

A note on all-bias designs with applications in spline regression models

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@rub.de

FAX: +49 234 3214 559

Viatcheslav B. Melas

St. Petersburg State University

Department of Mathematics

St. Petersburg, Russia

email: v.melas@pobox.spbu.ru

August 12, 2008

Abstract

If a model is fitted to empirical data, bias can arise from terms which are not incorporated in the model assumptions. As a consequence the commonly used optimality criteria based on the generalized variance of the estimate of the model parameters may not lead to efficient designs for the statistical analysis. In this note some general aspects of all-bias designs are presented, which were introduced in this context by Box and Draper (1959). We establish sufficient conditions such that a given design is an all-bias design and illustrate these in the special case of spline regression models. In particular our results generalize recent findings of Woods and Lewis (2006).

Keywords and Phrases: all-bias design, spline regression models, robust designs, quadrature formulas

AMS Subject Classification: 62J05, 62K99

1 Introduction

Linear regression models of the form

$$(1.1) \quad Y = \beta^T f_{(1)}(x) + \varepsilon$$

are widely used in statistical applications to predict a continuous response Y from the explanatory variable x . If n independent observations $(x_i, Y_i)_{i=1}^n$ are available, the model parameters are often estimated by the least squares technique, and a good experimental design advises the experimenter to choose the experimental conditions x_1, \dots, x_n such that statistical analysis can be performed most efficiently. Many criteria for such a choice are based on the minimization of the generalized variance of the estimators in the linear regression model (1.1) and include the well known D -, A - and E -optimality criterion [see e.g. Kiefer (1974), Silvey (1980) or Pukelsheim (1993)]. This approach has been criticized by many authors, because the “optimal” designs are constructed under a specific stated model assumption [namely the linear regression model (1.1)] and do not offer the opportunity for the experimenter to check if the model assumptions are violated [see e.g. Box and Draper (1959), Stigler (1971), Läuter (1974), Studden (1982) or Dette (1990), Pukelsheim and Rosenberger (1992) among many others]. Moreover, in some cases “classical” optimal designs do not reflect the particular goals of the experiment such as prediction or interpolation, because the criteria focus on a precise estimation of the parameters, and in linear models the parameters often have no specific interpretation.

Box and Draper (1959) demonstrated the importance of being able to take a possible bias into account for the construction of efficient experimental designs if the model assumptions are questionable. In their pioneering work these authors argued that a criterion minimizing the difference between the expected predicted response from the assumed model and the expected response from the “true” model would better reflect the demands of the experimenter. Nowadays such designs are known as all-bias designs and have found considerable interest in the statistical literature [see e.g. Ermakov and Sedunov (1974), Welch (1983), Galil and Kiefer (1977), Ermakov and Melas (1995), Yue and Hickernell (1999), Woods (2005) and Woods and Lewis (2006) among many others].

In the present paper we investigate some further aspects of all-bias designs. In Section 2 we introduce some basic terminology. We present sufficient conditions for all-bias designs, which relate the statistical problem of design construction to a problem from numerical analysis, which has a long history in mathematics: the determination of quadrature formulas for numerical integration [see e.g. de Boor (1978)]. It is demonstrated that the knots of such quadrature formulas yield all-bias designs, and consequently the results of this well established and elegant theory can be used for the construction of experimental designs. For this purpose we discuss an alternative method of estimation which is different from least squares estimation and commonly applied in series estimation [see e.g. Efromovich (1999), p. 121]. In Section 3 we illustrate the results of Section 2 in the context of spline regression models. In particular some of the recent findings of Woods and Lewis (2006) are put in the context of numerical quadrature and generalized in several directions.

2 All-bias designs and direct estimation

Consider the linear regression model (1.1), with the explanatory variable varying in the interval $[a, b]$, that is

$$(2.1) \quad E[Y(x)] = \beta^T f_{(1)}(x); \quad x \in [a, b]$$

where $f_{(1)}(x) = (f_1(x), \dots, f_{k_1}(x))^T$ denotes a vector of $k_1 \in \mathbb{N}$ linearly independent regression functions on the interval $[a, b]$, $\beta = (\beta_1, \dots, \beta_{k_1})^T$ is a vector of unknown parameters, and different observations are assumed to be uncorrelated with common variance $\sigma^2 > 0$. Following Box and Draper (1959) we consider the “true” model

$$(2.2) \quad E[Y(x)] = \beta^T f_{(1)}(x) + \gamma^T f_{(2)}(x),$$

where $\gamma = (\gamma_1, \dots, \gamma_{k_2})^T$ and $f_{(2)}(x) = (f_{k_1+1}(x), \dots, f_{k_1+k_2}(x))^T$ are a further vector of parameters and regression functions, respectively, such that the functions $f_1, \dots, f_{k_1+k_2}$ are linearly independent on the interval $[a, b]$. In other words $\gamma^T f_{(2)}(x)$ denotes the deviation of the “true” model from the model assumed by the experimenter. In what follows we denote by $\hat{Y}(x)$ a prediction of the response at experimental condition x using least squares estimation (an alternative estimate will be discussed below) and define

$$(2.3) \quad B = \frac{n}{\sigma^2} \int_a^b (E[\hat{Y}(x)] - E[Y(x)])^2 \lambda(x) dx$$

