Statistica Sinica **9**(1999), 1053-1069

ROBUST DESIGNS FOR FITTING LINEAR MODELS WITH MISSPECIFICATION

Rong-Xian Yue*† and Fred J. Hickernell*

**Hong Kong Baptist University and* †*Shanghai Normal University*

Abstract: This paper considers linear models with misspecification of the form $f(\boldsymbol{x}) = E(y|\boldsymbol{x}) = \sum_{j=1}^{p} \theta_j g_j(\boldsymbol{x}) + h(\boldsymbol{x})$, where $h(\boldsymbol{x})$ is an unknown function. We assume that the true response function f comes from a reproducing kernel Hilbert space and the estimates of the parameters θ_j are obtained by the standard least squares method. A sharp upper bound for the mean squared error is found in terms of the norm of h. This upper bound is used to choose a design that is robust against the model bias. It is shown that the continuous uniform design on the experimental region is the all-bias design. The numerical results of several examples show that all-bias designs perform well when some model bias is present in low dimensional cases.

Key words and phrases: Linear models with misspecification, model-robust designs, reproducing kernel Hilbert spaces.

1. Introduction

This paper considers the design problem for linear regression given by

$$y_i = \sum_{j=1}^{p} \theta_j g_j(\boldsymbol{x}_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

when there is deviation (misspecification or bias) from the assumed model form. Here the specified functions g_j are linearly independent, the $\boldsymbol{x}_i$ are n points drawn from the design region $\mathcal{X} \subset \mathcal{R}^s$. The ϵ_i are i.i.d. random errors with mean 0 and variance σ^2. We represent the true response by

$$\begin{aligned} f(\boldsymbol{x}) &= \textstyle\sum_{j=1}^{p} \theta_j g_j(\boldsymbol{x}) + h(\boldsymbol{x}), \\ y_i &= f(\boldsymbol{x}_i) + \epsilon_i, \quad i = 1, \ldots, n, \end{aligned} \tag{1.2}$$

where h is an unknown function from some class $\mathcal{H}$, which will be specified later. Since the bias h is unknown and may vary freely in $\mathcal{H}$, the designs must be chosen such that the fitted model provides an adequate approximation to a range of possible true models, i.e., is robust to the exact form of the true model in some sense. This is the fundamental goal of model-robust design.

The model-robust design problem has been studied by many authors, whose investigations differ in specification of the class $\mathcal{H}$, the design region, the regressors, and in the loss functions used. Box and Draper (1959) and Kiefer (1973) restrict their attentions to finite dimensional $\mathcal{H}$ and the least squares estimators or linear estimators. Huber (1975), Marcus and Sacks (1978), Pesotchinsky (1982), Li and Notz (1982), Li (1984) and Wiens (1990, 1992), and some others deal with infinite dimensional $\mathcal{H}$. Some of them take $\mathcal{H} = \{h : \ |h(\boldsymbol{x})| \leq \phi(\boldsymbol{x}), \ \boldsymbol{x} \in \mathcal{X}\}$ with various assumptions being made about ϕ. The designs constructed appear to be quite sensitive to the assumed form of ϕ. The others take $\mathcal{H} = \{h : \ \int_{\mathcal{X}} [h(\boldsymbol{x})]^2 d\boldsymbol{x} \leq c, \ \int_{\mathcal{X}} g_j(\boldsymbol{x}) h(\boldsymbol{x}) d\boldsymbol{x} = 0, \ j = 1, \ldots, p\}$, and use the least squares estimators. Here c is assumed known, and the second condition ensures the identifiability of the θ_j. One criticism of this specification is that only designs which are absolutely continuous on $\mathcal{X}$ have a finite loss. The reason is that such a $\mathcal{H}$ includes h which have an arbitrarily high, narrow spike above any points in $\mathcal{X}$. The review by Chang and Notz (1996) provides a good summary of the previous work in this subject.

In this paper, we assume that $\mathcal{H}$ is a reproducing kernel Hilbert space admitting a reproducing kernel $K(\boldsymbol{x}, \boldsymbol{w})$ and an inner product $\langle \cdot, \cdot \rangle$. The definition of a reproducing kernel is a function on $\mathcal{X} \times \mathcal{X}$ such that $K(\cdot, \boldsymbol{w}) \in \mathcal{H}$ for all $\boldsymbol{w} \in \mathcal{X}$, and

$$h(\boldsymbol{w}) = \langle h, K(\cdot, \boldsymbol{w}) \rangle, \quad \text{for all } h \in \mathcal{H} \text{ and all } \boldsymbol{x} \in \mathcal{X}. \tag{1.3}$$

We also assume that

$$\int_{\mathcal{X}} g_j(\boldsymbol{x}) h(\boldsymbol{x}) d\boldsymbol{x} = 0, \quad j = 1, \ldots, p, \quad \text{for any } h \in \mathcal{H}. \tag{1.4}$$

For the details of reproducing kernel Hilbert spaces, we refer to Aronszajn (1950), Saitoh (1988) and Wahba (1990). It can be shown that any reproducing property is symmetric in its arguments and positive definite:

$$K(\boldsymbol{x}, \boldsymbol{w}) = K(\boldsymbol{w}, \boldsymbol{x}) \quad \text{for all } \boldsymbol{x}, \boldsymbol{w} \in \mathcal{X},$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0 \quad \text{for all } a_i \in \mathcal{R} \text{ and any } \boldsymbol{x}_i \in \mathcal{X}.$$

As was done in the previous work mentioned above, we confine ourselves to the use of the least squares estimators both because these estimators do not depend on the type of deviation from the model, and because in the case of small deviations, the least squares estimators perform well, as was shown by Marcus and Sacks (1978) and Sacks and Ylvisaker (1978).

2. Development of Design Criteria

Assume the true response is given by (1.2), and the class $\mathcal{H}$ is as specified above. Let ξ_n be a sequence of n design points $\boldsymbol{x}_i$ in $\mathcal{X}$, $\boldsymbol{y}$ be the vector of n observations, y_i. Let $\boldsymbol{g}$ be the vector of p regressors g_j, and $\boldsymbol{X}$ be the design matrix, i.e., $\boldsymbol{X} = (\,\boldsymbol{g}(\boldsymbol{x}_1), \dots, \boldsymbol{g}(\boldsymbol{x}_n)\,)^\top$. Our consideration here is limited to the designs ξ_n with nonsingular information matrix $\boldsymbol{M} = \boldsymbol{X}^\top\boldsymbol{X}$. Then the least squares estimator of $\boldsymbol{\theta} = (\,\theta_1, \dots, \theta_p\,)^\top$ in the assumed model is $\hat{\boldsymbol{\theta}} = \boldsymbol{M}^{-1}\boldsymbol{X}^\top\boldsymbol{y}$. Let $f^*(\boldsymbol{x}) = \boldsymbol{\theta}^\top\boldsymbol{g}(\boldsymbol{x})$ and $\hat{f}(\boldsymbol{x}) = \hat{\boldsymbol{\theta}}^\top\boldsymbol{g}(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathcal{X}$. We consider the integrated mean squared error over the region $\mathcal{X}$:

$$R(\xi_n, h) = \int_{\mathcal{X}} E[\hat{f}(\boldsymbol{x}) - f^*(\boldsymbol{x})]^2 d\boldsymbol{x}. \tag{2.1}$$

Introducing the matrix and vector

$$\boldsymbol{G} = \int_{\mathcal{X}} \boldsymbol{g}(\boldsymbol{x})\boldsymbol{g}^\top(\boldsymbol{x})d\boldsymbol{x}, \quad \boldsymbol{h} = (\,h(\boldsymbol{x}_1), \dots, h(\boldsymbol{x}_n)\,)^\top, \tag{2.2}$$

the loss function $R(\xi_n, h)$ can be expressed as follows:

$$R(\xi_n, h) = \sigma^2 \mathrm{tr}\{\boldsymbol{M}^{-1}\boldsymbol{G}\} + \boldsymbol{h}^\top\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{G}\boldsymbol{M}^{-1}\boldsymbol{X}^\top\boldsymbol{h}. \tag{2.3}$$

The first term in (2.3) is the contribution from the variance error, and the second term is the contribution from the bias error. The following theorem provides a sharp upper bound for $R(\xi_n, h)$, all $h \in \mathcal{H}$.

