Robustness properties of minimally-supported Bayesian D-optimal designs for heteroscedastic models

Holger DETTE, Dale SONG and Weng Kee WONG

Key words and phrases: approximate designs; efficiency lower bounds; prior distributions.
AMS 1991 subject classifications: Primary 62K05; secondary 62J05.

ABSTRACT

Bayesian D-optimal designs supported on a fixed number of points were found by Dette and Wong (1998) for estimating parameters in a polynomial model when the error variance depends exponentially on the explanatory variable. This work provides optimal designs under a broader class of error variance structures and investigates the robustness properties of these designs to model and prior distribution assumptions. A comparison of the performance of the optimal designs relative to the popular uniform designs is also given. In addition, our results suggest that Bayesian D-optimal designs suported on a fixed number of points are more likely to be optimal among all designs if the prior distribution is symmetric and is concentrated around its mean.

RÉSUMÉ

Bayesian D-optimal designs supported on a fixed number of points were found by Dette and Wong (1998) for estimating parameters in a polynomial model when the error variance depends exponentially on the explanatory variable. This work provides optimal designs under a broader class of error variance structures and investigates the robustness properties of these designs to model and prior distribution assumptions. A comparison of the performance of the optimal designs relative to the popular uniform designs is also given. In addition, our results suggest that Bayesian D-optimal designs suported on a fixed number of points are more likely to be optimal among all designs if the prior distribution is symmetric and is concentrated around its mean.

1. INTRODUCTION

Heteroscedastic regression models are widely used in the biological and physical sciences, economics and engineering. Often, the heteroscedasticity is modeled as a function of the design variables and possibly, additional parameters. Common examples are modeling the error variance as a simple power function or an exponential function of the mean response. Some work on estimating such error variance structure includes Carroll and Ruppert (1982), Carroll (1982a, b), Davidian and Carroll (1987) and Engel (1992). Tests for the presence of heteroscedasticity in the model are given in Box and Hill (1974), Bickel (1978), Cook and Weisberg (1983), Carroll and Ruppert (1981), Jobson and Fuller (1980) and Dette and Munk (1998).

It appears that there is only a handful of papers that address the problem of designing an experiment when the error variance structure is known imprecisely. Schulz and Endrenyi (1983), Dasgupta, Mukhopadhyay and Studden (1992), and recently, Atkinson and Cook (1995) and Wiens (1998) are among the few who address this issue. Schulz and Endrenyi (1983) found Bayesian optimal designs for two nonlinear models when the error variance is a power of the mean. Dasgupta, Mukhopadhyay and Studden (1992) derived Bayesian optimal designs under a normal prior for the simple linear model when the error variance is a power function or an exponential function of the mean. Atkinson and Cook (1995) gave information matrices for various subsets of the parameters in a general heteroscedastic model and provided guidelines for finding optimal designs.

Dette and Wong (1995) considered optimal designs supported on a fixed number of points and found Bayesian optimal designs when the variance structure is partially known. When the mean regression function is a polynomial model of degree n, such optimal designs are found from the class of designs supported on n+1 points and they are called Bayesian D-optimal (n+1)-point designs. These Bayesian designs are optimal within the class of (n+1)-point designs and they may not be universally optimal, in the sense that they are also optimal within the class of all designs. Analytical Bayesian D-optimal (n+1)-point designs when the error variance depends exponentially on the design variable are given in Dette and Wong (1998).

The purpose of this paper is to find Bayesian D-optimal (n+1)-point designs for a broader class of error variance structures and study the sensitivities of Bayesian D-optimal (n+1)-point designs to model assumptions and the prior distribution. We also develop guidelines to ascertain if and when such designs are likely to be universally optimal. Ideally, we would like to use a design that is robust to both model and prior assumptions, and the design is efficient under a large class of designs.

Section 2 introduces our terminology and our setup. Section 3 describes the scope of our work and section 4 presents our numerical findings and a discussion of our results. Section 5 contains summary remarks and the appendix contains theoretical results used in the paper.

2. BACKGROUND: MODEL ASSUMPTIONS AND TERMINOLOGY

The statistical model of interest is given by

$$y = \sum_{j=0}^{n} \alpha_j \; x^j + e(x, \theta), \qquad\qquad x \in [0, b] \tag{1}$$

where y is the response and $e(x, \theta)$ is an unobservable error assumed to be normally distributed with mean zero and variance given by $1/\lambda(x, \theta)$. In the design literature, the interval $[0, b]$ is called the design space and $\lambda(x, \theta)$ is the efficiency function (Fedorov 1972, p. 72). The full set of parameters is denoted by $\beta^T = (\alpha^T, \theta)$ where θ is the vector of parameters in the efficiency function and $\alpha^T = (\alpha_0, \alpha_1, \ldots, \alpha_n)$ is the vector of structural parameters. We denote the regression function by $f'(x) = (1, x, x^2, ..., x^n)$ and it is assumed throughout that both n and b are given and finite. The goal here is to find an efficient design to estimate the vector of structural parameters α using prior information on the nuisance parameters θ. It is assumed that all observations are indepedent in the experiment.

In this paper, we assume that $\theta^T = (\theta_1, \theta_2)$ is a two-dimensional vector of unknown parameters and the form of $\lambda(x, \theta)$ can be adequately described by one of the following forms:

$$\begin{array}{lll} & (i) & \lambda_1(x, \theta) \quad \text{is a constant} \\ & (ii) & \lambda_2(x, \theta) = \theta_1 exp(-\theta_2 x), \quad \theta_1 > 0 \\ \text{and} & (iii) & \lambda_3(x, \theta) = x^{\theta_1}(b - x)^{\theta_2}, \quad \theta_1 \geq 0 \quad \text{and} \quad \theta_2 \geq 0. \end{array}$$

The first case corresponds to a homoscedastic model and it is a special case of the last two cases when $\theta_2 = 0$ in (ii) and when $\theta_1 = \theta_2 = 0$ in (iii). The exponential efficiency function $\lambda_2(x, \theta)$ was studied in Dette and Wong (1998) and $\lambda_3(x, \theta)$ is the additional class of efficiency functions to be studied here. Figure 1 shows several shapes of the efficiency functions for different values of θ. These variance structures are quite flexible and so they are useful in practice.

Throughout, we are concerned only with approximate designs. If the total number of observations is fixed at N at the onset of the experiment, then an approximate design ξ with mass ξ_i at $x_i \in [0, b]$ allocates roughly $N\xi_i$ observations at x_i, $i = 1, 2, \ldots, k$ subject to $N(\xi_1 + \xi_2 + \ldots + \xi_k) = N$. Additional background information on the approximate designs, and their advantages, can be found in design monographs, see Fedorov (1972) or Silvey (1980), for example.

