Robust designs for approximately polynomial regression

Shawn X. Liu\textsuperscript{a}, Douglas P. Wiens\textsuperscript{b,*}

\textsuperscript{a} Department of Mathematics, Physics and Engineering, Mount Royal College, Calgary, Alberta, Canada T3E 6K6
\textsuperscript{b} Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

Received 17 April 1995; received in revised form 11 December 1996

Abstract

We study designs for the regression model $E[Y|x] = \sum_{j=0}^{p-1} \theta_j x^j + x^p \psi(x)$, where $\psi(x)$ is unknown but bounded in absolute value by a given function $\phi(x)$. This class of response functions models departures from an exact polynomial response. We consider the construction of designs which are robust, with respect to various criteria, as the true response varies over this class. The resulting designs are shown to compare favourably with others in the literature. © 1997 Elsevier Science B.V.

\textit{AMS classifications:} primary 62K05; 62F35; secondary 62J05

Keywords: Bounded bias; Bounded variance; Approximate polynomial regression; Minimax designs

1. Introduction

We consider the construction of designs for polynomial regression which are robust to departures from the assumed response. Specifically, suppose that a researcher is to design an experiment, the results of which will require a regression analysis with fitted model

$$\hat{y}(x) = \sum_{j=0}^{p-1} \hat{\theta}_j x^j; \quad p \geq 2, \quad x \in [a - b, a + b] =: \mathcal{F}. \quad (1.1)$$

If the fitted response is exactly correct, and if, as is assumed in this paper, the errors are additive and uncorrelated, with common variance $\sigma^2$, then the least-squares estimates $\hat{\theta}_j$ are unbiased and possess well-known optimality properties with respect to their variances, especially when used in conjunction with a variance-minimizing design.

* Corresponding author.
Suppose however that the ‘true’ response is in fact only approximately polynomial in that for $x \in \mathcal{S}$ and some unknown function $\psi$ we have

$$E[Y|x] = \theta^T f(x) + \left(\frac{x - a}{b}\right)^p \psi \left(\frac{x - a}{b}\right),$$

(1.2)

where $\theta := (\theta_0, \ldots, \theta_p)^T$ and $f(x) := (1, x, x^2, \ldots, x^{p-1})^T$. In fitting (1.1), the experimenter is assuming, perhaps erroneously, that the second term in (1.2) may be ignored. This introduces possible bias into the least-squares estimates, hence into the fitted values (1.1). One might hope to reduce this bias, and simultaneously achieve small variances through a judicious choice of design points. It is the purpose of this paper to indicate ways of accomplishing these aims.

We shall assume

(A1) The function $\psi(x)$ is continuous, and $|\psi(x)| \leq \phi(x)$ on $[-1, 1]$, where $\phi(x)$ is a continuous, even function which is positive when $x \neq 0$.

(A2) The function $l(z) := z\phi(\sqrt{z})$ is convex for $z \in [0, 1]$, with $l(0) = 0$.

The continuity of $\psi$ at 0 ensures that the parameter $\theta$ in (1.2) is well defined. Note that (A1) and (A2) together require $l(z)$ to be non-decreasing on $[0, 1]$, so that then $l(z)$ and $\phi(z)$ are bounded above by $\phi(1)$.

Box and Draper (1959) made apparent the dangers of using regression designs which are optimal only when the assumed model is correct. By analyzing the relative importance of errors due to bias, and to variance, they found that very small departures from the model can eliminate any supposed gains resulting from the use of a design which minimizes variance alone. See Huber (1975, 1981) and Wiens (1990, 1992) for further work in this direction.

