A comparison of sequential and non-sequential designs for discrimination between nested regression models

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: holger.dette@ruhr-uni-bochum.de
FAX: +49 2 34 3214 559

Robert Kwiecien
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: robert.kwiecien@ruhr-uni-bochum.de

July 24, 2002

Abstract

Classical regression analysis is usually performed in two steps. In a first step an appropriate model is identified to describe the data generating process and in a second step statistical inference is performed in the identified model. In this paper we investigate various design strategies which take into account these different goals of the analysis for a class of nested models. A detailed comparison of sequential and non-sequential designs for model discrimination in polynomial and Fourier regression models is given and the difference between the two concepts are illustrated. It is demonstrated that non-sequential designs usually identify the "correct" model with a higher probability than the sequential methods. Additionally, the efficiencies of the non-sequential designs for the estimation of the parameters in the "correct" model are at least of the same size than the corresponding efficiencies of the sequential methods.

Keywords and Phrases: optimal design, robust design, discrimination design, sequential design, F-test, polynomial and Fourier regression

1 Introduction

Classical design theory assumes precise knowledge of the underlying model of the data generating process [see e.g. Silvey (1980)]. However, in many applications such knowledge is not available and the analysis of the data is usually performed in two steps. The data is firstly used to identify an appropriate model from a given class of competing models and the second step consists in the statistical analysis in the identified model (parameter estimation, prediction etc.). An optimal design for one task may be exceptionally inefficient for the other. Consider for example the regression model

$$Y = f(X) + \varepsilon, \tag{1.1}$$

where ε has a centered normal distribution. The real valued function f is assumed to belong to a given class of linear nested models, say

$$\mathcal{F} = \{g_1, \ldots, g_k\}, \tag{1.2}$$

where

$$g_j(x) = \sum_{i=1}^{\ell_j} \beta_{j,i} f_i(x), \qquad j = 1, \ldots, k, \tag{1.3}$$

are the competing nested models, $1 \leq \ell_1 < \ell_2 < \ldots < \ell_k$ and $f_1, \ldots, f_{\ell_k}$ are given and known regression functions.

Typically, the discrimination between these models is performed by a sequence of tests for the hypotheses

$$H_{0j} : f = g_{j-1} \quad \text{versus} \quad H_{1j} : f = g_j,$$

[see e.g. Anderson (1962)], while the inference in the identified model is usually done by classical methods for linear regression [see e.g. Seber (1977)]. In the literature there are essentially two proposals to determine efficient designs for the rather different objects of discrimination and inference in the identified model. Numerous authors propose to combine all aspects of interest in one design criterion [see e.g. Atkinson and Cox (1974), Läuter (1974), Dette (1990, 1994, 1995), Spruill (1990), Pukelsheim and Rosenberger (1993) among many others]. The resulting optimality criteria are called compound or composite optimality criteria and the corresponding design is non-sequential in the sense that all observations are taken at one stage. The second approach is based on sequential methods [see e.g. Andrews (1971), Montepiedra and Yeh (1998) or Biswas and Chaudhuri (2002)]. While statistical analysis based on data from a non-sequential design can essentially be carried out by standard methods, the analysis based on data from a sequential design is usually very difficult [see Silvey (1980)]. For example, it is not obvious that estimates in the identified model are consistent or (asymptotically) normal distributed with the appropriate variances and consequently tests or confidence intervals may not keep the required accuracy. The reason for these difficulties is that a sequential design procedure usually yields dependencies in the data and classical estimation and distribution theory is not directly applicable. Therefore most sequential designs for model discrimination and estimation are proposed on a basis of heuristic arguments, which roughly speaking ignore the dependencies caused by the sequential sampling [see e.g. Montepiedra and Yeh (1998, 2002)].
Recently, Biswas and Chaudhuri (2002) proposed in a remarkable paper a sequential strategy, where the object of the design is to select the correct model from the family of nested models as well as to efficiently estimate the parameters associated with that model. They showed that the corresponding tests used in the discrimination step keep their preassigned levels and that the sequential design is able to identify the correct model with a probability converging to one if the sample size tends to infinity. Moreover, the sequential design converges to the optimal design corresponding to the "true" model.

The purpose of the present note is to compare the sequential discrimination designs with the non-sequential discrimination designs and model robust designs introduced by Dette (1990, 1994 and 1995). In Section 2 we will briefly review the different concepts. Section 3 contains a detailed comparison of the different concepts for lower order polynomial and Fourier regression models. We investigate by a simulation study the rate of misspecifications and the efficiencies of the different designs in the identified model. It is demonstrated that non-sequential procedures usually yield substantially smaller rates of misspecification compared to the sequential procedures. Additionally, the efficiencies of the non-sequential designs for estimating the parameters in the identified model are not substantially worth than the corresponding efficiencies of the sequential designs (in many cases they are even better). Finally, some conclusions and recommendations for designing experiments with respect to the different goals of model identification and parameter estimation are given in Section 4. Our results demonstrate that non-sequential designs should usually be preferred in this context.

2 Different design strategies revisited

2.1 Non-sequential discrimination and model robust designs

The strategy of combining two or more aspects in one optimality criterion was firstly considered by Läuter (1974) [see also Dette (1990), Pukelsheim and Rosenberger (1993) among many other authors]. For computational reasons most authors use optimality criteria based on determinants and for the sake of transparency we will restrict our investigations to this type of criteria. To be precise let $X_j \in \mathbb{R}^{n\times \ell_j}$ denote the design matrix in the model g_j defined in (1.3) based on the total sample of n observations $(j = 1, \dots, k)$. Noting that the jth model contains ℓ_j unknown parameters Läuter (1974) proposed to choose the design points such that the geometric mean

$$\prod_{j=1}^{k} |X_j^T X_j|^{\lambda_j/\ell_j} \tag{2.1}$$

becomes maximal, where $|A|$ denotes the determinant of the matrix A. Here the quantities λ_j are nonnegative weights with sum 1 reflecting the experimenter's prior belief about the adequacy of the jth model. We call a design maximizing the function in (2.1) an *optimal model robust design* for the models $g_1, \dots, g_k$ [with respect to the prior $\lambda = (\lambda_1, \dots, \lambda_k)$]. In the case of Fourier regression and univariate and multivariate polynomial models optimal model robust approximate designs (in other words probability measures with finite support on the design space) have been determined by Lau and Studden (1985), Dette (1990) and Dette and Röder (1995), respectively.

For the purpose of model discrimination a different optimality criterion is appropriate. Note that an efficient discrimination between the model g_{j-1} and g_j requires a precise estimation of the "highest" coefficients $\beta_{j,\ell_{j-1}+1}, \dots, \beta_{j,\ell_j}$ corresponding to the functions $f_{\ell_{j-1}+1}, \dots, f_{\ell_j}$ in the jth model defined by (1.3). Define

$$K_j^T = [0 \mid I_{\ell_j-\ell_{j-1}}] \in \mathbb{R}^{\ell_j-\ell_{j-1}\times \ell_j}, \tag{2.2}$$

where $I_s \in \mathbb{R}^{s\times s}$ denotes the identity matrix, then it is well known [see e.g. Silvey (1980)] that the volume of the ellipsoid of concentration for the parameters $\beta_{j,\ell_{j-1}+1}, \dots, \beta_{j,\ell_j}$ is proportional to the determinant

$$|K_j^T (X_j^T X_j)^{-1} K_j|^{-1/2}.$$

Consequently, a good discrimination design for the models $g_1, \ldots, g_k$ should make these quantities as small as possible, and following Läuter (1974) we propose to choose the design such that

$$(2.3) \qquad \prod_{j=2}^{k} |K_j^T (X_j^T X_j)^{-1} K_j|^{\lambda_j/(\ell_j - \ell_{j-1})}$$

becomes maximal. A design maximizing the function in (2.3) is called *optimal discriminating design* for the models $g_1, \ldots, g_k$ [with respect to the prior $\lambda = (\lambda_1, \ldots, \lambda_k)$]. Optimal discriminating designs for polynomial and Fourier regression models have been determined by Spruill (1990), Dette (1994, 1995), Dette and Röder (1997) and Dette and Haller (1998). We finally note that there exists a sequence of step-wise F-tests corresponding to the criterion (2.3) [see Anderson (1962) for some optimality properties of this multiple decision procedure]. More precisely, starting with the model g_k the hypotheses

$$(2.4) \qquad H_{0j} : K_j^T \beta_j = \begin{pmatrix} \beta_{j,\ell_{j-1}+1} \\ \vdots \\ \beta_{j,\ell_j} \end{pmatrix} = 0 \qquad j = k, \ldots, 2$$

with $\beta_j^T = (\beta_{j,1}, \ldots, \beta_{j,\ell_j})^T$ are subsequently tested at a specified level, say $\alpha_j \in (0,1)$ and we select the model g_{j_0} for which the first test rejects the corresponding hypothesis. The test in the jth step is the classical F-test, which is based on the statistic

$$(2.5) \qquad F_j = \frac{RSS_{j-1} - RSS_j}{RSS_j} \cdot \frac{N - \ell_j}{\ell_j - \ell_{j-1}},$$

where RSS_j denotes the residual sum of squares based on a least squares fit in the model g_j from the total sample. Note that under the null-hypothesis (2.4) the statistic F_j has an F-distribution with $\ell_j - \ell_{j-1}$ and $N - \ell_j$ degrees of freedom, which allows a simple determination of the critical values. It is easy to see that for any design with determinant $\frac{1}{n}|X_k^T X_k| \geq c > 0$ in the model g_k, increasing sample size and levels converging to 0, this procedure identifies the "correct" model with a probability converging to 1. This remark applies in particular if the optimal model robust designs [maximizing the function (2.1)] or the optimal discrimination designs [maximizing the function (2.3)] are used in this selection procedure.