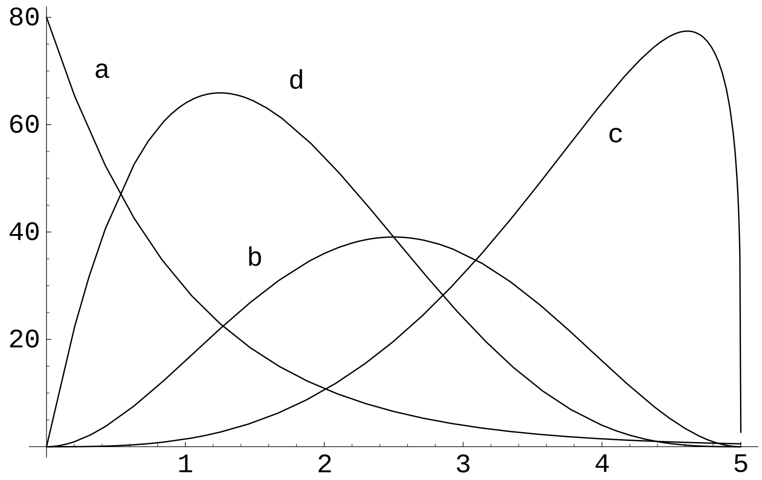

Figure 1: Different shapes of the efficiency functions on $[0, 5]$; (a) $\lambda_2(x, \theta)$ with $(\theta_1, \theta_2)^T = (80, 1)$ and $\lambda_3(x, \theta)$ with (b) $(\theta_1, \theta_2)^T = (2, 2)$, (c) $(\theta_1, \theta_2)^T = (3, 0.25)$ and (d) $(\theta_1, \theta_2)^T = (1, 3)$.

The usefulness of a design is gauged by its expected Fisher information matrix. When this matrix is multiplied by -1, its elements are means of the second derivatives of the log-likelihood function with respect to the parameters. For example, suppose model (1) holds with $\lambda_2(x, \theta)$ on $[0, b]$. A direct calculation shows the information matrix for a single observation at x is independent of α and is proportional to

$$I(\theta, x) = \begin{pmatrix} \frac{1}{\theta_1} I_1(\theta_2, x) & 0 \\ 0 & \frac{1}{2\theta_1^2} I_2(\theta_1, x) \end{pmatrix} ,$$

where $I_1(\theta_2, x) = exp(-\theta_2 x) f(x) f^T(x)$ and $I_2(\theta_1, x) = \begin{pmatrix} 1 & \theta_1 x \\ \theta_1 x & \theta_1^2 x^2 \end{pmatrix}$; see Dette and Wong (1998). This implies that the total information matrix of an approximate design ξ for estimating the full set of parameters β is proportional to

$$M(\theta, \xi) = \int_0^b \begin{pmatrix} \frac{1}{\theta_1} I_1(\theta_2, x) & 0 \\ 0 & \frac{1}{2\theta_1^2} I_2(\theta_1, x) \end{pmatrix} \xi(dx) = \begin{pmatrix} \frac{1}{\theta_1} M_1(\theta_2, \xi) & 0 \\ 0 & \frac{1}{2\theta_1^2} M_2(\theta_1, \xi) \end{pmatrix} ,$$

where $M_i(\theta_{3-i}, \xi) = \int_0^b I_i(\theta_3 - i, x)\, \xi(dx),\ i = 1, 2.$

A popular criterion for estimating model parameters is D-optimality, which seeks to maximize the determinant of the information matrix. For normally distributed errors, this criterion is equivalent to minimizing the volume of the confidence ellipsoid for the parameters of interest. When prior knowledge is available, the researcher elicits a prior distribution for the parameters and the design criterion is to maximize the logarithm of the determinant of the information matrix, after averaging it with the prior distribution. When the interest is on estimating the structural parameters α only, as is the case here, we first find the marginal distribution π of the nuisance parameters. For instance, if the efficiency function is $\lambda_2(x, \theta)$ and the support of the prior is the set B, the Bayesian D-optimality criterion for estimating α in model (1) is to maximize

$$\Phi(\xi) = \int_B log|M_1(\theta_2, \xi)| \, \pi(d\theta_2)$$

over the set of all designs on $[0, b]$. The resulting optimal design is called a Bayesian D-optimal for estimating α and they are increasingly used in the literature (Dasgupta, Mukhopadhyay and Studden 1992; Chaloner and Verdnelli 1995). Most of the results in this area are based on numerical work because it is notoriously difficult to find Bayesian optimal designs analytically. Frequently, equivalence theorems, such as the one given in the appendix, are used to find the optimal design, or to check the optimality of the design. For instance, to check if a design ξ is Bayesian D-optimal for estimating α in model (1) and the efficiency function is $\lambda_2(x, \theta)$, we plot the function on the left hand side of the inequality (2) versus values of x in $[0, b]$ and check if the conditions (2) in the appendix are met. If so, the design ξ is Bayesian D-optimal for estimating α; otherwise it is not.

The design problem for model (1) is simplified if the maximization is restricted to the set of all $(n + 1)$-point designs; the resulting design is called a Bayesian D-optimal $(n + 1)$-point design for estimating α. Dette and Wong (1998) found analytical Bayesian D-optimal $(n + 1)$-point designs for estimating α when the efficiency function is $\lambda_2(x, \theta)$. They found that these optimal designs depend only on n, b and the mean of the marginal distribution of θ_2. The technical arguments are based on the theory of canonical moments and continued fraction expansion. A nice illustration of this technique is given in Studden (1982), where he sought to obtain constrained optimal designs in homoscedastic polynomial models. In the appendix, we state a new result for constructing the Bayesian D-optimal $(n+1)$-point designs for estimating α in model (1) when the efficiency function is $\lambda_3(x, \theta)$. As will be shown, the Bayesian D-optimal $(n + 1)$-point designs also depend only on n, b and the means of the marginal distributions of θ_1 and θ_2. In what is to follow, we usually denote the means of the marginal distribution of θ_1 and θ_2 by E1 and E2 respectively.

3. SCOPE OF OUR STUDY

Except in a few simple instances like those described in Dette and Wong (1998), it seems a numerical study is the only feasible way to gain insight into the robustness properties of Bayesian D-optimal $(n+1)$-point designs. The plan here is to compute many such optimal designs for various situations and afterwards, deduce robustness properties of the designs. Specifically, we provide answers to the following questions:

(i) How robust are these Bayesian D-optimal $(n + 1)$-point designs to specification of the degree (n) of the polynomial ?

(ii) How do Bayesian D-optimal $(n + 1)$-point designs perform when the means of the marginal distributions of θ are misspecified?

(iii) How do Bayesian D-optimal $(n + 1)$-point designs perform under a variation of efficiency functions?

(iv) Are there general conditions that suggest Bayesian D-optimal $(n + 1)$-point designs are universally optimal ?