Theorem 1. *Suppose the true model is given by* (1.2). *Assume that the class of deviations, $\mathcal{H}$, is a reproducing kernel Hilbert space admitting the reproducing kernel $K(\boldsymbol{x}, \boldsymbol{w})$ and that $\mathcal{H}$ satisfies condition* (1.4). *Let $\boldsymbol{k}$ be the vector of n functions $K(\cdot, \boldsymbol{x}_i)$, $i = 1, \dots, n$, and let $\boldsymbol{K}$ be the matrix with entries $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $i, j = 1, \dots, n$. Define*

$$\boldsymbol{Q} = \boldsymbol{G}^{\frac{1}{2}}\boldsymbol{M}^{-1}\boldsymbol{X}^\top\boldsymbol{K}\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{G}^{\frac{1}{2}}. \tag{2.4}$$

Then

$$R(\xi_n; h) \leq \sigma^2 \mathrm{tr}\{\boldsymbol{M}^{-1}\boldsymbol{G}\} + \lambda_{\max}(\boldsymbol{Q})\|h^2\|, \tag{2.5}$$

where $\lambda_{\max}(\boldsymbol{Q})$ is the maximum eigenvalue of the matrix $\boldsymbol{Q}$. Let $\boldsymbol{v}_{\max}(\boldsymbol{Q})$ be the (*scaled*) *eigenvector of $\boldsymbol{Q}$ corresponding to $\lambda_{\max}(\boldsymbol{Q})$, and let*

$$W(\boldsymbol{x}) = \boldsymbol{v}^\top_{\max}(\boldsymbol{Q})\boldsymbol{G}^{\frac{1}{2}}\boldsymbol{M}^{-1}\boldsymbol{X}^\top\boldsymbol{k}(\boldsymbol{x}). \tag{2.6}$$

Then equality in (2.5) *is attained when h is a constant multiple of W.*

Proof. Comparing with (2.3) and setting $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{L}\boldsymbol{M}^{-1}\boldsymbol{X}^\top$, we need only show that

$$\boldsymbol{h}^\top \boldsymbol{A}\boldsymbol{h} \le \lambda_{\max}(\boldsymbol{Q})\|h\|^2. \tag{2.7}$$

First, using reproducing property (1.3) and the definition of $\boldsymbol{k}$, we have

$$\boldsymbol{h} = (\,\langle K(\cdot,\boldsymbol{x}_1)h\rangle, \ldots, \langle K(\cdot,\boldsymbol{x}_n)h\rangle\,)^\top = \langle \boldsymbol{k}h\rangle.$$

Put $\boldsymbol{B} = \boldsymbol{G}^{\frac{1}{2}}\boldsymbol{M}^{-1}\boldsymbol{X}^\top$ and $\boldsymbol{\zeta} = \boldsymbol{B}\boldsymbol{k}$. Then the left hand side in (2.7) can be written as

$$\boldsymbol{h}^\top \boldsymbol{A}\boldsymbol{h} = (\boldsymbol{B}\boldsymbol{h})^\top(\boldsymbol{B}\boldsymbol{h}) = \langle \boldsymbol{\zeta}^\top h\rangle\langle \boldsymbol{\zeta} h\rangle = \sum_{j=1}^{p}\langle \zeta_j h\rangle^2, \tag{2.8}$$

where the ζ_j are the components of $\boldsymbol{\zeta}$.

On the other hand, $h \in \mathcal{H}$ can be expressed as a sum of an element in $\operatorname{span}\{\zeta_1,\ldots,\zeta_p\}$ and its orthogonal complement. That is, $h = h_0 + h_1$ with $h_0 = \sum_{j=1}^{p} b_j\zeta_j \equiv \boldsymbol{b}^\top\boldsymbol{\zeta}$ and $\langle \zeta_j, h_1\rangle = 0$, $j = 1,\ldots,p$ for a set of some constants b_j. Let $\boldsymbol{e}_j$ be the the jth column vector of $\boldsymbol{I}_p$. Then for each j,

$$\langle \zeta_j, h\rangle = \sum_{\alpha=1}^{p} b_\alpha\langle \zeta_j, \zeta_\alpha\rangle = \boldsymbol{e}_j^\top(\boldsymbol{B}\boldsymbol{K}\boldsymbol{B}^\top)\boldsymbol{b} = \boldsymbol{e}_j^\top\boldsymbol{Q}\boldsymbol{b}. \tag{2.9}$$

Therefore we have

$$\|h\|^2 = \|h_0\|^2 + \|h_1\|^2 = \boldsymbol{b}^\top\langle \boldsymbol{\zeta}, \boldsymbol{\zeta}^\top\rangle\boldsymbol{b} + \|h_1\|^2 = \boldsymbol{b}^\top\boldsymbol{Q}\boldsymbol{b} + \|h_1\|^2,$$

which implies that

$$\boldsymbol{b}^\top\boldsymbol{Q}\boldsymbol{b} \le \|h\|^2, \tag{2.10}$$

equality holding when $h = \boldsymbol{b}^\top\boldsymbol{\zeta}$. On the other hand, applying (2.9) to (2.8), we obtain

$$\boldsymbol{h}^\top\boldsymbol{A}\boldsymbol{h} = \sum_{j=1}^{p}\boldsymbol{b}^\top\boldsymbol{Q}^\top\boldsymbol{e}_j\boldsymbol{e}_j^\top\boldsymbol{Q}\boldsymbol{b} = \boldsymbol{b}^\top\boldsymbol{Q}^2\boldsymbol{b}$$

since $\sum_{j=1}^{p}\boldsymbol{e}_j\boldsymbol{e}_j^\top = \boldsymbol{I}_p$. From the extremal properties of eigenvalues, it follows that

$$\boldsymbol{h}^\top\boldsymbol{A}\boldsymbol{h} \le \lambda_{\max}(\boldsymbol{Q})\boldsymbol{b}^\top\boldsymbol{Q}\boldsymbol{b},$$

equality holding when $\boldsymbol{b} = \boldsymbol{v}_{\max}(\boldsymbol{Q})$. The desired results then follow from (2.10).