as the average squared bias over the design interval $[a, b]$, which results from using the model (2.1) instead of the true model (2.2) for prediction. In equation (2.3) the function $\lambda : [a, b] \rightarrow \mathbb{R}$ denotes a non-negative weight function, normalized such that $\int_a^b \lambda(x) dx = 1$, and n is the size of the sample which is available for the prediction. If $a \leq x_1 \leq x_2 \leq \dots \leq x_n \leq b$ denote the predictors corresponding to these uncorrelated observations, then an all-bias design minimizes the expression B in (2.3) with respect to the points $\{x_1, \dots, x_n\}$. It was shown by Box and Draper (1959) that in the case of least squares estimation this minimization problem is equivalent to minimizing the expression

$$(2.4) \quad \frac{n}{\sigma^2} \gamma^T (M_{11}^{-1} M_{12} - \mu_{11}^{-1} \mu_{12})^T \mu_{11} (M_{11}^{-1} M_{12} - \mu_{11}^{-1} \mu_{12}) \gamma,$$

where

$$(2.5) \quad M_{ij} = \frac{1}{n} X_i^T X_j \in \mathbb{R}^{k_i \times k_j}; \quad i, j = 1, 2$$

$$(2.6) \quad X_i^T = [f_{(i)}(x_1), \dots, f_{(i)}(x_n)] \in \mathbb{R}^{k_i \times n}; \quad i = 1, 2$$

$$(2.7) \quad \mu_{ij} = \int_a^b f_i(x) f_j(x) \lambda(x) dx \in \mathbb{R}^{k_i \times k_j}; \quad i, j = 1, 2.$$

Let $\tilde{x}_1, \dots, \tilde{x}_k$ denote the distinct values among x_1, \dots, x_n and define \tilde{M}_{ij} as the matrix obtained from the definition (2.5) for the points $\tilde{x}_1, \dots, \tilde{x}_k$, then a straightforward calculation shows that

$$\tilde{M}_{11}^{-1} \tilde{M}_{12} = M_{11}^{-1} M_{12},$$

and consequently the value B of the all-bias criterion cannot be diminished by taking repeated observations (note that this fact is intuitively obvious, because the all-bias criterion refers to the bias and not to the variance). Therefore we assume in the following discussion $a \leq x_1 < x_2 < \dots < x_n \leq b$, and denote a design of this form by

$$(2.8) \quad \xi = \{x_1, \dots, x_n\}; \quad a \leq x_1 < x_2 < \dots < x_n \leq b.$$

The design ξ is called all-bias design for least squares estimation, if it minimizes the expression (2.3) [or equivalently (2.4)], where $\hat{Y}(x) = \hat{\beta}^T f_{(1)}(x)$ and $\hat{\beta}$ is the least square estimate in the assumed model (2.1). All-bias designs for other estimates are defined in a similar manner. Recall the definition $X_{(1)}^T = (f_i(x_j))_{i=1, \dots, k_1}^{j=1, \dots, n}$, then the least squares estimate for the parameter β in the linear regression model (2.1) is given by $\hat{\beta} = (X_{(1)}^T X_{(1)})^{-1} X_{(1)}^T Y$, where $Y = (Y_1, \dots, Y_n)^T$ denotes the vector of responses, and it is easy to see that a design $\xi = \{x_1, \dots, x_n\}$ is an all-bias design if and only if the identity

$$(2.9) \quad E[\hat{\beta}] = \beta + M_{11}^{-1} M_{12} \gamma = \beta$$

holds, where the matrices M_{11} and M_{12} are defined in (2.5) [see e.g. Ermakov and Sedunov (1974)]. The following example shows that in some cases the concept of all-bias designs can yield to designs which are not too useful for practical applications.

Example 2.1. Consider the polynomial regression model of degree $k_1 - 1$ for the model (2.1), that is $f_{(1)}(x) = (1, x, \dots, x^{k_1-1})^T$ and an extension by piecewise polynomials for (2.2), i.e. $f_{(2)}(x) = ((x - t_1)_+^{k_1-1}, \dots, (x - t_{k_2})_+^{k_1-1})^T$ where $a < t_1 < \dots < t_{k_2} < b$ are given knots and $\lambda(x) = I_{[a,b]}(x)/(b - a)$. Obviously we have for any design ξ satisfying $a \leq x_1 < \dots < x_n \leq t_1$ $f_{(2)}(x_j) = 0$ ($j = 1, \dots, n$), which implies $X_{(2)}^T = 0 \in \mathbb{R}^{k_2 \times n}$, $M_{12} = 0$, and consequently $E[\hat{\beta}] = \beta$. In other words: any design with predictors located in the interval $[a, t_1]$ is an all-bias design for the spline regression model.

Although the design found in Example 2.1 is not of particular importance from a practical point of view, it indicates that in some cases the all-bias criterion yield to designs with problematic properties for applications. The following result gives an alternative characterization of all-bias designs. A proof can be found in Ermakov and Melas (1995).

