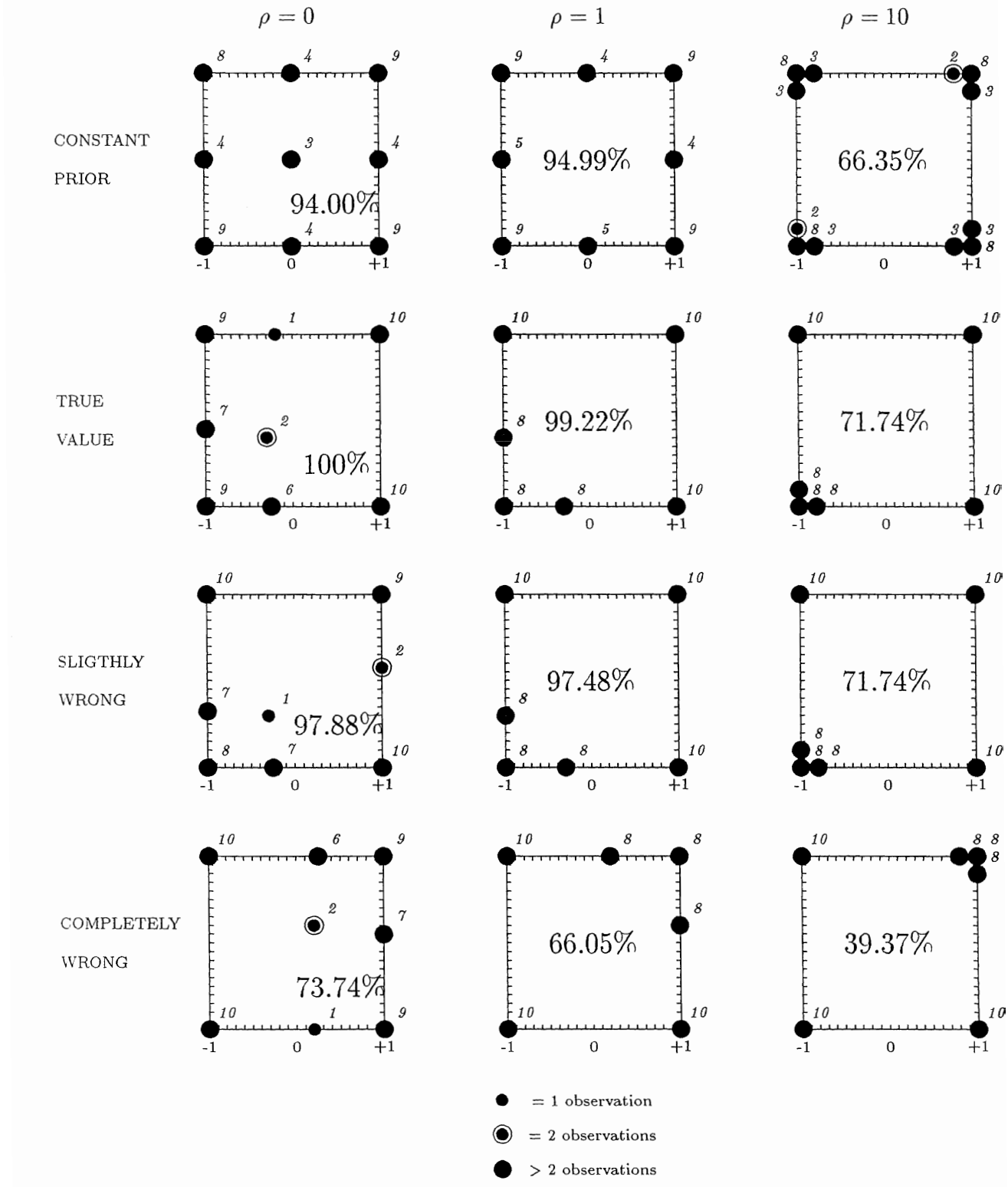

Figure 2: Optimal designs (54 observations; 21×21 grid)

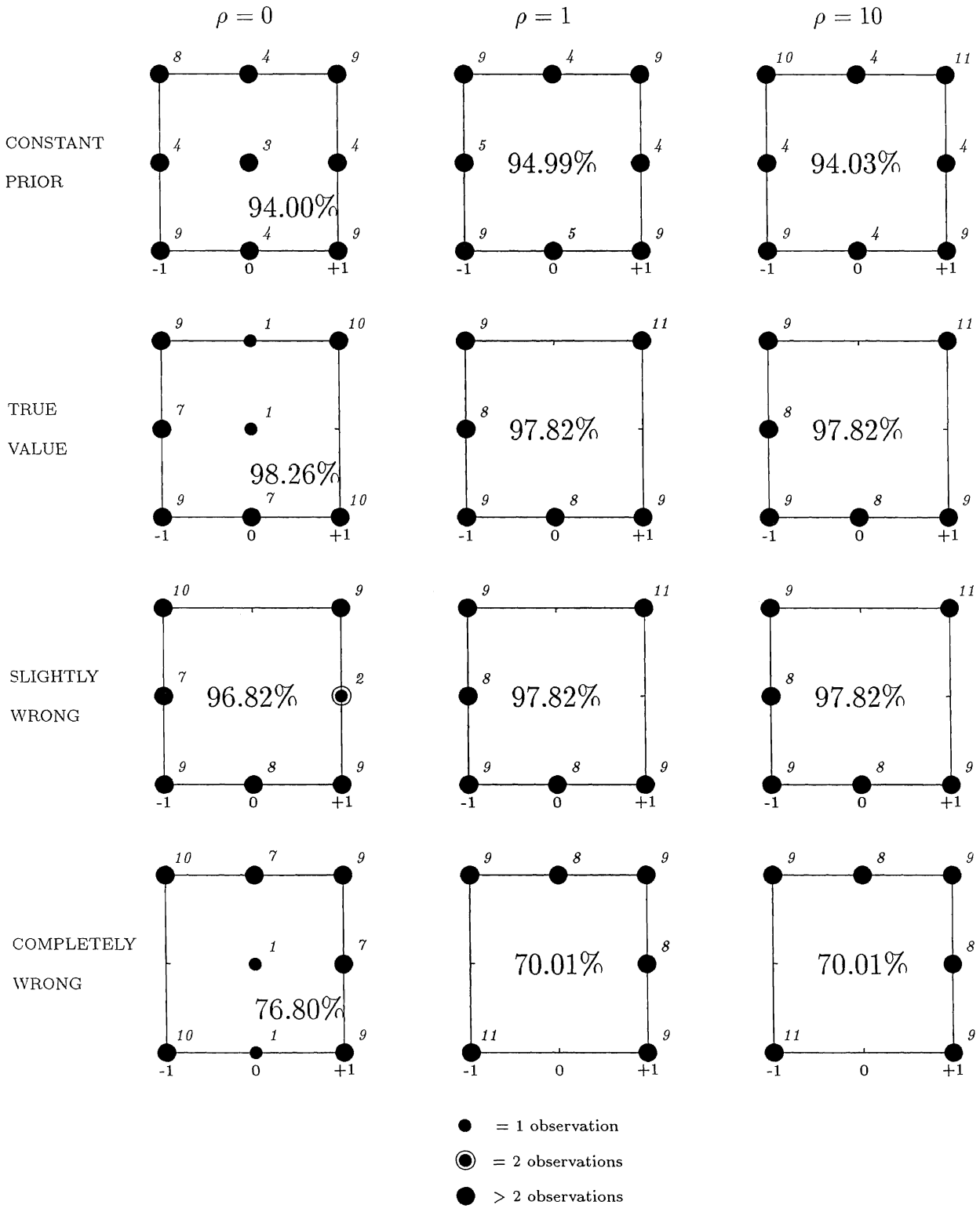

Figure 3: Optimal designs (54 observations; 3×3 grid)