Remark 1. We call the function $W(\cdot)$ defined in (2.6) a worst-case bias. From its definition,

$$\begin{aligned}\|W\|^2 = \langle W, W\rangle &= \boldsymbol{v}_{\max}^\top\boldsymbol{G}^{\frac{1}{2}}\boldsymbol{M}^{-1}\boldsymbol{X}^\top\langle \boldsymbol{k}, \boldsymbol{k}^\top\rangle\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{G}^{\frac{1}{2}}\boldsymbol{v}_{\max}\\ &= \boldsymbol{v}_{\max}^\top\boldsymbol{Q}\boldsymbol{v}_{\max} = \lambda_{\max}\boldsymbol{v}_{\max}^\top\boldsymbol{v}_{\max} = \lambda_{\max}(\boldsymbol{Q}),\end{aligned}$$

i.e., the quantity $\lambda_{\max}(\boldsymbol{Q})$ is the squared norm of the worst-case bias. The inequality (2.7) provides a *sharp* upper bound for the bias term in the right hand side of (2.3) for any $h \in \mathcal{H}$. This bound is composed of two parts: $\|h\|^2$, which can be viewed as a measure of the size of variation or fluctuation of the bias, and $\|W\|^2 = \lambda_{\max}(\boldsymbol{Q})$, which is a measure of the worst-case bias. This quantity depends only on the design points (and the specification of the reproducing kernel). At these points, the inequality (2.7) is similar to the quadrature error bound in the literature having the form

$$\left| \int_{\mathcal{X}} h(\boldsymbol{x}) d\boldsymbol{x} - \frac{1}{n} \sum_{i=1}^{n} h(\boldsymbol{x}_i) \right| \leq D(\xi_n) V(h).$$

See, for example, Niederreiter (1992) and Hickernell (1996).

Remark 2. The reproducing kernel $K(\cdot,\cdot)$ contains information about the worst possible bias $h(\boldsymbol{w})$ at a fixed point $\boldsymbol{w}$ in $\mathcal{X}$. Let $h^*_{\boldsymbol{w}}$ be the function in $\mathcal{H}$ that maximizes $h(\boldsymbol{w})$ over $\mathcal{H}$ subject to $\|h\| = 1$, i.e., $h^*_{\boldsymbol{w}} = \operatorname{argmax}_{\|h\|=1} h(\boldsymbol{w})$. This function is determined by the reproducing kernel as follows:

$$h^*_{\boldsymbol{w}}(\boldsymbol{x}) = [K(\boldsymbol{w},\boldsymbol{w})]^{-1/2} K(\boldsymbol{x},\boldsymbol{w}), \quad \boldsymbol{x} \in \mathcal{X}. \tag{2.11}$$

To see why, note that for any $h \in \mathcal{H}$ with $\|h\| = 1$, we have

$$h(\boldsymbol{w}) = \langle K(\cdot,\boldsymbol{w}) h \rangle \leq \|K(\cdot,\boldsymbol{w})\| \, \|h\| = \|K(\cdot,\boldsymbol{w})\| = [K(\boldsymbol{w},\boldsymbol{w})]^{1/2}$$

by (3.1), and the equality holds when $h = h^*_{\boldsymbol{w}} \propto K(\cdot,\boldsymbol{w})$. The value $K(\boldsymbol{w},\boldsymbol{w})$ is positive because the reproducing kernel is positive definite. By definition we then have $h^*_{\boldsymbol{w}}(\boldsymbol{x}) = c(\boldsymbol{w}) K(\boldsymbol{x},\boldsymbol{w})$ for some c depending only on $\boldsymbol{w}$. The value of $c(\boldsymbol{w})$ follows from the requirement that

$$\|h^*_{\boldsymbol{w}}\| = c(\boldsymbol{w}) \|K(\cdot,\boldsymbol{w})\| = c(\boldsymbol{w}) [K(\boldsymbol{w},\boldsymbol{w})]^{1/2} = 1.$$

Thus $c(\boldsymbol{w}) = [K(\boldsymbol{w},\boldsymbol{w})]^{-1/2}$ and (2.11) is proved. Furthermore, note that $h^*_{\boldsymbol{w}}(\boldsymbol{w}) = [K(\boldsymbol{w},\boldsymbol{w})]^{1/2}$, so that the reproducing kernel for $\mathcal{H}$ can be expressed in terms of the worst function $h^*_{\boldsymbol{w}}$ as follows:

$$K(\boldsymbol{x},\boldsymbol{w}) = h^*_{\boldsymbol{w}}(\boldsymbol{w}) h^*_{\boldsymbol{w}}(\boldsymbol{x}) = h^*_{\boldsymbol{x}}(\boldsymbol{x}) h^*_{\boldsymbol{x}}(\boldsymbol{w}).$$

In the next section we will display examples of $h^*_{\boldsymbol{w}}$ for various $\boldsymbol{w}$ in the one-dimension case.

Our criteria for choosing designs come from the upper bound for $R(\xi_n, h)$ given in (2.5). We define

$$\begin{aligned} J_{\mathrm{v}}(\xi_n) &= \operatorname{tr}\{\boldsymbol{M}^{-1}\boldsymbol{G}\}, \\ J_{\mathrm{b}}(\xi_n) &= \lambda_{\max}(\boldsymbol{Q}) = \lambda_{\max}\left(\boldsymbol{M}^{-1}\boldsymbol{G}\boldsymbol{M}^{-1}\boldsymbol{X}^{\top}\boldsymbol{K}\boldsymbol{X}\right), \end{aligned} \tag{2.12}$$

and

$$J(\xi_n; r) = rJ_{\mathrm{v}}(\xi_n) + (1-r)J_{\mathrm{b}}(\xi_n), \tag{2.13}$$

where

$$r = \frac{\sigma^2}{\|h\|^2 + \sigma^2}. \tag{2.14}$$

We call J_{v} and J_{b} a variance discrepancy and a bias discrepancy, respectively, while $J(\xi_n; r)$ is a weighted average of them. The upper bound for $R(\xi_n, h)$ can be written as $(\|h\|^2 + \sigma^2)J(\xi_n; r)$. It is clear that r in (2.14) is independent of the design ξ_n and reflects the relative proportion of the variance to bias. Values of r near 0 mean small variance error or serious bias, while values of r near 1 mean large variance error or small bias. Thus r can be understood as the prior belief of the experimenter as to the nature of the true response function. For a given value of r, $J(\xi_n; r)$ depends only on the design points, but not on the bias function h. The smaller the value of $J(\xi_n; r)$, the better the design ξ_n is. Therefore, for fitting the linear model with bias by using the least squares, we should choose a design such that $J(\xi_n; r)$ is as small as possible. The design that minimizes $J(\xi_n; r)$ for a given $r \in [0,1]$ is called compound optimal and denoted by ξ_n^r. In particular, ξ_n^r with $r = 0$ is called an all-bias design and ξ_n^r with $r = 1$ is called an all-variance design. It is known that the all-variance design is L-optimal if the assumed linear model is exactly correct (Atkinson and Donev (1992)). The meaning of the all-bias design is analogous to the uniform design introduced by Fang and Wang (1994).