These questions have practical importance. Polynomial models are widely used but the degree of the model is usually not known with certainty. There might also be disagreement on the form of the prior distribution and/or the values of the means of the marginal distributions of θ. It is thus desirable to find some answers to the above questions to understand the robustness properties of the optimal designs. The recommended design should perform reasonably well when the model assumptions are slightly changed. Question (iv) is relevant because we have confined our search for the optimal design to the set of all $(n+1)$-point designs and it is of interest to know if we have unduly restricted the search and strayed too far from the universally optimal design.

The performance of a design is measured by its efficiency relative to the optimal design under the assumed setup. For example, suppose model (1) holds, the efficiency function is $\lambda_2(x,\theta)$, and, as before, the goal is to estimate α using π as the marginal distribution of θ_2. The efficiency of an arbitrary design ξ for estimating α relative to the optimal design ξ^* is defined by $exp\{[\Psi(\xi)-\Psi(\xi^*)]/(n+1)\}$. Here the optimal design ξ^* can be the Bayesian D-optimal $(n+1)$-point design or the Bayesian D-optimal design, as appropriate. The former is used if we wish to study robustness properties of the Bayesian D-optimal $(n+1)$-point design (Questions (i), (ii) and (iii)), and the latter is used if we wish to ascertain if the Bayesian D-optimal $(n+1)$-point design is close to the Bayesian optimal design (Question (iv)). In practice, designs with efficiencies close to 1 are sought because they require fewer number of observations to achieve a given level of precision in the estimates. It is interesting to note that the Bayesian D-optimal $(n+1)$-point design for estimating α in model (1) depends only on the mean(s) of the marginal prior distribution(s) of θ but its efficiency depends on the marginal prior distributions.

We give some answers to the above questions by considering a simple setup so that the computation is manageable. There are many parameters to vary in the study; different design space (b), different polynomial degree (n), types of prior distributions and different means of the marginal prior distributions. We focus on low order polynomial with degree $n = 1$, 2, 3 and 4 because these are the most frequently used in practice. For simplicity, we consider priors indexed by a single parameter δ and they are equally-spaced discrete priors, symmetric about its mean E. The mass distribution for each prior has a form similar to that used in Cook and Nachtsheim (1982). The prior distributions take on the following form:

support	...	E-2δ	E-δ	E	E+δ	E+2δ	...
weight	...	$\frac{1}{2^{2\delta}}$	$\frac{1}{2^{\delta}}$	1	$\frac{1}{2^{\delta}}$	$\frac{1}{2^{2\delta}}$	...

The normalized prior distributions are uniquely determined by the number of support points and the values of E and δ. For instance, for a 3-point prior with E = δ =1, the prior distribution is supported at 0, 1 and 2 with mass 0.25, 0.5 and 0.25 respectively. Table 1 shows the types of priors used in our study; in each case, the prior π_2 is more concentrated around its mean when compared with the corresponding prior π_1.

To answer question (iv), we investigate distributional properties of the prior density (other than its mean) that may affect whether a Bayesian D-optimal $(n+1)$-point design is Bayesian D-optimal. In particular, we consider several prior distributions with the same means but have different spread and number of support points. Because Bayesian D-optimal designs are difficult to determine, we cannot directly evaluate the performance of the Bayesian D-optimal $(n+1)$-point design. In this case, the proximity of the Bayesian D-optimal $(n+1)$-point design to the (unknown) Bayesian D-optimal design is assessed by an efficiency lower bound. This lower bound is obtained from an equivalence theorem given in the appendix for an illustrative case when the efficiency function is $\lambda_2(x,\theta)$. A useful property of the bound is that it has a value of 1 if and only if the candidate design is Bayesian D-optimal.

Table 1: Priors with different shapes and number of support points used in our study.

	3-point priors		5-point priors		9-point priors	
	π_1	π_2	π_1	π_2	π_1	π_2
E	δ	δ	δ	δ	δ	δ
1	1	0.2	0.5	0.1	0.25	0.06
4	2	0.5	1.0	0.2	0.50	0.10
7	4	1.0	2.0	0.5	1.00	0.30

4. NUMERICAL RESULTS AND DISCUSSION

All the computations in this paper are accomplished using Mathematica^{TM} (Wolfram, 1988). There are 7 tables in this section, with the first set of tables (Tables 2 and 3) showing Bayesian D-optimal 2, 3, 4 and 5-point designs for estimating α in model (1) when $\lambda(x, \theta)$ is non-constant. The subsequent tables present results useful for answering the four questions in the order they were raised. Because of space consideration, we include only a sample of our numerical results.

The Bayesian D-optimal $(n+1)$-point designs are constructed as follows. When $\lambda(x, \theta)$ is constant, it is clear that the Bayesian D-optimal $(n+1)$-point designs for estimating α in model (1) coincide with the Bayesian D-optimal designs and they are the well known D-optimal designs described in Fedorov (1972, p. 88). The Bayesian D-optimal $(n+1)$-point designs for estimating α in model (1) with efficiency function $\lambda_2(x, \theta)$ are constructed from Theorem 3.1 in Dette and Wong (1998). This result is restated as Theorem 1 in the appendix for convenience. The Bayesian D-optimal $(n+1)$-point designs for estimating α in model (1) with $\lambda_3(x, \theta)$ are found from Theorem 3 in the appendix.

Table 2 and Table 3 show the support points of the Bayesian D-optimal 2, 3, 4 and 5-point designs for estimating α when $\lambda(x, \theta)$ is non-constant and $b = 1$ or 5. Table 2 and 3 show, respectively, the optimal designs when the efficiency function is $\lambda_2(x, \theta)$ or $\lambda_3(x, \theta)$. The priors used here are degenerate distributions. In both tables, the first column lists the values of the means of the marginal prior distributions of θ, and the support points of the Bayesian D-optimal 2, 3, 4 and 5-point designs are shown on the same row. All optimal designs are equally supported in accordance with the theory. Note that in Table 3, we can deduce the support points of the Bayesian D-optimal $(n+1)$-point design on $[0, b]$ for estimating α from the corresponding optimal design on $[0, 1]$. This is because of the invariance of the design problem under linear transformations; see the proof of Theorem 3 in the appendix.

Table 4 and Table 5 display the efficiencies of the Bayesian D-optimal $(n+1)$-point designs when $b = 1$ and the degree of the polynomial is misspecified. The asssumed degree is n' and the true degree is n. The Bayesian $(n'+1)$-point D-optimal designs $\xi^*_{n',E}$ and Bayesian $(n+1)$-point D-optimal designs $\xi^*_{n,E}$ are first constructed for $4 = n' > n = 1$. When $n' > n$, the robustness of the Bayesian $(n'+1)$-optimal design $\xi^*_{n',E}$ relative to $\xi^*_{n,E}$ is measured by $exp\{(\Phi(\xi^*_{n',E}) - \Phi(\xi^*_{n,E}))/(n+1)\}$. The results when $n' < n$ or $n' = n$ are not displayed because the efficiencies of the Bayesian D-optimal $(n+1)$-point design for these situations are 0 and 1 respectively.