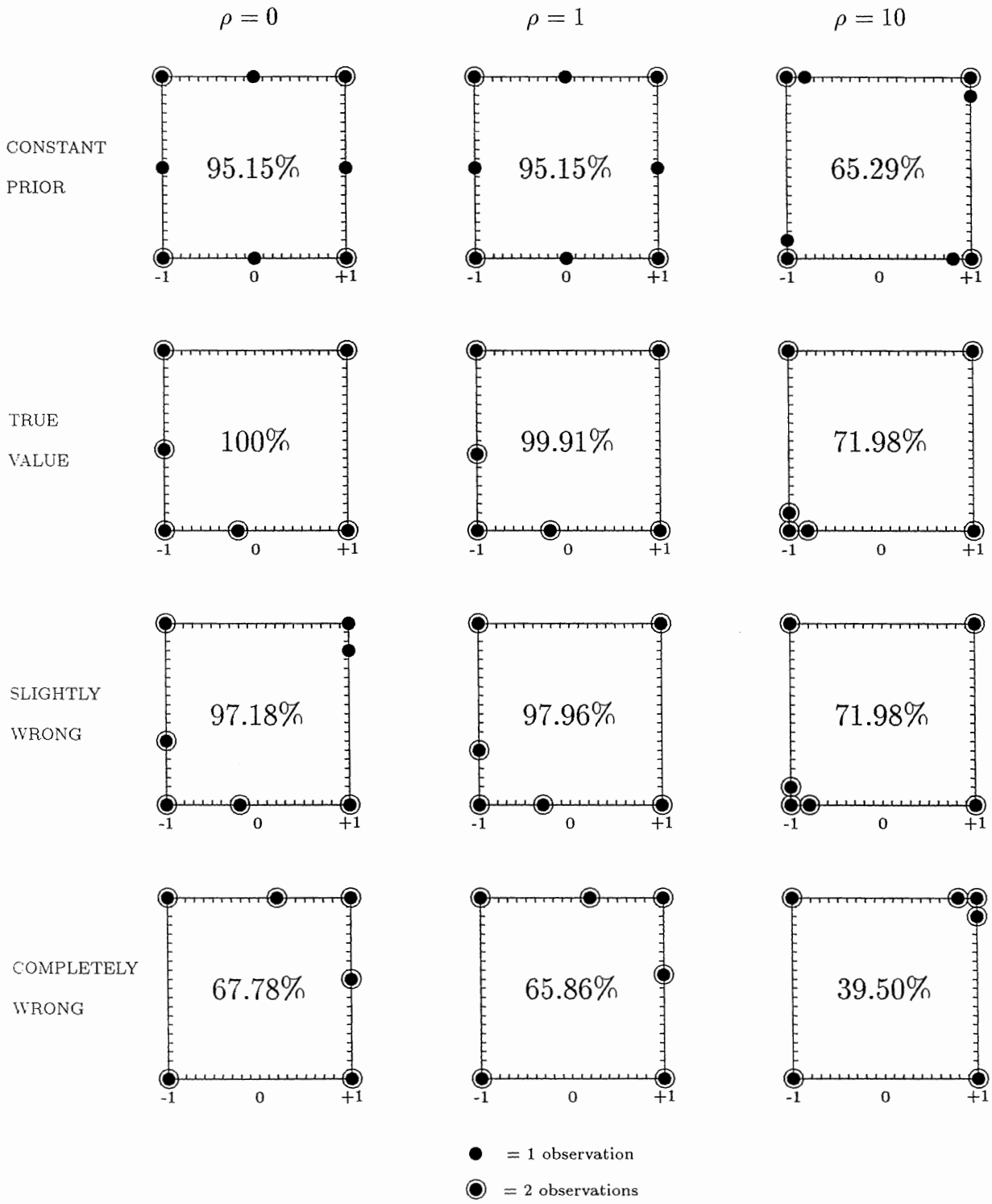

Figure 4: Optimal designs (12 observations; 21×21 grid)

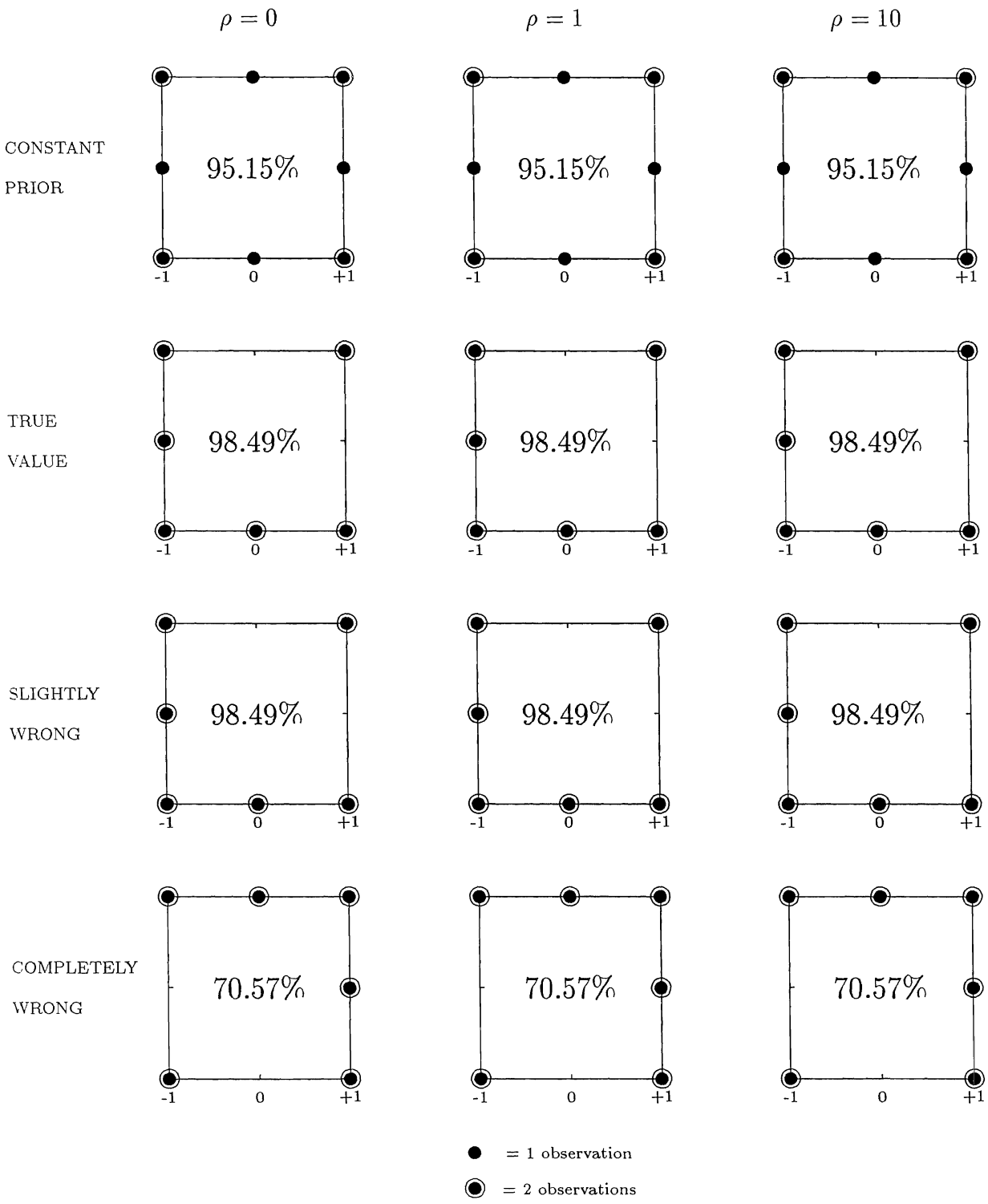

Figure 5: Optimal designs (12 observations; 3×3 grid)

4 Estimation of mean and variance functions

Under the assumption of normally distributed and homogeneous error terms, using maximum likelihood estimators, or equivalently ordinary least squares estimators, for mean function estimation is generally agreed upon. Since the variance-covariance matrix of the estimators depends on the inverse of the information matrix, the D-optimality criterion is intuitively appealing. No doubt therefore exists as to the relationship between resulting designs and efficiency of mean function estimation.

When it comes to estimating the mean and variance function simultaneously, there is far less agreement on which estimation method to use and therefore on which design to choose. As the maximum-likelihood estimator for the variance does not behave well, one has developed several other estimation techniques of which we will describe some in the sequel. The purpose of the second part of this paper is to measure to what extent the semi-Bayesian D-optimal designs developed in the previous section are fit to allow for efficient mean and variance function estimation in combination with these estimation methods.

This rest of this section will provide the reader with a brief overview of the estimation methods we used. In the next section, we describe the results of our simulation study.

4.1 Variance function estimation based on squared residuals

These methods are basically iterative procedures where one starts with an initial estimate for $\boldsymbol{\beta}$ which is used to compute the residuals $(y_i - \mathbf{f}^T(\mathbf{x})\hat{\boldsymbol{\beta}})^2$. These residuals are then

used to estimate γ with one of the methods described in this section. The estimates for γ are used in a weighted least squares algorithm to improve the estimates for β. This process can be iterated using the new estimate for β. Typically, only a small number of iterations is needed for convergence. These methods are often more efficient than those using squared variances (see section 4.2), especially when the number of replications at the design points is small. Their major disadvantage is that they are unreliable if the response model is possibly misspecified which explains why they are less often used in practice than could be expected from their statistical properties.