Further, in order to compare the behaviour of different designs, such as the all-variance and all-bias designs described above, we define the efficiency of a design, $\tilde{\xi}_n$, as follows:

$$\mathrm{Eff}(\tilde{\xi}_n; r) = \frac{\min_{\xi_n} J(\xi_n; r)}{J(\tilde{\xi}_n; r)}. \tag{2.15}$$

Introducing the concept of a design measure ξ, a probability measure on $\mathcal{X}$ with mass n^{-1} on each points $\boldsymbol{x}_i$, we write

$$n^{-1}\boldsymbol{M} = \int_{\mathcal{X}} \boldsymbol{g}(\boldsymbol{x})\boldsymbol{g}^{\top}(\boldsymbol{x})d\xi(\boldsymbol{x}) \equiv \boldsymbol{M}_{\xi},$$

$$n^{-2}\boldsymbol{X}^{\top}\boldsymbol{K}\boldsymbol{X} = \int_{\mathcal{X}\times\mathcal{X}} \boldsymbol{g}(\boldsymbol{x})\boldsymbol{g}^{\top}(\boldsymbol{w})K(\boldsymbol{x},\boldsymbol{w})d\xi(\boldsymbol{x})d\xi(\boldsymbol{w}) \equiv \boldsymbol{P}_{\xi}.$$

Then the bias discrepancy J_{b} can be expressed by $J_{\mathrm{b}}(\xi) = \lambda_{\max}(\boldsymbol{M}_{\xi}^{-1}\boldsymbol{G}\boldsymbol{M}_{\xi}^{-1}\boldsymbol{P}_{\xi})$. Note that this expression is well defined for any probability measure ξ on $\mathcal{X}$ with nonsingular information matrix $\boldsymbol{M}_{\xi}$. It is clear that $J_{\mathrm{b}}(\xi) \geq 0$ for any ξ, equality holding when ξ is the uniform measure on $\mathcal{X}$. This implies that the uniform design on $\mathcal{X}$ is an all-bias design in the view of approximate designs. However, for finite n it has not yet proved possible to obtain any analytic minimization for $J_{\mathrm{b}}(\xi_n)$ in the general case.

Before ending this section, we give a construction of a reproducing kernel K for $\mathcal{H}$ which satisfies condition (1.4). We suppose that the underlying true response function $f(\cdot)$ comes from a reproducing kernel Hilbert space, $\mathcal{F}$, admitting a reproducing kernel K_0. Then for a set of linearly independent functions $g_1, \ldots, g_p$ from $\mathcal{F}$, we have the following theorem.

Theorem 2. *Suppose that the Hilbert space $\mathcal{F}$ has a reproducing kernel K_0 with $\int_{\mathcal{X}} K_0(\boldsymbol{x}, \boldsymbol{x}) d\boldsymbol{x} < \infty$, and that $\mathcal{F}$ is a direct sum of span$\{g_1, \ldots, g_p\}$ and $\mathcal{H}$ which satisfies condition* (1.4)*, i.e.,*

$$\mathcal{F} = span\{g_1, \ldots, g_p\} \oplus \mathcal{H}, \quad \int_{\mathcal{X}} \boldsymbol{g}(\boldsymbol{x}) h(\boldsymbol{x}) d\boldsymbol{x} = \mathbf{0}, \quad h \in \mathcal{H}. \tag{2.16}$$

Let $\boldsymbol{\eta}$ be a vector of p functions $\int_{\mathcal{X}} g_j(\boldsymbol{x}) K_0(\boldsymbol{x}, \cdot) d\boldsymbol{x}$, $j = 1, \ldots, p$, and let $\boldsymbol{\Psi}$ be a $p \times p$ matrix whose (i, j)th entry is $\int_{\mathcal{X} \times \mathcal{X}} g_i(\boldsymbol{x}) g_j(\boldsymbol{w}) K_0(\boldsymbol{x}, \boldsymbol{w}) d\boldsymbol{x} d\boldsymbol{w}$. Define

$$K(\boldsymbol{x}, \boldsymbol{w}) = K_0(\boldsymbol{x}, \boldsymbol{w}) - \boldsymbol{\eta}^{\top}(\boldsymbol{x}) \boldsymbol{\Psi}^{-1} \boldsymbol{\eta}(\boldsymbol{w}). \tag{2.17}$$

Then such K is the reproducing kernel for the subspace $\mathcal{H}$.

Proof. For the fixed $g_j \in \mathcal{F}$, consider the following linear functionals on $\mathcal{F}$:

$$T_j(f) = \int_{\mathcal{X}} g_j(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}, \quad f \in \mathcal{F},\ j = 1, \ldots, p. \tag{2.18}$$

Note that

$$\begin{aligned} |T_j(f)| &= \left| \int_{\mathcal{X}} \langle g_j K_0(\cdot, \boldsymbol{x}) \rangle \, \langle f K_0(\cdot, \boldsymbol{x}) \rangle d\boldsymbol{x} \right| \\ &\le \int_{\mathcal{X}} \|g_j\| \, \|K_0(\cdot, \boldsymbol{x})\|^2 \, \|f\| d\boldsymbol{x} = \int_{\mathcal{X}} K_0(\boldsymbol{x}, \boldsymbol{x}) d\boldsymbol{x} \, \|g_j\| \, \|f\|. \end{aligned}$$

It follows from the assumption on K_0 that the T_j are bounded. Let η_j be the representers for T_j:

$$T_j(f) = \langle \eta_j f \rangle, \quad f \in \mathcal{F},\ j = 1, \ldots, p. \tag{2.19}$$

We have

$$\eta_j(\boldsymbol{x}) = \langle \eta_j, K_0(\cdot, \boldsymbol{x}) \rangle = T_j(K_0(\cdot, \boldsymbol{x})) = \int_{\mathcal{X}} g_j(\boldsymbol{w}) K_0(\boldsymbol{w}, \boldsymbol{x}) d\boldsymbol{w}, \tag{2.20}$$

which means that the jth component of $\boldsymbol{\eta}$ in (2.17) is the representer η_j. Further, from (2.18) and (2.19), we have

$$\langle \eta_i, \eta_j \rangle = T_i(\eta_j) = \int_{\mathcal{X}} g_i(\boldsymbol{x}) \eta_j(\boldsymbol{x}) d\boldsymbol{x} = \int_{\mathcal{X} \times \mathcal{X}} g_i(\boldsymbol{x}) g_j(\boldsymbol{w}) K_0(\boldsymbol{x}, \boldsymbol{w}) d\boldsymbol{x} d\boldsymbol{w},$$

which means that the (i,j)th entry of $\boldsymbol{\Psi}$ in (2.17) is $\langle \eta_i, \eta_j\rangle$. We show that $\eta_1,\ldots,\eta_p$ are linearly independent. Suppose the equality $\boldsymbol{a}^\top\boldsymbol{\eta} \equiv a_1\eta_1(\boldsymbol{x}) + \cdots + a_p\eta_p(\boldsymbol{x}) = 0$ holds for a set of constants a_j. Then $\boldsymbol{0} = \langle \boldsymbol{a}^\top\boldsymbol{\eta}, \boldsymbol{g}^\top\rangle = \boldsymbol{a}^\top\langle \boldsymbol{\eta}, \boldsymbol{g}^\top\rangle = \boldsymbol{a}^\top\boldsymbol{G}$, the last equality holding due to (2.18), (2.19) and the definition of $\boldsymbol{G}$ in (2.2). But $\boldsymbol{G}$ is nonsingular, so $\boldsymbol{a} = \boldsymbol{0}$ and the $\eta_1,\ldots,\eta_p$ are linearly independent. Hence $\boldsymbol{\psi}$ is nonsigular.