2.2 A sequential strategy

In this paragraph we briefly describe a promising sequential approach, which was recently proposed by Biswas and Chaudhuri (2002). The method starts with a convex combination of the D-optimal designs for the individual models and this design is updated in several steps. To be precise, let ξ_j denote the approximate [in the sense of Kiefer (1974)] D-optimal design for the regression model g_j $(j = 1, \ldots, k)$ and assume that $N = m_0 + m_1 + \ldots + m_s$ experiments are permitted. Let $\alpha_i^{(0)} = \frac{1}{k}$ $(i = 1, \ldots, k)$ and consider as the design for the first stage

$$(2.6) \qquad \xi^{(0)} := \sum_{i=1}^{k} \alpha_i^{(0)} \xi_i$$

the uniform mixture of the D-optimal designs for the models $g_1, \ldots, g_k$. The first m_0 observations are chosen at experimental conditions sampling randomly from the design $\xi^{(0)}$, which defines the initial design. This design is sequentially updated by $s \geq 0$ steps in the following way. For $r = 1, \dots, s$ the sample of $m_0 + \dots + m_{r-1}$ observation is used to test subsequently the hypotheses

$$H_{0j} : K_j^T \beta_j = 0 \quad \text{versus} \quad H_{1j} : K_j^T \beta_j \neq 0 \tag{2.7}$$

($j = k, \dots, 2$) where the matrix K_j is defined in (2.2) and $\beta_j = (\beta_{j,1}, \dots, \beta_{j,\ell_j})^T$ is the vector of parameters in the model g_j. Note that the hypothesis H_{0j} is valid if and only if the parameters $\beta_{j,\ell_{j-1}+1}, \dots, \beta_{j,\ell_j}$ in the model g_j vanish simultaneously, which means that the model g_{j-1} should be preferred instead of g_j. The hypothesis H_{0j} will be rejected with some specified level α_r^j if $T_r^j > c_r^j$, where the statistic T_r^j is defined by

$$T_r^j = \frac{\sum_{i=0}^{r-1} \left\{ RSS_{j-1}^{(i)} - RSS_j^{(i)} \right\}}{\sum_{i=0}^{r-1} RSS_j^{(i)}} \cdot \frac{\sum_{i=0}^{r-1} (m_i - \ell_j)}{r(\ell_j - \ell_{j-1})}, \tag{2.8}$$

and $RSS_j^{(i)}$ denotes the residual sum of squares based on a least squares fit in the model g_j from the m_i observations in the ith step ($i = 0, \dots, r-1$). Note that in contrast to the classical F-statistic defined in (2.5) these sums are calculated seperately for each sample of m_i observations ($i = 0, \dots, r-1$). It was shown by Biswas and Chaudhuri (2002) that the critical values c_r^j can be obtained as the α_r^j quantiles of an F-distribution with $r(\ell_j - \ell_{j-1})$ and $\sum_{i=0}^{r} (m_i - \ell_j)$ degrees of freedom.
The design for the next m_r observations is then defined as follows. Let j_0 denote the first index for which the null-hypotheses $H_{0k}, \dots H_{0j_0+1}$ are accepted and the null-hypothesis H_{0j_0} is rejected (if all tests accept the corresponding null-hypothesis we put $j_0 = 1$), then the model g_{j_0} is selected and the design is updated by

$$\xi^{(r)} = \sum_{i=1}^{k} \alpha_i^{(r)} \xi_i, \tag{2.9}$$

where the new weights $\alpha_i^{(r)}$ are defined by

$$\alpha_i^{(r)} = \begin{cases} \left(\alpha_i^{(r-1)} + \frac{1}{(k+r-1)} \right) / \sum & \text{if } i = j_0 \\ \alpha_i^{(r-1)} / \sum & \text{if } i \neq j_0 \end{cases}$$

and $\sum$ is a normalizing constant defined by the condition $\sum_{i=1}^{k} \alpha_i^{(r)} = 1$. The next m_r design points are then generated from the design $\xi^{(r)}$.
This procedure is repeated to obtain $m_0 + \dots + m_s = N$ observations, where the observations at the rth stage are taken by sampling randomly from the design $\xi^{(r)}$. Finally, the sequence of tests for the hypotheses $H_{0k}, H_{0k-1}, \dots, H_{02}$ based on the statistics $T_{s+1}^k, \dots, T_{s+1}^2$ is performed for the total sample and the model g_{j_0} is chosen for which the corresponding test rejects the null-hypothesis for the first time (in other words: $T_{s+1}^i \leq c_{s+1}^i; i = k, \dots, j_0 + 1; T_{s+1}^{j_0} > c_{s+1}^{j_0}$). It was shown by Biswas and Chaudhuri (2002) that under appropriate (asymptotic) assumptions on $N, s, m_0, \dots, m_s$ and α_r^j ($r = 0, \dots, s, j = 2, \dots, k$) this procedure identifies the "correct" model with probability converging to one. Moreover, the information matrix of the design $\xi^{(s)}$ in the identified model , say g_{j_0}, converges to the information matrix of the D-optimal design for this model, say M_{j_0}, and the corresponding parameter estimate in the identified model has an asymptotic normal distribution with asymptotic mean $\beta_{j_0} \in \mathbb{R}^{\ell_{j_0}}$ and covariance matrix $\frac{1}{N} M_{j_0}^{-1}$ [see Biswas and Chaudhuri (2002) for more details].

2.3 A two stage procedure

Recently, Montepiedra and Yeh (2002) proposed a two stage procedure for the construction of efficient designs for model discrimination and estimation of the parameters in the identified model. The total sample of N observations is splitted into two parts, say $n_0 + n_1 = N$. In a first step n_0 observations are taken according to a design, which maximizes the criterion (2.3) in order to apply an efficient design for model discrimination. The data obtained from this design is used to select the appropriate model among $g_1, \dots, g_k$, say g_{j_0}. Here the sequence of F-tests defined in (2.5) is used for the discrimination step. For the second step let $X_{j_0}(n_0) \in \mathbb{R}^{n_0 \times \ell_{j_0}}$ denote the design matrix of the design from the first step in the model g_{j_0}, then the next n_1 observations are taken such that the determinant

$$|f \cdot X_{j_0}^T(n_0)X_{j_0}(n_0) + (1-f) \cdot X_{j_0}^T(n_1)X_{j_0}(n_1)| \tag{2.10}$$

becomes maximal, where $X_{j_0}(n_1)$ is the design matrix in the model g_{j_0} obtained from these new observations and $f \in (0,1)$ is a weight to be determined by the experimenter. After all $N = n_0 + n_1$ observations have been collected from the two stage design, the model selection step will be repeated using the total sample and the parameters in the identified model are estimated. Although this procedure is intuitively appealing, it has not been shown so far that this will yield a consistent procedure in a strong statistical sense. For example, the (F)-tests in the final step will usually not keep the preassigned level [see Section 3 for more details] and it is not obvious that for an increasing sample size (in both steps) the "correct" model is identified with a probability converging to 1. We finally note that the construction of the design maximizing the criterion (2.10) in the two stage procedure is closely related to the Bayesian design problem for linear regression models and for this reason not an easy problem [see e.g. Chaloner (1984) and Pukelsheim (1993), p. 275].

3 A comparison of sequential and non-sequential designs

Usually efficiencies [see Pukelsheim (1993), p. 132] are used to compare different designs, but due to the randomness of the sequential procedures these efficiencies cannot be calculated. In this section we apply a different approach to evaluate the performance of the different designs, which is based on a detailed simulation study of the sequential and non-sequential procedures and takes into account the two different goals of the design of experiment. To be precise we simulated 10.000 times a specific scenario for the designs described in Section 2. In each simulation we investigate the performance of the different designs and discrimination strategies, where the level of the corresponding F-tests was always taken as 5%. These simulations are used for two purposes:

- The calculation of the rate of a correct identification of the underlying model (we simply count how often the "correct" model is chosen in the 10.000 trials).
- The determination of the D-efficiency

  $$|X_{j_0}^T X_{j_0}|^{1/\ell_j}$$

  in the identified model. This calculation is no problem for a non-sequential design. For the sequential designs we used the 10.000 trials to define an "averaged" sequential design and this design is used for the efficiency calculations (for more details see Section 3.1 and 3.2).

We considered two cases of regression models, the cubic polynomial model and the Fourier-regression model of degree 2, for which the D-optimal designs in the nested submodels (which are needed in the sequential procedures) and the optimal model robust and optimal discrimination designs are known from the literature. For the sake of simplicity we restrict ourselves to the approximate design case and in our simulation study the experimental conditions are chosen randomly from the corresponding approximate designs.