Different priors are used for different efficiency functions for model (1) to assess the impact of the misspecification of the degree of the polynomial. When the efficiency function is $\lambda_2(x, \theta)$, the

Table 2: Support points of equally weighted Bayesian D-optimal $(n+1)$ point designs for estimating α in model (1) with $n = 1$, 2, 3 or 4, $b = 1(5)$ and the efficiency function is $\lambda_2(x, \alpha)$. The prior for θ_2 is degenerate at $\theta_2 = \mathrm{E}$.

$n = 1$ and $b = 1$ ($b = 5$)

	x_1	x_2
E=1	0(0)	1(2)
E=4	0(0)	0.5(0.5)
E=7	0(0)	0.286(0.286)

$n = 2$ and $b = 1$ ($b = 5$)

	x_1	x_2	x_3
E=1	0(0)	0.439(1.268)	1(4.732)
E=4	0(0)	0.293(0.317)	1(1.183)
E=7	0(0)	0.181(0.181)	0.676(0.676)

$n = 3$ and $b = 1$ ($b = 5$)

	x_1	x_2	x_3	x_4
E=1	0(0)	0.245(0.782)	0.688(2.629)	1(5)
E=4	0(0)	0.174 (0.234)	0.567(0.826)	1(1.934)
E=7	0(0)	0.129(0.134)	0.451(0.472)	1(1.108)

$n = 4$ and $b = 1$ ($b = 5$)

	x_1	x_2	x_3	x_4	x_5
E=1	0(0)	0.156(0.535)	0.469(1.769)	0.808(3.535)	1(5)
E=4	0(0)	0.117(0.186)	0.380(0.643)	0.736(1.433)	1(2.739)
E=7	0(0)	0.091(0.106)	0.306(0.367)	0.645(0.819)	1(1.565)

Table 3: Support points of equally weighted Bayesian D-optimal $(n+1)$ point designs for estimating α in model (1) with $n = 1, 2, 3$ or 4, $b = 1(5)$ and the efficiency function is $\lambda_3(x, \alpha)$. The prior for $\theta = (\theta_1, \theta_2)$ is degenerate at $(\theta_1, \theta_2) =$ (E1, E2).

$n = 1$ and $b = 1$ $(b = 5)$		
	x_1	x_2
E1=0, E2=0	0(0)	1(5)
E1=0.5, E2=3	0.063(0.313)	0.483(2.413)
E1=3, E2=0.5	0.517(2.587)	0.937(4.686)

$n = 2$ and $b = 1$ $(b = 5)$			
	x_1	x_2	x_3
E1=0.5, E2=3	0.036(0.178)	0.292(1.461)	0.672(3.361)
E1=0.5, E2=0.5	0.067(0.335)	0.500(2.500)	0.933(4.665)
E1=3, E2=0.5	0.328(1.639)	0.708(3.539)	0.964(4.822)

$n = 3$ and $b = 1$ $(b = 5)$				
	x_1	x_2	x_3	x_4
E1=0.5, E2=3	0.023 (0.115)	0.195(0.974)	0.479(2.395)	0.777(3.884)
E1=1.0, E2=1	0.069(0.347)	0.330(1.650)	0.670(3.350)	0.931(4.653)
E1=3.0, E2=0.5	0.223(1.116)	0.521(2.605)	0.805(4.026)	0.977(4.885)

$n = 4$ and $b = 1$ $(b = 5)$					
	x_1	x_2	x_3	x_4	x_5
E1=0.5,E2=3	0.016(0.081)	0.139(0.694)	0.354(1.770)	0.608(3.041)	0.839(4.196)
E1=3.0,E2=0.5	0.161(0.804)	0.392(1.958)	0.646(3.230)	0.861(4.306)	0.984(4.920)
E1=3.0, E2=3	0.115(0.576)	0.290(1.448)	0.500(2.500)	0.711(3.552)	0.885(4.424)

prior distributions are as described at the end of section 3 with 3, 5 or 9 support points and the marginal prior means are 1, 4 or 7. The first line in the Table 4 gives the efficiencies of the optimal designs constructed under the prior distribution π_1 and the second line gives the corresponding efficiencies for the prior distribution π_2. When the efficiency function is $\lambda_3(x,\theta)$, the priors are degenerate at the marginal means E1 and E2. Table 5 lists the efficiencies of the optimal designs when each value of E1 or E2 is either 0, 0.5, 1 or 3.

Table 4: Efficiencies of the Bayesian $(n'+1)$-point D-optimal design on $[0,1]$ assuming the true model is (1) of degree n, $n = 1, 2, 3$ and $\lambda_3(x,\theta) = \theta_1 exp(-\theta_2 x)$. The numbers on the first (second) line are the results for the priors of the form π_1 (π_2) with different prior means (E2) and different number of support points (s).

		$n=1$			$n=2$			$n=3$		
n'	E2	s=3	s=5	s=9	s=3	s=5	s=9	s=3	s=5	s=9
	1	0.837	0.835	0.834						
	1	0.828	0.828	0.828						
2	4	0.828	0.824	0.821						
	4	0.796	0.794	0.794						
	7	0.766	0.774	0.775						
	7	0.753	0.751	0.752						
	1	0.794	0.793	0.792	0.872	0.871	0.870			
	1	0.788	0.788	0.788	0.866	0.866	0.866			
3	4	0.800	0.798	0.795	0.902	0.901	0.899			
	4	0.778	0.776	0.776	0.886	0.885	0.885			
	7	0.663	0.670	0.670	0.882	0.895	0.897			
	7	0.650	0.649	0.650	0.864	0.862	0.863			
	1	0.776	0.774	0.773	0.833	0.832	0.831	0.900	0.899	0.899
	1	0.770	0.770	0.769	0.829	0.829	0.829	0.896	0.896	0.896
4	4	0.780	0.777	0.775	0.866	0.864	0.863	0.909	0.907	0.829
	4	0.759	0.758	0.758	0.854	0.853	0.853	0.898	0.898	0.898
	7	0.632	0.639	0.639	0.862	0.871	0.872	0.927	0.934	0.935
	7	0.620	0.619	0.619	0.847	0.846	0.846	0.915	0.914	0.915

There are a few immediate implications from the tables. They suggest that the performance of the Bayesian D-optimal $(n+1)$-point designs are quite robust when the polynomial degree is not too misspecified. The efficiencies are roughly between 0.62 and 0.92 in Table 4 and between 0.66 to 0.90 in Table 5, with greater misspecification resulting in greater loss in efficiency. The efficiencies are nearly the same when the number of support points in the prior distribution is changed but overall they tend to vary when the means of the marginal distributions are changed. Table 4 also suggests that a more spread out prior distribution results in lower efficiencies when misspecification occurs. Overall, the Bayesian D-optimal $(n+1)$-point designs are quite robust