It is natural to quantify the loss through the mean-squared error of $\hat{\theta}$. The bias and variance components of the mean-squared error are most conveniently described in terms of the design measure – the empirical distribution function of the design points $x_1, \ldots, x_n$. Denote this measure by $\xi$ and define

$$B(\xi) = n^{-1} \sum_{i=1}^{n} f(x_i)f^T(x_i) = \int_{\mathcal{S}} f(x)f^T(x) d\xi(x),$$

$$b(\xi, \psi) = n^{-1} \sum_{i=1}^{n} f(x_i) \left(\frac{x_i - a}{b}\right)^p \psi \left(\frac{x_i - a}{b}\right)$$

$$= \int_{\mathcal{S}} f(x) \left(\frac{x - a}{b}\right)^p \psi \left(\frac{x - a}{b}\right) d\xi(x).$$

Then the bias vector and covariance matrix are

$$E[\hat{\theta} - \theta] = B^{-1}(\xi)b(\xi, \psi),$$

$$\text{COV}[\hat{\theta}] = \frac{\sigma^2}{n} B^{-1}(\xi),$$
so that the mean-squared error matrix $E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T]$ has determinant

$$|\text{MSE}(\xi, \psi)| = \left(\frac{\sigma^2}{n}\right)^p \left(1 + \frac{n}{\sigma^2} b^T(\xi, \psi)B^{-1}(\xi)b(\xi, \psi)\right)/|B(\xi)|. \quad (1.3)$$

We will consider three separate criteria, each leading to the construction of optimally robust designs:

(C1) **Bounded Bias:** We seek to minimize the size – as measured by the determinant – of the covariance matrix, subject to bounding the normalized bias. Equivalently,

Choose $\xi$ to maximize $|B(\xi)|$, subject to $\sup_{\psi} b^T(\xi, \psi)B^{-1}(\xi)b(\xi, \psi) \leq k^2$

for a given constant $k^2$.

(C2) **Bounded Variance:** We seek to minimize the maximum bias, subject to bounding the variance. Equivalently,

Choose $\xi$ to minimize $\sup_{\psi} b^T(\xi, \psi)B^{-1}(\xi)b(\xi, \psi)$, subject to $|B(\xi)| \geq c^2 b^{p(p-1)}$

for a given constant $c^2$.

(C3) **Minimax:** We seek to minimize the maximum size of the MSE matrix, viz.,

Choose $\xi$ to minimize $\sup_{\psi} |\text{MSE}(\xi, \psi)|$.

Note that if $\xi_{ab}$ is a design measure on $S$, $\xi_{01}(x) = \xi_{ab}(a + bx)$ is the induced measure on $[-1, 1]$, and $C$ is the triangular matrix of coefficients defined by $f(a + bx) = Cf(x)$, then

$$b(\xi_{ab}, \psi) = Cb(\xi_{01}, \psi), \quad B(\xi_{ab}) = CB(\xi_{01})C^T,$$

$$|B(\xi_{ab})| = b^{p(p-1)}|B(\xi_{01})|.$$ It follows that if $\xi_{01}$ is an optimal design on $[-1, 1]$ with respect to any of the above criteria, then $\xi_{ab}$ is an optimal design on $[a - b, a + b]$. Without loss of generality we shall henceforth take $a = 0$, $b = 1$, $\phi(1) = 1$.

We adopt here the methods of approximate design theory. That is, we search for optimal measures $\xi$ in the class $\Xi$ of all probability measures on $S$. These turn out to be discrete, and supported on $p$ points, but with atoms which are typically not integer multiples of $n^{-1}$. Optimal rounding methods for implementing such designs are discussed in Pukelsheim (1993, Ch. 12).

Because the optimal designs are supported on only $p$ points, they afford no opportunity to assess the fit of the model. As a practical guide, we suggest designing for a slightly higher degree than is in fact anticipated, and then carrying out appropriate hypothesis tests.

Pesotchinsky (1982) obtained minimax designs under departures similar to those in (1.2), but for the case $f(x) = (1, x_1, \ldots, x_p)^T$, so that the fitted response surface is a hyperplane. See Marcus and Sacks (1976), Sacks and Ylvisaker, (1978), Li and Notz,
(1982), Li (1984) and Liu (1994) for other instances of designs for departures as in (1.2). Comparisons with some specific designs of Stigler (1971) and Pukelsheim and Rosenberger (1993) are presented in Section 4.