It remains to be shown is that the function K given in (2.17) satisfies (1.3). For any fixed $\boldsymbol{w} \in \mathcal{X}$, from (2.18), (2.19) and (2.17) we have for each j

$$\begin{aligned}\int_{\mathcal{X}} g_j(\boldsymbol{x})K(\boldsymbol{x},\boldsymbol{w})d\boldsymbol{x} &= \langle \eta_j, K(\cdot,\boldsymbol{w})\rangle = \langle \eta_j, K_0(\cdot,\boldsymbol{w})\rangle - \langle \eta_j, \boldsymbol{\eta}^\top\rangle\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}(\boldsymbol{w})\\ &= \eta_j(\boldsymbol{w}) - \boldsymbol{e}_j^\top\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}(\boldsymbol{w}) = \boldsymbol{0},\end{aligned}$$

where $\boldsymbol{e}_j$ is as in (2.9). It follows that $K(\cdot,\boldsymbol{w}) \in \mathcal{H}$ for any $\boldsymbol{w} \in \mathcal{X}$. Furthermore, for any $h \in \mathcal{H} \subset \mathcal{F}$,

$$\langle h, K(\cdot,\boldsymbol{w})\rangle = \langle h, K_0(\cdot,\boldsymbol{w})\rangle - \langle h, \boldsymbol{\eta}^\top\rangle\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}(\boldsymbol{w}) = h(\boldsymbol{w}),$$

since $\langle h, \eta_j\rangle = \int_{\mathcal{X}} g_j(\boldsymbol{x})h(\boldsymbol{x})d\boldsymbol{x} = 0$ for $j = 1,\ldots,p$. Therefore, K is the reproducing kernel for $\mathcal{H}$.

Remark 3. The result in Theorem 2 is different from that of Wahba (1978, Section 3). As the condition (1.4) is not assumed there.

3. Illustrative Examples

In this section we present some numerical results on the all-bias, all-variance and compound optimal designs for several models. Throughout, the design region is the unit cube in $\mathcal{R}^s$, i.e., $\mathcal{X} = [0,1]^s$. A generic point in $[0,1]^s$ is denoted by $\boldsymbol{x} = (x^1,\ldots,x^s)^\top$, and the point $\boldsymbol{x}_i$ is written as $\boldsymbol{x}_i = (x_i^1,\ldots,x_i^s)^\top$. Suppose that the underlying true response functions, represented by

$$f(\boldsymbol{x}) = \boldsymbol{\theta}^\top\boldsymbol{g}(\boldsymbol{x}) + h(\boldsymbol{x}) \quad \text{with} \quad \int_{[0,1]^s} \boldsymbol{g}(\boldsymbol{x})h(\boldsymbol{x})d\boldsymbol{x} = \boldsymbol{0},$$

come from the space $\mathcal{F} = \otimes^s\mathcal{W}_m$, the tensor product of $\mathcal{W}_m$ with itself s times. Here, $\mathcal{W}_m$ is the so-called Sobolev-Hilbert space defined by

$$\mathcal{W}_m:\ \mathcal{W}_m[0,1] = \left\{f:\ f, f', \ldots, f^{(m-1)} \text{ absolutely continuous, } f^{(m)} \in \mathcal{L}_2[0,1]\right\}.$$

The positive integer m indicates the degree of smoothness of functions in this space. It is known that $\mathcal{W}_m$ is a Hilbert space with reproducing kernel (Wahba (1990), Section 10.2)

$$K_0^*(x,w) = \sum_{l=0}^m \frac{1}{(l!)^2}B_l(x)B_l(w) + \frac{(-1)^{m+1}}{(2m)!}B_{2m}(\{x-w\}), \quad (x,w) \in [0,1]^2$$

under the square norm

$$\|f\|^2 = \sum_{l=0}^{m-1} \left[\int_0^1 f^{(l)}(x)dx\right]^2 + \int_0^1 \left[f^{(m)}(x)\right]^2 dx, \quad f \in \mathcal{W}_m.$$

Here B_l is the lth Bernoulli polynomial, and $\{\,\cdot\,\}$ is the fractional part of a real number, i.e., $\{t\} = t \pmod 1$. It follows that the tensor product space $\mathcal{F}$ is a Hilbert space with reproducing kernel $K_0(\boldsymbol{x}, \boldsymbol{w}) = \prod_{k=1}^{s} K_0^*(x^k, w^k)$. Then for a given regressor vector $\boldsymbol{g}(\boldsymbol{x})$ the reproducing kernel for the subspace $\mathcal{H}$ is well defined by (2.17). For simplicity we only consider the case $m = 1$.

Example 1. *Location model with bias.* In the true model (3.1), $\boldsymbol{g}(\boldsymbol{x}) \equiv 1$. In this case, the reproducing kernel for $\mathcal{H}$ is

$$K(\boldsymbol{x}, \boldsymbol{w}) = \prod_{k=1}^{s} \left[1 + B_1(x^k)B_1(w^k) + \frac{1}{2}B_2(\{x^k - w^k\})\right] - 1.$$

We first give the norm of a function defined in this subspace. In the one-dimensional case, $s = 1$, the square norm of $h \in \mathcal{H}$ is $\|h\|^2 = \int_0^1 [h'(x)]^2 dx$, since here $\int_0^1 h(x)dx = 0$. The function h_w^* defined in (2.11) is displayed in Figure 1 for each given value of $w \in [0, 1]$.

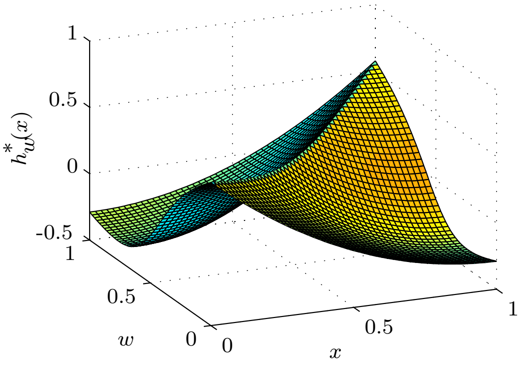


Figure 1. The plots of $h_w^*(x)$ given in (2.11) for the one-dimensional location model with bias given in Example 1.

For the multi-dimensional case, the norm of a function in $\mathcal{H}$ is more complicated and is best defined using the ANOVA decomposition of a function. For $u \subseteq S \equiv \{1, \ldots, s\}$, let $|u|$ denote the cardinality of u, and $\bar{u}$ denote the complement $S - u$. By $[0, 1)^u$ we denote the $|u|$-dimensional unit cube involving the coordinates in u, by $\boldsymbol{x}^u$ we denote the coordinate projection of $\boldsymbol{x}$ onto $[0, 1)^u$, and $d\boldsymbol{x}^u = \prod_{k \in u} dx^k$. The ANOVA decomposition of $h \in \mathcal{H}$ takes the form

$h(\boldsymbol{w}) = \sum_{u\subseteq S} h_u(\boldsymbol{x})$ (see Owen (1992)), where the terms h_u are defined recursively starting with $h_\emptyset = \int_{[0,1)^s} h(\boldsymbol{x})d\boldsymbol{x} = 0$, and using the rule

$$h_u(\boldsymbol{x}) = \int_{[0,1)^{\bar{u}}} \Big[h(\boldsymbol{x}) - \sum_{v\subset u} h_v(\boldsymbol{x})\Big] d\boldsymbol{x}^{\bar{u}}, \quad \emptyset \neq u \subseteq S.$$

Then the square norm of h is given by (Hickernell (1996))

$$\|h\|^2 = \sum_{u\subseteq S} \int_{[0,1)^s} \left[\frac{\partial^{|u|} h_u}{\partial \boldsymbol{x}^u}\right]^2 d\boldsymbol{x} = \sum_{u\subseteq S} \int_{[0,1)^u} \left[\int_{[0,1)^{\bar{u}}} \frac{\partial^{|u|} h}{\partial \boldsymbol{x}^u} d\boldsymbol{x}^{\bar{u}}\right]^2 d\boldsymbol{x}^u.$$

Consider the designs for this model. Clearly, the criterion $J(\xi_n; r)$ in (2.13) depends on the design points only through $J_{\rm b}$. For the one-dimensional case, $s = 1$, straightforward calculations give

$$\begin{aligned} J_{\rm b}(\xi_n) &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \Big[B_1(x_i)B_1(x_j) + \frac{1}{2}B_2(\{x_i - x_j\})\Big] \\ &= \frac{1}{12n^2} + \frac{1}{n}\sum_{i=1}^n \left[x_{(i)} - \frac{2i-1}{2n}\right]^2, \end{aligned}$$

where $x_{(1)} \leq \cdots \leq x_{(n)}$ are the order statistics of $x_1, \ldots, x_n$. It is clear that the minimizer of $J_{\rm b}(\xi_n)$ in (3.4) is $\left\{\frac{2i-1}{2n},\ i = 1, \ldots, n\right\}$.