Lemma 2.2. *A design $\xi = \{x_1, \dots, x_n\}$ is an all-bias design if and only if the condition*

$$(2.10) \quad S(\xi)u = u$$

holds for all vectors of the form $u = (u_1, \dots, u_{k_1}, 0, \dots, 0)^T \in \mathbb{R}^{k_1+k_2}$, where the matrix $S(\xi)$ corresponds to the orthogonal projection onto the subspace.

$$(2.11) \quad \mathcal{L} = \text{span} \left\{ \begin{pmatrix} f_{(1)}(x_1) & \dots & f_{(1)}(x_n) \\ f_{(2)}(x_1) & \dots & f_{(2)}(x_n) \end{pmatrix} \right\}.$$

Example 2.3. Consider the polynomial regression model of degree $k_1 - 1$, that is $f_{(1)}(x) = (1, x, \dots, x^{k_1-1})^T$ and its extension to a model of degree k_1+k_2-1 , i.e. $f_{(2)}(x) = (x^{k_1}, \dots, x^{k_1+k_2-1})^T$ [see Stigler (1971) or Studden (1982)]. Any design $\xi = \{x_1, \dots, x_{k_1+k_2}\}$ with $k_1 + k_2$ distinct points yields $\mathcal{L} = \mathbb{R}^{k_1+k_2}$ and as a consequence $S(\xi) = I_{k_1+k_2}$. Therefore condition (2.10) is obviously satisfied and any design with $k_1 + k_2$ different points is an all-bias design.

In practical applications the all-bias designs derived in Example 2.1 and 2.3 are of no practical interest. Similarly, the concept of all-bias designs does not help substantially to discriminate between competing designs for polynomial regression models. In the following discussion we would like to develop a formal construction of particular all-bias designs which avoids solutions of this type. The basic idea is to introduce an alternative method of estimating the coefficients in the linear regression model (2.1). For these estimates the problem of determining all-bias designs becomes equivalent to the problem of determining quadrature formulas for numerical integration, which has no trivial solutions of the form described in Example 2.1 and 2.3. Finally we show that the resulting non-trivial all-bias design are also all-bias if least squares estimation is used.

For this purpose we define $\psi_1, \psi_2, \dots, \psi_{k_1+k_2}$ as the functions obtained by orthonormalizing the regression functions $f_1, f_2, \dots, f_{k_1+k_2}$ with respect to the measure $\lambda(x)dx$, that is

$$(2.12) \quad \int_a^b \psi_i(x)\psi_j(x)\lambda(x)dx = \delta_{ij}; \quad i, j = 1, \dots, k_1 + k_2,$$

where δ_{ij} denotes Kronecker's delta. Obviously, if $\psi_{(1)}(x) = (\psi_1(x), \dots, \psi_{k_1}(x))^T$, $\psi_{(2)}(x) = (\psi_{k_1+1}(x), \dots, \psi_{k_1+k_2}(x))^T$, there exist vectors $\bar{\beta} \in \mathbb{R}^{k_1}$ and $\bar{\gamma} \in \mathbb{R}^{k_2}$ such that

$$(2.13) \quad \mu(x) = \beta^T f_{(1)}(x) + \gamma^T f_{(2)}(x) = \bar{\beta}^T \psi_{(1)}(x) + \bar{\gamma}^T \psi_{(2)}(x),$$

and it is easy to see that the change of the basis does not change the value B defined in (2.3).

For a given design $\xi = \{x_1, \dots, x_n\}$ we define an alternative class of estimates, which we call "direct" estimates, i.e.

$$(2.14) \quad \tilde{\beta} = \sum_{j=1}^n \delta_j \psi_{(1)}(x_j) Y_j,$$

where $Y_j = Y(x_j)$ denotes the j th observation under experimental condition x_j ($j = 1, \dots, n$), and the real coefficients $\delta_1, \dots, \delta_n$ will be specified later. The estimate of the form (2.14) is motivated from series estimation in nonparametric regression [see Efromovich (1999), p. 121] where an L^2 -approximation of the function μ in (2.13) is constructed using the regression functions of model (2.1), that is

$$(2.15) \quad \mu^*(x) = \sum_{i=1}^{k_1} \left(\int_a^b \mu(x) \psi_i(x) \lambda(x) dx \right) \psi_i(x) = \bar{\beta}^T \psi_{(1)}(x).$$

The minimal value

$$(2.16) \quad \min_{a_1, \dots, a_{k_1}} \int_a^b [\mu(x) - \sum_{j=1}^{k_1} a_j \psi_j(x)]^2 \lambda(x) dx$$

corresponding to μ^* is given by $\|\bar{\gamma}\|_2^2$ (which follows by a standard calculation). For an arbitrary estimate of the form (2.14) it follows that the L^2 -distance of the difference between the expected predicted response and the expected response of the “true” model is given by

$$(2.17) \quad \bar{B} = \int_a^b \left[\mu(x) - \sum_{i=1}^{k_1} \left(\sum_{j=1}^n \mu(x_j) \psi_i(x_j) \delta_j \right) \psi_i(x) \right]^2 \lambda(x) dx \geq \|\bar{\gamma}\|_2^2,$$

where the symbol \bar{B} reflects the fact that a direct estimate has been used for prediction. A simple calculation shows that there is equality in (2.16) if and only if the identities

$$(2.18) \quad \int_a^b \psi_i(x) \psi_j(x) \lambda(x) dx = \sum_{\ell=1}^n \psi_i(x_\ell) \psi_j(x_\ell) \delta_\ell$$