3.1 Discrimination designs for the cubic regression model

Consider the cubic regression model

$$g_3(x) = \beta_{30} + \beta_{31}x + \beta_{32}x + \beta_{32}x^2 + \beta_{33}x^3; \tag{3.1}$$

on the interval $[-1, 1]$. The corresponding nested models are the linear and quadratic regression model, i.e.

$$g_1(x) = \beta_{10} + \beta_{11}x; \quad g_2(x) = \beta_{20} + \beta_{21}x + \beta_{22}x^2. \tag{3.2}$$

In this case we have $k = 3, \ell_1 = 2, \ell_2 = 3, \ell_3 = 4$ and the matrix K_j in (2.3) reduces to the $(j+1)$th unit vector in $\mathbb{R}^{j+1}$ $(j = 2, 3)$.
The D-optimal designs are known to have equal masses at the point $-1, 1$ (for the linear regression model g_1), $-1, 0, 1$ (for the quadratic regression function g_2) and $-1, -1/\sqrt{5}, 1/\sqrt{5}, 1$ (for the cubic model g_3), respectively [see Pukelsheim (1993), p. 217-218)]. Our first simulations serve for a comparison of different partitions in the sequential scheme of Biswas and Chaudhuri (2002) (see Section 2.2). For the sake of brevity we only report the results for the sample size $N = 100$, simulations for different sample sizes ($N = 50$ and $N = 200$) showed a very similar picture. We considered the partitions

$$\begin{array}{lll} m_0 & = \ldots = m_9 = 10 & (A) \\ m_0 & = \ldots = m_4 = 20 & (B) \\ m_0 & = m_1 = m_2 = 30, m_3 = 10 & (C) \\ m_0 & = m_1 = 50 & (D) \end{array} \tag{3.3}$$

for the 100 observations. Figures 3.1a and 3.1b show the rate of correct specification if the cubic or quadratic model is "correct", respectively. The model under consideration was

$$g_3(x) = 1 + x + x^2 + b_{33}x^3 \quad (b_{33} \neq 0)$$

in the cubic case and

$$g_2(x) = 1 + x + b_{22}x^2$$

for the quadratic case. If the linear model $g_1(x) = 1 + b_{11}x$ is the "correct" one, all designs in (3.3) behave similary and for this reason the results are not depicted. Note that the rate of correct specification is a function of the parameter of the absolute value of the highest coefficient, similar to the power function of the corresponding F-test. For example if the "correct" model is quadratic but the coefficient of x^2 is small in absolute value the model can hardly be distinguished from the linear model. From Figure 3.1a and 3.1b we observe that the rate of correct specification of the sequential procedure of Biswas and Chaudhuri (2002) is improved, if large sample sizes are chosen in the first steps of the sequential procedure. In other words: the allocation scheme (D) yields

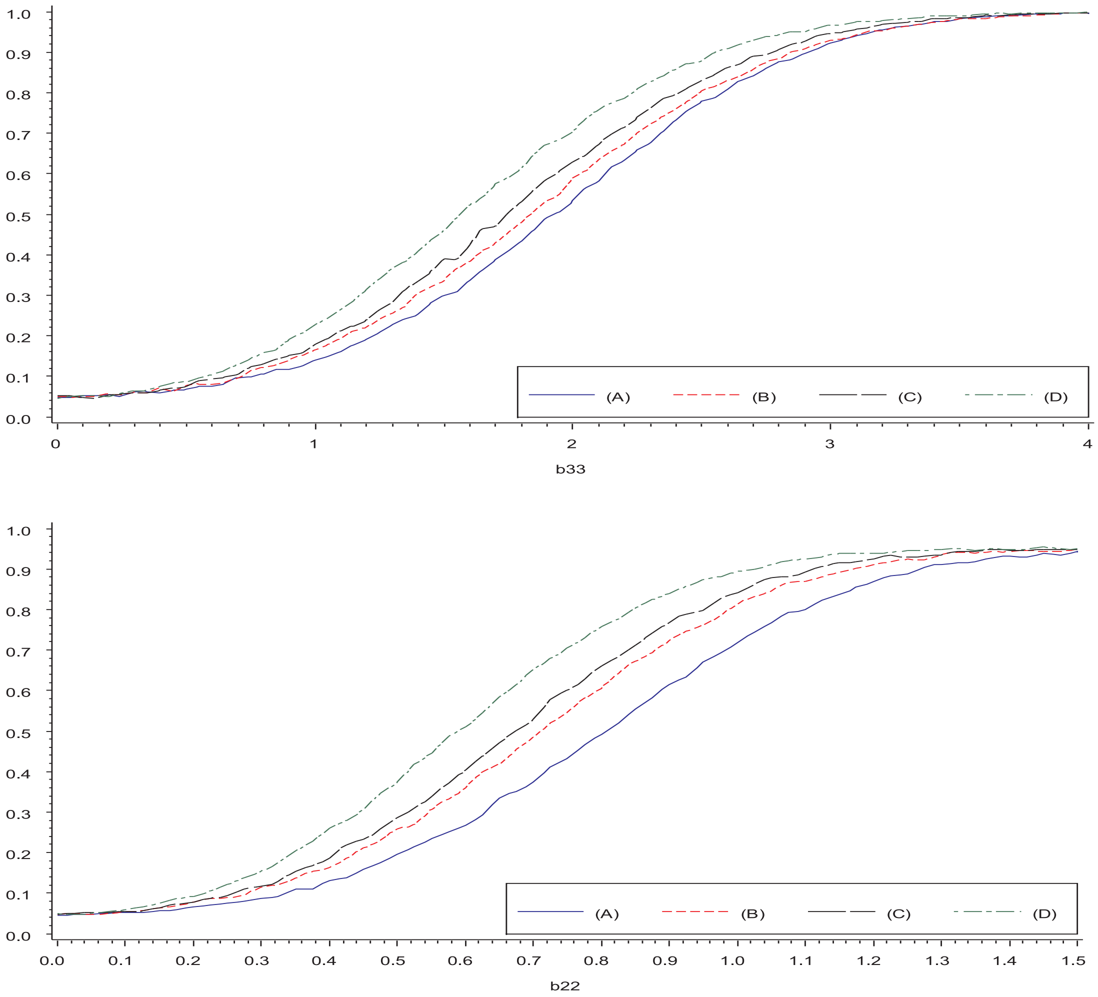


Figure 3.1: *Simulated probabilities of correct specification for the sequential designs (3.3) proposed by Biswas and Chaudhuri (2002) in the polynomial regression model (3.1). Figure 3.1a (upper panel): cubic model is "correct"; Figure 3.1b (lower panel): quadratic model is "correct".*

the uniformly best rates of "correct" model specification, independently of the "correct" model. This performance can even be improved by choosing all 100 observations in the first step which does not yield a sequential procedure. This at a first glance strange behaviour can be explained by the definition of the statistic (2.8) which is used in the different F-tests of the sequential discrimination procedure. The residual sums of squares $RSS_j^{(i)}$ are calculated separately for each subsample and then added, which yields to a substantial loss of power, if the sample sizes for the initial sequential stages are too small.

The results for the corresponding efficiencies of the "averaged" designs are presented in Figure 3.2a and 3.2b. From Figure 3.2a we observe the opposite behaviour for the efficiencies in the cubic case. Here the sequential design (A) yields the best efficiency (in the average) and the performance is usually improved by using finer partitions. On the other hand, if the quadratic model is correct, the sequential design with the finest partition has only a better performance if the parameter β_{22} is relatively small, for moderate values of β_{22}, the design (C) is preferrable, while

