KATHOLIEKE
UNIVERSITEIT
LEUVEN

DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

RESEARCH REPORT 9925

EFFICIENT D-OPTIMAL DESIGNS UNDER MULTIPLICATIVE HETEROSCEDASTICITY

by
L. TACK
P. GOOS
M. VANDEBROEK

D/1999/2376/25

Efficient $\mathcal{D}$-optimal Designs under Multiplicative Heteroscedasticity

L. Tack
P. Goos
M. Vandebroek
Katholieke Universiteit Leuven, Belgium

In optimum design theory designs are constructed that maximize the information on the unknown parameters of the response function. The major part deals with designs optimal for response function estimation under the assumption of homoscedasticity. In this paper, optimal designs are derived in case of multiplicative heteroscedasticity for either response function estimation or response and variance function estimation by using a Bayesian approach. The efficiencies of Bayesian designs derived with various priors are compared to those of the classic designs with respect to various variance functions. The results show that any prior knowledge about the sign of the variance function parameters leads to designs that are considerably more efficient than the classic ones based on homoscedastic assumptions.

1 Introduction

Delivering reliable, high quality products and processes at low cost has become the key to survival in today's global economy. Driven by the need to compete on cost and performance, many quality-conscious organizations are increasingly focusing on the optimization of product design. This reflects the realization that quality can not be achieved economically through inspection. Designing in quality is cheaper than trying to inspect and re-engineer it after the product hits the production floor, or worse, after it gets to the customer. Experimental design is a powerful quality improvement technique.

Atkinson (1996) gives a selection of the existing literature on experimental design theory. The origins of mathematical work on the design of experiments go back at least to Smith (1918) who found optimal designs for a series of single-factor polynomial models. Fisher (1960) gives a mathematical treatment of the determination of designs for some nonlinear models. The first extended presentation of the ideas of optimum experimental design was in Kiefer (1959). A brief history of statistical work on optimum experimental design is given by Wynn (1985) and the subject continues to develop, recently at an increasing rate. The two most cited books in the field are from Fedorov (1972) and from Silvey (1980).

The recent book by Pukelsheim (1993) emphasizes connections with convex programming, whereas that of Atkinson and Donev (1992) stresses more statistical aspects. One major strand of development since Kiefer (1959) includes the various equivalence theorems for optimal designs. Kiefer and Wolfowitz (1960) published the equivalence theorem between $\mathcal{G}$- and $\mathcal{D}$-optimality the following year. Another is the construction of algorithms for numerical calculation of both exact and continuous designs.

Since the late 1950s Taguchi (Taguchi, 1986 and Taguchi, 1987) has introduced several new statistical tools and concepts of quality improvement that depend heavily on statistical theory for design of experiments. He 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. This so-called robust design is a method for making a manufacturing process less sensitive to manufacturing variations and is extensively described by Huele (1998). Because it reduces variation by limiting the influence of sources of variation, not by controlling them, robust design is a cost-efficient technique for improving process quality. Only recently, industrial statisticians became aware of the fact that they can no longer concern themselves only with the expected value of the response of interest. Instead they must also consider the response variability. An interesting approach assumes that the response variance, or a suitable transformation thereof, may be well approximated by a linear model in the independent variables. Separate linear models for the response and variance structure are estimated. Separate models have the advantage of providing the analyst with a better scientific understanding of the total process. The analyst is better able to see what levels of the independent factors will lead to acceptable response values as well as acceptable variability.

The essence of robust design is to reduce variation of product's or process's functional characteristics. These functional characteristics are basic, measurable quantities that determine how well the final product functions. According to a manufacturing process, the functional characteristics are usually measurements that can be made on the incomplete product soon after each specific step.
There are three types of variables that affect functional characteristics: control parameters, signal factors and sources of noise. Control parameters are the controllable process variables of which the operating standards can be specified by the process engineers. Signal factors are those factors whose levels are set by the user or operator to attain the target performance or to express the intended output. For example, the steering angle is a signal factor for the steering mechanism of an automobile. In contrast, sources of noise are the variables that are impossible or expensive to control. They, in turn, cause variations in the product's functional characteristics. External noise factors are those factors external to the product such as load conditions, temperature, humidity, dust, supply voltage, vibrations due to nearby machinery, human errors in operating the product, etc. Noise can also be induced by manufacturing imperfection. This causes product parameter variation from unit to unit and is inevitable in a manufacturing environment. A third type of noise is called deterioration and refers to the fact that the values of performance characteristics

of a product sold may change as time passes by. In practice it is often possible to take advantage of certain control factors, which can be adjusted, to achieve the desired functional relationship between the signal factors and the response. One such commonly encountered factor is a scaling factor. The gearing ratio in the steering mechanism of a car is an example of a scaling factor. The objective of robust design is to find those control parameter settings where noise has a minimal effect on the functional characteristics. The key idea is to reduce functional characteristic sensitivity by making the process insensitive to noise rather than by controlling the sources of noise. To attain this objective, the control parameters are systematically varied in an experiment and the effect of noise is measured for each experimental run. Finally, the results are used to predict which control parameter settings will make the process insensitive to noise. As an example, Kackar and Shoemaker (1986) applied these principles to the improvement of the process for making optical fibers.

According to the Taguchi philosophy, the three major steps in designing a quality product are system design, parameter design and tolerance design. System design is the process of applying scientific and engineering knowledge to produce a basic functional prototype design. The next step, parameter design, is an investigation conducted to identify the settings of design parameters that optimize the performance characteristics and reduce the sensitivity of the engineering design to the sources of variation. Four operational steps complete this robust design method. In the first step, functional characteristics, control parameters, signal factors and sources of noise are determined. In the second step the experiment is planned by determining the levels of the control parameters and the signal factors. Conducting the experiment is done in the third step and the obtained results are used to predict the improved control parameter settings by minimizing the variance. To achieve this, the robust design method uses a statistical measure of performance called signal-to-noise (S/N) ratio. This S/N-ratio has been criticized for being a performance statistic of ambiguous nature that may deliver unreliable results. Some authors argue that it is better to separate performance statistics for mean and variance, since sometimes variability changes as a function of the mean value, while for other response variables, average and variability appear to be independent. Often one runs a confirmation experiment in the fourth step to check the prediction. Afterwards, the process of determining tolerances around the nominal settings identified in the parameter design process, is called tolerance design and is only required if robust design can not produce the required performance without costly special components or high process accuracy. Unal and Daen (1991) apply these steps to the optimization process of a heat exchanger. Phadke and Dehnad (1987) developed a two stage approach whereby the signal-to-noise ratio is maximized at first and then the performance is brought on target by special adjustment parameters.

This paper will focus on step two of the parameter design stage: planning the experiment. Most literature on optimal designs assumes errors with constant variance. The case of multiplicative heteroscedasticity for either response function estimation or response and variance function estimation is considered here. The resulting designs depend on the

unknown parameters of the variance function and Bayesian experimental design theory provides design criteria reflecting prior knowledge of these parameters. The purpose of this paper is to give a more thorough analysis of the impact of the prior information on the efficiencies of Bayesian designs. Various priors are used and the efficiencies of the resulting designs are compared to the classic ones, based on homoscedastic assumptions, with respect to various variance functions.