For multi-dimensional case, $s > 1$,

$$J_{\rm b}(\xi_n) = -1 + \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\prod_{k=1}^s \Big[1 + B_1(x_i^k)B_1(x_j^k) + \frac{1}{2}B_2(\{x_i^k - x_j^k\})\Big].$$

For finite n, there is no analytic solution for this minimization problem, and some numerical optimization methods must be used. For small n and small s, we can find the minimizer, $\xi_n^{\rm b}$, of $J_{\rm b}(\xi_n)$ over $[0,1]^s$ by a constrained optimization routine in MATLAB. Some of these results with $s = 2$ are shown in Figure 2 by circles ($\circ$). On the other hand, we also consider the following family of n-point designs in $[0,1]^s$:

$$\mathcal{P}_{\rm equ}(s,n) \equiv \left\{\{\boldsymbol{x}_i\}_{i=1}^n : \quad x_1^k, \ldots, x_n^k \text{ is a permutation of } \frac{1}{2n}, \frac{3}{2n}, \ldots, \frac{2n-1}{2n}\right\}.$$

A design in $\mathcal{P}_{\rm equ}(s,n)$ implies that the possible levels for each factor are to be equi-distributed in [0,1]. The designs, $\xi_{n,\rm equ}^{\rm b}$, that minimize $J_{\rm b}(\xi_n)$ over $\mathcal{P}_{\rm equ}(2,n)$ are shown in Figure 2 by dots ($\cdot$). It is observed that the designs, $\xi_n^{\rm b}$ and $\xi_{n,\rm equ}^{\rm b}$, are quite similar and uniformly scattered over the region $[0,1]^2$.

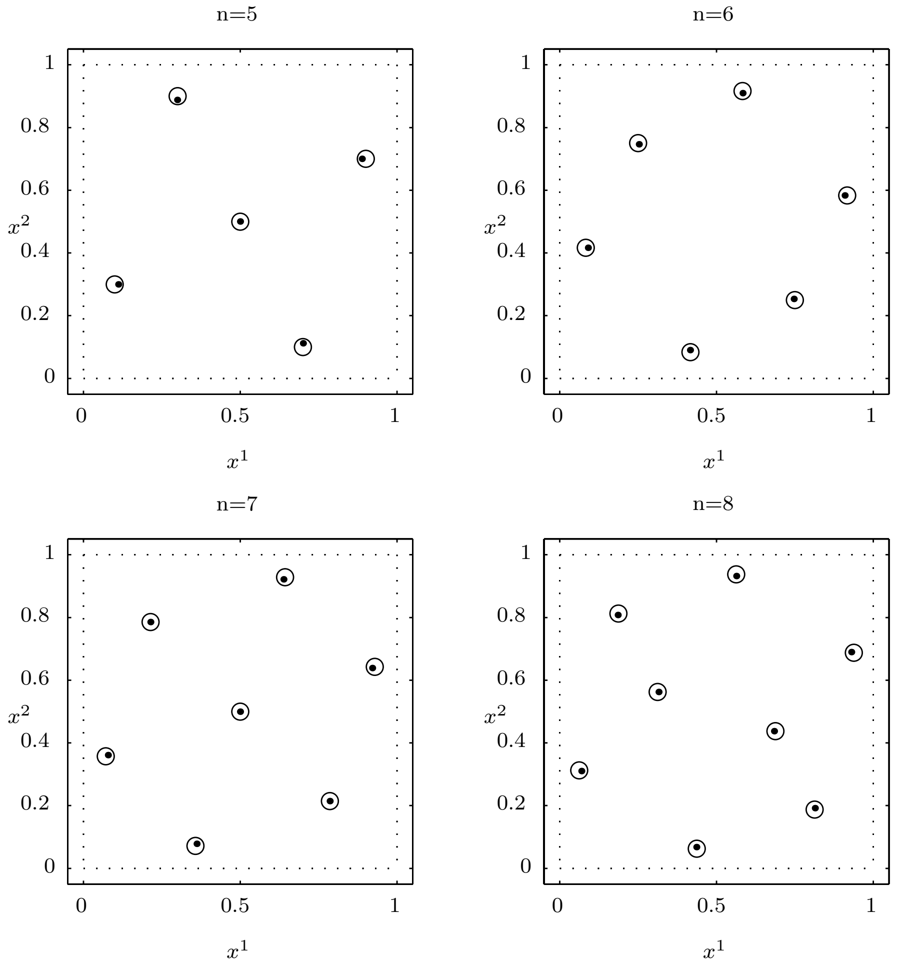


Figure 2. The designs ξ_n^{b} ($\circ$) and $\xi_{n,\mathrm{equ}}^{\mathrm{b}}$ ($\cdot$) for the two-dimensional location model with bias given in Example 1.

Example 2. First-degree model with bias. For $\boldsymbol{x} = (x^1, \ldots, x^s)^\top$ in $[0,1]^s$ we set

$$\Phi_k(\boldsymbol{x}) = 2\sqrt{3}B_1(x^k), \quad k = 1, \ldots, s.$$

The regressor vector here is $\boldsymbol{g}(\boldsymbol{x}) = (\,1, \Phi_1(\boldsymbol{x}), \ldots, \Phi_s(\boldsymbol{x})\,)^\top$. We then have $\boldsymbol{G} = \boldsymbol{I}_{s+1}$. From (2.17) we find the reproducing kernel for $\mathcal{H}$:

$$K(\boldsymbol{x}, \boldsymbol{w}) = \prod_{k=1}^{s} \left[1 + B_1(x^k)B_1(w^k) + \frac{1}{2}B_2(\{x^k - w^k\})\right] - 1 - \frac{5}{6}\sum_{k=1}^{s} B_{1,3}(x^k)B_{1,3}(w^k),$$

where $B_{1,3}(\cdot)$ is defined as

$$B_{1,3}(\cdot) = B_1(\cdot) - 2B_3(\cdot). \tag{3.1}$$

Under this subspace with $s = 1$, the functions $h^*_w(x)$ for various w are shown in Figure 3.

The compound optimal design ξ^r_n that minimizes $J(\xi_n; r)$ over $[0,1]^s$ for a given $r \in [0,1]$ can be found numerically. Figure 4 shows the 8-run designs with $s = 1$ and $s = 2$ corresponding to $r = 0$ ($\circ$), 0.1 ($\cdot$), $\ldots$, 0.9 ($\cdot$) and 1($\times$), respectively. The efficiencies of all-variance and all-bias designs calculated from (2.15) are also shown in this figure. It is observed that the points of all-bias designs ξ^{b}_n ($\circ$) are uniformly scattered over the regions [0,1] and $[0,1]^2$, respectively. The points of all-variance designs ξ^{v}_n ($\times$) are located at the ends of [0,1] for $s = 1$, and at the vertices of the cube for $s = 2$. As r increases from 0 to 1, one proceeds from all-bias designs to all-variance designs. For higher dimensional cases, the efficiencies of all-bias designs are compared numerically with those of all-variance designs. The results with $n = 12$ and $s = 3, 4$ are also shown in Figure 4.