2. General theory

We first show that \( \sup_\psi |\text{MSE}(\xi, \psi)| \) is attained at one of at most two possible functions \( \psi \). Define

\[
 b_i(\xi) = \int_S f(x) x^p (\text{sign} x)^i |\psi(x)| d\xi(x), \quad i = 1, 2,
\]

\[
 T_i(\xi) = \begin{cases} 
 b_i^T(\xi) B^{-1}(\xi) b_i(\xi) & \text{if } |B(\xi)| > 0, \\
 \infty & \text{if } |B(\xi)| = 0.
\end{cases}
\]

\[
 T(\xi) = \begin{cases} 
 \max_{i=1,2} T_i(\xi) & \text{if } \psi(0) = 0, \\
 T_i(\xi) & \text{if } \psi(0) > 0.
\end{cases}
\]

It follows from Lemma 2 of Wiens (1993) that \( T(\xi) \) is a convex functional of \( \xi \in \Xi \).

**Theorem 2.1.** For any \( \xi \in \Xi \) with \( |B(\xi)| > 0 \),

\[
 \max_\psi b^T(\xi, \psi) B^{-1}(\xi) b(\xi, \psi) = T(\xi).
\]

**Proof.** Fix a \( \xi \) with \( |B(\xi)| > 0 \). Denote by \( \Psi \) the set of functions \( \psi \) satisfying assumption (A1). Note that the functional \( A(\psi) := b^T(\xi, \psi) B^{-1}(\xi) b(\xi, \psi) \) is quadratic, hence convex, on the convex set \( \Psi \). Thus, if \( \psi \in \Psi \) is not an extremal, say \( \psi = (1 - \alpha) \psi_1 + x \psi_2 \) for \( \alpha \in (0, 1) \) and \( \psi_1, \psi_2 \in \Psi \), then \( A(\psi) \leq \max(A(\psi_1), A(\psi_2)) \). It follows that to determine \( \sup \{ A(\psi) : \psi \in \Psi \} \), we need only consider extremals of \( \Psi \). But these extremals have \( |\psi(x)| = \psi(x) \). In order that such an extremal be continuous, we must have \( \psi(x) = \psi(x) \psi(x) \) or \( \psi(x) = -\psi(x) \) if \( \psi(0) > 0 \), while if \( \psi(0) = 0 \) we could also have \( \psi(x) = -\text{sign}(x) \psi(x) \) or \( \psi(x) = -\text{sign}(x) \psi(x) \). The result follows upon noting that \( A(\psi) = A(-\psi) \).

The Bounded Bias problem may now be phrased as

**BB:** Find \( \xi_k := \arg \max_{\xi} \{ |B(\xi)| : T(\xi) \leq k^2 \} \)

and the Bounded Variance problem as

**BV:** Find \( \xi_c := \arg \min_{\xi} \{ T(\xi) : |B(\xi)| \geq c^2 \} \).

From (C3) and (1.3) the Minimax problem is

**M:** Find \( \xi_k \), where

\[
 k^* := \arg \min_{k} \left( \frac{\sigma^2}{n} \right)^p \left( 1 + \frac{n}{\sigma^2} T(\xi_k) \right)^{\frac{1}{p}} |B(\xi_k)|. \tag{2.1}
\]

The minimax design may also be obtained from the optimal bounded variance design, in a similar fashion. Throughout the remainder of this paper we shall largely concentrate
on (BB). The analogous results for (BV) and (M) are given in Liu and Wiens (1994); some special cases are detailed in Sections 3.1 and 3.2 below.