hold for all $i = 1, \dots, k_1$; $j = 1, 2, \dots, k_1 + k_2$, or equivalently

$$(2.19) \quad \int_a^b f_i(x) f_j(x) \lambda(x) dx = \sum_{\ell=1}^n f_i(x_\ell) f_j(x_\ell) \delta_\ell$$

for all $i = 1, \dots, k_1$; $j = 1, 2, \dots, k_1 + k_2$ (note that the functions $\{\psi_i | i = 1, \dots, k_1 + k_2\}$ and $\{f_i | i = 1, \dots, k_1 + k_2\}$ generate the same space). Condition (2.19) means that there exists a quadrature formula with knots x_1, \dots, x_n , which integrates the functions $\{f_i f_j | i = 1, \dots, k_1; j = 1, \dots, k_1 + k_2\}$ exactly [see e.g. de Boor (1978), Engels (1980) or Gizzetti and Ossicini (1970)]. The following result now relates the problem of determining all-bias designs to the problem of constructing quadrature formulas.

Proposition 2.4. *Consider the linear regression model (2.1) and its extension (2.2), where direct estimates of the form (2.14) are used for estimating the parameters corresponding to $f_{(1)}$. A design $\xi = \{x_1, \dots, x_n\}$ is an all-bias design (i.e. it minimizes the quantity \bar{B} defined in (2.16)*

if and only if there exists weights $\delta_1, \dots, \delta_n$ such that the quadrature formula with knots x_1, \dots, x_n and weights $\delta_1, \dots, \delta_n$ is exact for the functions $\{f_i(x)f_j(x) \mid i = 1, \dots, k_1; j = 1, \dots, k_1 + k_2\}$.

Note that Proposition 2.4 refers to the application of direct estimators of the form (2.14), and in the following discussion we will explore the relations between estimators of the form (2.14) and the least square estimates. For this purpose it is useful to recall the concept of approximate designs. Following Kiefer (1974) an approximate optimal design is defined as a probability measure η with finite support, say x_1, \dots, x_n , and positive weights w_1, \dots, w_n at these points satisfying $\sum_{i=1}^n w_i = 1$ (here we assume that the x_i are distinct and ordered). The weights w_i represent the relative proportion of total observations taken at the points x_i ($i = 1, \dots, n$). If N observations can be made, the experimenter takes approximately $r_i \approx Nw_i$ observations at each x_i ($i = 1, \dots, n$), such that $\sum_{i=1}^n r_i = N$. The information matrix of an approximate design η in the extended regression model (2.2) is defined by

$$(2.20) \quad M(\eta) = \sum_{i=1}^n w_i f(x_i) f^T(x_i).$$

If \bar{Y}_i denotes the mean of the observations taken at experimental condition x_i ($i = 1, \dots, n$), the direct estimate is defined by

$$(2.21) \quad \tilde{\beta} = \sum_{j=1}^n \delta_j \psi_{(1)}(x_j) \bar{Y}_j.$$

Theorem 2.5. *In the linear regression models (2.1) and (2.2) the following assertions are correct:*

- (i) *If least squares estimation is used in model (2.1), then condition (2.18) is a sufficient condition such that the design $\xi = \{x_1, \dots, x_n\}$ is an all bias design.*
- (ii) *If direct estimates are used in model (2.1), then condition (2.18) is a necessary and sufficient such that the design $\xi = \{x_1, \dots, x_n\}$ is an all-bias design.*
- (iii) *The least squares estimate for the orthonormalized regression model $\beta^T \psi_{(1)}(x)$ coincides with the direct estimate (2.14) if condition (2.18) is satisfied with $\delta_i = \frac{1}{n}$ $i = 1, \dots, n$. In this case we have $\bar{B} = \|\bar{\gamma}\|_2^2$ and*

$$E \left\{ \int_a^b [\mu(x) - \tilde{\beta}^T \psi_{(1)}(x)]^2 \lambda(x) dx \right\} = \|\bar{\gamma}\|_2^2 + \frac{k_2}{n} \sigma^2$$

- (iv) *If η is an approximate design with weights $w_i = r_i/N$ at the points x_i ($i = 1, \dots, n$), then the direct estimator (2.21) and the least squares estimate coincide if the design $\xi = \{x_1, \dots, x_n\}$ satisfies (2.18).*

Proof. The first part has been established in Ermakov and Sedunov (1974). In order to prove (ii) define $\beta' = E[\tilde{\beta}]$, where $\tilde{\beta}$ is the direct estimate of $\bar{\beta}$ defined in (2.14), then a straightforward calculation shows that

$$(2.22) \quad \bar{B} = \int_a^b [\mu(x) - \beta' \psi_{(2)}(x)]^2 \lambda(x) dx = \|\bar{\gamma}\|_2^2 + \|\bar{\beta} - \beta'\|_2^2 .$$

Consequently, we have $\bar{B} = \|\bar{\gamma}\|_2^2$ if $\bar{\beta} = \beta'$, which implies (observing (2.14))

$$(2.23) \quad \bar{\beta} = \beta' = \sum_{j=1}^n \mu(x_j) \psi_{(1)}(x_j) \delta_j = G \begin{pmatrix} \bar{\beta} \\ \bar{\gamma} \end{pmatrix} ,$$

where the matrix G is given by

$$(2.24) \quad G = \begin{pmatrix} \sum_{j=1}^n \psi_\ell(x_j) \psi_i(x_j) \delta_j \end{pmatrix}_{\substack{\ell=1, \dots, k_1+k_2 \\ i=1, \dots, k_1}} .$$

This implies that $G = (I_{k_1} \ ; \ 0)$, where I_{k_1} is the $k_1 \times k_1$ unit matrix and 0 is the $k_1 \times k_2$ matrix with all entries equal to 0. Consequently condition (2.18) is necessary and sufficient.