2 Bayesian $\mathcal{D}$-optimal designs

The section starts with the formulation of the assumed heteroscedastic model for both response and variance function estimation. Henceforth y denotes the response of interest and $\mathbf{x}$ and $\mathbf{z}$ are the $(p \times 1)$ and $(q \times 1)$ vectors of control variables presumed to influence the response function 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. Both $\mathbf{f}(\mathbf{x})$ and $\mathbf{g}(\mathbf{z})$ contain an intercept. With $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ the $(p_r \times 1)$ and $(q_v \times 1)$ vectors of unknown parameters, the following heteroscedastic model is assumed:

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

Note that $v > 0$ is a known twice-continuously differentiable function. The error term ε is standardized such that $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{VAR}(\varepsilon) = 1$ which yields the following mean and variance functions:

$$\begin{aligned} \mathrm{E}(y) &= \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}, & (2)\\ \mathrm{VAR}(y) &= v[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]. & (3) \end{aligned}$$

The optimum design problem is concerned with the selection of the number of observations at the different design points at which to observe the response y in order to provide maximal information on $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ or some subset thereof. One of the many optimality criteria is the widely spread $\mathcal{D}$-optimality criterion in which designs with the greatest determinant of the total information matrix for the parameters will be preferred. With α the significance level, the $100(1-\alpha)\%$ confidence region for all elements of $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$ is an ellipsoid in the p_r or q_v dimensional space respectively, of which the volume is inversely proportional to the square root of the determinant of the information matrix. In consequence, the $\mathcal{D}$-optimality criterion minimizes the volume of the ellipsoid representing the confidence region. The per observation information matrix on $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ equals

$$\mathbf{I}(\mathbf{x}, \mathbf{z}) = -\mathrm{E}\begin{bmatrix} \frac{\partial^2 \log \mathbf{L}}{\partial \boldsymbol{\beta}^2} & \frac{\partial^2 \log \mathbf{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}} \\ \frac{\partial^2 \log \mathbf{L}}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}} & \frac{\partial^2 \log \mathbf{L}}{\partial \boldsymbol{\gamma}^2} \end{bmatrix}, \tag{4}$$

where $\mathbf{L}$ is the likelihood function, written as

$$\mathbf{L}(y, \mathbf{x}, \mathbf{z} | \boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{\sqrt{v[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]}\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \frac{[y - \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}]^2}{v[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]} \right\}. \tag{5}$$

Substituting the expression of the likehood function (5) into the expression of the per observation matrix (4) yields

$$\mathbf{I}(\mathbf{x},\mathbf{z}) = \begin{bmatrix} \frac{\mathbf{f}(\mathbf{x})\mathbf{f}^T(\mathbf{x})}{v[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]} & 0 \\ 0 & \frac{1}{2}\left\{\frac{v'[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]}{v[\mathbf{g}^T(\mathbf{z})\boldsymbol{\gamma}]}\right\}^2 \mathbf{g}(\mathbf{z})\mathbf{g}^T(\mathbf{z}) \end{bmatrix}, \tag{6}$$

where v' indicates the first derivative of v with respect to $\boldsymbol{\gamma}$. After summation of the per observation matrices (6) over all N observations, the total information in the design $\{\mathbf{x},\mathbf{z}\}_{i=1}^N$ equals

$$\sum_{i=1}^N \mathbf{I}(\mathbf{x}_i,\mathbf{z}_i) = \begin{bmatrix} \sum_{i=1}^N \frac{\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)}{v[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]} & 0 \\ 0 & \sum_{i=1}^N \frac{1}{2}\left\{\frac{v'[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]}{v[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]}\right\}^2 \mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i) \end{bmatrix}. \tag{7}$$

The determinant of the information matrix (7) becomes

$$\left|\sum_{i=1}^N \mathbf{I}(\mathbf{x}_i,\mathbf{z}_i)\right| = \left|\sum_{i=1}^N \frac{\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)}{v[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]}\right| \times \frac{1}{2}\left|\sum_{i=1}^N \left\{\frac{v'[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]}{v[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]}\right\}^2 \mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i)\right|. \tag{8}$$

Considering the case of *multiplicative heteroscedasticity* by replacing function v by the exponential function, expression (8) simplifies to

$$\left|\sum_{i=1}^N \mathbf{I}(\mathbf{x}_i,\mathbf{z}_i)\right| = \left|\sum_{i=1}^N \frac{\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)}{\exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}]}\right| \times \frac{1}{2}\left|\sum_{i=1}^N \mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i)\right|. \tag{9}$$

Unfortunately, the determinant of the information matrix depends on the unknown vector $\boldsymbol{\gamma}$. Aitkin (1987) avoids this problem by using the previous collected data to estimate $\boldsymbol{\gamma}$.

Another possibility is a Bayesian approach to design optimality. An extensive review of developments in Bayesian design theory can be found in Chaloner and Verdinelli (1995) and Broemeling (1985). Chaloner (1984) derived Bayesian design optimality criteria for the linear model when the parameters of the mean function are supposed to follow a normal distribution. Chaloner and Larntz (1989) extended these findings to nonlinear models. DuMouchel and Jones (1994) proposed Bayesian $\mathcal{D}$-optimal designs with reduced dependence on regressor specification and Atkinson and Cook (1995) supposed a discrete prior distribution for $\boldsymbol{\gamma}$. Vining and Schaub (1996) made use of a multivariate normal distribution $N(\boldsymbol{\gamma}_0, \rho I)$ as a prior for $\boldsymbol{\gamma}$ and proposed a semi-Bayesian approach in that they maximized the determinant of the expected information matrix. Vandebroek and Goos (1997) studied the impact of the mean $\boldsymbol{\gamma}_0$ of the prior distribution and of the uncertainty ρ about this mean for that approach. The approach of Vining and Schaub (1996) is susceptible to discussion in that they used a semi-Bayesian approach by maximizing the determinant of the expected information matrix rather than using the correct criterion of maximization of the expected determinant of the information matrix. Furthermore, they omitted the parameter ρ in the determinant of the expected information matrix

for no obvious reason. Maximization of the expected determinant of the information matrix is elaborated in our paper. It is worth mentioning that both approaches give rather different results. In the remainder of this exposition a discrete approximation of the multivariate normal distribution $N(\boldsymbol{\gamma}_0, \rho I)$ is used as prior distribution for $\boldsymbol{\gamma}$. This discrete approximation contains g support points $\boldsymbol{\gamma}_{0j}$ and corresponding probabilities $p_\rho(\boldsymbol{\gamma}_{0j})$, with $j \in \{1, \ldots, g\}$. The expected determinant of the information matrix (9) then becomes equal to

$$\sum_{j=1}^{g} p_\rho(\boldsymbol{\gamma}_{0j}) \times \left| \sum_{i=1}^{N} \frac{\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)}{\exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}_{0j}]} \right| \times \frac{1}{2} \left| \sum_{i=1}^{N} \mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i) \right| . \tag{10}$$