By Theorem 2.3 below, we have that $T(\xi) \leq E[|X|^2+]$ if $|B(\xi)| > 0$. Thus, one can make $T(\xi)$ arbitrarily small while retaining a non-singular covariance matrix $|B(\xi)|$, for instance, by employing a $p$-point design with a sufficient amount of mass at 0. It follows that $\{ |B(\xi)|: T(\xi) \leq k^2 \} \neq \emptyset$ for every $k > 0$. Standard arguments now show that there exists a solution $\xi_k$ to (BB) for every $k > 0$, and then the convexity of $-\log |B(\xi)|$ and of $T(\xi)$ may be used to show that $\xi_k$ lies in the subclass $\Xi$ of symmetric members of $\Xi$, i.e. those $\xi$ satisfying $\xi(x) = 1 - \xi((-x)^-) / \xi$ for all $x \in \mathcal{S}$.

The next result shows that the upper bound $k^2$ is attained by $T(\xi_k)$, as long as $k$ is small enough that $\xi_k$ does not solve the unconstrained problem of maximizing $|B(\xi)|$ unconditionally. Let $\xi_D$ be the solution to this unconstrained problem -- i.e. the $D$-optimal design -- and define $k_D^2 = T(\xi_D)$.

**Lemma 2.2.** If $k^2 < k_D^2$, then $T(\xi_k) = k^2$.

**Proof.** Suppose for contradiction that $T(\xi_k) < k^2 < k_D^2$. Let

$$\omega = \max_{x \in \mathcal{S}} f^T(x)B^{-1}(\xi_k)f(x),$$

and suppose that $\xi_* \in \Xi$ places all mass on those points at which $f^T(x)B^{-1}(\xi_k)f(x) = \omega$. Put $\xi_* = (1-x)\xi_k + x\xi_*$, $0 \leq x \leq 1$. Then, as $x \to 0$, we have $\xi_* \to \xi_k$, $T(\xi_*) \to T(\xi_k)$, and $|B(\xi_*)| \to |B(\xi_k)|$. Thus $T(\xi) \leq k^2$ for sufficiently small $\alpha > 0$. Put $\tau(x) = -\log |B(\xi_*)|$. Then $\tau$ is differentiable, with

$$\tau'(0) = -\int_S [f^T(x)B^{-1}(\xi_k)f(x)]d(\xi_*(x) - \xi_k(x))$$

$$= \int_S [f^T(x)B^{-1}(\xi_k)f(x) - \omega]d\xi_k(x)$$

$$< 0,$$

unless

$$\xi_k(\{ x: f^T(x)B^{-1}(\xi_k)f(x) = \omega \}) = 1. \tag{2.2}$$

But if (2.2) holds, then for any $\xi_* \in \Xi$ we have $\tau'(0) \geq 0$, so that the convex functional $-\log |B(\xi)|$ is minimized unconditionally by $\xi_k$. Then $\xi_k$ solves the unconstrained problem, whence $|B(\xi_k)| = |B(\xi_D)|$ and $k^2 = k_D^2$, a contradiction.

Thus $\tau'(0) < 0$, and so for sufficiently small $\alpha > 0$, $T(\xi_\alpha) \leq k^2$ and $|B(\xi_\alpha)| > |B(\xi_k)|$, contradicting the optimality of $\xi_k$. \(\square\)

The following result gives a useful interpretation, in terms of the solution to an associated least-squares regression, of $T(\xi)$. Define

$$m(x) = x^p \phi(x), \quad m_i(x) = (\text{sign } x)^{i-1}m(x),$$

where $p$ is the regression order.
\[
\hat{\theta}_i(\xi) = \arg\min_{\theta} E_{\xi}\{m_i(X) - \theta^T f(X)\}^2,
\]
\[
\hat{m}_i(x; \xi) = \hat{\theta}_i(\xi) f(x).
\]

Denote by \( \Xi_p \) the subset of \( \Xi_X \) whose members are supported on exactly \( p \) points.