4.1.1 Generalized Least Squares estimation

Given a preliminary estimator $\hat{\beta}_*$, the generalized least squares estimator $\hat{\gamma}_{GLS}$ maximizes in γ the normal log-likelihood function $l(\hat{\beta}_*, \gamma)$ where

$$l(\beta, \gamma) = -\sum_{i=1}^{N} \log \sqrt{v[\mathbf{g}^T(\mathbf{z}_i)\gamma]} - \sum_{i=1}^{N} \frac{[y_i - \mathbf{f}^T(\mathbf{x}_i)\beta]^2}{2v[\mathbf{g}^T(\mathbf{z}_i)\gamma]}$$

Since we assume multiplicative heteroscedasicity, we have

$$l(\beta, \gamma) = -\frac{1}{2}\sum_{i=1}^{N} \mathbf{g}^T(\mathbf{z}_i)\gamma - \sum_{i=1}^{N} \frac{[y_i - \mathbf{f}^T(\mathbf{x}_i)\beta]^2}{2\exp[\mathbf{g}^T(\mathbf{z}_i)\gamma]}$$

Taking derivatives with respect to the different components of γ, $\hat{\gamma}_{GLS}$ is the solution, assuming it exists, to the equations

$$\sum_{i=1}^{N} \left(1 - \frac{[y_i - \mathbf{f}^T(\mathbf{x}_i)\hat{\beta}]^2}{\exp[\mathbf{g}^T(\mathbf{z}_i)\gamma]}\right) \mathbf{g}(\mathbf{z}_i) = \mathbf{0} \tag{5}$$

These equations have the form of a set of p_v normal equations. The dependence on the design is contained within the vector $\mathbf{g}(\mathbf{z}_i)$. From this set of equations, we learn that the

scale parameter can be expressed in function of all other parameters

$$\hat{\gamma}_1 = \log \frac{1}{N} \sum_{i=1}^{N} \frac{[y_i - \mathbf{f}^T(\mathbf{x}_i)\hat{\boldsymbol{\beta}}]^2}{\sum_{j=2}^{p_v} \gamma_j z_{ji}} \quad \text{or} \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} \frac{[y_i - \mathbf{f}^T(\mathbf{x}_i)\hat{\boldsymbol{\beta}}]^2}{\sum_{j=2}^{p_v} \gamma_j z_{ji}} \tag{6}$$

Substituting this expression in (5) yields a set of $p_v - 1$ normal equations, providing estimates $\hat{\gamma}_i$ $(i = 2, .., p_v)$. The estimate for the scale parameter can then be calculated from (6). Our results will be based on the solutions of the sets of $p_v - 1$ equations, since the results obtained in this way are more stable than through solving the equivalent set of p_v equations.

4.1.2 Adapted GLS

One objection to generalized least squares is that it does not take into account the loss of degrees of freedom resulting from the preliminary estimation of $\boldsymbol{\beta}$. Therefore applying generalized least squares yields biased estimates, the magnitude of the bias depending on the ratio p_r/N which is relatively large in most designed experiments.

A simple way to take into account the loss of degrees of freedom consists of adapting the formula for the scale parameter as follows

$$\hat{\gamma}_1 = \log \frac{1}{N - p_r} \sum_{i=1}^{N} \frac{[y_i - \mathbf{f}^T(\mathbf{x}_i)\hat{\boldsymbol{\beta}}]^2}{\sum_{j=2}^{p_v} \gamma_j z_{ji}} \tag{7}$$

and proceeding as in (4.1.1).

4.1.3 Restricted maximum likelihood estimation

Another method to account for the loss of degrees of freedom relies on Bayesian ideas. This method, called restricted maximum likelihood (REML), is elaborated by Patterson and

Thompson (1971) and by Harville (1977). The restricted maximum likelihood estimator turns out to be equivalent with a generalized least squares estimator corrected for the effect of leverage and this is the approach we adopt here.

Jobson and Fuller (1980) showed that the expected value of the squared residuals obtained by a weighted least squares procedure is approximately given by

$$\mathrm{E}\ [y_i - \mathbf{f}^T(\mathbf{x}_i)\hat{\boldsymbol{\beta}}]^2 \approx [1 - h_{ii}(\boldsymbol{\gamma})]\ \exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]$$

The leverage values $h_{ii}(\boldsymbol{\gamma})$ are the diagonal elements of the $N \times N$ hat matrix $\boldsymbol{H}$

$$\boldsymbol{H}(\boldsymbol{\gamma}) = \boldsymbol{X}_*(\boldsymbol{X}_*^T\boldsymbol{X}_*)^{-1}\boldsymbol{X}_*^T$$

with $\boldsymbol{X}_*(\boldsymbol{\gamma})$ the $N \times p_r$ matrix with ith row the vector

$$\left(\frac{\mathbf{f}^T(\mathbf{x}_i)}{\sqrt{\exp\sum_{j=2}^{p_v}\gamma_j z_{ji}}}\right)$$

To account for the loss in degrees of freedom from estimating $\boldsymbol{\beta}$, the suggestion was to equate

$$\sum_{i=1}^{N}\frac{[y_i - \mathbf{f}^T(\mathbf{x}_i)\hat{\boldsymbol{\beta}}]^2}{\exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]}\mathbf{g}(\mathbf{z}_i)$$

to its expectation, leading to

$$\sum_{i=1}^{N}\frac{[y_i - \mathbf{f}^T(\mathbf{x}_i)\hat{\boldsymbol{\beta}}]^2}{\exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]}\mathbf{g}(\mathbf{z}_i) = \sum_{i=1}^{N}\mathbf{g}(\mathbf{z}_i)[1 - h_{ii}(\boldsymbol{\gamma})] \tag{8}$$

Note that $\boldsymbol{H}$ is an idempotent matrix. Therefore the sum of the leverage values $h_{ii}(\boldsymbol{\gamma})$ equals the rank of $\boldsymbol{H}$, here p_r. Making use of this property enables us to derive from (8) the same expression for the scale parameter as with the adapted generalized least squares method.

4.2 Variance function estimation based on sample standard deviations

Frequently, experimenters replicate the response at certain design points, allowing them to calculate sample standard deviations at each replicated setting of the predictor variables. As was mentioned earlier, the main advantage of this approach is that sample standard deviations contain valuable information about the variance function even if the model for the means is incorrectly specified.

Assume each design point i $(i = 1, ..M)$ is replicated r_i times, such that $\sum_1^M r_i = N$. The sample variance at the ith replicated design point is computed as

$$s_i^2 = \frac{1}{r_i - 1} \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_i)^2$$

where $\bar{y}_i$ is the average response at the ith design point.

In order to estimate the variance function parameters, transformations of the sample standard deviations are used. Davidian (1990) gives an excellent overview of the transformations used in practice. Davidian and Carroll (1987) define a general class of estimators for γ based on transformations of the sample standard deviations s_i as follows:

$$\sum_{i=1}^{M} \frac{\partial M_i(\gamma, \bar{y}_i)}{\partial \gamma_j} \left[\frac{T(s_i) - M_i(\gamma, \bar{y}_i)}{V_i(\gamma, \bar{y}_i)} \right] = 0 \quad \text{for } j = 1 \ldots q_v \tag{9}$$

with $T(s_i)$ the transformation function, $M_i(\gamma, \bar{y}_i)$ the variance function model and $V_i(\gamma, \bar{y}_i)$ the *weights* corresponding to the design points. We will consider two special cases of this general class of estimators that arise naturally from the assumption of multiplicative heteroscedasticity.

4.2.1 Regressing $\log s_i^2$ on $\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}$

From the variance function $\text{VAR}(Y) = \exp[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]$ we obtain the linear regression model in the sample variances

$$\log s_i^2 = \mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma} + \nu_i$$

which can be fitted by standard regression software. Unfortunately this regression model has several deficiencies. It can be shown (see Davidian (1990)) that the expected values of ν_i differ from zero and that their variances depend on the number of replications at each design point. More specifically

$$\begin{aligned} \text{E}(\nu_i) &= \Psi(\frac{1}{2}r_i - \frac{1}{2}) - \log(\frac{1}{2}r_i - \frac{1}{2}) && (10) \\ \text{VAR}(\nu_i) &= \Psi'(\frac{1}{2}r_i - \frac{1}{2}) && (11) \end{aligned}$$

with Ψ and Ψ' the digamma and trigamma functions. To get an idea of the magnitude of the problem, table 1 gives the values of equations (10) and (11) for some small values of r_i. From these results it is apparent that —if no precautions are taken— the estimate of the intercept will be biased and for unequal replications, weighted instead of unweighted least squares should be used to get efficient estimators.