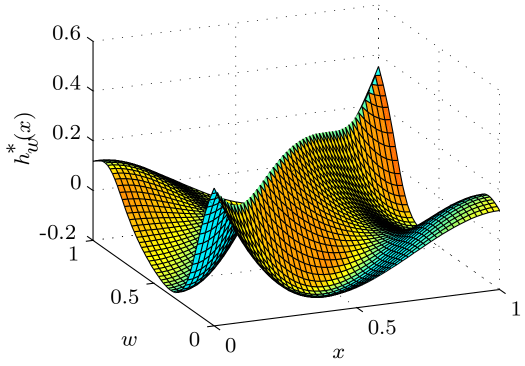
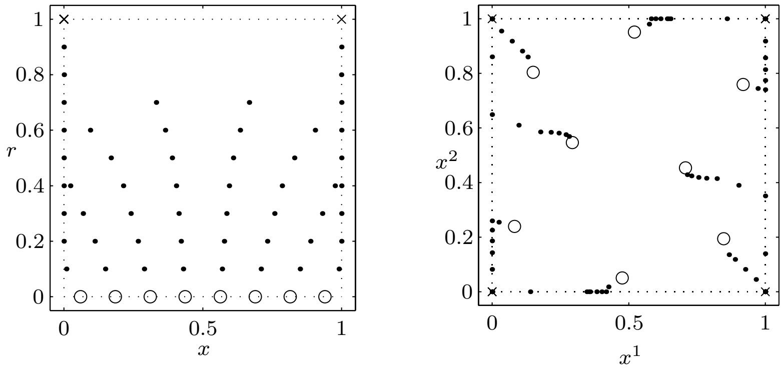


Figure 3. The plots of $h^*_w(x)$ given in (2.11) for the one-dimensional first-degree model with bias given in Example 2.

(i)



(ii)

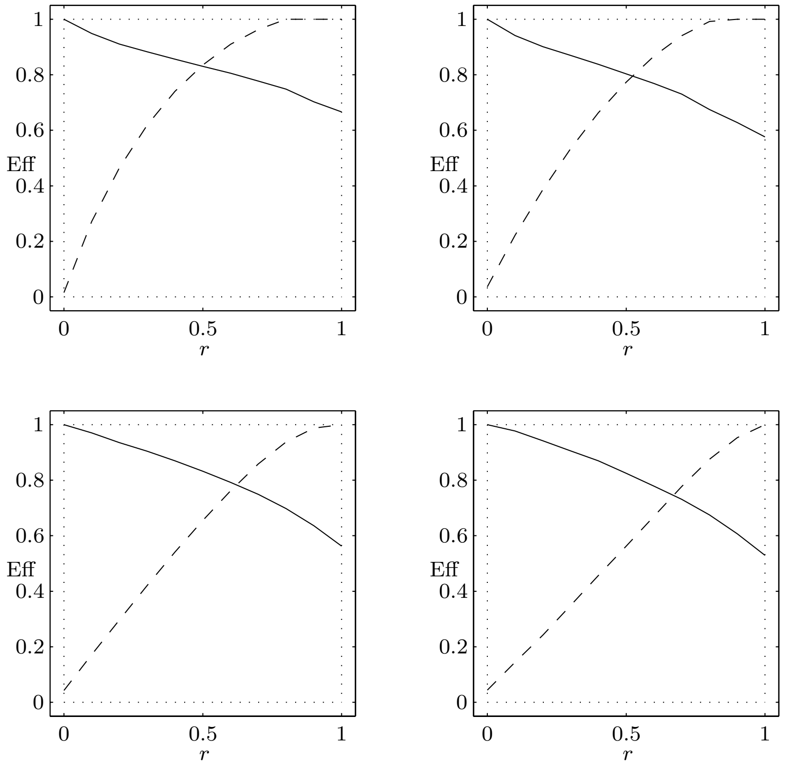


Figure 4. (i) The designs for the first-degree model with bias given in Example 2 with $s = 1$ and $s = 2$: ξ_n^r $(\cdot)$, ξ_n^{b} $(\circ)$ and ξ_n^{v} $(\times)$. (ii) The efficiencies of the all-bias designs and all-variance designs: $\mathrm{Eff}(\xi_n^{\mathrm{b}}; r)$ (*solid*), $\mathrm{Eff}(\xi_n^{\mathrm{v}}; r)$ (*dashed*).

The efficiency of the designs ξ_n^{b} is not so bad even without misspecification in the model. We also observe that the efficiency of the all-variance design decreases somewhat as the dimension increases, if misspecification is present. The efficiency of the all-bias design also changes a bit as the dimension increases.

Example 3. Second-degree model with bias. For $\boldsymbol{x} = (x^1, \ldots, x^s)^\top$ in $[0,1]^s$ we let Φ_k be the same as in Example 2, and let

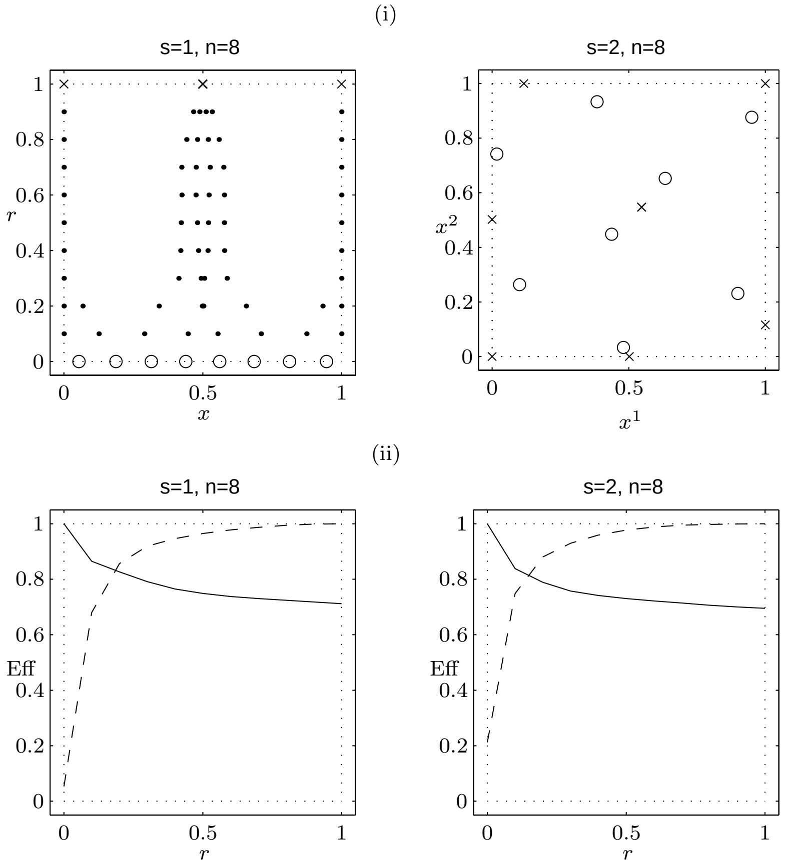
$$\Phi_{kl}(\boldsymbol{x}) = \begin{cases} \Phi_k(\boldsymbol{x})\Phi_l(\boldsymbol{x}),\ k < l, \\ 6\sqrt{5}B_2(x^k),\ k = l. \end{cases}$$

The regressor vector here is $\boldsymbol{g} = (1, \Phi_1, \ldots, \Phi_s, \Phi_{12}, \ldots, \Phi_{s-1,s}, \Phi_{11}, \ldots, \Phi_{ss})^\top$.