For a proof of part (iii) we use the definition of $Z_{(1)}^T = (f_i(x_j))_{i=1, \dots, k_1}^{j=1, \dots, n}$. If $\xi = \{x_1, \dots, x_n\}$ denotes a design such that (2.18) is satisfied with $\delta_j = \frac{1}{n}$ ($j = 1, \dots, n$) then $Z_{(1)}^T Z_{(1)} = nI_{k_1}$. This implies for the least squares estimate in the orthonormalized regression model

$$(2.25) \quad \hat{\beta} = (Z_{(1)}^T Z_{(1)})^{-1} Z_{(1)}^T Y = \frac{1}{n} Z_{(1)}^T Y = \frac{1}{n} \sum_{j=1}^n \psi_{(1)}(x_j) Y_j = \tilde{\beta} .$$

Finally the representation of the mean squared error follows by a straightforward calculation observing the orthonormality of the functions ψ_i with respect to the measure $\delta(x) dx$ and (2.17). \square

The main implication of Theorem 2.5 is the following. Any quadrature formula satisfying (2.17) generates an all bias design for least squares and direct estimation. Moreover, from the basic properties of numerical integration it follows that trivial solutions as indicated by Example 2.1 and 2.3 can be avoided by using all-bias designs derived from quadrature formulas. Moreover, for quadrature formulas with equal weight the least squares and direct estimate coincide. Thus from the well-known results from numerical integration one can easily derive non-trivial all-bias designs, and we will illustrate this in the following section for spline regression models.

3 Spline regression models

Spline models represent an attractive class of regression models because of their ability of approximating a curve with different degrees of smoothness at different locations [see e.g. De Boor (1978) or Diercx (1995)]. Several authors have investigated optimal design problems for spline regression models using optimality criteria based on the predicted variance [see e.g. Studden and Van Arman (1969), Studden (1971), Murty (1971), Park (1978), Kaishev (1989) or Heiligers (1998) among others]. Recently Woods (2005) and also Woods and Lewis (2006) considered the problem of constructing all-bias designs for maximally smooth splines, and we will illustrate potential applications of the discussion in Section 2 for these models.

To be precise we consider the models (2.1) and (2.2) where the vectors

$$(3.1) \quad f_{(1)}^T(x) = (1, x, \dots, x^{d-1}, (x - t_1)_+^{d-1}, \dots, (x - t_\ell)_+^d)^T,$$

$$(3.2) \quad f_{(2)}^T(x) = ((x - u_1)_+^{d-1}, \dots, (x - u_k)_+^{d-1})^T,$$

correspond to a maximally smooth spline of degree $d - 1$, the design interval is given by $[a, b] = [-1, 1]$ and the pairwise different knots $t_1, \dots, t_\ell, u_1, \dots, u_k$ satisfy $a < t_1 < \dots < t_\ell < b$; $a < u_1 < \dots < u_k < b$. The motivation for considering models of this type stems from the fact that in practical applications the number of knots cannot be fixed in advance [see Woods and Lewis (2006)]. We consider the all-bias criterion (2.3) with weight function $\lambda(x) = \frac{1}{2}I_{[-1,1]}(x)$. Recently Woods and Lewis (2006) determined all-bias designs in this context using the condition (2.18) with $\delta_j = 1/n$ ($j = 1, \dots, n$). In the following we will present more general results about all-bias designs for spline regression models using quadrature formulas which have not necessarily equal weights.

For this purpose we note that $k_1 = d + \ell, k_2 = k$ and define by

$$(3.3) \quad \lambda_0 = a < \lambda_1 < \lambda_2 < \dots < \lambda_{\ell+k} < \lambda_{\ell+k+1} = 1$$

the ordered knots $t_1, \dots, t_\ell, u_1, \dots, u_k$. For a given design $\xi = \{x_1, \dots, x_n\}$ we consider the numbers n_j of points located in the interval $[\lambda_{j-1}, \lambda_j]$ ($j = 1, \dots, \ell + k + 1$), that is

$$(3.4) \quad \begin{aligned} a = \lambda_0 &\leq x_1 < x_2 < \dots < x_{n_1} \leq \lambda_1 \\ \lambda_1 &\leq x_{n_1+1} < \dots < x_{n_1+n_2} \leq \lambda_2 \\ &\vdots \\ \lambda_{\ell+k} &\leq x_{n_1+\dots+n_{\ell+k+1}} < \dots < x_{n_1+\dots+n_{\ell+k+1}} \leq \lambda_{\ell+k+1} = b \end{aligned}$$

(note that $n_1 + \dots + n_{\ell+k+1} = n$) and obtain the following result.