When the design is built up with l different design points i and design point i is replicated n_i times, the expected determinant in expression (10) can be rewritten as

$$\sum_{j=1}^{g} p_\rho(\boldsymbol{\gamma}_{0j}) \times \left| \sum_{i=1}^{l} \frac{\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)n_i}{\exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}_{0j}]} \right| \times \frac{1}{2} \left| \sum_{i=1}^{l} \mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i)n_i \right| . \tag{11}$$

The computation of Bayesian $\mathcal{D}$-optimal designs for both response and variance function estimation is based on maximization of expression (11). When only the response function is of interest, the computation is based on maximization of

$$\sum_{j=1}^{g} p_\rho(\boldsymbol{\gamma}_{0j}) \times \left| \sum_{i=1}^{l} \frac{\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)n_i}{\exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}_{0j}]} \right| . \tag{12}$$

Note that $\sum_{i=1}^{l} n_i = N$. The degree of uncertainty attached to the prior is expressed by parameter ρ with increasing ρ indicating higher uncertainty. Each value of ρ yields another discrete prior distribution for $\boldsymbol{\gamma}$ with probabilities $p_\rho(\boldsymbol{\gamma}_{0j})$ for $j \in \{1, \ldots, g\}$. Imposing the condition $\rho = 0$ is nothing else than stating the parameter vector $\boldsymbol{\gamma}$ is known in advance with certainty. High values of ρ correspond to an almost uniform discrete prior distribution for the parameter vector $\boldsymbol{\gamma}$.
The classic homoscedastic approach corresponds to choosing the mean $\boldsymbol{\gamma}_0$ of the prior distribution equal to $(1\ 0\ \cdots\ 0)^T$ and $\rho = 0$. The experiment is then planned assuming a constant variance over the region of interest, as can be seen from expression (3).

In the next sections, $\mathcal{D}$-optimal designs will be computed for different prior distributions for $\boldsymbol{\gamma}$ and the efficiencies of the obtained designs will be compared with those of classic designs, in order to stipulate whether any prior information leads to considerably more efficient designs.

3 Efficiencies under multiplicative heteroscedasticity

The degree to which the experimental goals are achieved with the number of observations allotted is measured by the efficiency. The $\mathcal{D}$-efficiency of an arbitrary design, represented

by a measure ξ referring to the chosen design points and their corresponding number of replicates, is defined in Atkinson and Donev (1992) as

$$D_{\text{eff}} = \left(\frac{|M(\xi)|}{|M(\xi^*)|} \right)^{\frac{1}{p}}, \tag{13}$$

where M denotes the information matrix, ξ^* stands for a measure referring to the $\mathcal{D}$-optimal design and p is the number of parameters in the model. Taking the ratio of the determinants in (13) to the power $1/p$ results in an efficiency measure which is proportional to design size, irrespective of the dimension of the model. This means for instance that two replicates of a design measure for which $D_{\text{eff}} = \frac{1}{2}$ would be as efficient as one replicate of the optimum measure.

Denote the number of replications of design point i for the $\mathcal{D}$-optimal design under heteroscedasticity and for the $\mathcal{D}$-optimal design in case of homoscedasticity respectively as $n_i(\boldsymbol{\gamma_0}, \rho)$ and n_i^c, where c refers to the constant variance. Remark that for the Bayesian design the numbers of observations depend on the prior mean $\boldsymbol{\gamma_0}$ and the degree of uncertainty ρ. For sake of briefness, from this time on, the dependence of n_i on $\boldsymbol{\gamma_0}$ and ρ will no longer be rendered.

According to expression (11), the computation of classic homoscedastic $\mathcal{D}$-optimal designs for both response and variance function estimation is based on maximization of

$$\left| \sum_{i=1}^{l} \frac{\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i) n_i^c}{\exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma_0}]} \right| \times \frac{1}{2} \left| \sum_{i=1}^{l} \mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i) n_i^c \right| \tag{14}$$

and when attention is restricted to only response function estimation, the computation is based on maximization of

$$\left| \sum_{i=1}^{l} \frac{\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i) n_i^c}{\exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma_0}]} \right|. \tag{15}$$

The efficiency measure of the computed Bayesian $\mathcal{D}$-optimal design for a prior distribution of the parameter vector $\boldsymbol{\gamma}$ relative to the classic design and for a particular variance function vector $\boldsymbol{\gamma}_*$ equals

$$\left(\frac{\left| \sum_{i=1}^{l} \frac{\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i) n_i}{\exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}_*]} \right| \times \frac{1}{2} \left| \sum_{i=1}^{l} \mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i) n_i \right|}{\left| \sum_{i=1}^{l} \frac{\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i) n_i^c}{\exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}_*]} \right| \times \frac{1}{2} \left| \sum_{i=1}^{l} \mathbf{g}(\mathbf{z}_i)\mathbf{g}^T(\mathbf{z}_i) n_i^c \right|} \right)^{\frac{1}{p_r + q_v}}. \tag{16}$$

Expression (16) measures how well the Bayesian $\mathcal{D}$-optimal design derived with prior mean $\boldsymbol{\gamma_0}$ and degree of uncertainty ρ (numerator) performs relative to the classic design (denominator), for a particular variance function vector $\boldsymbol{\gamma}_*$ that is unknown in practice and assessed by the discrete distribution with prior mean $\boldsymbol{\gamma_0}$.

When we confine ourselves to response function estimation, the efficiency measure (16) becomes equal to

$$\left(\frac{\left| \sum_{i=1}^{l} \frac{\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i) n_i}{\exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}_*]} \right|}{\left| \sum_{i=1}^{l} \frac{\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i) n_i^c}{\exp[\mathbf{g}^T(\mathbf{z}_i)\boldsymbol{\gamma}_*]} \right|} \right)^{\frac{1}{p_r}}. \tag{17}$$

This paper deals with the question whether any prior knowledge about the variance function leads to considerably more efficient designs compared to the classic ones where constant variance is assumed. Henceforth, for reasons of simplification, $\boldsymbol{x} = \boldsymbol{z}$ is supposed. In other words, the factors possibly influencing the mean and variance function are the same. Besides, only two factor cases will be reported of which the design region is of a discrete form and restricted to a 3×3 grid on $\chi = [-1, 1]^2$. The adopted mean function is the full-second order polynomial defined by

$$\mathbf{f}^T(\mathbf{x}) = (1 \; x_1 \; x_2 \; x_1^2 \; x_2^2 \; x_1 x_2). \tag{18}$$

This paper both deals with first order variance functions without interaction and first order variance functions augmented with an interaction term $x_1 x_2$. The corresponding polynomials are respectively defined as

$$\mathbf{g}^T(\mathbf{x}) = (1 \; x_1 \; x_2), \tag{19}$$

$$\mathbf{g}^T(\mathbf{x}) = (1 \; x_1 \; x_2 \; x_1 x_2). \tag{20}$$

$\mathcal{D}$-optimal designs will be computed for different prior distributions for $\boldsymbol{\gamma}$ with prior mean $\boldsymbol{\gamma_0}$ and degree of uncertainty ρ. Special attention will be given to the case where constant variance is supposed and no possible misspecification ($\rho = 0$) of the variance function parameters is taken into account. The vector representations for the parameters of the first order variance function and the parameters of the first order variance function with interaction effect are respectively:

$$\boldsymbol{\gamma}^T = (\; 1 \quad \gamma^{(1)} \quad \gamma^{(2)} \;), \tag{21}$$

$$\boldsymbol{\gamma}^T = (\; 1 \quad \gamma^{(1)} \quad \gamma^{(2)} \quad \gamma^{(3)} \;). \tag{22}$$

The prior means are represented as $\boldsymbol{\gamma_0}^T = (1 \; \gamma_0^{(1)} \; \gamma_0^{(2)})$ or $\boldsymbol{\gamma_0}^T = (1 \; \gamma_0^{(1)} \; \gamma_0^{(2)} \; \gamma_0^{(3)})$ depending on whether an interaction term is present or not.