4.2.2 Regressing s_i^2 on $\exp[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]$

If one is not confined to fitting linear models for the variance, one can fit the non-linear model

$$s_i^2 = \exp[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]$$

r_i	$\mathrm{E}(\nu_i)$	$\mathrm{VAR}(\nu_i)$
2	-0.635	1.234
3	-0.288	0.411
4	-0.182	0.234

Table 1: mean and variances of ν_i

which corresponds to solving the system of equations in (9) with

$$T(s_i) = s_i^2, M_i(\boldsymbol{\gamma}, \bar{y}_i) = \exp[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}], V_i(\boldsymbol{\gamma}, \bar{y}_i) = \frac{2\exp[2\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]}{r_i - 1}$$

Since

$$\mathrm{E}(s_i^2) = \exp[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]$$

regardless of the distribution of the error term ϵ, the resulting equations are unbiased estimating equations for the variance function parameters $\boldsymbol{\gamma}$ if $N \to \infty$ and $\sigma \to 0$.

4.3 Asymptotic efficiencies of variance function estimation

Davidian and Carroll (1987) and Davidian (1990) provide asymptotic relative efficiencies for variance function estimation by means of transformations of residuals and by means of transformations of sample standard deviations.

From Davidian (1990), it is obvious that, under the assumption of normal error distributions, using the transformation $T(s_i) = \log s_i^2$ is worse than choosing $T(s_i) = s_i^2$ in case the number of replications is small. However, the more the error distribution deviates from the normal, the better perform the former transformations.

Davidian and Carroll (1987) point out that for the small amount of replication found in practice, using sample variances may entail a loss in efficiency compared to using squared residuals. The asymptotic variance ratios when squared residuals are used instead of sample variances is $\frac{(r-1)2}{2+(r-1)2}$ for equal replications. For instance, with design points replicated twice, using squared residuals is double as efficient as using sample variances ! Davidian and Carroll (1987) and Carroll and Ruppert (1988) note that using squared residuals might cause outliers, leading to considerably degraded performance. On the other hand, for logarithm methods based on sample variances, it is of crucial importance to omit the smallest few variances for the same reason.

5 Evaluation of estimation methods and optimal designs

For each design derived on the 21 × 21 grid, we performed 1000 simulations using the following mean and variance functions:

$$\mathrm{E}(Y) = 100 + 10x_1 - 10x_2 - 5x_1^2 + 5x_2^2 + 2.5x_1x_2 \tag{12}$$

$$\mathrm{VAR}(Y) = \exp(1 + x_1 + \frac{1}{2}x_2) \tag{13}$$

Recall that (13) is the variance structure we defined as the right prior in section 3. In order to avoid computational difficulties in the estimation of variance function parameters, we generated responses with squared residuals larger than 0.001. From these simulations, the variance function parameters were estimated using each of the five methods described

in the previous section. Starting from these estimates, we estimated the mean function parameters using weighted least squares. Detailed simulation results can be found in tables A3, A4, A5 and A6 in the appendix. We will comment on the most important results here. Firstly, we will compare the estimates based on the designs resulting from the assumption of a constant variance. Next, we analyze whether knowledge of the true variance parameters when determining the optimal design yields significant better estimates.

5.1 Using designs based on a constant prior

The most important outcomes are those related to the designs derived under the constant variance assumption. We will concentrate on 54-trial designs. Table A3 of the appendix contains the average estimates, their variances and the average squared deviations from the true parameters for each of the three variance function parameters separately. Figure 6 plots the average squared deviations over the three variance function parameters for the five estimation procedures and the three designs. Remember that for $\rho = 0$, we found the classical designs. For $\rho = 1$ and $\rho = 10$, some uncertainty is attached to the assumption of homoscedasticity, and more weight is given to variance estimation. In figure 7 the corresponding average squared deviations from the true parameters over the six parameters of the mean function are shown.

In figure 6 we can see that for the methods based on residuals the variance function estimation improves considerably as ρ increases. The pattern is very similar for all three methods, but the GLS results are worse than those of the adapted GLS and the REML.

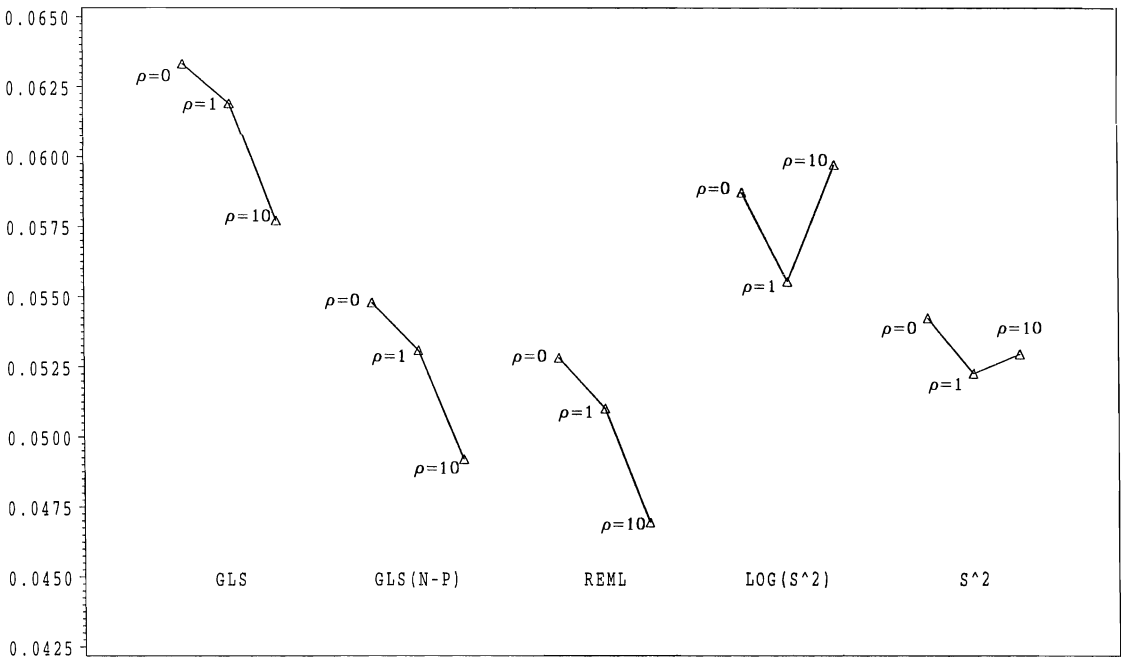

Figure 6: constant prior: average squared deviations of $\hat{\gamma}$

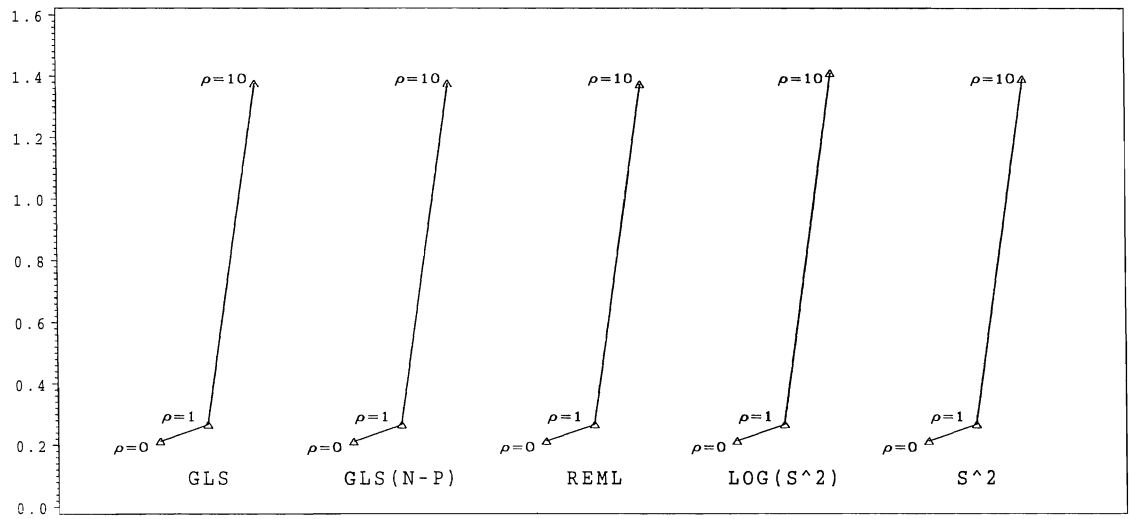

Figure 7: constant prior: average squared deviations of $\hat{\beta}_{WLS}$

From table A3 we learn that this difference is almost completely due to a bad estimation of the intercept. If we do not take into account the intercept — which is acceptable if the goal is to minimize the variance — GLS and adapted GLS perform equally well and slightly worse than REML.

As predicted by Davidian (1990), the methods based on sample variances are not as good as the methods based on residuals. Moreover, variance function estimation does not even improve with increasing ρ, indicating that the D-optimality criterion might not be appropriate when using sample variances.