Theorem 3.1. Consider a design $\xi = \{x_1, \dots, x_n\}$ of the form (3.4) such that for each $j = 1, \dots, \ell + k + 1$ the points ($n_0 = 0$)

$$(3.5) \quad x_{n_0+\dots+n_{j-1}+1} < x_{n_0+\dots+n_{j-1}+2} < \dots < x_{n_0+\dots+n_j}$$

define a quadrature formula which integrates polynomials of degree $2d - 2$ exactly on the interval $[\lambda_{j-1}, \lambda_j]$. The design ξ is an all-bias design for the spline regression model (3.1) and (3.2).

Proof. Note that on the interval $[\lambda_{j-1}, \lambda_j]$ ($j = 1, \dots, \ell + k + 1$) the functions $f_i(x)f_j(x)$ are polynomials of degree $2d - 2$. Because each set of design points $\{x_{n_0+\dots+n_{j-1}+1}, \dots, x_{n_0+\dots+n_j}\}$ corresponds to a quadrature formula, which integrates polynomials of degree $2d - 2$ exactly, there exists weights $\delta_1, \dots, \delta_n$ such that for each $j = 1, \dots, \ell + k + 1$ the identity

$$(3.6) \quad \int_{\lambda_{j-1}}^{\lambda_j} f_i(x)f_h(x)dx = \sum_{s=n_0+\dots+n_{j-1}+1}^{n_0+\dots+n_j} \delta_s f_i(x_s)f_h(x_s); \quad i, h = 1, \dots, k_1 + k$$

is satisfied. This implies

$$(3.7) \quad \int_{-1}^1 f_i(x)f_h(x)dx = \sum_{s=1}^n \delta_s f_i(x_s)f_h(x_s); \quad i, h = 1, \dots, k_1 + k,$$

and shows that condition (2.18) is satisfied. The assertion now follows from Theorem 2.5. \square

Theorem 3.2. Let z_1, \dots, z_d denote the roots of the d th Legendre polynomial orthogonal with respect to the measure $\frac{1}{2} dx$ on the interval $[-1, 1]$, define $n_1 = \dots = n_{\ell+k+1} = d, n = (\ell+k+1) d$, then the design $\xi = \{x_1, \dots, x_n\}$ with

$$\begin{aligned} x_1 &= \frac{\lambda_1 - \lambda_0}{2} z_1 + \frac{\lambda_1 + \lambda_0}{2}, \dots, x_d = \frac{\lambda_1 - \lambda_0}{2} z_d + \frac{\lambda_1 + \lambda_0}{2} \\ x_{d+1} &= \frac{\lambda_2 - \lambda_1}{2} z_1 + \frac{\lambda_2 + \lambda_1}{2}, \dots, x_{2d} = \frac{\lambda_2 - \lambda_1}{2} z_d + \frac{\lambda_2 + \lambda_1}{2} \\ &\vdots \\ x_{(k+\ell)d+1} &= \frac{\lambda_{k+\ell+1} - \lambda_{k+\ell}}{2} z_1 + \frac{\lambda_{k+\ell+1} + \lambda_{k+\ell}}{2}, \dots, x_{(k+\ell+1)d} = \frac{\lambda_{k+\ell+1} - \lambda_{k+\ell}}{2} z_d + \frac{\lambda_{k+\ell+1} + \lambda_{k+\ell}}{2} \end{aligned}$$

is an all-bias design for the spline regression model (3.1) and (3.2).

Proof. It is well known [see e.g. Ghizetti and Ossicini (1970)] that the roots z_1, \dots, z_d of the d th Legendre polynomial define a quadrature formula, which integrates polynomials of degree $2d - 2$ on the interval $[-1, 1]$ exactly, i.e. there exists weights $\delta_1, \dots, \delta_d$ such that the identities

$$(3.8) \quad \frac{1}{2} \int_{-1}^1 t^j dt = \sum_{s=1}^d z_s^j \delta_s$$

are satisfied for $j = 1, \dots, 2d - 2$. A straightforward calculation now shows that for any $a < b$ the transformed points $\bar{z}_i = \frac{b-a}{z}z_i + \frac{b+a}{2}$ ($i = 1, \dots, d$) define a quadrature formula on the interval $[a, b]$, which integrates polynomials of degree $2d - 2$ exactly. Consequently, for each $j = 1, \dots, \ell + k + 1$ the points $x_{(j-1)d+1}, \dots, x_{jd}$ defined in Theorem 3.2 correspond to a quadrature formula, which integrates polynomials of degree $2d - 2$ exactly on the interval $[\lambda_{j-1}, \lambda_j]$. The assertion now follows from Theorem 3.1. \square

Example 3.3. We consider the case $d = 2, \ell = k = 1$ and $t_1 = -\frac{3}{7}, u_1 = \frac{1}{7}$ [see Woods and Lewis (2006), Section 5]. In this case we obtain for the regression functions in (3.1) and (3.2)

$$(3.9) \quad f_{(1)}^T(x) = (1, x, (x + \frac{3}{7})_+); f_{(2)}(x) = (x - \frac{1}{7})_+.$$