The $\mathcal{D}$-optimal designs are calculated for various values of ρ and the efficiency is measured relative to the classic $\mathcal{D}$-optimal design, for particular variance function vectors $\boldsymbol{\gamma_*}$ from which the components $\gamma_*^{(1)}$, $\gamma_*^{(2)}$ and $\gamma_*^{(3)}$ vary between -3 and +3. As already said, increasing ρ means introducing larger variance in the multivariate distribution $N(\boldsymbol{\gamma_0}, \rho I)$ and corresponds to increased uncertainty about the prior mean. Suppose the discrete prior distribution for a variance function without interaction contains $g = w^2$ support points $\boldsymbol{\gamma_{0j}}$ with $j \in \{1, \ldots, w^2\}$ on a square $w \times w$ grid. However, the discrete prior distribution for a variance function with interaction contains $g = w^3$ support points $\boldsymbol{\gamma_{0j}}$ with $j \in \{1, \ldots, w^3\}$ on a cubic $w \times w \times w$ grid. Furthermore, w is supposed to be an odd number such that $j = \lceil \frac{g}{2} \rceil$ corresponds with prior mean $\boldsymbol{\gamma_0}$. The discrete approximation

to the multivariate normal density function can be written as

$$f(\boldsymbol{\gamma}_{0j};\boldsymbol{\gamma}_0,\rho I)=\frac{1}{(2\pi\rho)^{\frac{q_v}{2}}\exp\left[\frac{1}{2\rho}\sum_{k=0}^{q_v-1}\left(\gamma_{0j}^{(k)}-\gamma_0^{(k)}\right)^2\right]}, \tag{23}$$

where $\gamma_0^{(k)}$ is the $(k+1)$th component of prior mean $\boldsymbol{\gamma}_0$ and $\gamma_{0j}^{(k)}$ is the $(k+1)$th component of $\boldsymbol{\gamma}_{0j}$. In accordance with the previous notations, $\gamma_0^{(0)}=1$ and $\gamma_{0j}^{(0)}=1$ for all j. The probability $p_\rho(\boldsymbol{\gamma}_{0s})$ attached to support point s, with $s\in\{1,\ldots,g\}$, can be calculated as

$$p_\rho(\boldsymbol{\gamma}_{0s})=\frac{f(\boldsymbol{\gamma}_{0s};\boldsymbol{\gamma}_0,\rho I)}{\sum_{j=1}^{g}f(\boldsymbol{\gamma}_{0j};\boldsymbol{\gamma}_0,\rho I)}, \tag{24}$$

taking into account that $\sum_{j=1}^{g}p_\rho(\boldsymbol{\gamma}_{0j})$ must add up to one. Simplifying and rewriting the above expression for $\boldsymbol{\gamma}_{0s}$ equal to the prior mean $\boldsymbol{\gamma}_0$ gives

$$p_\rho(\boldsymbol{\gamma}_0)=\frac{1}{\sum_{j=1}^{g}\left\{\exp\left[d^2\left(\boldsymbol{\gamma}_{0j},\boldsymbol{\gamma}_0\right)\right]\right\}^{-\frac{1}{2\rho}}}, \tag{25}$$

where $d(\boldsymbol{\gamma}_{0j},\boldsymbol{\gamma}_0)$ denotes the Euclidean distance between vector $\boldsymbol{\gamma}_{0j}$ and prior mean $\boldsymbol{\gamma}_0$. Because ρ is a rather theoretical concept and hard to interpret, equation (25) enables the industrial statistician to translate the more practical and meaningful probability $p_\rho(\boldsymbol{\gamma}_0)$ to the degree of uncertainty ρ. Note that in industrial environments $p_\rho(\boldsymbol{\gamma}_0)$ often can be estimated by former experience. Relationship (25) is shown in Table 1 and Figure 1.

ρ	$p_\rho(\boldsymbol{\gamma}_0)$	ρ	$p_\rho(\boldsymbol{\gamma}_0)$	ρ	$p_\rho(\boldsymbol{\gamma}_0)$	ρ	$p_\rho(\boldsymbol{\gamma}_0)$
0.000	1.00	0.030	0.48	0.1	0.143	1	0.029
0.010	0.96	0.035	0.41	0.2	0.074	2	0.024
0.015	0.83	0.040	0.36	0.3	0.053	3	0.023
0.020	0.68	0.045	0.32	0.4	0.044	4	0.022
0.025	0.56	0.050	0.29	0.5	0.038	∞	0.020

Table 1: Relation Between ρ and $p_\rho(\boldsymbol{\gamma}_0)$

Note that the case $\rho=\infty$ corresponds to a uniform prior distribution with $p_\rho(\boldsymbol{\gamma}_0)=\frac{1}{g}$.

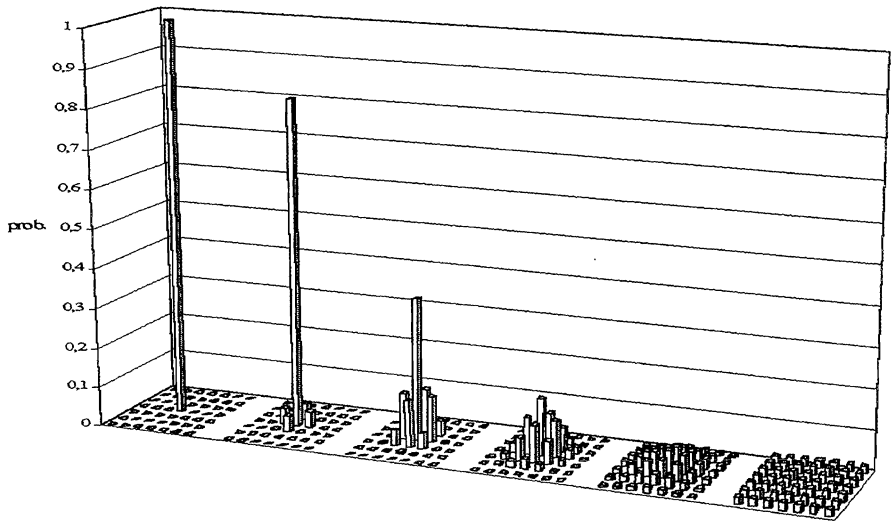


Figure 1: Discrete Prior Distributions for $\rho \in \{0; 0.015; 0.04; 0.1; 0.3; 3\}$

4 Discussion of results

This section deals with the calculation of $\mathcal{D}$-optimal designs for various prior distributions which are compared with the classic designs that assume homoscedasticity. The $\mathcal{D}$-optimal designs are presented for $N = 36$ observations but the results can be generalized to other numbers of observations. Different values for w were investigated but the results do not change significantly with varying w.