Table 5: Efficiencies of the Bayesian $(n'+1)$-point D-optimal design on $[0,1]$ when the true model is (1) of degree n, $n=1, 2, 3$, $\lambda_3(x,\theta)=x^{\theta_1}(b-x)^{\theta_2}$ and the priors for θ_1 and θ_2 are degenerate at E1 and E2 respectively.

n	n'	E1	E2=0	E2=0.5	E2=1	E2=3
		0	0.817	0.815	0.805	0.782
1	2	0.5	0.815	0.817	0.810	0.790
		1	0.805	0.810	0.805	0.790
		3	0.782	0.790	0.790	0.785
		0	0.775	0.773	0.753	0.706
1	3	0.5	0.773	0.777	0.762	0.722
		1	0.753	0.762	0.753	0.721
		3	0.706	0.722	0.721	0.709
		0	0.756	0.754	0.727	0.664
1	4	0.5	0.754	0.761	0.741	0.685
		1	0.727	0.741	0.727	0.683
		3	0.664	0.685	0.683	0.664
		0	0.865	0.866	0.863	0.852
2	3	0.5	0.866	0.868	0.865	0.855
		1	0.863	0.865	0.863	0.855
		3	0.852	0.855	0.855	0.852
		0	0.828	0.830	0.822	0.796
2	4	0.5	0.830	0.833	0.827	0.804
		1	0.822	0.827	0.822	0.803
		3	0.796	0.804	0.803	0.794
		0	0.895	0.896	0.895	0.888
3	4	0.5	0.896	0.897	0.896	0.890
		1	0.895	0.896	0.895	0.890
		3	0.888	0.890	0.890	0.888

with respect to the model misspecifications; if the misspeciation of the degree is not off by more than unity, the efficieny lost is at most 25% for all the cases considered in the table and, this amount decreases as the true degree n increases. Our suggestion is that if we are uncertain about the degree of the polynomial, the optimal design should be constructed for the polynomial model of the highest anticipated degree.

Table 6 and Table 7 display the efficiencies of the Bayesian $(n+1)$-point D-optimal design, n = 1, 2, 3 and 4, when the prior mean of the marginal distribution of θ_2 is misspecified to be E$'$. Model (1) is assumed with $\lambda_2(x, \theta)$ as the efficiency function on $[0, 1]$ (Table 6) and $[0, 5]$ (Table 7). The prior distributions have the form of π_1 with three possible marginal means (E2 = 1, 4, 7) for θ_2. Corresponding results for $\lambda_3(x, \theta)$ are more complicated and are omitted.

Table 6 and Table 7 assume that the efficiency function is $\lambda_2(x, \theta)$ but the mean of the prior distribution E2 for θ_2 is misspecificed and $b = 1$ or $b = 5$. It is clear that the effect of misspecifying E2 can be significant; if E2 = E is the true value and E2 = E$'$ was assumed, the drop in efficiency can be as large as 99% (see Table 7 when E$'$ = 1 and E = 7). The corresponding decline in efficiency when a smaller design space is used ($b = 1$) is much reduced; for the same case just discussed, the drop is only about 21%. Although the design problem is not scale invariant, this difference is intuitively clear, because a misspecification of the mean of the prior distribution on the interval $[0, 5]$ corresponds to a 5 times larger misspecification on the interval $[0, 1]$. Thus for relatively large design spaces, extra care is necessary with respect to the choice of the prior distribution; this problem is less cruical if the design space is not too large.

There also appears to be no effect of the number of support points in the prior distribution on the performance of the optimal designs when the prior mean is misspecified. The tables show that the efficiencies are usually increased when the number of support points of the prior distributions is also increased.

Table 8 reports the efficiencies of the Bayesian 2 and 3-point D-optimal designs on $[0, 1]$ when model (1) holds and the efficiency function is misspecified. For space consideration, we display only the efficiencies of the Bayesian D-optimal 2 and 3-point designs when $\lambda_2(x, \theta)$ is the true efficiency function and $\lambda_3(x, \theta)$ is wrongly assumed. Only degenerate prior distributions are used; the prior distribution for θ_2 in $\lambda_2(x, \theta)$ is degenerate at E = 1, 4 or 7 and the prior distributions for θ in $\lambda_3(x, \theta)$ are degenerate at E1 = 0, 0.5, 1 or 3 and E2 = 0, 0.5, 1 or 3.

The results suggest that loses from a misspecified efficiency function can vary substantially. We note that if the roles of $\lambda_2(x, \theta)$ and $\lambda_3(x, \theta)$ are reversed, the misspecification becomes very costly. Most of the cells in the table are now zero because the information matrices of the Bayesian D-optimal $(n+1)$-point designs are now singular due to the location of the support points of the optimal designs and the form of the efficiency function $\lambda_3(x, \theta)$ (see Table 2 and 3).

Table 9 shows the values of the efficiency lower bounds of the Bayesian D-optimal $(n'+1)$-point designs for the model (1) and $\lambda_2(x, \theta)$ is the efficiency function. When there is no misspecification in the degree of the polynomial, i.e. $n = n'$, the values of the efficiency lower bounds are unity for the priors described in Table 1. This means that the Bayesian D-optimal $(n+1)$-point design is also Bayesian D-optimal. It is evident from the table that the number of support points in the prior distributions does not influence whether the Bayesian D-optimal $(n+1)$-point design is Bayesian D-optimal. The table does not give the exact efficiencies but only a lower bound for the efficiency of the Bayesian D-optimal $(n'+1)$-point designs. So if $n = 3$, $n' = 4$, s = 5 and E2 = 4, the Bayesian D-optimal 4-point design has at least 82.0% efficiency under two types of prior distributions π_1 and π_2. The exact efficiency cannot be calculated because the Bayesian D-optimal design is not known, but the actual efficiency could be substantially larger; see Dette (1995) for an example, where the lower bound and the exact efficiency can be calculated.

It is instructive to consider two new symmetric priors of θ_2 about its mean E2 with a larger spread: π_3 is supported at 0, 4 and 8 with mass at 4 equal to 0.6, and π_4 is equally supported at

Table 6: Relative efficiency of the Bayesian $(n+1)$-point D-optimal design on $[0,1]$ for model (1) with $n = 1$, 2, 3 and $\lambda_2(x, \theta)$ as the efficiency function. The prior distribution has the form π_1 with s = 3, 5 or 9 points. The true prior mean for θ_2 is E2 = E but the assumed mean is E$'$.