Then we have $\boldsymbol{G}=\boldsymbol{I}_p$, $p=1+2s+s(s-1)/2$. The reproducing kernel for $\mathcal{H}$ is

$$\begin{aligned}
&K(\boldsymbol{x},\boldsymbol{w})\\
&=\prod_{k=1}^{s}\left[1+B_1(x^k)B_1(w^k)+\frac{1}{2}B_2(\{x^k-w^k\})\right]-1-\frac{5}{6}\sum_{k=1}^{s}B_{1,3}(x^k)B_{1,3}(w^k)\\
&\quad-\frac{25}{36}\sum_{1\le j<k\le s}B_{1,3}(x^j)B_{1,3}(x^k)B_{1,3}(w^j)B_{1,3}(w^k)-\frac{105}{2}\sum_{k=1}^{s}B_4(x^k)B_4(w^k),
\end{aligned}$$

where $B_{1,3}(\cdot)$ was defined in (3.1). The compound optimal designs can be found numerically. For each $r=0$, 0.1, $\ldots$, 0.9 and 1, we find the 8-run compound design ξ_n^r that minimizes $J(\xi_n;r)$ over $[0,1]^s$ for $s=1$ and $s=2$, respectively. Some of these results are shown in Figure 5. The behaviours of ξ_n^{b} and ξ_n^{v} are similar to those in Example 2. Figure 5 also shows the efficiencies of the designs for this model when $(s,n)=(3,12)$ and $(s,n)=(4,18)$, respectively. It seems that the efficiency of the all-bias design decreases as the dimension increases if model misspecification is small, and the efficiency of the all-variance design decreases with dimension if misspecification is large.


(i)

(ii)

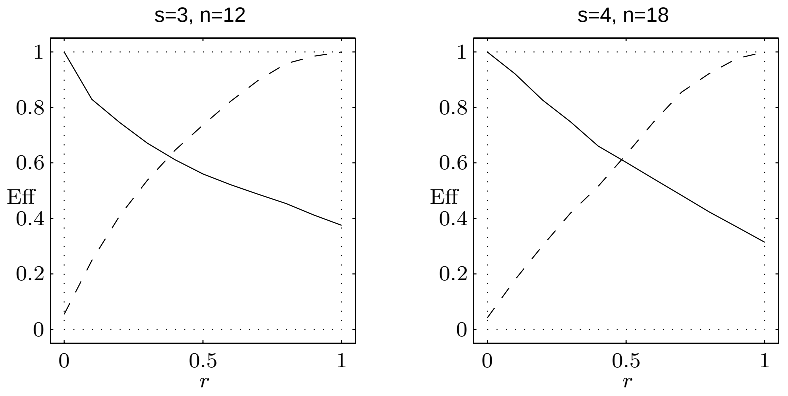


Figure 5. (i) The designs for the second-degree model with bias given in Example 3 with $s = 1$ and $s = 2$: ξ_n^r $(\cdot)$, ξ_n^{b} $(\circ)$ and ξ_n^{v} $(\times)$. (ii) The efficiencies of the all-bias designs and all-variance designs: $\mathrm{Eff}(\xi_n^{\mathrm{b}}; r)$ (*solid*), $\mathrm{Eff}(\xi_n^{\mathrm{v}}; r)$ (*dashed*).

4. Summary

We have considered the design problem for fitting linear models with misspecification. It is assumed that the bias is a deterministic but unknown function in a reproducing kernel Hilbert space. The model parameters are estimated by standard least squares and the criterion for choosing designs is developed in terms of a sharp upper bound for the integrated mean squared error of the estimates. This criterion is a weighted average of the so-called bias discrepancy and variance discrepancy. The bias discrepancy is derived by using a reproducing kernel Hilbert space approach.

The dependence of the efficiency of the all-variance and all-bias designs on the number of experiments, the order of the model and the dimension remains to be fully explored. However, numerical results for some specific cases suggest certain trends. As expected, the all-variance design has a high efficiency when the misspecification is small and the all-bias design has a high efficiency when the misspecification is large. The efficiency of the all-variance becomes small when the misspecification is large, and this trend is worse as the dimension increases. The efficiency of the all-bias design is moderate, even when no misspecification is present, however it too decays with increasing dimension.

Whereas the all-variance designs depend strongly on the form of the model, the all-bias designs are relatively independent of the model, spreading the design points evenly over the experimental domain. If one does not know the model a priori, one might use any all-bias design to perform the experiment, and then use standard variable selection techniques to identify the appropriate model.

Acknowledgement

This research is partially supported by Hong Kong RGC Grant 94-95/38.

References

Aronszajn, M. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68**, 337-404.

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

Box, G. E. P. and Draper, N. R. (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.* **54**, 622-654.

Chang, Y. J. and Notz, W. I. (1996). Model Robust designs. *Handbook of Statistics* (Edited by S. Ghosh and C. R. Rao) **13**, 1055-1098. Elsevier.

Fang, K. T. and Wang, Y. (1994). *Number-Theoretic Methods in Statistics.* Chapman and Hall, London.

Hickernell, F. J. (1996). Quadrature error bounds with applications to lattice rules. *SIAM, J. Num. Anal.* **33**, 1995-2016.

Huber, P. (1975). Robustness and designs. In *A Survey of Statistical Design and Linear Models.* (Edited by Srivastava), 287-303. North Holland, Amsterdam.

Kiefer, J. (1973). Optimal designs for fitting multiresponse surfaces. *Symposium on Multivariate Analysis III*, 287-297. Academic Press, New York.

Li, K. C. and Notz, W. (1982). Robust designs for nearly linear regression. *J. Statist. Plann. Inference* **6**, 135-151.

Li, K. C. (1984). Robust regression designs when the design space consists of finitely many points. *Ann. Statist.* **12**, 269-282.

Marcus, M. B. and Sacks, J. (1978). Robust designs for regression problems. *Statistical Decision Theory and Related Topics, II*, 245-268. Academic Press, New York.

Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods.* SIAM, Philadelphia.

Owen, A. B. (1992). Randomized orthogonal arrays for computer experiments, integration and visualization. *Statist. Sinica* **2**, 439-452.

Pesotchinsky, L. (1982). Optimal robust designs: linear regression in R^k. *Ann. Statist.* **10**, 511-525.

Sacks, J. and Ylvisaker, D. (1978). Linear estimation for approximately linear models. *Ann. Statist.* **6**, 1122-1137.

Saitoh, S. (1988). *Theory of Reproducing Kernels and Its Applications.* Longman Scientific and Technical, Essex, England.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. R. Statist. Soc. Ser. B* **40**, 364-372.

Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia.

Wiens, D. P. (1990). Robust minimax designs for multiple linear regression. *Linear Algebra Appl.* **127**, 327-340.

Wiens, D. P. (1992). Minimax designs for approximately linear regression. *J. Statist. Plann. Inference* **31**, 353-371.

College of Mathematical Science, Shanghai Normal University, 100 Guilin Road, Shanghai 200234, China.

E-mail: rxyue@online.sh.cn

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong.
E-mail: fred@hkbu.edu.hk

(Received January 1998; accepted December 1998)