From figure 7 we learn that the mean function estimation deteriorates with increasing ρ, independently from the method used to fit the variance function.

These outcomes suggest it might be useful to generate semi-Bayesian designs for strictly positive ρ if the variance function is of interest, even if one has no idea whatsoever about the variance function parameters. Undoubtedly, it is of crucial importance to choose an appropriate estimation technique.

5.2 The benefit of knowing the variance function

In order to assess the benefit of knowing the true variance function parameters, as in equation (13), we will compare the simulation results from the design based on the right prior and on the assumption of constant variance. Figure 8 is similar to figure 6 in that it compares the average squared deviations of variance function parameters, but now based on the constant prior and the right prior designs. For estimation based on residuals,

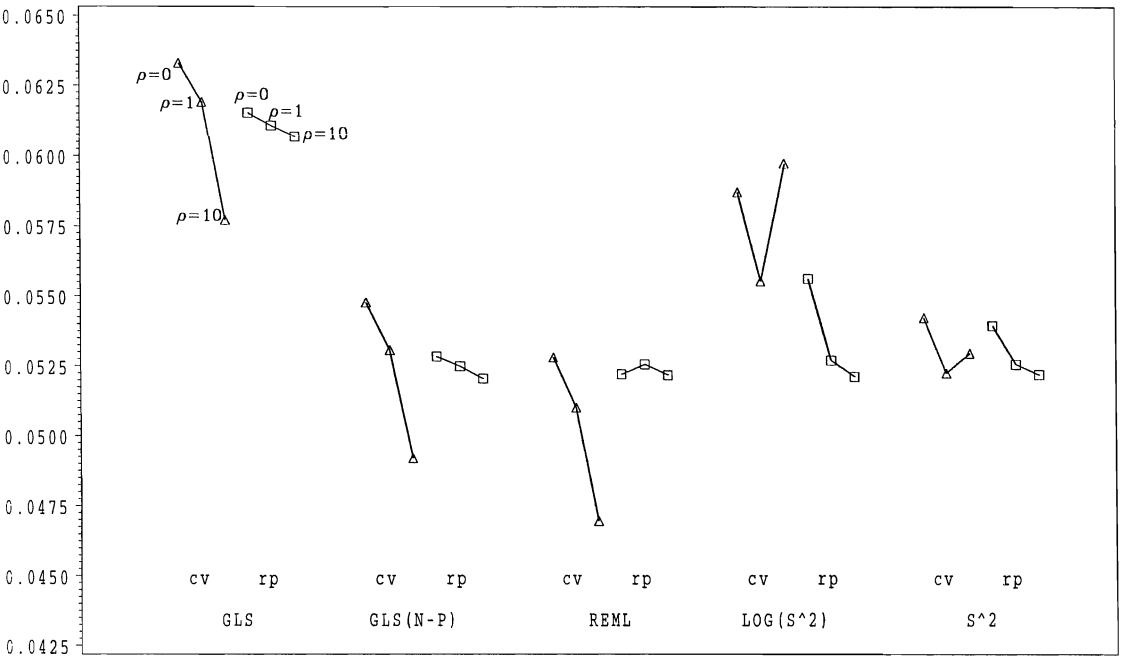

Figure 8: constant prior versus right prior: average squared deviations of $\hat{\gamma}$

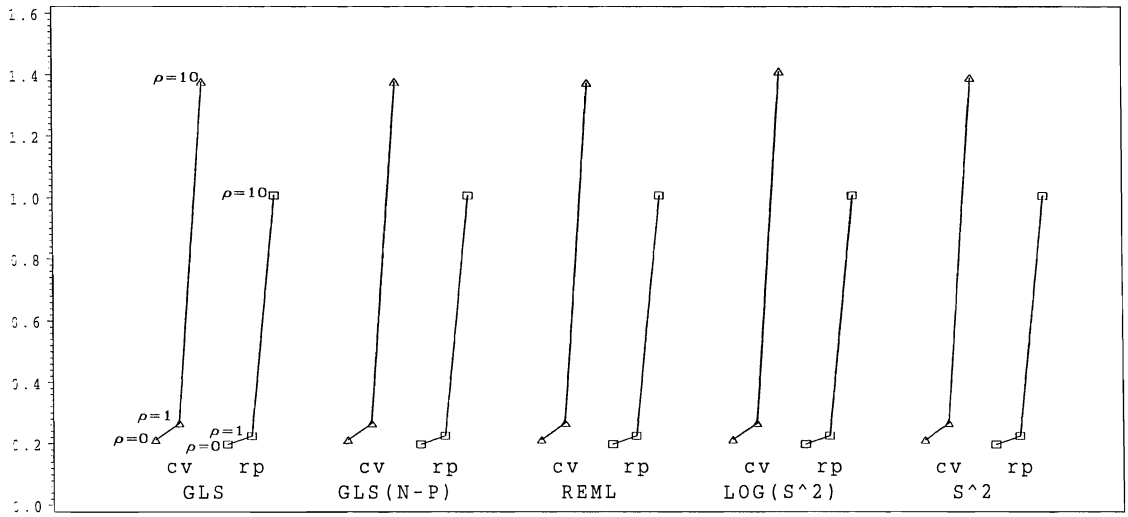

Figure 9: constant prior versus right prior: average squared deviations of $\hat{\beta}_{WLS}$

we see immediately that the results for the designs based on the right prior are far less sensitive to the value of ρ than the results for the designs derived under the assumption of homoscedasticity. The design derived under the assumption of constant variance with $\rho = 10$ even outperforms its right prior counterparts. Attaching more weight to variance function estimation is thus of no great use if one already has a fairly good idea about the variance structure. As to the methods using sample variances, it is less clear how the variance function estimation can be improved. Neither increasing ρ nor choosing a better prior seems to yield outcomes comparable to those based on residuals.

From figure 9 it is clear that the impact on mean function estimation is nearly the same for designs derived under the right prior and under constant variance.

5.3 Final remarks

In tables A2 to A6 in the appendix, the reader can find the results for the slightly wrong and the completely wrong priors. As could be expected from the similarity between the designs, using a prior that is slightly wrong is nearly as good as using the right prior. For the designs derived starting from a completely wrong prior, the results for the variance function are not dramatically worse than the results for the three other priors used. However, in this case the response function estimation deteriorates dramatically since most observations are taken at points with high variance.

The mean function was also fitted using ordinary least squares method. The results are very similar to those of the weighted least squares method for small ρ but the OLS-estimators are only half as efficient as the WLS-estimations for $\rho = 10$. The better fit of

the variance function can not compensate for the bad choice of design points.

We performed a similar comparison based on 12 observations (see the designs on pages 13, 14,15 and 16) but as the designs based on a constant prior have some non-replicated design points, only the methods based on residuals could be used here. It turns out that the results exhibit the same patterns as those for 54 observations. Of course, the estimates are much worse than those based on 54 observations. These findings can be verified in tables A4 and A6.

6 Conclusions

We derived semi-Bayesian D-optimal designs for several multivariate normal prior distributions on the parameters of the variance function. For discrete experimental regions, the designs resulting from different priors are almost equivalent except when the prior is really misspecified. For continuous experimental regions, the D-optimal design depends to a much larger extent on the importance that is attached to variance function estimation. With respect to estimation efficiency, the designs based on a constant prior perform very well compared to the designs based on the right prior. Moreover, if variance estimation is important and will be based on residuals, these designs allow one to balance out effectively the efficiency of mean and function estimation by increasing ρ.

It turns out that the D-optimality criterion is not really suited to fix the design if variance function estimation will be based on the sample variances in replicated design points. A possible explanation is that the optimality criterion does not take into account that

replicated design points yield only one sample variance each and as such have the same weight in the variance function estimation. When residuals are used, replicated design points contribute as much residuals as their number of replications, which makes the D-optimality criterion more suited for residual based estimation.

Weighted least squares yield satisfactory estimates for the mean function parameters and the efficiency does not depend much on the choice of the design.

Acknowledgement

We would like to thank Marie Davidian and Gillian Raab for their helpful comments.