$n = 1$		E=1			E=4			E=7	
E$'$	s=3	s=5	s=9	s=3	s=5	s=9	s=3	s=5	s=9
1	1	1	1	0.736	0.736	0.736	0.287	0.287	0.287
4	0.642	0.642	0.642	1	1	1	0.827	0.827	0.827
7	0.408	0.408	0.408	0.877	0.877	0.877	1	1	1
$n = 2$		E=1			E=4			E=7	
E$'$	s=3	s=5	s=9	s=3	s=5	s=9	s=3	s=5	s=9
1	1	1	1	0.924	0.924	0.927	0.656	0.656	0.658
4	0.935	0.935	0.945	1	1	1	0.821	0.821	0.814
7	0.477	0.477	0.524	0.788	0.788	0.821	1	1	1
$n = 3$		E=1			E=4			E=7	
E$'$	s=3	s=5	s=9	s=3	s=5	s=9	s=3	s=5	s=9
1	1	1	1	0.929	0.929	0.931	0.755	0.755	0.762
4	0.931	0.931	0.941	1	1	1	0.939	0.939	0.938
7	0.780	0.780	0.806	0.944	0.944	0.954	1	1	1
$n = 4$		E=1			E=4			E=7	
E$'$	s=3	s=5	s=9	s=3	s=5	s=9	s=3	s=5	s=9
1	1	1	1	0.942	0.942	0.944	0.788	0.788	0.794
4	0.942	0.942	0.951	1	1	1	0.943	0.943	0.942
7	0.793	0.793	0.818	0.945	0.945	0.955	1	1	1

Table 7: Relative efficiency of the Bayesian $(n+1)$-point D-optimal design on $[0,5]$ for model (1) with $n = 1, 2, 3$ and $\lambda_2(x,\theta)$ as the efficiency function. The prior distribution has the form π_1 with s = 3, 5 or 9 points. The true prior mean for θ_2 is E2 = E but the assumed mean is E′.

$n=1$		E=1			E=4			E=7	
E′	s=3	s=5	s=9	s=3	s=5	s=9	s=3	s=5	s=9
1	1	1	1	0.199	0.199	0.199	0.017	0.017	0.017
4	0.529	0.529	0.529	1	1	1	0.827	0.827	0.827
7	0.337	0.337	0.337	0.877	0.877	0.877	1	1	1
$n=2$		E=1			E=4			E=7	
E′	s=3	s=5	s=9	s=3	s=5	s=9	s=3	s=5	s=9
1	1	1	1	0.040	0.040	0.048	0	0	0
4	0.280	0.280	0.352	1	1	1	0.683	0.683	0.681
7	0.113	0.113	0.159	0.769	0.769	0.810	1	1	1
$n=3$		E=1			E=4			E=7	
E′	s=3	s=5	s=9	s=3	s=5	s=9	s=3	s=5	s=9
1	1	1	1	0.095	0.095	0.106	0.001	0.001	0.001
4	0.182	0.182	0.238	1	1	1	0.565	0.565	0.561
7	0.047	0.047	0.073	0.675	0.675	0.729	1	1	1
$n=4$		E=1			E=4			E=7	
E′	s=3	s=5	s=9	s=3	s=5	s=9	s=3	s=5	s=9
1	1	1	1	0.178	0.178	0.189	0.003	0.003	0.003
4	0.169	0.169	0.221	1	1	1	0.467	0.467	0.463
7	0.028	0.028	0.045	0.592	0.592	0.657	1	1	1

Table 8: Efficiencies of the Bayesian 2(3)-point D-optimal design on $[0,1]$ when the true model is (1) for $n=1(2)$ with $\lambda_2(x,\theta)$ but $\lambda_3(x,\theta)$ was wrongly assumed with a degenerate prior at θ_2 at E = 1, 4 or 7. the prior distributions for θ in $\lambda_3(x,\theta)$ are degenerate at E1 = 0, 0.5, 1 or 3 and E2 = 0, 0.5, 1 or 3.

E=1				
	E2=0	E2=0.5	E2=1	E2=3
E1=0	1(0.990)	0.884(0.877)	0.789(0.768)	0.540(0.462)
E1=0.5	0.724(0.785)	0.707(0.742)	0.672(0.682)	0.527(0.462)
E1=1	0.564(0.629)	0.583(0.623)	0.577(0.594)	0.498(0.443)
E1=3	0.296(0.301)	0.335(0.331)	0.357(0.345)	0.378(0.330)
E=4				
	E2=0	E2=0.5	E2=1	E2=3
E1=0	1(0.860)	0.878(0.901)	0.955(0.901)	0.977(0.763)
E1=0.5	0.395(0.578)	0.520(0.645)	0.613(0.679)	0.767(0.662)
E1=1	0.252(0.405)	0.346(0.472)	0.425(0.516)	0.604(0.560)
E1=3	0.089(0.137)	0.125(0.174)	0.159(0.206)	0.278(0.287)
E=7				
	E2=0	E2=0.5	E2=1	E2=3
E1=0	1(0.574)	0.463(0.710)	0.615(0.811)	0.938(0.968)
E1=0.5	0.114(0.326)	0.203(0.431)	0.297(0.520)	0.592(0.728)
E1=1	0.060(0.200)	0.109(0.275)	0.166(0.344)	0.389(0.544)
E1=3	0.014(0.048)	0.025(0.071)	0.038(0.094)	0.109(0.191)

Table 9: Values of the efficiency lower bounds of the Bayesian $(n'+1)$-point D-optimal designs for estmating α in model (1) when the efficiency function is $\lambda_2(x,\theta) = \theta_1 exp(-\theta_2 x)$ and $1 = n < n'$ $= 4$. The first(second) line in each block shows the values of the efficiency lower bounds when the priors have the form $\pi_1(\pi_2)$, each with s = 3, 5 or 9 support points and marginal mean E2 = 1, 4 or 7.

		$n=1$			$n=2$			$n=3$		
n'	E2	s=3	s=5	s=9	s=3	s=5	s=9	s=3	s=5	s=9
2	1	0.773	0.772	0.771						
2	1	0.768	0.768	0.768						
2	4	0.717	0.716	0.716						
2	4	0.710	0.709	0.709						
2	7	0.702	0.777	0.704						
2	7	0.699	0.698	0.699						
3	1	0.701	0.700	0.700	0.810	0.809	0.809			
3	1	0.699	0.699	0.699	0.808	0.808	0.808			
3	4	0.613	0.613	0.612	0.791	0.791	0.790			
3	4	0.610	0.610	0.610	0.787	0.787	0.787			
3	7	0.567	0.569	0.569	0.775	0.777	0.778			
3	7	0.565	0.565	0.565	0.773	0.773	0.773			
4	1	0.667	0.667	0.667	0.731	0.730	0.730	0.839	0.838	0.838
4	1	0.665	0.665	0.665	0.729	0.730	0.729	0.837	0.837	0.837
4	4	0.559	0.559	0.559	0.691	0.690	0.690	0.908	0.830	0.829
4	4	0.557	0.557	0.557	0.689	0.689	0.689	0.827	0.827	0.827
4	7	0.504	0.505	0.505	0.664	0.665	0.665	0.822	0.824	0.824
4	7	0.502	0.502	0.502	0.663	0.663	0.663	0.821	0.821	0.821

1, 4 and 7. Consider also the prior π_1 in Table 1 with three support points at 2, 4 and 6 (i.e. δ = 2 and mean equal to 4). The variance of π_3 is 6.4, the variance of π_4 is 6.0 and the variance of π_1 is 1.33. So, π_3 is a bit more spread out than π_4 which in turn has a much wider spread than π_1. Because all these priors have the same marginal mean E2 = 4, the Bayesian D-optimal $(n+1)$-point designs are identical. However, the Bayesian D-optimal $(n+1)$-point designs can lose their universal D-optimality if the priors are more spread out. This is seen in Table 10 where the values of the efficiency lower bounds for the more spread out priors π_3 and π_4 are now less than unity when $n = n' = 1$, 2. This implies the Bayesian D-optimal $(n+1)$-point designs for the linear and quadratic models are no longer Bayesian D-optimal, even though the models are correctly specified. As pointed out earlier, this is not the case if the less spread out prior π_1 is used.