**Theorem 2.3.** (i) If \(|B(\xi)| > 0\) then for \( i = 1, 2, \):
\[
(E_{\xi}[m_i(X)])^2 \leq T_i(\xi) = E_{\xi}[\hat{m}_i^2(X; \xi)] = E_{\xi}[m_i(X)\hat{m}_i(X; \xi)] \leq E_{\xi}[m^2(X)].
\]
(ii) If \( \xi \in \Xi_p \) then \(|B(\xi)| > 0\) and for \( i = 1, 2, \):
\[
P_z(\hat{m}_i(X; \xi) = m_i(X)) = 1,
\]
so that
\[
T_1(\xi) = T_2(\xi) = T(\xi) = E_{\xi}[m^2(X)].
\]

**Proof.** We calculate that \( \hat{\theta}_i(\xi) = B^{-1}(\xi)b_i(\xi) \), from which the first two equalities in (i) follow. The lower bound is then the Cauchy–Schwarz inequality together with the identity \( E_{\xi}[\hat{m}_i(X; \xi)] = E_{\xi}[m_i(X)] \). The upper bound also follows from the Cauchy–Schwarz inequality:
\[
T^2_i(\xi) \leq E_{\xi}[m_i^2(X)]E_{\xi}[\hat{m}_i^2(X; \xi)] = E_{\xi}[m^2(X)]T_i(\xi).
\]
For (ii) we use the fact that if \( \xi \in \Xi_p \) has support points \( \{x_j\}_{j=1}^p \) then we have the representation \( B(\xi) = FA\tilde{F}^T \), where \( F \) and \( \tilde{A} \) are non-singular, \( \tilde{F} \) has columns \( f(x_j) \), and \( \tilde{A} = \text{diag}(\xi(\{x_1\}), \ldots, \xi(\{x_p\})) \). We then calculate that \( \hat{m}_i(x_j; \xi) = m_i(x_j) \). \( \square \)

**Corollary 2.4.** \( k^2 = E_{\xi_0}[m^2(X)] \).

**Corollary 2.5.** Define \( \mu_j = E_{\xi}[X^{2j}] \). If \(|B(\xi)| > 0\) then \( \mu_1^{p-2}I^2(\mu_1) \leq T(\xi) \leq \mu_{p-2} \).

**Proof.** Put \( Z = X^2 \). Then \( m(Z) = Z^{p/2-1}l(Z) \) and by Theorem 2.3
\[
T(\xi) \leq E_{\xi}[Z^{p-2}l^2(Z)] \leq E_{\xi}[Z^{p-2}] = \mu_{p-2}.
\]
For the lower bound first assume that \( p \geq 4 \). Then
\[
T(\xi) \geq (E_{\xi}[m_1(X)])^2 = (E_{\xi}[Z^{p/2-1}l(Z)])^2 \geq (E_{\xi}[Z])^{p-2}l^2(E_{\xi}[Z]) = \mu_1^{p-2}I^2(\mu_1),
\]
where we have applied Jensen’s inequality to the convex function \( z^{p/2-1}l(z) \). For \( p = 3 \), a direct calculation of \( T_1(\xi) \) followed by Jensen’s inequality gives
\[
T(\xi) \geq T_1(\xi) = \frac{(E_{\xi}[Zl(Z)])^2}{\mu_1} \geq \frac{(\mu_1l(\mu_1))^2}{\mu_1} = \mu_1l^2(\mu_1);\]
for $p = 2$ we have

$$T(\xi) \geq T_1(\xi) = (E_\xi[l(Z)])^2 \geq l^2(\mu_1).$$

The solution to the unconstrained design problem is known to belong to $\Xi_p$. It is natural to conjecture that this is also the case for (BB). The following theorem gives conditions under which this is indeed the case. In the special cases of Section 3, we shall proceed by verifying these conditions and then considering the resulting numerical problem (2.6). A referee has conjectured that a bounded bias design necessarily belongs to $\Xi_p$; this conjecture remains open.