Appendix

Table A1: Efficiency of 54-run designs for mean and variance function estimation

54 observations							
		21×21 grid			3×3 grid		
		$\rho = 0$	$\rho = 1$	$\rho = 10$	$\rho = 0$	$\rho = 1$	$\rho = 10$
cv	mean	92.22	92.26	51.75	92.22	92.26	90.19
	var	97.67	100.70	109.11	97.67	100.70	102.21
rp	mean	100.00	98.23	59.42	97.10	96.17	96.17
	var	100.00	101.21	104.59	100.62	101.19	101.19
sw	mean	96.68	95.54	59.42	95.16	96.17	96.17
	var	100.35	101.49	104.59	100.24	101.19	101.19
cw	mean	63.32	52.15	24.16	67.10	58.33	58.33
	var	100.00	101.21	104.59	100.62	101.19	101.19

Table A2: Efficiency of 12-run designs for mean and variance function estimation

12 observations							
		21×21 grid			3×3 grid		
		$\rho = 0$	$\rho = 1$	$\rho = 10$	$\rho = 0$	$\rho = 1$	$\rho = 10$
cv	mean	91.59	91.59	50.01	91.59	91.59	91.59
	var	102.71	102.71	111.30	102.71	102.71	102.71
rp	mean	100.00	99.88	60.31	97.56	97.56	97.56
	var	100.00	99.97	102.55	100.37	100.37	100.37
sw	mean	96.88	96.95	60.31	97.56	97.56	97.56
	var	97.77	100.01	102.55	100.37	100.37	100.37
cw	mean	55.80	53.46	24.52	59.17	59.17	59.17
	var	100.00	99.97	102.55	100.37	100.37	100.37

Table A3: Simulation results for variance function estimation from 54-run designs[1]

GENERALIZED LEAST SQUARES

	$\rho = 0$									$\rho = 1$									$\rho = 10$								
	γ_1			γ_2			γ_3			γ_1			γ_2			γ_3			γ_1			γ_2			γ_3		
	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR
cv	0.071	0.853	0.209	0.061	1.053	0.249	0.057	0.538	0.241	0.072	0.850	0.207	0.059	1.055	0.242	0.056	0.544	0.237	0.071	0.853	0.209	0.052	1.055	0.227	0.050	0.547	0.224
rp	0.073	0.850	0.211	0.054	1.035	0.233	0.058	0.521	0.242	0.077	0.854	0.222	0.054	1.025	0.234	0.052	0.522	0.229	0.080	0.854	0.227	0.052	1.024	0.230	0.050	0.520	0.225
sw	0.076	0.854	0.220	0.052	1.033	0.229	0.056	0.528	0.237	0.078	0.854	0.222	0.054	1.026	0.235	0.052	0.521	0.228	0.080	0.854	0.227	0.052	1.024	0.230	0.050	0.520	0.225
cw	0.075	0.849	0.214	0.053	1.027	0.231	0.059	0.548	0.244	0.072	0.853	0.211	0.052	1.022	0.230	0.051	0.527	0.228	0.068	0.872	0.209	0.046	1.008	0.216	0.046	0.519	0.215

ADAPTED GENERALIZED LEAST SQUARES

	$\rho = 0$									$\rho = 1$									$\rho = 10$								
	γ_1			γ_2			γ_3			γ_1			γ_2			γ_3			γ_1			γ_2			γ_3		
	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR
cv	0.046	0.971	0.209	0.061	1.050	0.249	0.057	0.542	0.241	0.045	0.967	0.207	0.059	1.052	0.242	0.055	0.541	0.237	0.046	0.971	0.209	0.052	1.056	0.227	0.050	0.546	0.224
rp	0.046	0.970	0.211	0.054	1.036	0.233	0.058	0.539	0.242	0.051	0.979	0.222	0.055	1.049	0.235	0.052	0.544	0.229	0.052	0.986	0.227	0.053	1.052	0.230	0.051	0.549	0.226
sw	0.050	0.978	0.220	0.052	1.030	0.229	0.056	0.534	0.237	0.051	0.980	0.222	0.055	1.049	0.235	0.052	0.546	0.228	0.052	0.986	0.227	0.053	1.052	0.230	0.051	0.549	0.226
cw	0.048	0.972	0.214	0.053	1.006	0.231	0.059	0.529	0.244	0.046	0.978	0.211	0.053	0.998	0.230	0.052	0.549	0.226	0.044	1.001	0.211	0.049	0.982	0.218	0.048	0.492	0.218

RESTRICTED MAXIMUM LIKELIHOOD

	$\rho = 0$									$\rho = 1$									$\rho = 10$								
	γ_1			γ_2			γ_3			γ_1			γ_2			γ_3			γ_1			γ_2			γ_3		
	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR
cv	0.046	0.972	0.209	0.059	1.027	0.244	0.054	0.529	0.238	0.048	0.970	0.207	0.056	1.027	0.238	0.052	0.528	0.230	0.046	0.973	0.209	0.049	1.024	0.222	0.046	0.525	0.217
rp	0.047	0.967	0.211	0.053	1.024	0.231	0.057	0.517	0.240	0.052	0.970	0.222	0.054	1.017	0.234	0.052	0.514	0.229	0.054	0.969	0.227	0.052	1.016	0.230	0.050	0.513	0.225
sw	0.047	0.960	0.209	0.054	1.012	0.233	0.057	0.518	0.240	0.048	0.962	0.211	0.053	1.020	0.231	0.052	0.507	0.228	0.057	0.960	0.231	0.050	1.022	0.226	0.052	0.505	0.227
cw	0.052	0.959	0.219	0.056	1.040	0.239	0.054	0.537	0.234	0.047	0.969	0.212	0.053	1.032	0.231	0.052	0.537	0.229	0.044	0.988	0.200	0.046	1.017	0.217	0.046	0.527	0.218

LOG S^2

	$\rho = 0$									$\rho = 1$									$\rho = 10$								
	γ_1			γ_2			γ_3			γ_1			γ_2			γ_3			γ_1			γ_2			γ_3		
	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR
cv	0.054	1.011	0.232	0.062	0.996	0.249	0.060	0.499	0.246	0.052	1.012	0.228	0.059	0.996	0.242	0.056	0.500	0.237	0.060	1.011	0.244	0.062	0.996	0.250	0.057	0.498	0.240
rp	0.052	1.014	0.228	0.055	1.007	0.234	0.060	0.495	0.245	0.051	1.012	0.226	0.055	1.000	0.235	0.052	0.498	0.226	0.056	1.012	0.231	0.053	1.001	0.230	0.050	0.498	0.224
sw	0.058	1.012	0.241	0.057	0.999	0.240	0.059	0.502	0.243	0.052	1.012	0.227	0.055	1.000	0.235	0.052	0.498	0.227	0.054	1.012	0.231	0.053	1.001	0.230	0.050	0.498	0.224
cw	0.052	1.015	0.229	0.055	0.991	0.233	0.060	0.798	0.246	0.048	1.014	0.218	0.053	0.993	0.231	0.052	0.500	0.228	0.047	1.018	0.221	0.049	0.996	0.225	0.047	0.502	0.222

S^2

	$\rho = 0$									$\rho = 1$									$\rho = 10$								
	γ_1			γ_2			γ_3			γ_1			γ_2			γ_3			γ_1			γ_2			γ_3		
	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR
cv	0.048	0.969	0.214	0.059	1.025	0.245	0.055	0.529	0.237	0.048	0.969	0.213	0.056	1.026	0.239	0.053	0.528	0.231	0.052	0.966	0.222	0.055	1.025	0.236	0.052	0.528	0.229
rp	0.050	0.967	0.218	0.053	1.024	0.232	0.059	0.516	0.244	0.052	0.970	0.222	0.054	1.017	0.234	0.052	0.514	0.229	0.054	0.969	0.227	0.052	1.016	0.230	0.050	0.513	0.225
sw	0.057	0.965	0.232	0.055	1.018	0.237	0.057	0.518	0.239	0.052	0.970	0.223	0.054	1.018	0.235	0.052	0.513	0.228	0.054	0.969	0.227	0.052	1.016	0.230	0.050	0.513	0.225
cw	0.051	0.969	0.221	0.053	1.028	0.231	0.057	0.537	0.241	0.047	0.969	0.211	0.052	1.032	0.230	0.052	0.537	0.229	0.048	0.969	0.215	0.049	1.033	0.224	0.049	0.539	0.223

[1] Average squared deviations of estimates from true parameter values (SQ DEV), average estimates (AVG) and their variances (VAR) are shown. Computations were made for designs derived under constant prior (cv), under the right prior (rp), under a slightly wrong (sw) and under a completely wrong (cw) prior.