Table 10 also suggests that if a Bayesian D-optimal $(n+1)$-point design is not universally optimal for the polynomial of degree n, the Bayesian D-optimal k-point design is not likely to be universally optimal for the model of degree k if $k = n-1$. This is because in each of the cases in Table 10, the Bayesian D-optimal k-point design is also Bayesian D-optimal when $k = 3$ but not if k is 1 or 2. This rule can be helpful if we wish to ascertain if a Bayesian D-optimal $(n+1)$-point design is also Bayesian D-optimal because the computation for finding a Bayesian D-optimal $(n+1)$-point design can be laborious. In summary, our results suggest that if concentrated priors are used with more mass concentrated around its mean, the Bayesian D-optimal $(n+1)$-point design is more likely to be universally optimal for the models considered here.

Table 10: First rows: Efficiencies of Bayesian D-optimal $(n'+1)$-point designs on $[0,1]$ relative to the Bayesian D-optimal $(n+1)$-point designs on $[0,1]$ using priors π_3 and π_4. Second rows: Values of the efficiency lower bounds of the Bayesian D-optimal $(n'+1)$-point designs when the true model is of degree n and the assumed degree is n'.

	prior distribution π_3				prior distribution π_4			
n'	$n=1$	$n=2$	$n=3$	$n=4$	$n=1$	$n=2$	$n=3$	$n=4$
1	1				1			
	0.741				0.863			
2	0.947	1.000			0.951	1.000		
	0.740	0.892			0.742	0.936		
3	0.895	0.967	1.000		0.891	0.965	1.000	
	0.627	0.804	1.000		0.625	0.805	1.000	
4	0.865	0.922	0.953	1.000	0.862	0.917	0.951	1.000
	0.568	0.699	0.839	1.000	0.567	0.699	0.839	1.000

We now compare the performance of Bayesian D-optimal $(n+1)$-point designs with uniform designs. Table 11 lists the efficiencies of uniform designs supported on 2, 3 and 4 points on $[0,1]$ and $[0,5]$ relative to the Bayesian D-optimal $(n+1)$-point designs when $\lambda_2(x,\theta)$ is the efficiency function and $n = 1$, 2 and 3. The priors used in the comparison have the form of π_1 with the mean of the marginal prior distribution of θ_2 equal to E2 = 1, 4 or 7 and are supported at s = 3, 5, or 9 points. The relative efficiencies do not change much when the number of support points in the prior distribution is changed and so we display only the efficiencies when s = 5.

For $\lambda_2(x, \theta)$, the error variance is increasing over the design space; this increase is more dramatic especially when the marginal prior mean of θ_2, E2, is large. The Bayesian D-optimal $(n+1)$-point design puts more support points near 0 while keeping the range of its support as large as possible. As can be seen from the table, the uniform design generally has lower efficiency if E2 is large (E2=7, for example). In addition, when the design space is enlarged from $[0, 1]$ to $[0, 5]$, we observe there is generally a substantial drop in the efficiencies of the uniform designs, with some approaching zero. This implies that Bayesian D-optimal $(n+1)$-point designs can outperform uniform designs by a wide margin when the design space is large.

Table 11: Efficiencies of a 2-point uniform design (Δ_2), 3-point uniform design (Δ_3) and 4-point uniform design (Δ_4) on $[0, 1]$ and $[0, 5]$ relative to the Bayesian D-optimal $(n+1)$-point design for model (1) for $n = 1$, 2 or 3 and $\lambda_2(x, \theta)$ holds. The priors have the form π_1 with s = 5 support points and the mean of the marginal distribution of θ_2 is E2 = 1, 4 or 7.

		$n = 1$		$n = 2$		$n = 3$	
	E2	$b = 1$	$b = 5$	$b = 1$	$b = 5$	$b = 1$	$b = 5$
	1	1.000	0.558				
Δ_2	4	0.736	0.001				
	7	0.287	0.000				
	1	0.841	0.848	0.990	0.796		
Δ_3	4	0.849	0.061	0.860	0.007		
	7	0.593	0.003	0.574	0.000		
	1	0.769	0.847	0.907	0.948	0.951	0.782
Δ_4	4	0.801	0.163	0.896	0.066	0.841	0.023
	7	0.618	0.023	0.784	0.001	0.650	0.000

5. SUMMARY

We investigated the robustness properties of the Bayesian D-optimal $(n + 1)$-point designs for the polynomial regression model when the error variance depends on the explanatory variable. Our main interest was to find efficient designs for estimating the structural parameters using prior information of the nuisance parameters in the efficiency function. Our results provided some insight into the Bayesian D-optimal $(n+1)$-point designs for the polynomial regression model. In essence, these designs are robust if the degree is not completely misspecified. The designs have moderate performance if the means of the marginal distributions are nearly correct. This is especially so if the design space is small, or if the design space is large. Our results also suggest that the designs are less efficient if the variance of the prior distribution is increasing and they can perform much better than the widely used uniform designs when the design space is large. We conclude with a note that if there is interest in estimating all the model parameters β or a subset of the parameters of θ, a similar approach may be used; some analytical results are given in Dette and Wong (1998).

6. APPENDIX

Here, we provide information for carrying out the numerical study. Details and proofs for Theorems 1 and 2 are available in Dette and Wong (1998).

THEOREM 1. *Suppose model (1) holds on the design space $[0,b]$ and the efficiency function is $\lambda_2(x,\theta) = \theta_1 exp(-\theta_2 x),\ \theta_1 > 0$. Let $E = E_\pi(\theta_2) > 0$ and $z_{max,n}$ denote the largest zero of the nth Laguerre polynomial $L_n(x)$.*

a. *If $bE \geq z_{max,n}$, the Bayesian D-optimal $(n+1)$-point design for estimating α has equal mass at the zeros of the polynomial $xL_n(xE)$.*

b. *If $bE < z_{max,n}$, the Bayesian D-optimal $(n+1)$-point design for estimating α is determined by its canonical moments $p_1, p_2, ..., p_{2n-1} \in (0,1), p_{2n} = 1$, which solve the system of equations: $(q_0 = 0)$*

$$\begin{aligned}\frac{n-k+1}{p_{2k-1}} - \frac{n-k+1}{q_{2k-1}} - bE_{q_{2k-2}} + bE_{p_{2k}} &= 0,\ k = 1,\dots,n \\ \frac{n-k+1}{p_{2k}} - \frac{n-k+1}{q_{2k}} - bE_{q_{2k-1}} + bE_{p_{2k+1}} &= 0,\ k = 1,\dots,n-1.\end{aligned}$$

Moreover, this design has equal mass at its support points.