The Legendre polynomial of degree 2 is given by $x^2 - \frac{1}{3}$ with roots $-\frac{1}{\sqrt{3}}$ and $\frac{1}{\sqrt{3}}$. From Theorem 3.2 we obtain that the design $\xi = \{x_1, \dots, x_6\}$ with

$$(3.10) \quad \begin{aligned} x_1 &= \frac{1}{7}(-5 - \frac{2}{\sqrt{3}}), & x_2 &= \frac{1}{7}(-5 + \frac{2}{\sqrt{3}}), \\ x_3 &= \frac{1}{7}(-1 - \frac{2}{\sqrt{3}}), & x_4 &= \frac{1}{7}(-1 + \frac{2}{\sqrt{3}}), \\ x_5 &= \frac{1}{7}(4 - \frac{3}{\sqrt{3}}), & x_6 &= \frac{1}{7}(4 + \frac{3}{\sqrt{3}}) \end{aligned}$$

is an all-bias design. Note that in contrast to the all-bias design derived by Woods and Lewis (2006) this design requires only 6 points, while the design derived by these authors has 7 points.

We can now use results from numerical integration to derive numerous further all-bias designs. For example, if $d = 3$ it follows from De Boor (1978), that the quadrature formula with weights $5/18, 4/9, 5/18$ at the points $-\sqrt{3/5}, 0, \sqrt{3/5}$ integrates polynomials of degree 4 exactly. The corresponding all-bias design for the model (3.1) and (3.2) with $d = 1$ or $d = 2, \ell = k = 1, t_1 = -3/7$ and $u_1 = 1/7$ is given by

$$(3.11) \quad \begin{aligned} x_1 &= \frac{1}{7}(-5 - 2\sqrt{\frac{3}{5}}), & x_2 &= -\frac{3}{7}, & x_3 &= \frac{1}{7}(-5 + 2\sqrt{\frac{3}{5}}), \\ x_4 &= \frac{1}{7}(-1 - 2\sqrt{\frac{3}{5}}), & x_5 &= \frac{1}{7}, & x_6 &= \frac{1}{7}(-1 + 2\sqrt{\frac{3}{5}}), \\ x_7 &= \frac{1}{7}(4 - 3\sqrt{\frac{3}{5}}), & x_8 &= \frac{4}{7}, & x_9 &= \frac{1}{7}(4 + 3\sqrt{\frac{3}{5}}). \end{aligned}$$

The main differences between the results of this paper and the results of Woods and Lewis (2006) is that the last named authors restrict their considerations to designs satisfying (2.18) with equal

weights $\delta_1 = \dots = \delta_n = 1/n$. As illustrated in Section 2 this corresponds to the determination of a quadrature formula with equal weights, which integrates polynomials of degree $2d - 2$ exactly. By relaxing the condition of equal weights we can derive a large number of alternative all-bias designs. Moreover, the discussion presented so far allows us a simple calculation of the all-bias designs using the theory of quadrature which has very well been developed for a long time [see e.g. de Boor (1978)].

Theorem 3.4. *Consider the polynomial spline model (3.1) and (3.2), where $n_j = (\lambda_j - \lambda_{j-1})n/2 \in \mathbb{N}$ ($j = 1, \dots, \ell + k + 1$). If for each $j = 1, \dots, \ell + k + 1$ the knots $q_1^{(j)}, \dots, q_{n_j}^{(j)}$ correspond to a quadrature formula, which integrates polynomials of degree $2d - 2$ exactly, then the design $\xi = \{x_1, \dots, x_n\}$ with*

$$\begin{aligned} x_1 &= \frac{\lambda_1 - \lambda_0}{2} q_1^{(1)} + \frac{\lambda_1 + \lambda_0}{2}, \dots, x_{n_1} = \frac{\lambda_1 - \lambda_0}{2} q_{n_1}^{(1)} + \frac{\lambda_1 + \lambda_0}{2} \\ x_{n_1+1} &= \frac{\lambda_2 - \lambda_1}{2} q_1^{(2)} + \frac{\lambda_2 + \lambda_1}{2}, \dots, x_{n_1+n_2} = \frac{\lambda_2 - \lambda_1}{2} q_{n_2}^{(2)} + \frac{\lambda_2 + \lambda_1}{2} \\ &\vdots \\ x_{n_1+\dots+n_{\ell+k+1}} &= \frac{\lambda_{\ell+k+1} - \lambda_{\ell+k}}{2} q_1^{(\ell+k+1)} + \frac{\lambda_{\ell+k+1} + \lambda_{\ell+k}}{2}, \dots, \\ & x_{n_1+\dots+n_{\ell+k+1}} = \frac{\lambda_{\ell+k+1} - \lambda_{\ell+k}}{2} q_{n_{\ell+k+1}}^{(\ell+k+1)} + \frac{\lambda_{\ell+k+1} + \lambda_{\ell+k}}{2} \end{aligned}$$

satisfies condition (2.18) with equal weights $\delta_1 = \dots = \delta_n = 1/n$ and is an all-bias design.