4.1 First order variance functions without interaction

In this section our attention is restricted to first order variance functions with the polynomial expansion given in expression (19). Several degrees of uncertainty ρ increasing from 0 to 4 were investigated. Only those degrees of uncertainty by which changes in the computed $\mathcal{D}$-optimal design arised will be discussed. By this, five degrees of uncertainty about the mean of the prior distribution are mentioned, namely $\rho \in \{0; 0.3; 0.5; 1; 3\}$.
The classic $\mathcal{D}$-optimal designs assume constant variance ($\gamma_0^{(1)} = \gamma_0^{(2)} = 0$ in (21)) and do not take into account any misspecification ($\rho = 0$). For response function estimation, the classic $\mathcal{D}$-optimal designs are computed according to expression (15) and for response and variance function estimation, they are computed according to expression (14). These classic $\mathcal{D}$-optimal designs are displayed in Figure 1.1 of Appendix 1. The numbers indicate the number of replications of the different design points in the $\mathcal{D}$-optimal design. The

horizontal axis refers to the factor levels of x_1, whereas the vertical axis refers to the factor levels of x_2.

Different prior means were investigated but only a few examples are taken in for discussion (Table 2).

Prior	$\gamma_0^{(1)}$	$\gamma_0^{(2)}$	$\gamma_0^{(3)}$
Mean [1]	0.400	0.230	-
Mean [2]	0.033	0.033	-
Mean [3]	0.400	0.230	0.200

Table 2: Prior Means

Suppose that due to prior knowledge, one assumes that the variance function vector $\boldsymbol{\gamma}$ has mean $\boldsymbol{\gamma_0} = (1\ 0.400\ 0.230)^T$ (mean [1] in Table 2). The corresponding $\mathcal{D}$-optimal designs for response function estimation and for response and variance function estimation are shown in Figure 1.2. These designs are computed by maximizing expression (12) and (11) respectively. The $\mathcal{D}$-optimal designs for response function estimation and for response and variance function estimation depend on the uncertainty attached to the prior mean. For increased ρ the observations have the tendency to move from the corner points or the centre to the midpoints of the outer sides of the design region. From these figures it can be seen that when also the variance function is of interest, the observations tend to shift towards the cornerpoints of the design region. By this, more information for variance function estimation is obtained because a first order variance function is assumed. The same graphical insight can be observed for other prior means. The efficiency plots in Appendix 2 display the efficiency ratios for response function estimation and for response and variance function estimation for various vectors $\boldsymbol{\gamma}_*$ with components belonging to [-3,3], according to expression (17) and expression (16) respectively. Figure 2.1a shows that the Bayesian $\mathcal{D}$-optimal design for prior mean [1] is more efficient than the classic one (efficiency ratio (17) larger than one) for variance function vectors $\boldsymbol{\gamma}_*$ with components $\gamma_*^{(1)}$ and $\gamma_*^{(2)}$ that have the same sign as those of the prior mean, i.e. both components positive. Maximal efficiency improvements of about 9.6% were observed. When uncertainty about the prior mean grows, the surface across which the Bayesian design is more efficient expands (Figure 2.1b and Figure 2.1c).
The importance of such efficiency improvements can be seen when deriving the relationship between efficiency improvement and reduction in the volume of the confidence ellipsoid of the parameter estimates. It was already stated that the $100(1-\alpha)\%$ confidence region for the parameter estimates is an ellipsoid of which the volume V is inversely proportional to the square root of the determinant of the information matrix of a design with measure ξ:

$$V_\xi \propto \frac{1}{\sqrt{|M(\xi)|}}. \tag{26}$$

Furthermore, for two design measures ξ_1 and ξ_2 it follows that, according to expression (13),

$$|M(\xi_1)| = \left(D_{\text{eff}}\right)^p \times |M(\xi_2)| \,. \tag{27}$$

From (26) and (27):

$$V_{\xi_1} \propto \left(D_{\text{eff}}\right)^{-\frac{p}{2}} \times V_{\xi_2}, \tag{28}$$

or the volume V_{ξ_1} is $\left(D_{\text{eff}}\right)^{-\frac{p}{2}}$ times smaller than V_{ξ_2} if $\left(D_{\text{eff}}\right)^{-\frac{p}{2}} > 1$. Note that for response function estimation and for response and variance function estimation $p = 6$ and $p = 9$ respectively, for the polynomial expansions given in (18) and (19). The parameter p equals 10 when an interaction term is introduced in the polynomial expansion of the variance function. Table 3 shows the reduction in volume of the confidence region for different values of D_{eff}:

D_{eff}	$p = 6$	$p = 9$	$p = 10$
1.02	0.06	0.09	0.09
1.04	0.11	0.16	0.18
1.06	0.16	0.23	0.25
1.08	0.21	0.29	0.32
1.10	0.25	0.35	0.38
1.12	0.29	0.40	0.43
1.14	0.33	0.45	0.48
1.16	0.36	0.49	0.52
1.18	0.39	0.53	0.56
1.20	0.42	0.56	0.60

Table 3: Reduction $\left(1 - \left(D_{\text{eff}}\right)^{-\frac{p}{2}}\right)$ in Volume of Confidence Region

Reductions of 20% and more are easily attainable with limited efficiency improvements. The efficiency plots for the $\mathcal{D}$-optimal designs for mean [1] for response and variance function estimation are shown in Figure 2.2. Again, the Bayesian design performs better (efficiency ratio (16) larger than one) for variance function vectors $\boldsymbol{\gamma}_*$ with positive components $\gamma_*^{(1)}$ and $\gamma_*^{(2)}$. Efficiency improvements of about 12.7% are observable in Figure 2.2a.

The next question to be addressed is whether the above mentioned conclusions also apply to prior means close to the origin, i.e. when prior knowledge indicates that the prior mean only differs little from $(1\ 0\ 0)^T$ representing constant variance. This will be investigated on the basis of prior mean $(1\ 0.033\ 0.033)^T$ (mean [2] in Table 2). The corresponding $\mathcal{D}$-optimal designs are given in Figure 1.3. Remark that when only the response function

is of interest, the $\mathcal{D}$-optimal design for this prior mean and $\rho = 0$ equals that of the classic design in Figure 1.1a. In case of increased ρ (Figure 1.3b), the Bayesian design again outperforms the classic one for variance function components with the same sign as those of prior mean [2].
Figure 2.3 displays the efficiency plots for prior mean [2] for both response and variance function estimation. These plots indicate that even for small prior means the Bayesian design outperforms the classic one for a wide range of variance function vectors $\boldsymbol{\gamma}_*$, in our case those with positive components $\gamma_*^{(1)}$ and $\gamma_*^{(2)}$. The maximal improvement is about 7.6%. When ρ grows, the surface across which the relative efficiency is larger than one enlarges.