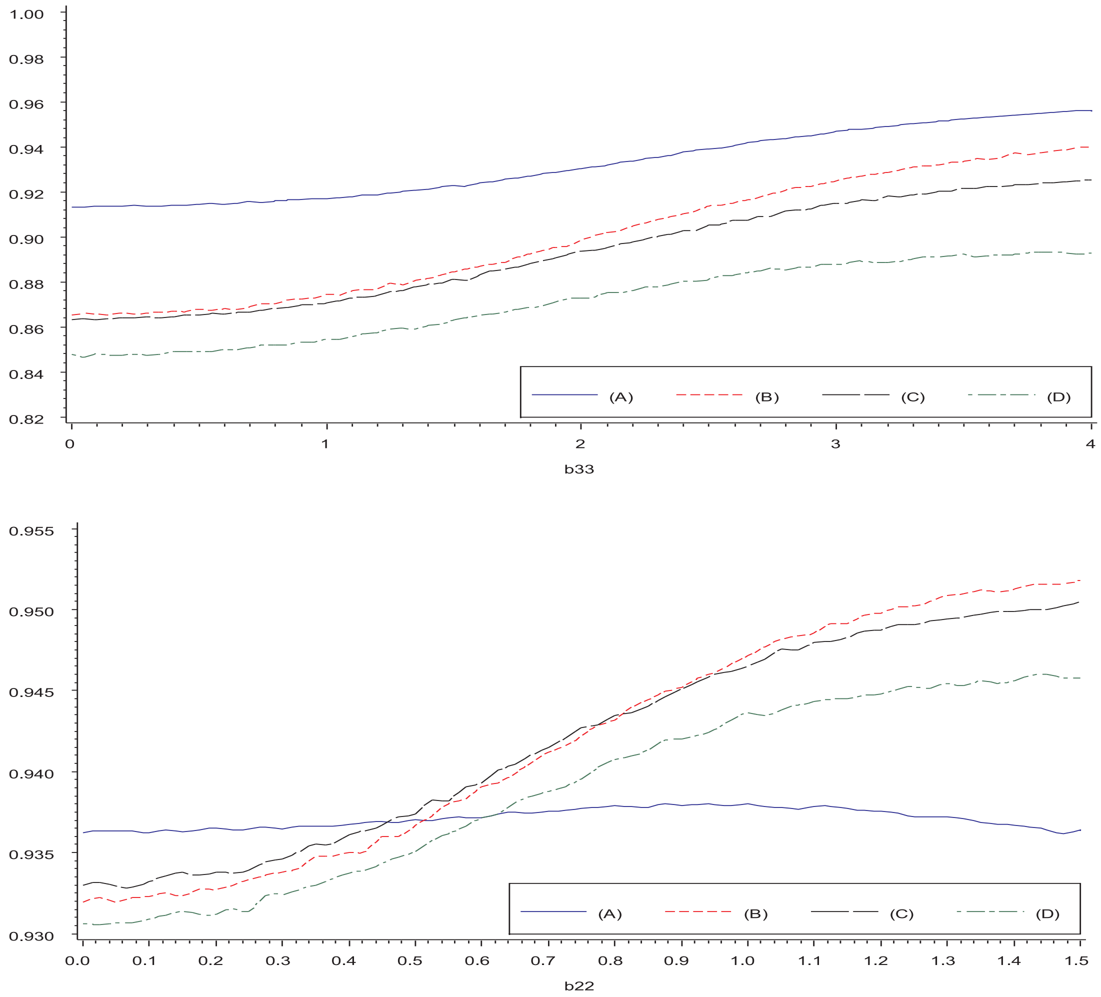


Figure 3.2: *Simulated D-efficiencies of the sequential designs (3.3) proposed by Biswas and Chaudhuri (2002) in the identified polynomial regression model (3.1). Figure 3.2a (upper panel): cubic model is "correct"; Figure 3.2b (lower panel): quadratic model is "correct".*

for large values of the parameter β_{22} the design (B) yields the best efficiencies [see Figure 3.2b]. Finally, in the linear case we observe a slight loss in (averaged) efficiency if the finest partition is used. More precisely, the efficiencies do not vary substantially with the parameter β_{11}, the design (A) in (3.3) has D-efficiency 0.82 and the designs (B), (C), (D) have D-efficiency 0.86. However, it should be remarked that the differences between the efficiencies of the different designs are relatively small compared to the differences between the probabilities of correct specification. The relative D-efficiencies of the different designs in the cubic, quadratic and linear models differ by at most 7%, 2%, 6%, respectively. On the other hand, we observe more substantial differences between the probabilities of correct specification in the cubic and quadratic model (provided that the parameter β_{33} of the highest coefficient is not too large or too small). For example, in the cubic model the probabilities of correct specification between the best and worst sequential design can differ between 63% and 32% if the parameter β_{33} varies in the interval $[1, 2]$. Similary, in the quadratic case we observe a relative difference between 76% and 25% if the parameter β_{22}

varies in the interval $[0.3, 1]$ (see Figure 3.1b). In general the probabilities of correct specification are substantially more sensitive with respect to the sample sizes in the different stages of the procedure of Biswas and Chandhuri (2002) than the efficiencies. Consequently, the sequential design proposed by Biswas and Chaudhuri (2002) should be constructed such that the sample sizes m_0, m_1 for the initial steps are not too small.

In a second step of our study we performed a similar simulation study for the two stage designs proposed by Montepiedra and Yeh (2002). As design for the first stage these authors proposed the design maximizing the function in (2.3) for the weights $\lambda_2 = \lambda_3 = 1/2$, which can be obtained from the results in Dette (1994) and is given by

$$\xi_{\text{Disc}} = \begin{pmatrix} -1 & -0.40825 & 0.40825 & 1 \\ 0.2 & 0.3 & 0.3 & 0.2 \end{pmatrix} \tag{3.4}$$

(note that we do not test the coefficient in the linear model). We considered three possible allocations of the $N = 100 = n_0 + n_1$ observations, namely

$$\begin{array}{ll} n_0 = n_1 = 50 & (E) \\ n_0 = 75, \ n_1 = 25 & (F) \\ n_0 = 25, \ n_1 = 75 & (G) \end{array} \tag{3.5}$$

where the design for the second stage is obtained by maximizing the function defined by (2.10) in the corresponding approximate setup with weight $f = n_0/(n_0 + n_1)$ [see Montediedra and Yeh (2002) for more details]. In the case of polynomial models these results can be calculated by an application of the theory of canonical moments [see Dette and Studden (1997)]. These calculations are omitted for the sake of brevity. For example in the situation (E) we obtain the designs

$$\xi_{E,3} = \begin{pmatrix} -1 & -0.4645 & 0.4645 & 1 \\ 0.2983 & 0.2017 & 0.2017 & 0.2983 \end{pmatrix}$$

$$\xi_{E,2} = \begin{pmatrix} -1 & 0 & 1 \\ 0.395 & 0.210 & 0.395 \end{pmatrix} \quad \xi_{E,1} = \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

for the second stage if the cubic, quadratic or linear model has been chosen in the first stage, respectively. Similary, in the case (F) and (G) these designs are given by

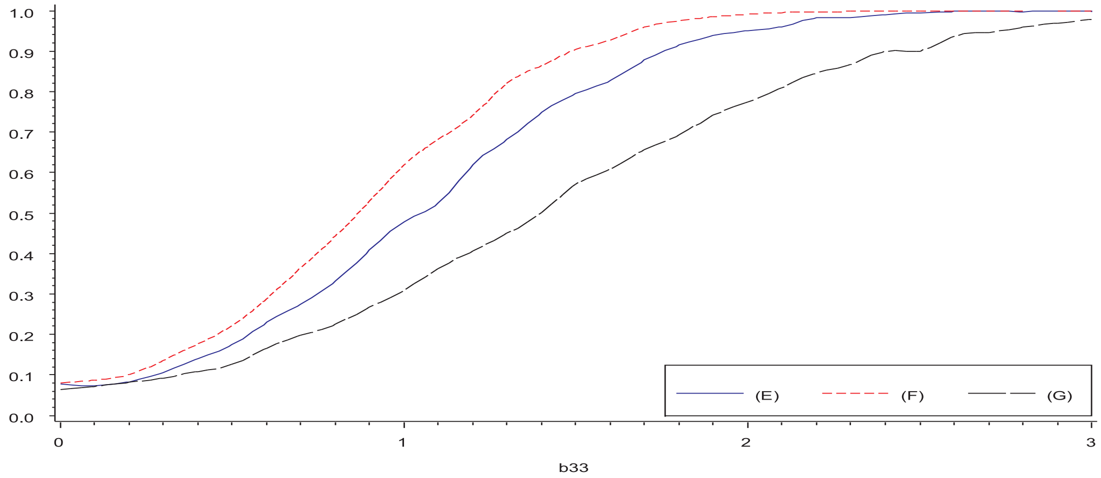
$$\xi_{F,3} = \begin{pmatrix} -1 & -0.4784 & 0.4784 & 1 \\ 0.3925 & 0.1075 & 0.1075 & 0.3925 \end{pmatrix}$$

$$\xi_{F,2} = \xi_{F,1} = \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

and

$$\xi_{G,3} = \begin{pmatrix} -1 & -0.4544 & 0.4544 & 1 \\ 0.2663 & 0.2337 & 0.2337 & 0.2663 \end{pmatrix}$$

$$\xi_{G,2} = \begin{pmatrix} -1 & 0 & 1 \\ 0.3540 & 0.2920 & 0.3540 \end{pmatrix}, \quad \xi_{G,1} = \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$