Proof. The proof follows by similar arguments as presented in the proof of Theorem 3.2 and is therefore omitted. □

Example 3.5. Some quadrature formulas with equal weight can be found in Engels (1980), p. 58. For example in the situation considered in Example 3.3, it follows from these results, that the points $-1/\sqrt{3}$ and $1/\sqrt{3}$ correspond in fact to a quadrature formula with equal weights, which integrates polynomials of degree 2 exactly. The case considered by Woods and Lewis (2006) [see Example 3.3] can be obtained by using a quadrature with equal weights and two points at the intervals $[-1, -\frac{3}{7}]$, $[-\frac{3}{7}, \frac{1}{7}]$ and three points on the interval $[\frac{1}{7}, 1]$. Similarly, if $d = 3$ (cubic spline) it follows from Engels (1980), that the points

$$q_1 = -\sqrt{\frac{1}{15}(5 - 2\sqrt{5})} = -q_3 \approx -0.1876; \quad -\sqrt{\frac{1}{15}(5 + 2\sqrt{5})} \approx -0.7947$$

define a quadrature formula with equal weights, which integrates polynomials of degree 4 exactly [see Engels (1980), p. 58], and all-bias designs for models with knots satisfying $\lambda_j - \lambda_{j-1} = 8/n$ are obtained easily.

Acknowledgements. The support of the Deutsche Forschungsgemeinschaft (SFB 475, "Komplexitätsreduktion in multivariaten Datenstrukturen") is gratefully acknowledged. The work by V.B.Melas was partly supported by the Russian Foundation of Basic Research (project 04-01-00519). The work of H. Dette was supported in part by the DAAD and by an NIH grant award IR01GM072876:01A1. The authors are also grateful to Martina Stein, who typed parts of this paper with considerable technical expertise.

References

- C. De Boor (1978). A practical guide to splines. Applied Mathematical Sciences. Vol. 27. New York-Heidelberg-Berlin: Springer-Verlag. XXIV.
- G.E.P. Box, D.R. Draper (1959). A basis for the selection of a response surface design. Journal of the American Statistical Association 54, 622-654.
- H. Dette (1990). A generalization of D - and D_1 -optimal designs in polynomial regression. Ann. Statist. 18, 1784 - 1805.
- P. Dierckx (1995). Curve and surface fitting with splines. Monographs on Numerical Analysis. Oxford: Clarendon Press.
- S. Efromovich (1999). Nonparametric curve estimation. Springer, N.V.
- H. Engels (1980). Numerical quadrature and cubature. Academic Press, N.Y.
- S.M. Ermakov, V.B. Melas (1995). Design and Analysis of Simulation Experiments. Kluwer Academic Publishers, London.
- S.M. Ermakov, E.V. Sedunov (1974). Unbiased design and regression experiment analysis in a finite dimensional function space. Vestnik Leningrad Universitita 1, 11-21
- Z. Galil, J. Kiefer (1977). Comparison of Box-Draper and D -optimum designs for experiments with mixtures. Technometrics 19, 441-444.
- A. Ghizetti, A. Ossicini (1970). Quadrature formulae. Birkhäuser, Basel.
- B. Heiligers (1998). E-optimal designs for polynomial spline regression. J. Stat. Plann. Inference 75, 159-172.
- V. K. Kaishev (1989). Optimal experimental designs for the B-spline regression. Comput. Stat. Data Anal. 8, 39-47.
- J. C. Kiefer (1974). General equivalence theory for optimum designs (approximate theory). Ann. Statist. 2, 849-879.
- E. Läuter (1974). Experimental design for a class of models. Mathematische Operationsforschung und Statistik 5, 379-398.

- V. N. Murty (1971). Optimal designs with a Tchebycheffian spline regression function. *Ann. Math. Stat.* 42, 643-649.
- S. H. Park (1978). Experimental designs for fitting segmented polynomial regression models. *Technometrics* 20, 151-154.
- F. Pukelsheim (1993). *Optimal Design of Experiments*. Wiley, N.Y.
- F. Pukelsheim, J.L. Rosenberger (1993). Experimental designs for model discrimination. *Journal of the American Statistical Association* 98, 642-649.
- S.D. Silvey (1980). *Optimal Design*. Chapman and Hall, London.
- S. Stigler (1971). Optimal experimental designs for polynomial regression. *Journal of the American Statistical Association* 66, 311-318.
- W. J. Studden (1971). Optimal designs and spline regression. In *Optimizing methods in statistics (Proc. Sympos., Ohio State Univ., Columbus, Ohio)*, 63-76. Academic Press, New York.
- W.J. Studden (1982). Some robust type D -optimal designs in polynomial regression. *Journal of the American Statistical Association* 8, 916-921.
- W. J. Studden, D. J. Van Arman (1969). Admissible designs for polynomial spline regression. *Ann. Math. Statist.*, 40:1557-1569.
- W.J. Welch (1983). A mean squared error criterion for the design of experiments. *Biometrika* 70, 205-213.
- D.C. Woods (2005). Designing experiments under random contamination with application to polynomial spline regression. *Statistica Sinica*, 15, 619-635.
- D. Woods, S. Lewis (2006). All-bias designs for polynomial spline regression models. *Aust. N.Z.J. Stat.* 48, 49 -58.
- R.X Yue, F.Y. Hickernell (1999). Robust designs for fitting linear models with misspecification. *Stat. Sinica* 9, 1053-1069.