We conclude that even for prior means close to the origin and thus representing nearly constant variance, Bayesian $\mathcal{D}$-optimal designs can be constructed that are significantly more efficient than the classic ones for a wide range of variance function vectors $\boldsymbol{\gamma}_*$ of which the components $\gamma_*^{(1)}$ and $\gamma_*^{(2)}$ have the same sign as the components of the prior mean. This range even grows larger when uncertainty about the prior mean is taken into account. But the closer the prior mean to the origin, the smaller the maximal efficiency improvements.

4.2 First order variance functions with interaction

In this section a first order variance function with interaction term x_1x_2 is used as given in (20). This corresponds to a vector for the variance function parameters composed of four components as in (22). Again, several degrees of uncertainty ρ increasing from 0 to the one representing a nearly uniform prior distribution were investigated. Only those degrees of uncertainty at which changes in the computed $\mathcal{D}$-optimal design arised, will be discussed. Five degrees of uncertainty about the mean of the prior distribution are mentioned, namely $\rho \in \{0; 0.015; 0.02; 0.04; 1\}$.
The classic $\mathcal{D}$-optimal designs, in case of $\boldsymbol{\gamma_0}^T = (1\ 0\ 0\ 0)$ and $\rho = 0$, are shown in Figure 1.4. The question is whether the above derived rules for first order variance functions without interaction also hold for variance functions with an interaction term.

As an example, the $\mathcal{D}$-optimal designs for prior mean $(1\ 0.4\ 0.23\ 0.2)^T$ (mean [3] in Table 2) are shown in Figure 1.5. Figure 2.4 displays the efficiency plots for response function estimation and $\rho \in \{0, 0.015, 0.02\}$ for various values of the interaction effect $\gamma_*^{(3)}$. One establishes that, similar to the first order case without interaction, the Bayesian design performs better than the classic one for components $\gamma_*^{(1)}$ and $\gamma_*^{(2)}$ of $\boldsymbol{\gamma}_*$ with the same sign as the components $\gamma_0^{(1)}$ and $\gamma_0^{(2)}$ of the prior mean. The above rule holds for any interaction effect $\gamma_*^{(3)}$ (even if it is opposite signed to the interaction effect $\gamma_0^{(3)}$ of the prior mean) and especially comes true when the interaction effect $\gamma_*^{(3)}$ is in the neighbourhood of the interaction effect $\gamma_0^{(3)}$ of the prior mean. Maximal efficiency improvements of about 10% were observed.

Increased uncertainty about the prior mean - reflected by rised ρ - again leads to expanded surfaces across which the efficiency ratio becomes larger than one. Other interaction effects, estimation of both response and variance function and prior means close to the origin gave similar results.

4.3 Summary of results

The computational results show that for any prior knowledge about the mean of the variance function vector, the Bayesian $\mathcal{D}$-optimal designs perform better than the classic ones for a wide range of variance function parameters $\gamma_*^{(1)}$ and $\gamma_*^{(2)}$ that have the same sign as the respective components $\gamma_0^{(1)}$ and $\gamma_0^{(2)}$ of the prior mean. This phenomenon even strengthens when uncertainty about the prior mean is increased, reflected by augmented ρ. The farther away the prior mean from the origin, the larger the maximal relative efficiency improvement. Efficiency improvements of more than 12% are observed, resulting in reductions in the volume of the confidence ellipsoid of at least 30% and up to 40% in case of both response and variance function estimation.

The case of first order variance functions with an interaction effect is also considered. Incorporation of prior knowledge leads to Bayesian designs that outperform the classic ones for components $\gamma_*^{(1)}$ and $\gamma_*^{(2)}$ of the variance function vector with the same sign as the components $\gamma_0^{(1)}$ and $\gamma_0^{(2)}$ of the prior mean and for interaction effects $\gamma_*^{(3)}$ in the neighbourhood of the interaction effect $\gamma_0^{(3)}$ of the prior mean. The surfaces of efficiency ratio larger than one enlarge when uncertainty about the prior mean - reflected by increased ρ - is taken into account.

5 Conclusion

This paper analyzes the benefit of prior knowledge about the variance function parameters in $\mathcal{D}$-optimal design construction. For discrete approximations of multivariate normal distributions of the variance function parameters, $\mathcal{D}$-optimal designs are derived for response function estimation and both response and variance function estimation. The efficiencies of the computed designs are compared with those obtained by the assumption of constant variance, the so-called homoscedastic classic designs. The effect of the presence of an interaction term in the variance function on the design efficiency is also studied. It turns out that any prior knowledge about the sign of the variance function parameters leads to $\mathcal{D}$-optimal designs that outperform the classic ones for a wide range of actual variance function parameters. This phenomenon is even stronger when uncertainty about the prior mean is taken into account. In practice, this means that incorporating knowledge about the sign of the parameters of the variance function leads to designs that outperform the designs based on assumptions of constant variance.

Appendix 1. $\mathcal{D}$-optimal Designs

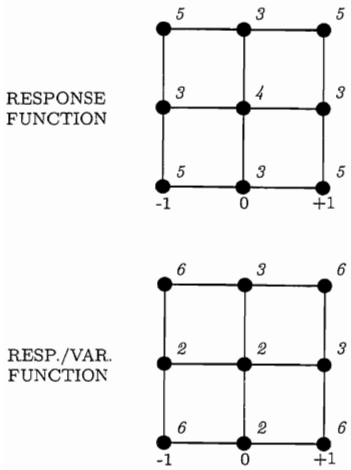


Figure 1.1: Classic $\mathcal{D}$-optimal Designs (Constant Variance and $\rho = 0$), No Interaction

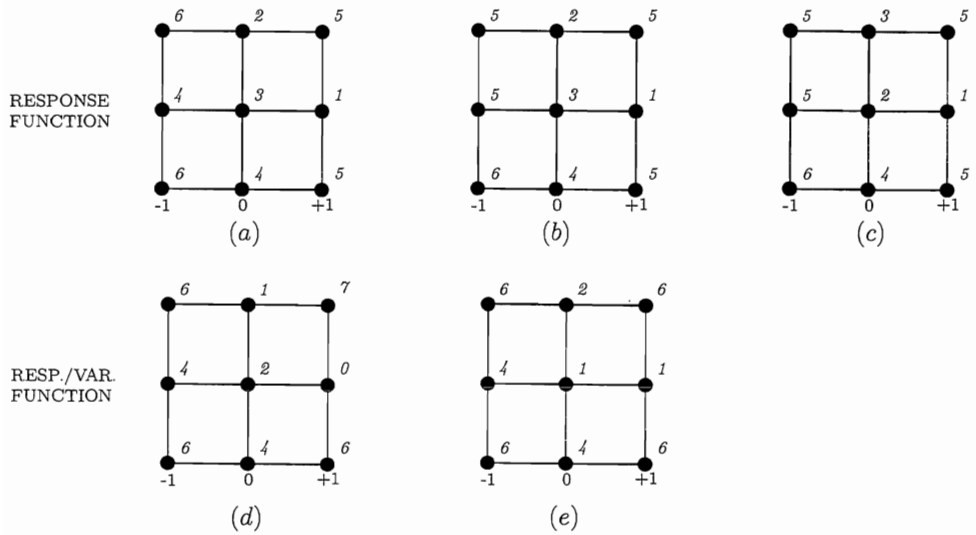


Figure 1.2: $\mathcal{D}$-optimal Designs for Prior Mean [1]
(a) $\rho = 0$ (b) $\rho \in \{0.3; 0.5; 1\}$ (c) $\rho = 3$ (d) $\rho = 0$ (e) $\rho \in \{0.3; 0.5; 1; 3\}$