Table A4: Simulation results for variance function estimation from 12-run designs[2]

GENERALIZED LEAST SQUARES

	$\rho = 0$									$\rho = 1$									$\rho = 10$								
	γ_1			γ_2			γ_3			γ_1			γ_2			γ_3			γ_1			γ_2			γ_3		
	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR
cv	1.308	0.195	0.624	0.651	1.046	0.821	0.695	0.747	0.890	1.308	0.195	0.624	0.651	1.046	0.821	0.695	0.747	0.890	1.190	0.266	0.634	0.579	0.888	0.720	0.514	0.641	0.751
rp	1.733	0.088	0.715	0.614	1.089	0.809	0.614	0.652	0.824	1.747	0.088	0.721	0.618	1.094	0.813	0.616	0.656	0.826	1.872	0.075	0.763	0.586	1.072	0.786	0.576	0.634	0.795
sw	1.605	0.132	0.699	0.584	1.090	0.790	0.603	0.676	0.821	1.741	0.096	0.726	0.625	1.103	0.820	0.608	0.658	0.821	1.872	0.075	0.763	0.586	1.072	0.786	0.576	0.634	0.795
cw	1.580	0.117	0.674	0.477	1.171	0.724	0.531	0.816	0.771	1.564	0.123	0.673	0.472	1.168	0.720	0.505	0.800	0.749	1.486	0.149	0.665	0.409	1.051	0.652	0.441	0.766	0.697

ADAPTED GENERALIZED LEAST SQUARES

	$\rho = 0$									$\rho = 1$									$\rho = 10$								
	γ_1			γ_2			γ_3			γ_1			γ_2			γ_3			γ_1			γ_2			γ_3		
	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR
cv	0.443	0.883	0.625	0.654	1.032	0.819	0.691	0.738	0.885	0.443	0.883	0.625	0.654	1.032	0.819	0.691	0.738	0.885	0.414	0.959	0.632	0.566	0.828	0.685	0.527	0.638	0.760
rp	0.528	1.000	0.727	0.629	1.469	0.816	0.646	0.994	0.829	0.532	1.017	0.735	0.635	1.483	0.814	0.649	0.993	0.831	0.524	1.137	0.755	0.635	1.471	0.818	0.714	1.031	0.817
sw	0.524	1.115	0.762	0.656	1.462	0.846	0.778	1.194	0.858	0.519	1.054	0.736	0.644	1.464	0.830	0.680	1.057	0.834	0.524	1.137	0.755	0.656	1.471	0.818	0.714	1.031	0.817
cw	0.489	1.026	0.707	0.729	0.837	0.788	0.724	0.547	0.867	0.484	1.042	0.708	0.734	0.840	0.791	0.728	0.858	0.872	0.420	0.990	0.648	0.718	0.741	0.729	0.614	0.460	0.770

RESTRICTED MAXIMUM LIKELIHOOD

	$\rho = 0$									$\rho = 1$									$\rho = 10$								
	γ_1			γ_2			γ_3			γ_1			γ_2			γ_3			γ_1			γ_2			γ_3		
	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR
cv	0.381	0.955	0.604	0.489	0.768	0.602	0.431	0.551	0.669	0.381	0.955	0.604	0.489	0.768	0.602	0.431	0.551	0.669	0.402	1.017	0.639	0.441	0.717	0.533	0.331	0.491	0.672
rp	0.624	0.769	0.691	0.572	1.013	0.759	0.531	0.607	0.755	0.640	0.769	0.691	0.536	1.013	0.759	0.531	0.607	0.755	0.596	0.616	0.670	0.586	0.834	0.694	0.502	0.420	0.704
sw	0.627	0.897	0.682	0.569	0.791	0.725	0.504	0.435	0.708	0.569	0.637	0.662	0.548	0.851	0.726	0.419	0.494	0.648	0.596	0.616	0.670	0.509	0.834	0.694	0.502	0.420	0.704
cw	0.684	2.939	2.442	0.544	1.248	0.784	0.527	0.848	0.767	0.577	0.785	0.668	0.535	1.232	0.777	0.540	0.854	0.777	1.237	10.268	9.604	0.610	1.166	0.818	0.557	0.583	0.714

LOG S^2

	$\rho = 0$									$\rho = 1$									$\rho = 10$								
	γ_1			γ_2			γ_3			γ_1			γ_2			γ_3			γ_1			γ_2			γ_3		
	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR
cv																											
rp	0.957	0.643	0.808	0.727	1.184	0.901	0.786	0.759	0.958	0.970	0.643	0.814	0.734	1.194	0.909	0.794	0.773	0.966	1.061	0.640	0.857	0.769	1.224	0.936	0.844	0.784	0.999
sw										0.994	0.641	0.825	0.732	1.199	0.908	0.799	0.749	0.963	1.061	0.640	0.857	0.769	1.224	0.936	0.844	0.784	0.999
cw	0.937	0.654	0.803	0.714	1.488	0.922	0.767	1.024	0.953	0.949	0.653	0.809	0.720	1.490	0.926	0.774	1.034	0.961	1.014	0.643	0.836	0.775	1.490	0.969	0.876	1.120	1.042

S^2

	$\rho = 0$									$\rho = 1$									$\rho = 10$								
	γ_1			γ_2			γ_3			γ_1			γ_2			γ_3			γ_1			γ_2			γ_3		
	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR	SQ DEV	AVG	VAR
cv																											
rp	0.957	0.643	0.808	0.727	1.184	0.901	0.786	0.759	0.958	0.970	0.643	0.814	0.734	1.194	0.909	0.794	0.773	0.966	1.061	0.640	0.857	0.769	1.224	0.936	0.844	0.784	0.999
sw										0.994	0.641	0.825	0.732	1.199	0.908	0.799	0.749	0.963	1.061	0.640	0.857	0.769	1.224	0.936	0.844	0.784	0.999
cw	0.937	0.654	0.803	0.714	1.488	0.922	0.767	1.024	0.953	0.949	0.653	0.809	0.720	1.490	0.926	0.774	1.034	0.961	1.014	0.643	0.836	0.775	1.490	0.969	0.876	1.120	1.042

[2]Average squared deviations of estimates from true parameter values (SQ DEV), averages estimates (AVG) and their variances (VAR) are shown. Computations were made for designs derived under constant prior (cv), under the right prior (rp), under a slightly wrong (sw) and under a completely wrong (cw) prior.

Table A5: Simulation results for mean function estimation from 54-run designs[3]

GENERALIZED LEAST SQUARES

	$\rho = 0$						$\rho = 1$						$\rho = 10$					
	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6
cv	0.3701	0.0986	0.0979	0.3522	0.2391	0.1134	0.5847	0.0935	0.0950	0.4244	0.2873	0.1100	4.0806	0.0893	0.0872	2.1935	1.7167	0.0935
rp	0.3658	0.1164	0.1135	0.2885	0.1841	0.1152	0.4980	0.1208	0.1199	0.2884	0.1981	0.1241	3.2439	0.1209	0.1199	1.2191	1.2214	0.1241
sw	0.4177	0.1205	0.1300	0.2773	0.1989	0.1339	0.5584	0.1208	0.1199	0.2884	0.2383	0.1241	3.2439	0.1209	0.1199	1.2191	1.2214	0.1241
cw	1.2874	0.1152	0.1070	0.8197	1.0420	0.1094	4.2179	0.1404	0.1406	1.9280	1.8756	0.1368	50.0972	0.1274	0.1306	19.7868	17.0493	0.1282

ADAPTED GENERALIZED LEAST SQUARES

	$\rho = 0$						$\rho = 1$						$\rho = 10$					
	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6
cv	0.3702	0.0986	0.0979	0.3522	0.2391	0.1134	0.5850	0.0935	0.0950	0.4246	0.2874	0.1100	4.0802	0.0893	0.0872	2.1943	1.7160	0.0935
rp	0.3655	0.1164	0.1137	0.2888	0.1840	0.1153	0.4982	0.1208	0.1199	0.2884	0.1981	0.1241	3.2438	0.1208	0.1199	1.2192	1.2212	0.1241
sw	0.4177	0.1205	0.1300	0.2773	0.1989	0.1338	0.5582	0.1208	0.1199	0.2884	0.2383	0.1240	3.2438	0.1208	0.1199	1.2192	1.2212	0.1241
cw	1.2870	0.1151	0.1070	0.8186	1.0425	0.1094	4.2155	0.1403	0.1405	1.9275	1.8747	0.1367	49.7459	0.1274	0.1306	19.6584	16.9993	0.1283

RESTRICTED MAXIMUM LIKELIHOOD

	$\rho = 0$						$\rho = 1$						$\rho = 10$					
	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6
cv	0.3706	0.0986	0.0982	0.3517	0.2401	0.1133	0.5870	0.0937	0.0949	0.4253	0.2887	0.1098	4.0697	0.0895	0.0872	2.1868	1.7183	0.0935
rp	0.3665	0.1163	0.1136	0.2889	0.1848	0.1150	0.4980	0.1210	0.1200	0.2889	0.1982	0.1241	3.1828	0.1209	0.1207	1.2071	1.1618	0.1235
sw	0.4128	0.1143	0.1207	0.2629	0.2065	0.1204	0.5084	0.1100	0.1103	0.2887	0.2104	0.1132	3.1828	0.1209	0.1207	1.2071	1.1618	0.1235
cw	1.3262	0.1097	0.0965	0.8150	1.0096	0.1076	4.2222	0.1406	0.1409	1.9340	1.8826	0.1371	50.0569	0.1266	0.1306	19.8281	17.1213	0.1277