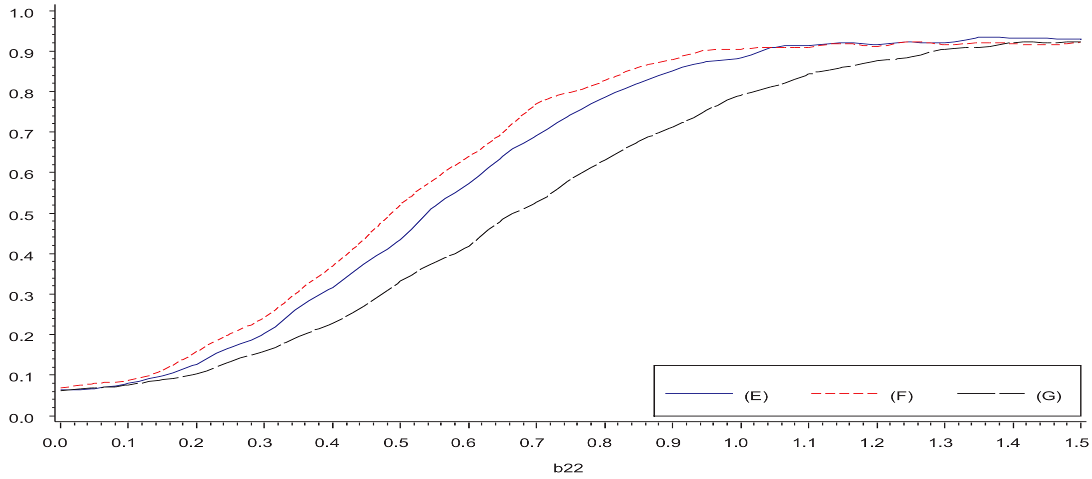


Figure 3.3: *Simulated probabilities of correct specification for the two stage designs (3.5) proposed by Montepiedra and Yeh (2002) in the polynomial regression model (3.1). Figure 3.3a (upper panel): cubic model is "correct"; Figure 3.3b (lower panel): quadratic model is "correct".*

respectively. Figure 3.3a and 3.3b show the rates of "correct" specification of the underlying model for the three two stage designs specified by (3.5), if the cubic or quadratic model are "correct", respectively. The probabilities in the linear case are not depicted because the differences between the two stage designs are negligible. The corresponding efficiencies in the "correct" model are depicted in Figure 3.4a and 3.4b. We observe that the probabilities of correct specification for the quadratic model are increasing with the sample size for the first step, which corresponds to intuition [see Figure 3.3a and 3.3b]. In the cubic model we observe relative differences between 76% and 58% if the parameter β_{33} varies in the interval $[0.5, 1.5]$. The design (F) which allocates a larger sample size to the first case is the best, while the design (G) is the worst. A comparison of the efficiencies also yields substantial differences between the two stage designs in the cubic model [see Figure 3.4a]. The efficiencies differ by not more than 12% if $\beta_{33} \leq 1$. In the quadratic model we observe only relative differences of approximately 6% [see Figure 3.4b]. Note that the design (F) yields the best efficiencies in the cubic model, provided that the parameter of the cubic

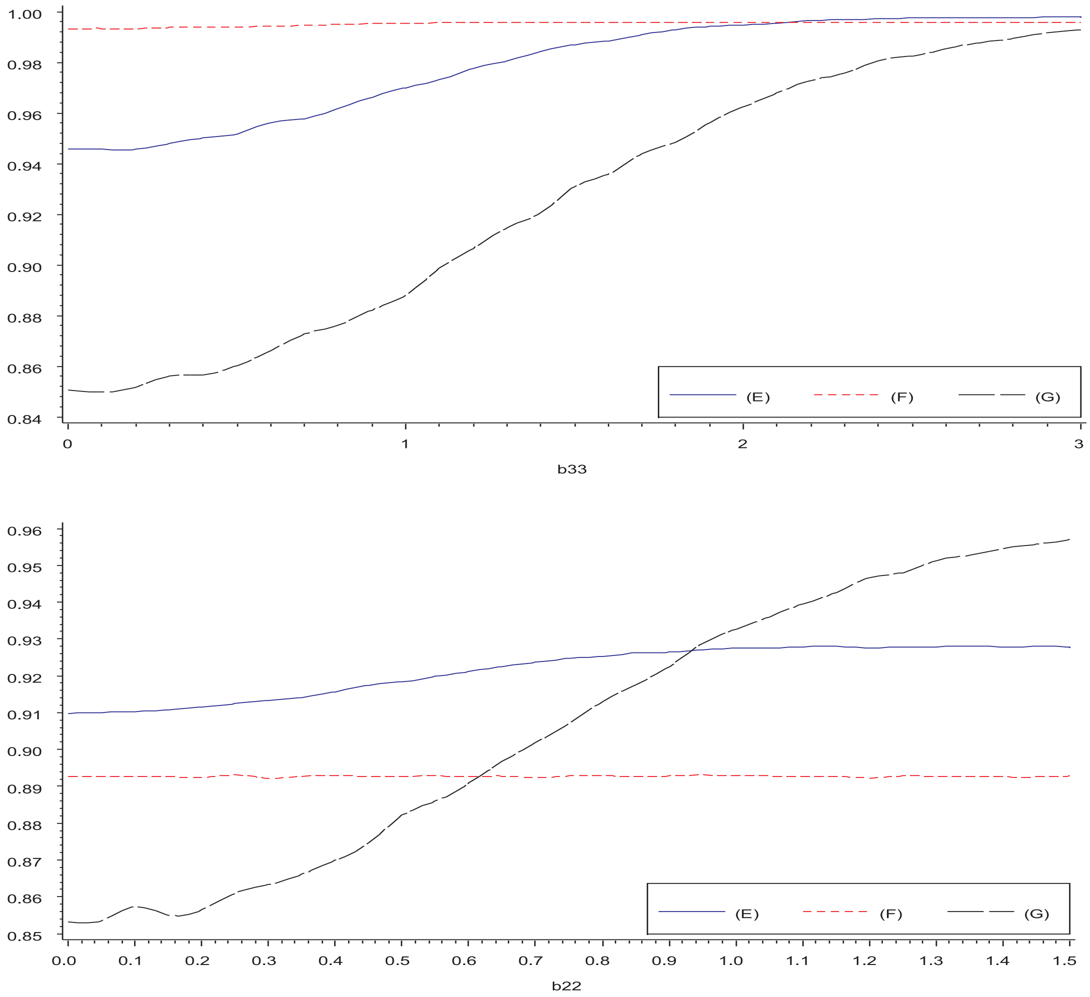


Figure 3.4: *Simulated D-efficiencies for the two stage designs (3.5) proposed by Montepiedra and Yeh (2002) in the identified polynomial regression model (3.1). Figure 3.4a (upper panel): cubic model is "correct"; Figure 3.4b (lower panel): quadratic model is "correct".*

term is not too large. In the quadratic model the design (E) is the best but the design (F) is only 5% worse. If the linear model is "correct", the D-efficiencies of the designs (E), (F) and (G) are approximately 0.85, 0.78 and 0.91, respectively. Thus taking 3/4 of the total observations in the first stage and only 1/4 of the observations in the second stage seems to be a good strategy for the two stage design of Montepiedra and Yeh (2002). Finally, we remark that the corresponding F-tests do not keep the preassigned levels, if the designs are constructed by the two stage procedure of Montepiedra and Yeh (2002). For example, in the cubic and quadratic model the simulated level varies between 7% and 9%, while the preassigned level is 5%.

We now compare the sequential designs with the non-sequential discrimination and model robust designs proposed in Section 2. For the sake of brevity and transparency we compare from each sequential approach only one design (which is in our opinion the best) and take one optimal model robust and one optimal discrimination design. For the sequential design proposed by Biswas and Chaudhuri (2002) we used the design (D) defined in (3.3) with only two stages, because the impact of these designs on the probability of correct specification is more severe than on the efficiencies with respect to estimation in the identified model. For the two stage design proposed by Montepiedra and Yeh (2002) we used the design (F) in (3.5) which uses 75 % of the observations in the first step. Moreover, in order to obtain a fair comparison the nominal level of the F-test was

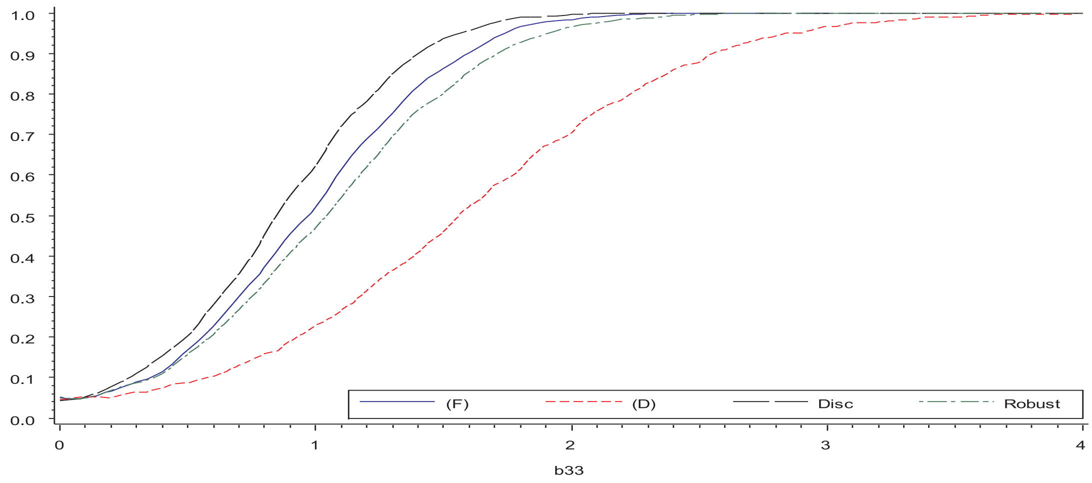


Figure 3.5: *Simulated probabilities of correct specification for sequential and non-sequential designs in the polynomial regression model (3.1). The sequential design (D) of Biswas and Chaudhuri (2002) is defined in (3.3), the two stage design (F) proposed by Montepiedra and Yeh (2002) is given in (3.5). The optimal model robust design and optimal discrimination design are defined by (3.6) and (3.4), respectively. Figure 3.5a (upper panel): cubic model is "correct"; Figure 3.5b (lower panel): quadratic model is "correct".*