THEOREM 2. *Under the setup in Theorem 1, the design ξ is Bayesian D-optimal for estimating α if and only if*

$$\int_B exp(-\theta_2 x) f^T(x) M_1^{-1}(\theta_2,\xi) f(x)\pi(d\theta_2) \leq n+1 \qquad \forall\ x \in [0,b]. \tag{2}$$

Moreover, equality holds in the inequality if x is a support point of the Bayesian D-optimal design. If the design ξ is not Bayesian D-optimal, the proximity of ξ to the Bayesian D-optimal design ξ^* can be quantified by a lower bound for $exp\{(\Phi(\xi) - \Phi(\xi^*))/(n+1)\}$. This bound is given by

$$exp\{(\Phi(\xi) - \Phi(\xi^*))/(n+1)\} \geq \frac{n+1}{\sup_{x\in[0,b]} \int_B exp(-\theta_2 x) f^T(x) M_1^{-1}(\theta_2,\xi) f(x)\pi(d\theta_2)}\ .$$

THEOREM 3. *Suppose model (1) holds on $[0,b]$ and the efficiency function is $\lambda_3(x,\theta) = x^{\theta_1}(b-x)^{\theta_2},\ \theta^T = (\theta_1,\theta_2),\ \theta_1 \geq 0$ and $\theta_2 \geq 0$. If $\pi(\theta_1,\theta_2)$ is the prior distribution for θ, the Bayesian D-optimal $(n+1)$-point design for estimating α puts equal masses at the $n+1$ zeros of the polynomial $P_{n+1}^{(\nu-1,\mu-1)}((2x/b)-1)$. Here $P_{n+1}^{(\alpha,\beta)}(t)$ is the $(n+1)$th Jacobi polynomial orthogonal with respect to the measure $(1-x)^\alpha(1+x)^\beta dx$ and the parameters (μ,ν) are the means of the first and second marginal distributions of θ respectively, i.e.*

$$\mu = \int_0^\infty \int_0^\infty \theta_1\ d\pi(\theta_1,\theta_2) \qquad and \qquad \nu = \int_0^\infty \int_0^\infty \theta_2\ d\pi(\theta_1,\theta_2)\ .$$

PROOF. The design problem is invariant under linear transformations, and in particular, to the transformation from x to $2x/b-1$. After the transformation, the design problem is equivalent to finding a Bayesian D-optimial $(n+1)$-point design on $[-1,1]$ for estimating α in model (1) with $\lambda_3(x,\theta)$. The solution to the latter problem was given in Theorem 3.4 of Dette and Wong (1998).

ACKNOWLEDGEMENTS

The research of Wong is partially supported by a NIH grant R29 AR44177-01A1. The work of H. Dette was supported by the SFB 475 (Komplexitätsreduktion in multivariaten Datenstrukturen).

REFERENCES

Atkinson, A. C. and Cook, R. D. (1995). D-optimal designs for heteroscedastic linear models. *Journal of the American Statistician Association*, 90, 204-212.

Box, G. E. P. and Hill, W. J. (1974). Correcting inhomogeneity of variance with power transformation weighting. *Technometrics*, 16, 385-389.

Carroll, R. J. (1982a). Adapting for heteroscedasticity in linear models. *Annals of Statistics*, 10, 1224-1233.

Carroll, R. J. (1982b). Robust estimation in certain heteroscedastic linear models when there are many parameters, *Journal of Statistical Planning and Inference*, 7, 1-12.

Carroll, R. J. and Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. *Annals of Statistics*, 10, 429-441.

Chaloner, K. and Verdnelli, I. (1995). Bayesian Experimental Design: A review. *Statistical Science*, 10, 237-304.

Cook, R. D. and Nachtsheim, C. J. (1982). Model robust, linear-optimal designs. *Technometrics*, 24, 49-54.

Cook, R. D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, 70, 1-10.

Dasgupta, A., Mukhopadhyay, S. and Studden, W. J. (1992). Compromise designs in heteroscedastic linear models. *Journal of Statistical Planning and Inference*, 32, 363-384.

Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistician Association*, 82, 1079-1091.

Dette, H. (1995). A lower bound for efficiencies with applications. *In Research Developments in Probability and Statistics: Festschrift in Honor of Madan Puri on the occasion of his 65th birthday, Ed. E. Brunner and M. Denker*. Utrecht: VSP International Science.

Dette, H. and Munk, A. (1998). Testing heterosedasticity in nonparamteric regression. *Journal of the Royal Statistical Society, Series B*, 60, 693-708.

Dette, H. and Studden, W. J. (1997). The theory of canonical moments with Applications in Statistics, Probability and Analysis. *John Wiley & Sons Inc.*

Dette, H. and Wong, W. K. (1995). Bayesian optimal designs for models with partially specified heteroscedastic structure. *The Annals of Statistics*, 24, 2108-2127.

Dette, H. and Wong, W. K. (1998). Bayesian D-Optimal designs on a fixed number of points for heteroscedastic polynomial models. *Biometrika*, 85, 869-882.

Engel, J. (1992). Modelling variation in industrial experiments. *Applied Statistics*, 41, 579-593.

Fedorov, V. V (1972). Theory of optimal experiments. Translated and edited by W. J Studden and E. M Klimko, New York; *Academic Press.*

Schulz, M. and Endrenyi, L. (1983). Design of experiments for estimating parameters with unknown heterogeneity of the error variance. *Proceedings of the American Statistical Association Computing Section*, 177-181.

Silvey, S. D. (1980). Optimal design. *Chapman and Hall. London.*

Studden, W. J. (1982). Some robust-type D-optimal designs in polynomial regression. *Journal of the American Statistician Association*, 77, 916-921.

Wiens, D. (1998). Minimax robust designs and weights for approximately specified regression models with heteroscedastic errors. *Journal of the American Statistician Association*, 93, 1440-1 1450.

Wolfram, S. (1988). Mathematica^{TM}; A system for doing Mathematics by computer. *Addison-Wesley Publishing Company.*

Fakultät und Institut für Mathematik
44780 Bochum, Germany

Dale SONG: `e-mail ???`
Scirex Corporation
225 E Lake Street, Bloomingdale
IL 60108

Weng Kee WONG: `wkwong@sunlab.ph.ucla.edu`
Department of Biostatistics
10833 Le Conte Ave, UCLA, Los Angeles
CA 90095-1772