LOG S^2

	$\rho = 0$						$\rho = 1$						$\rho = 10$					
	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6
cv	0.3726	0.0987	0.0979	0.3536	0.2401	0.1128	0.5867	0.0938	0.0951	0.4240	0.2884	0.1102	4.2105	0.0893	0.0871	2.2482	1.7341	0.0935
rp	0.3655	0.1166	0.1138	0.2902	0.1841	0.1152	0.4980	0.1208	0.1199	0.2884	0.1981	0.1241	3.2448	0.1208	0.1199	1.2194	1.2215	0.1240
sw	0.4183	0.1205	0.1300	0.2766	0.1993	0.1341	0.5582	0.1208	0.1199	0.2884	0.2383	0.1241	3.2448	0.1208	0.1199	1.2194	1.2215	0.1240
cw	1.2929	0.1153	0.1072	0.8252	1.0410	0.1096	4.2166	0.1404	0.1405	1.9278	1.8756	0.1368	48.8438	0.1233	0.1266	19.3137	16.5177	0.1241

S^2

	$\rho = 0$						$\rho = 1$						$\rho = 10$					
	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6
cv	0.3721	0.0988	0.0981	0.3518	0.2412	0.1130	0.5885	0.0937	0.0950	0.4253	0.2896	0.1100	4.1380	0.0895	0.0870	2.2171	1.7254	0.0934
rp	0.3663	0.1164	0.1137	0.2896	0.1847	0.1152	0.4981	0.1210	0.1200	0.2889	0.1981	0.1241	3.2449	0.1210	0.1199	1.2205	1.2213	0.1240
sw	0.4183	0.1206	0.1301	0.2774	0.2005	0.1341	0.5583	0.1210	0.1199	0.2888	0.2383	0.1240	3.2449	0.1210	0.1199	1.2205	1.2213	0.1240
cw	1.2950	0.1153	0.1071	0.8185	1.0501	0.1095	4.2187	0.1404	0.1405	1.9300	1.8799	0.1368	50.0354	0.1276	0.1307	19.8403	17.0830	0.1280

[3]Average squared deviations of estimates from true parameter values are shown. Computations were made for designs derived under constant prior (cv), under the right prior (rp), under a slightly wrong (sw) and under a completely wrong (cw) prior.

Table A6: Simulation results for mean function estimation from 12-run designs[4]

GENERALIZED LEAST SQUARES

	$\rho = 0$						$\rho = 1$						$\rho = 10$					
	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6
cv	3.6144	0.4595	0.4645	2.2744	1.9862	0.6138	3.6144	0.4595	0.4645	2.2744	1.9862	0.6138	34.5865	0.4437	0.4031	2.3737	12.8807	0.5812
rp	1.9198	0.5969	0.6007	1.2452	0.7236	0.6030	1.9650	0.5990	0.5973	1.2079	0.7182	0.6017	11.6705	0.5830	0.5857	4.3372	4.5310	0.6190
sw	2.2422	0.6166	0.7210	1.1840	0.9453	0.7297	2.2153	0.6110	0.5980	1.2043	0.8775	0.6068	11.6705	0.5830	0.5857	4.3372	4.5310	0.6190
cw	12.5709	0.6281	0.6010	5.5177	6.3776	0.6090	15.1928	0.6247	0.5993	6.9969	6.8364	0.6106	111.8051	0.4795	0.5147	62.3049	27.3827	0.4973

ADAPTED GENERALIZED LEAST SQUARES

	$\rho = 0$						$\rho = 1$						$\rho = 10$					
	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6
cv	3.7052	0.4625	0.4608	2.3236	2.0332	0.6053	3.7052	0.4625	0.4608	2.3236	2.0332	0.6053	35.0041	0.4453	0.3955	22.1375	13.1093	0.5720
rp	1.9202	0.5966	0.5991	1.2426	0.7207	0.6046	1.9752	0.6055	0.5980	1.2106	0.7169	0.6056	12.7084	0.6135	0.5927	5.0144	4.4533	0.6149
sw	2.1139	0.6146	0.6810	1.2264	0.8885	0.6860	2.2098	0.6095	0.5964	1.2045	0.8733	0.6092	12.7084	0.6135	0.5927	5.0144	4.4533	0.6149
cw	12.4741	0.6232	0.5695	5.4462	6.3896	0.5820	14.7438	0.6232	0.5707	6.7240	6.7814	0.5841	119.7037	0.4782	0.4919	68.6952	28.8027	0.4943

RESTRICTED MAXIMUM LIKELIHOOD

	$\rho = 0$						$\rho = 1$						$\rho = 10$					
	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6
cv	3.7949	0.4501	0.4331	2.3871	1.9361	0.5825	3.7949	0.4501	0.4331	2.3871	1.9361	0.5825	31.6244	0.4036	0.4043	19.7007	13.2561	0.5209
rp	1.7836	0.6025	0.5487	1.2030	0.7897	0.5795	2.0269	0.6075	0.5899	1.2287	0.7202	0.5791	13.4757	0.6247	0.5629	5.0459	4.9826	0.6162
sw	2.0201	0.5679	0.6486	1.2126	0.8826	0.6489	2.1812	0.5631	0.6026	1.0955	0.9859	0.6123	13.4757	0.6247	0.5629	5.0459	4.9826	0.6162
cw	11.9520	0.6145	0.5850	5.4433	6.0919	0.5704	14.4362	0.6135	0.5740	6.9257	6.5178	0.5677	114.5017	0.6259	0.5267	67.6699	28.7860	0.4976

LOG S^2

	$\rho = 0$						$\rho = 1$						$\rho = 10$					
	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6
cv																		
rp	1.8261	0.5675	0.5335	1.2208	0.7286	0.5974	1.9608	0.5741	0.5431	1.2072	0.7226	0.6063	11.7216	0.5910	0.6376	4.9042	4.2115	0.6005
sw							2.0719	0.5652	0.5346	1.2078	0.8544	0.5939	11.7267	0.6145	0.6187	4.9122	4.2345	0.5983
cw	12.1473	0.6028	0.6465	5.5729	5.8896	0.6509	14.3190	0.5770	0.6132	7.0297	6.1786	0.6615	123.2922	0.5338	0.9803	67.6566	28.9695	0.5780

S^2

	$\rho = 0$						$\rho = 1$						$\rho = 10$					
	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6	β_1	β_2	β_3	β_4	β_5	β_6
cv																		
rp	1.8261	0.5675	0.5335	1.2208	0.7286	0.5974	1.9608	0.5741	0.5431	1.2072	0.7226	0.6063	11.7216	0.5910	0.6376	4.9042	4.2115	0.6005
sw							2.0719	0.5652	0.5346	1.2078	0.8544	0.5939	11.7267	0.6145	0.6187	4.9122	4.2345	0.5983
cw	12.1473	0.6028	0.6465	5.5729	5.8896	0.6509	14.3190	0.5770	0.6132	7.0297	6.1786	0.6615	123.2922	0.5338	0.9803	67.6566	28.9695	0.5780

[4] Average squared deviations of estimates from true parameter values are shown. Computations were made for designs derived under constant prior (cv), under the right prior (rp), under a slightly wrong (sw) and under a completely wrong (cw) prior.

References

Atkinson,A.C. and Cook, R.D. (1995), "D-Optimum Designs for Heteroscedastic Linear Models," *Journal of the American Statistical Association*, 90, 204–212.

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

Carroll, R.J. 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.

Harville, D. (1977), "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems," *Journal of the American Statistical Association*, 79, 302–308.

Jobson, J.D. and Fuller, W.A. (1980), "Least Squares Estimation When the Covariance Matrix and Parameter Vector Are Funtionally Related," *Journal of the American Statistical Association*, 75, 176–181.

Kiefer, J. and Wolfowitz, J. (1959), "Optimum Designs in Regression Problems," *Annals of Mathematical Statistics*, 30, 271–294.

Mays, D.P. and Easter, S.M. (1997), "Optimal Response Surface Designs in the Presence of Dispersion Effects," *Journal of Quality Technology*, 29, 59–70.

Patterson, H.D. and Thompson, R. (1971), "Recovery of Inter-block Information When Block Sizes Are Unequal," *Biometrika*, 5, 545–554.

Raab, G.M. (1981), "Estimation of a Variance Function, with Application to Immunoassay," *Applied Statistics*, 30, 32–40.

Vining, G. and Schaub, D. (1996) "Experimental Designs for Estimating Both Mean and Variance Functions," *Journal of Quality Technology*, 28, 135–147.