adjusted such that the effective level of the resulting tests obtained from the design of Montepiedra and Yeh (2002) is precisely 5% (see the discussion in the previous paragraph). As model robust design we chosed the design

$$\xi_{\text{Robust}} = \begin{pmatrix} -1 & -0.4011 & 0.4011 & 1 \\ 0.3194 & 0.1806 & 0.1806 & 0.3194 \end{pmatrix}, \tag{3.6}$$

which maximizes the criterion (2.1) for the uniform prior $\beta_1 = \beta_2 = \beta_3 = \frac{1}{3}$ [see Dette and Studden (1997), p. 192]. The discrimination design is the design ξ_{Disc} defined in (3.4).
If the linear model is the "correct" one the differences between the four designs are negligible and therefore not depicted. Figure 3.5a and 3.5b show the probabilities of correct specification in the cubic and quadratic model, respectively. We observe that the optimal discrimination, model robust and the two stage design of Montepiedra and Yeh (2002) have a substantially better performance than the sequential procedure of Biswas and Chaudhuri (2002). Note that we chosed the best design (D) in (3.3) with respect to the criterion of correct specification [see Figure 3.1a] and all other designs in (3.3) have a worse performance. Comparing the remaining three designs we observe that the optimal discrimination design ξ_{Disc} ist the best, while the two stage design of Montepiedra and Yeh (2002) is better than the model robust design ξ_{Robust}. The picture in the quadratic model is similar. There is only a minor difference between the optimal discrimination design ξ_{Disc} and the model robust design ξ_{Robust}. The two stage design of Montepiedra and Yeh (2002) has a slightly worse performance but it has a substantial better rate of correct specification then the sequential design of Biswas and Chaudhuri (2002) [see Figure 3.5b]. We mention again that - in contrast to the model robust design, the optimal discrimination design and the sequential design of Biswas and Chaudhuri (2002) - the two stage design of Montepiedra and Yeh (2002) does not keep its preassigned level $\alpha = 5\%$, because it uses dependent observations in the sequence of classical F-tests. In our simulation study the nominal level was chosen substantially smaller than 5% such that the effective (simulated) level is precisely 5%.
A comparison of the efficiencies of the four designs can be found in Figure 3.6a and 3.6b for the cubic and quadratic model, respectively. In the cubic polynomial the design of Biswas and Chaudhuri (2002) is about 10 % less efficient than the discrimination, model robust and two stage design, while the relative efficiencies of the lastnamed designs differ only about 3%. In the quadratic model [see Figure 3.6b] the model robust design and the sequential design of Biswas and Chaudhuri (2002) yield the best efficiencies, while the two stage design of Montepiedra and Yeh (2002) is 4% – 5 % worse and the discrimination design ξ_{Disc} has the worst performance with a loss of efficiency of approximately 11% compared to the best case.
In summary we found an obvious ranking with respect to the criterion of correct model identification, but there seems to be no clear winner with respect to the efficiency criterion. The sequential designs of Biswas and Chaudhuri (2002) have the lowest probabilities of correct identification of the underlying model. This difference can be substantial. For example, if $\beta_{33} = 0.5$ or $\beta_{22} = 0.5$ the probability of correct specification is about 166% or 33% larger for the discrimination design in cubic or quadratic model, respectively. The two stage design of Montepiedra and Yeh (2002) has a similar performance as the optimal discrimination and the optimal robust design but a better rate of correct identification than the sequential design of Biswas and Chaudhuri (2002). A comparison of the efficiencies in the identified model shows a loss of about 12 % for the design of Biswas and Chaudhuri (2002) if the cubic model is "correct" and a loss of about 11 % for the discrimination design if the quadratic model is "correct". The efficiencies of the remaining three designs do not differ essentially in both cases. From these results we recommend the optimal model robust design ξ_{Robust} for statistical inference in the cubic model, because it presents a reasonable compromise between the two different goals of model identification and precise estimation in the identified model. Additionally, the corresponding F-tests keep the preassigned level in contrast to the two stage procedure of Montepiedra and Yeh (2002), which is in this sense not reliable. If model identification is considered as the more important goal, the optimal discrimination design ξ_{Disc} could be used alternatively. This design yields slightly better rates of correct specification, a

more efficient inference in the identified model if the cubic model is "correct", but a less efficient inference in the identified model if the quadratic regression is the "correct" polynomial.

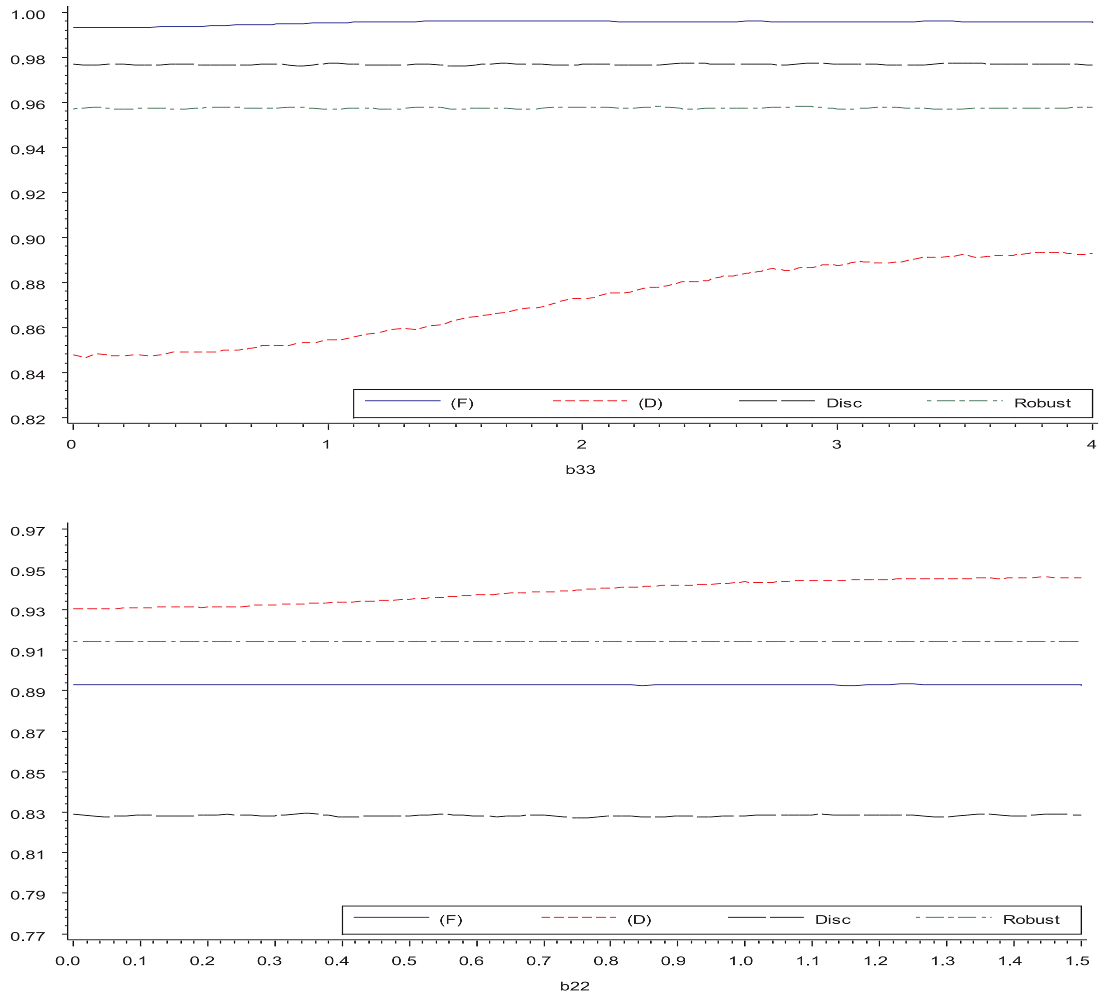


Figure 3.6: *Simulated D-efficiencies for sequential and non-sequential designs in the identified polynomial regression model (3.1). The sequential design (D) of Biswas and Chaudhuri (2002) is defined in (3.3), the two stage design (F) proposed by Montepiedra and Yeh (2002) is given in (3.5). The optimal model robust design and optimal discrimination design are defined by (3.6) and (3.4), respectively. Figure 3.6a (upper panel): cubic model is "correct"; Figure 3.6b (lower panel): quadratic model is "correct".*

3.2 Discrimination designs for Fourier regression models

In order to investigate if the results of the previous section are representative we investigate a second example, for which the model robust, discrimination and D-optimal designs are also known. The model under consideration is the quadratic trigonometric regression model

$$g_4(x) = a_{40} + a_{41}\sin x + a_{42}\cos x + a_{43}\sin(2x) + a_{44}\cos(2x);\;\; x \in [-\pi, \pi] \tag{3.7}$$

and the nested models are given by

$$g_3(x) = a_{30} + a_{31}\sin x + a_{32}\cos x + a_{33}\sin(2x) \tag{3.8}$$
$$g_2(x) = a_{20} + a_{21}\sin x + a_{22}\cos x \tag{3.9}$$
$$g_1(x) = a_{10} + a_{11}\sin x. \tag{3.10}$$