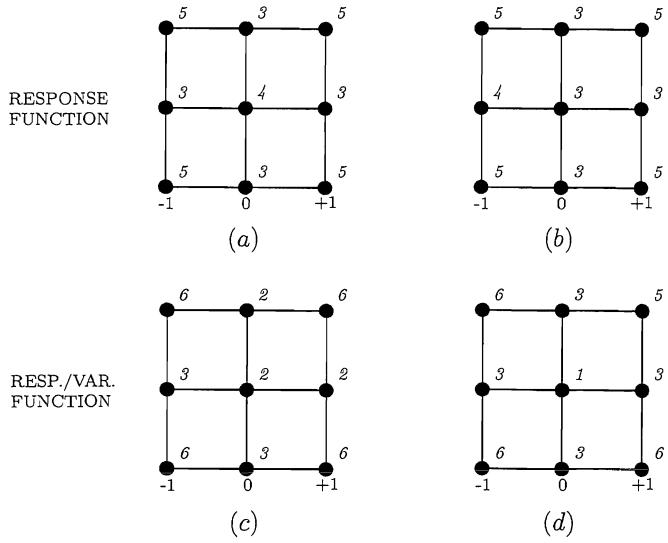


Figure 1.3: $\mathcal{D}$-optimal Designs for Prior Mean [2]
(a) $\rho = 0$ (b) $\rho \in \{0.3; 0.5; 1; 3\}$ (c) $\rho \in \{0; 0.3; 0.5\}$ (d) $\rho \in \{1; 3\}$

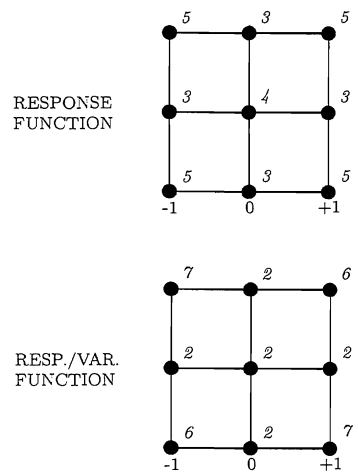


Figure 1.4: Classic $\mathcal{D}$-optimal Designs (Constant Variance and $\rho = 0$), Interaction

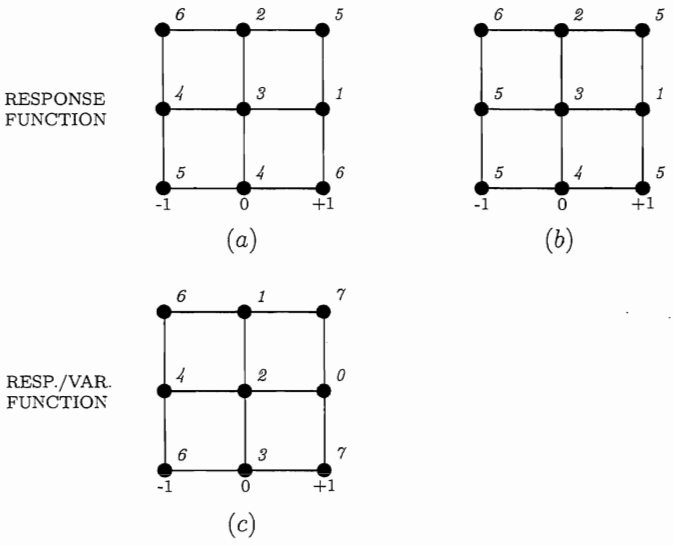
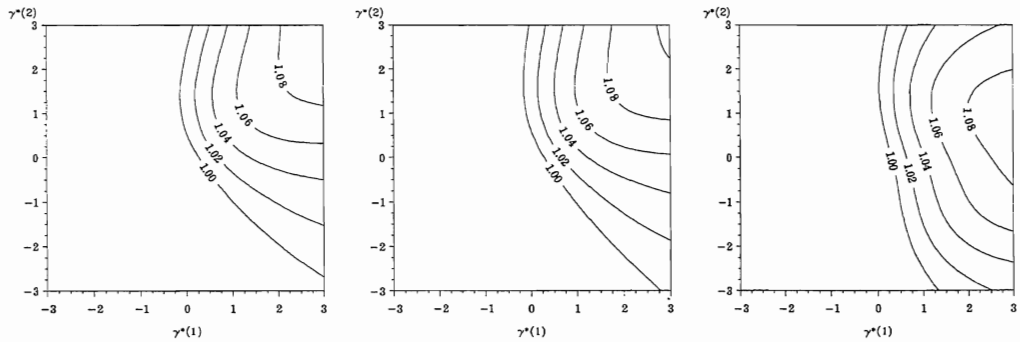


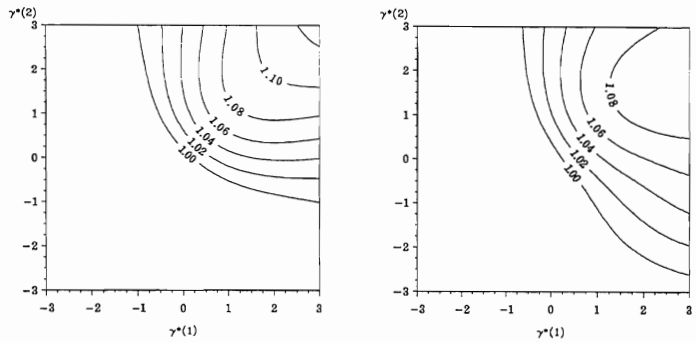
Figure 1.5: $\mathcal{D}$-optimal Designs for Prior Mean [3], $\gamma_0^{(3)} = 0.2$
(a) $\rho \in \{0; 0.015; 0.02\}$ (b) $\rho \in \{0.04; 1\}$ (c) $\rho \in \{0; 0.015; 0.02; 0.04; 1\}$

Appendix 2. Efficiency Plots



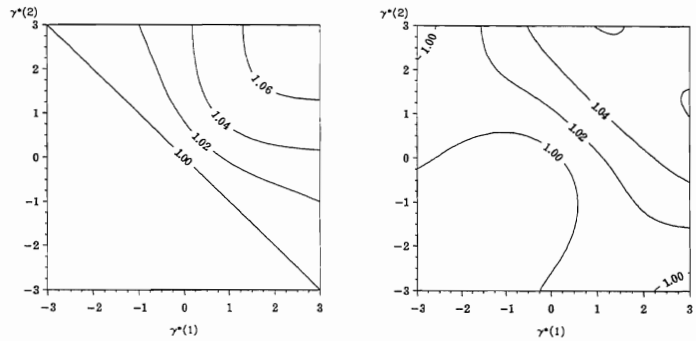
(a) $\rho = 0$ (b) $\rho \in \{0.3; 0.5; 1\}$ (c) $\rho = 3$