The D-optimal designs can be found in Pukelsheim (1993) for the models g_2 and g_4 and are obtained by similar methods for the models g_1 and g_3. These designs have equal masses at the points

$$\begin{array}{ll} (g_1) & -\frac{\pi}{2}, \frac{\pi}{2} \\[1em] (g_2) & -\frac{2}{3}\pi, 0, \frac{2}{3}\pi \\[1em] (g_3) & -\frac{3}{4}\pi, -\frac{1}{4}\pi, \frac{1}{4}\pi, \frac{3}{4}\pi \\[1em] (g_4) & -\frac{4}{5}\pi, -\frac{2}{5}\pi, 0, \frac{2}{5}\pi, \frac{4}{5}\pi\,. \end{array}$$

The model robust designs corresponding to the criterion (2.1) have been found by Lau and Studden (1985) and for the uniform prior an model robust design is given by

$$\xi_{\text{Robust}} = \begin{pmatrix} -2.4294 & -1.2530 & 0 & 1.2530 & 2.4294 \\ 0.1965 & 0.2252 & 0.1566 & 0.2252 & 0.1965 \end{pmatrix}. \tag{3.11}$$

Optimal discrimination designs corresponding to the criterion (2.3) have been determined by Dette and Haller (1998) and for the uniform prior an optimal discrimination design is given by

$$\xi_{\text{Disc}} = \begin{pmatrix} -\pi & -\frac{3}{5}\pi & -\frac{1}{5}\pi & \frac{1}{5}\pi & \frac{3}{5}\pi & \pi \\ \frac{1}{10} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{10} \end{pmatrix}. \tag{3.12}$$

We did not include the two stage designs of Montepiedra and Yeh (2002) in our comparison for two reasons. On the one hand it was shown in Section 3.1 that the F-tests obtained from these designs do not keep the preassigned levels, on the other hand the determination of the designs for the second stage maximizing (2.10) is a non-trivial optimization problem, except in the case of low order univariate polynomial models (this actually restricts the general application of this method in practice). For the sequential method of Biswas and Chaudhuri (2002) we considered two cases

$$\begin{array}{ll} (F1) & m_0 = \ldots = m_4 = 20 \\[1em] (F2) & m_0 = 50,\, m_1 = 50 \end{array} \tag{3.13}$$

Figure 3.7a – 3.7c show the corresponding probabilities of correct specification if the model g_4, g_3 or g_2 is the underlying model. We observe again substantially larger probabilities of correct identification for the non-sequential designs. For example, if the model g_4 is "correct", $a_{44} = 0.3$ or

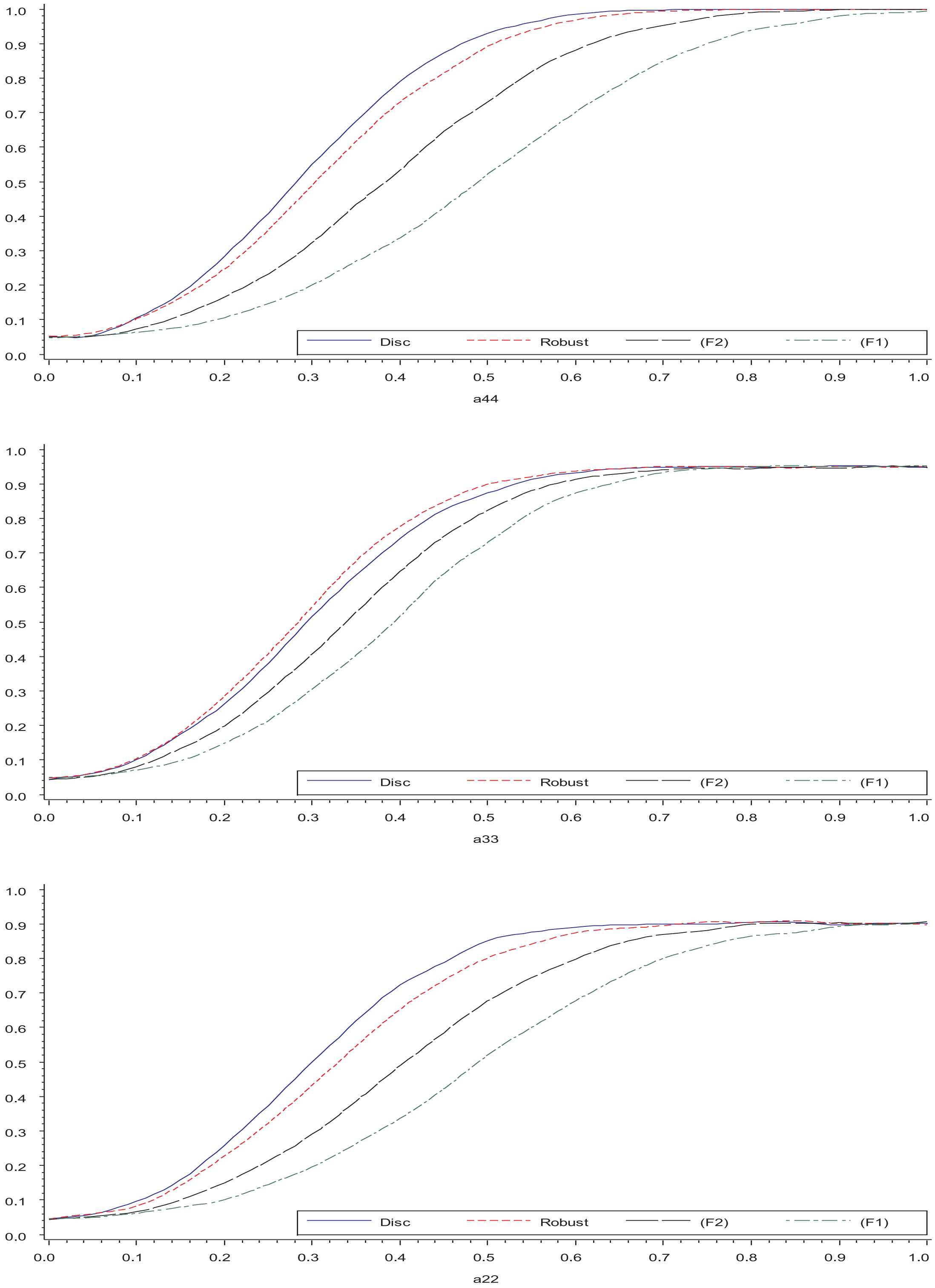


Figure 3.7: *Simulated probabilities of correct specification for sequential and non-sequential designs in the trigonometric regression model (3.7). The sequential designs (F1) and (F2) of Biswas and Chaudhuri (2002) are defined in (3.13), the optimal model robust design is given by (3.11) and the optimal discrimination design is defined in (3.12). Figure 3.7a (upper upper panel): the model (3.7) is "correct"; Figure 3.7b (middle panel): the model (3.8) model is "correct"; Figure 3.7c (lower panel): the model (3.9) is "correct".*

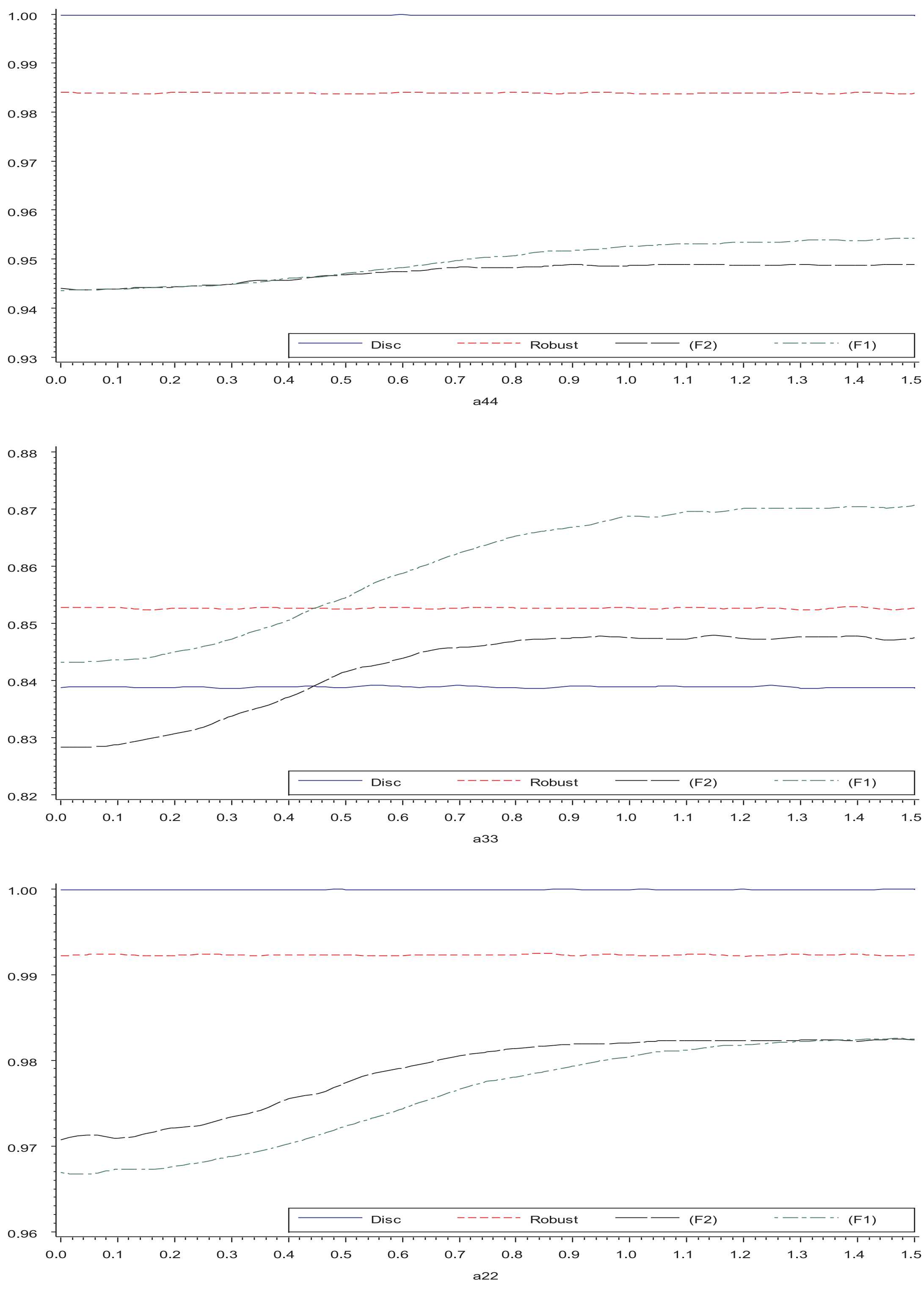


Figure 3.8: *Simulated D-efficiencies for sequential and non-sequential designs in the identified trigonometric regression model (3.7). The sequential designs (F1) and (F2) of Biswas and Chaudhuri (2002) are defined in (3.13), the optimal model robust design is given by (3.11) and the optimal discrimination design is defined in (3.12). Figure 3.8a (upper left panel): the model (3.7) is "correct"; Figure 3.8b (right panel): the model (3.8) model is "correct"; Figure 3.8c (lower panel): the model (3.9) is "correct".*

0.5 the best non-sequential design yields a rate of correct specification, which is 70% or 27% larger than the rate obtained from the best sequential design. It is interesting to note that this difference is smaller (but still substantial) in the model g_3 (38% if $a_{33} = 0.3$ or 11% if $a_{33} = 0.5$) and of larger size in the model g_2 (155% if $a_{22} = 0.3$ or 64% if $a_{22} = 0.5$). However, in all considered cases the non-sequential designs performed substantially more reliable than the sequential procedure of Biswas and Chaudhuri (2002) with respect to the criterion of correct model identification. A comparison of the efficiencies for the four designs is presented in Figure 3.8a – 3.8c corresponding to the cases where the model g_4, g_3 and g_2 is the "correct" regression model, respectively. Here we observe much smaller differences between the sequential and non-sequential designs. For example, in the model g_4 the discrimination design ξ_{Disc} defined in (3.12) is the best [because it is in fact also D-optimal for the quadratic trigonometric regression, see Pukelsheim (1993)] but the sequential designs are at most 6% less efficient [see Figure 3.8a]. In the model g_3 the performance of the different designs depends on the parameter a_{32} but the efficiencies differ by at most 4% [see Figure 3.8b]. Similary, in the model g_2 the discrimination design ξ_{Disc} is again D-optimal, but the other designs are at most 3.5% less efficient [see Figure 3.8c]. In summary, our simulation results show that for the trigonometric regression model the optimal discrimination and model robust (non-sequential) designs have a substantially better performance than the sequential designs proposed by Biswas and Chaudhuri (2002). The differences between the efficiencies are negligible, while the non-sequential designs are substantially more reliable with respect to a correct identification of the underlying model.

4 Conclusions

In this paper we presented a numerical comparison of different design strategies for two different goals in regression models

- identification of an appropriate model
- efficient estimation of the parameters in the identified model

Sequential and non-sequential design strategies are investigated in a polynomial and trigonometric regression model. The two stage designs proposed by Montepiedra and Yeh (2002) provide reasonable probabilities of correct identification of the underlying model and reasonable efficiencies for the estimation of the parameters in the identified model. However, due to the dependencies of the observations obtained from sequential sampling, the corresponding F-tests for the identification of the model do not keep the preassigned level and the statistical properties of the estimator in the underlying model are not clear in general. Additionally, the determination of the design for the second step seems to be difficult, because it is equivalent to an optimal design problem in a Bayesian linear regression model.

The sequential strategy proposed by Biswas and Chaudhuri (2002) yields to a sequence of (modified) F-tests which keep the preassigned level. Moreover, for large sample sizes the "correct" model is identified with high probability, the sequential design is close to the D-optimal design for the "correct" model and the estimates obtained from this design are consistent and asymptotically normal. Although these properties are interesting from a theoretical (asymptotic) point of view, our numerical results show that for finite sample sizes these designs cannot be recommended for applications because of their low rate of correct identification of the underlying model. The two stage designs of Montepiedra and Yeh (2002) and the non-sequential discrimination and model robust designs of Dette (1990, 1994) yield substantially higher rates of correct identification of the underlying model and have (at least) comparable efficiencies for the estimation of the parameters in the underlying "correct" model (in many cases they are even more efficient). The poor performance of the sequential designs proposed by Biswas and Chaudhuri (2002) can be explained by the fact that a modified version of the classical F-test is used in the identification steps. This modification is required to keep the preassigned level of the corresponding tests from the dependent data, and usually causes a substantial loss of power. The non-sequential discrimination and model robust designs yield better rates of identification of the "correct" model than the two stage designs introduced by Montepiedra and Yeh (2002) and comparable efficiencies in the identified model. Moreover the non-sequential designs yield a sequence of F-tests for the identification step, which keep the preassigned levels and the calculation is usually simpler because it can be performed by standard algorithms [see e.g. Läuter (1974) or Atkinson and Donev (1992)]. For these reasons the application of optimal discrimination or optimal model robust designs is recommended for the statistical inference in (nested) regression models.

Acknowledgements. The work of H. Dette was supported by the Deutsche Forschungsgemeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen). The authors would like to thank Isolde Gottschlich who typed numerous versions of the paper with considerable technical expertise and A. Biswas and G. Montepiedra for making their work available to us before publication.

References

T.W. Anderson (1962). The choice of the degree of a polynomial regression as a multiple decision problem. Ann. Math. Statist. 33, 255-265.

A.C. Atkinson, D.R. Cox (1974). Planning experiments for discriminating between models. J.R. Statist. Soc. B, 36, 321-348.

A.C. Atkinson, A.N. Donev (1992). Optimum Experimental Design. Oxford: Clarendon Press.

D.F. Andrews (1971). Sequentially designed experiments for screening out bad models. Biometrika 58, 427-432.

A. Biswas, P. Chaudhuri (2002). An efficient design for model discrimination and parameter estimation in linear models. Biometrika, to appear. Tech. Report ASD/2000/8. http://www.isical.ac.in/ asu/pub.html

K. Chaloner (1984). Optimal Bayesian experimental design for linear models. Ann. Statist. 12, 283-300.

H. Dette (1990). A generalization of D- and D_1-optimal designs in polynomial regression. Ann. Statist. 18, 1784-1801.

H. Dette (1994). Discrimination designs for polynomial regression on a compact interval. Ann. Statist. 22, 890-904.

H. Dette (1995). Optimal designs for identifying the degree of a polynomial regression. Ann. Statist. 23, 1248-1267.

H. Dette, G. Haller (1998). Optimal discriminating designs for Fourier regression. Ann. Statist. 26, 1496-1521.

H. Dette, I. Röder (1995). Designs for multivariate regression with missing terms. Scand. J. Statist. 23, 195-203.

H. Dette, I. Röder (1997). Optimal discrimination designs for multi-factor models. Ann. Statist. 25, 1161-1175.

H. Dette, W.J. Studden (1997). The Theory of Canonical Moments with Applications in Statistics, Probability and Analysis. Wiley, N.Y.

J. Kiefer (1974). General equivalence theory for optimum designs (approximate theorey). Ann. Statist. 2, 849-879.

E. Läuter (1974). Experimental design for a class of models. Math. Oper. Forsch. Statist. 5, 379-398.

T.S. Lau, W.J. Studden (1985). Optimal designs for trigonometric and polynomial regression using canonical moments. Ann. Statist. 13, 383-394.

G. Montepiedra, A.B. Yeh (1998). A two stage strategy for the construction of D-optimal experimental designs. Commun. in Statistics, Simul. and Comput., 27, 277-402.

G. Montepiedra, A.B. Yeh (2002). Two stage designs for estimation and identification of polynomial models. Technical Report, Department of Applied Statistics and Operations Research. Bowling Green State University.

F. Pukelsheim (1993). Optimal Design of Experiments. Wiley, N.Y.

F. Pukelsheim, J.L. Rosenberger (1993). Experimental designs for model discrimination. J. Amer. Stat. Assoc. 88, 642-649.

G.A.F. Seber (1977). Linear Regression Analysis. Wiley, N.Y.

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

M.G. Spruill (1990). Good designs for testing the degree of a polynomial mean. Sankhyā, Ser. B, 52, 67-74.