Figure 2.1: Efficiency plots for Prior Mean [1], Response Function



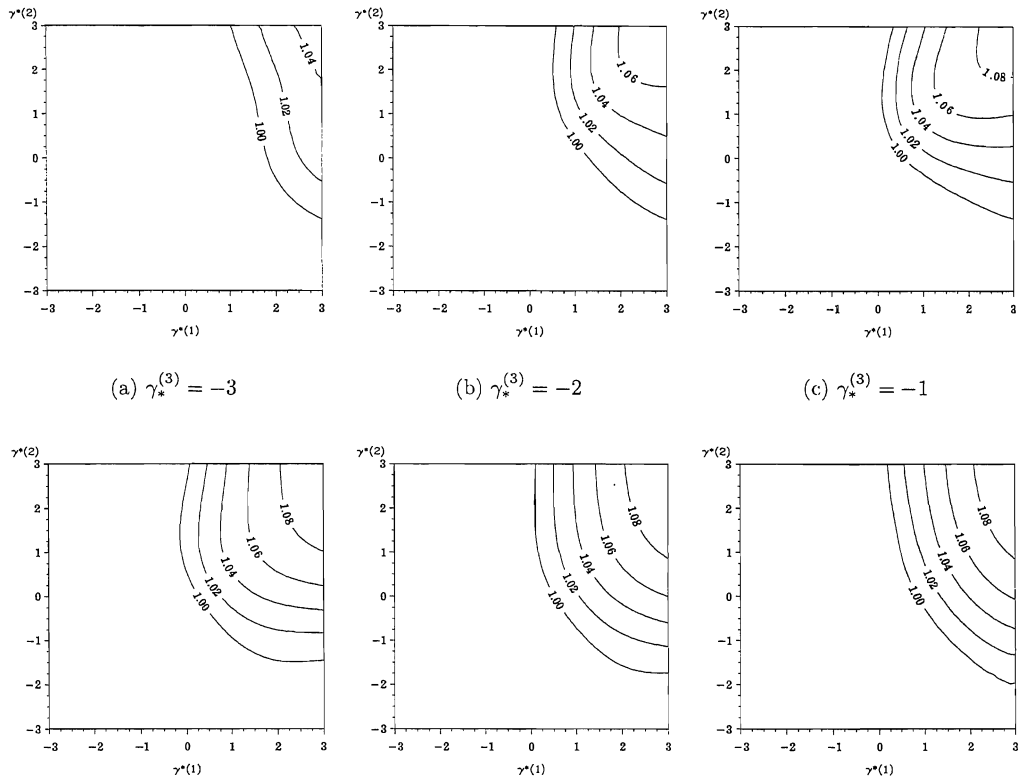
(a) $\rho = 0$ (b) $\rho \in \{0.3; 0.5; 1; 3\}$

Figure 2.2: Efficiency plots for Prior Mean [1], Response and Variance



(a) $\rho \in \{0; 0.3; 0.5\}$ (b) $\rho \in \{1; 3\}$

Figure 2.3: Efficiency plots for Prior Mean [2], Response and Variance



(a) $\gamma_*^{(3)} = -3$ (b) $\gamma_*^{(3)} = -2$ (c) $\gamma_*^{(3)} = -1$

(d) $\gamma_*^{(3)} = 1$ (e) $\gamma_*^{(3)} = 2$ (f) $\gamma_*^{(3)} = 3$

Figure 2.4: Efficiency plots for Prior Mean [3], Response Function, $\rho \in \{0; 0.015; 0.02\}$

Acknowledgments

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

References

Aitkin, M. (1987). "Modeling Variance Heterogeneity in Normal Linear Regression Using GLIM". *Applied Statistics* 36, pp. 332-339.

Atkinson, A.C. (1996). "The Usefulness of Optimum Experimental Designs". *Journal of the Royal Statistical Society, Series B* 58, No. 1, pp. 59-76.

Atkinson, A.C. and Cook, R.D. (1995). "$\mathcal{D}$-optimal Designs for Heteroscedastic Linear Models". *Journal of the American Statistical Association* 90, pp. 204-212.

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

Broemeling, L.D. (1985). *Bayesian Analysis of Linear Models.* Marcel Dekker, Inc, New York.

Chaloner, K. (1984). "Optimal Bayesian Experimental Design for Linear Models". *Annals of Statistics* 12, pp. 283-300.

Chaloner, K. and Larntz, K. (1989). "Optimal Bayesian Design Applied to Logistic Regression Experiments". *Journal of Statistical Planning and Inference* 21, pp. 191-208.

Chaloner, K. and Verdinelli, I. (1995). "Bayesian Experimental Design: A Review". University of Minnesota technical report.

Dumouchel, W. and Jones, B. (1994). "A Simple Bayesian Modification of D-optimal Designs to Reduce Dependence on an Assumed Model". *Technometrics* 36, pp. 37-47.

Fedorov, V.V. (1972). *Theory of Optimal Experiments.* Academic Press, New York.

Fisher, R.A. (1960). *The Design of Experiments*, 7th. ed. Oliver and Boyd, Edinburgh.

Huele, A.F. (1998). *Statistical Robust Design.* Ph. D. Thesis, Amsterdam.

Kackar, R.N. and Shoemaker, A.C. (1986). "Robust design: a Cost-Effective Method for Improving Manufacturing Processes". *AT & T Technical Journal* 65, pp. 39-50.

Kiefer, J. (1959). "Optimum experimental designs". *Journal of the Royal Statistical Society, Series B* 21, pp. 272-319.

Kiefer, J. and Wolfowitz, J. (1960). "Optimum Designs in Regression Problems". *Annals of Mathematical Statistics* 30, pp. 271-294.

Phadke, M.S. and Dehnad, K. (1987). "Optimization of Product and Process Design for Quality and Cost". *ASQC Quality Congress Transactions*, pp. 330-334.

Pukelsheim, F. (1993). *Optimal Design of Experiments.* Wiley, New York.

Silvey, S.D. (1980). *Optimum Design.* Chapman and Hall, London.

Smith, K. (1918). "On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations". *Biometrika* 12, pp. 1-85.

Taguchi, G. (1986). *Introduction to Quality Engineering: Designing Quality into Products and Processes.* Asian Productivity Organization, Tokyo.

Taguchi, G. (1987). *System of Experimental Design*, Vol. 1 and 2. Asian Productivity Organization, Tokyo.

Unal, R. and Daen, E.B. (1991). "Design for Cost and Quality: the Robust design Approach". *Journal of Parametrics* 11, pp. 73-93.

Vandebroek, M. and Goos, P. (1997). "Semi-Bayesian $\mathcal{D}$-optimal Designs and Estimation Procedures for Mean and Variance Functions". Research Report No. 9745, DTEW, K.U.Leuven.

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

Wynn, H.P. (1985). *Jack Kiefer's contributions to experimental design.* Jack Carl Kiefer Collected Papers Vol. III. Wiley, New York, pp. xvii-xxiv.

Key Words: *Bayesian Design, $\mathcal{D}$-optimality, Experimental Design, Heteroscedasticity*