Let $\Xi(\mu_{p-2})$ denote those members of $\Xi_S$ with fixed even moments

$$E_\xi[X^{2j}] = \mu_j, \quad j = 0, \ldots, p-2, \quad (\mu_0 = 1),$$

and denote by $\Xi(\mu_{p-2}, k)$ those members with, as well, $T(\xi) \leq k^2$. The tightness of $\Xi(\mu_{p-2}, k)$ ensures that $\sup_{\Xi(\mu_{p-2}, k)} E_\xi[X^{2(p-1)}]$ is attained. Similarly, the extrema of $T(\xi)$ are attained in $\Xi(\mu_{p-2})$.

We say that a vector $\mu_{p-2}$ is admissible for (BB) if

$$\min_{\Xi(\mu_{p-2})} T(\xi) \leq k^2 \leq \max_{\Xi(\mu_{p-2})} T(\xi).$$

(2.4)

Note that if $k^2 < \min_{\Xi(\mu_{p-2})} T(\xi)$ then $\Xi(\mu_{p-2}, k) = \emptyset$, whereas (by Lemma 2.2) if $k^2 > \max_{\Xi(\mu_{p-2})} T(\xi)$ then $\xi_k$ cannot belong to $\Xi(\mu_{p-2}, k)$. Our search for the solution to (BB) can then be restricted to those sets $\Xi(\mu_{p-2}, k)$ for which $\mu_{p-2}$ is admissible.

**Theorem 2.6.** (i) Let $k^2 < k^2_D$. Suppose that, whenever $\mu_{p-2}$ is admissible for (BB), there is a $p$-point design $\xi_{\mu_{p-2}, k} \in \Xi(\mu_{p-2}, k) \cap \Xi_p$ satisfying

$$E_\xi[X^{2(p-1)}] = \max_{\Xi(\mu_{p-2}, k)} E_\xi[X^{2(p-1)}].$$

(2.5)

Then the solution to the Bounded Bias problem is $\xi_k = \xi_{\mu_{p-2}, k}$, where

$$\mu_{p-2}^* = \arg \max \{ |B(\xi_{\mu_{p-2}, k})| : \mu_{p-2} \text{ is admissible for (BB)} \}. \tag{2.6}$$

(ii) If $k^2 < k^2_D$ and $\phi(x) \equiv 1$ then there exists $\xi_{\mu_{p-2}, k} \in \Xi(\mu_{p-2}, k) \cap \Xi_p$ satisfying (2.5).

(iii) If $k^2 \geq k^2_D$ then the optimal Bounded Bias design is $\xi_D$.

**Proof.** Only the case $k^2 < k^2_D$ requires a proof. (i) Suppose that $\mu_{p-2}$ is admissible. Note that (2.3) fixes all of the elements of $B(\xi)$ except that element $- E_\xi[X^{2(p-1)}]$ — in its lower right-hand corner. Subject to (2.3), $|B(\xi)|$ is then an increasing function of $E_\xi[X^{2(p-1)}]$ and so by (2.5) is maximized by $\xi_{\mu_{p-2}, k}$. Thus $\xi_{\mu_{p-2}, k}$ is the optimal distribution in $\Xi(\mu_{p-2}, k)$ and it remains only to maximize $|B(\xi_{\mu_{p-2}, k})|$ over admissible vectors $\mu_{p-2}$.

We prove (ii) by first showing that if $\mu_{p-2}$ is admissible then there exists $\xi_{\mu_{p-2}, k} \in \Xi(\mu_{p-2}) \cap \Xi_p$ with

$$T_1(\xi_{\mu_{p-2}, k}) = k^2.$$ 

(2.7)
By Theorem 2.3 (ii) we have that \( \xi_{\mu, p-2, k} \in \Xi(\mu_{p-2}, k) \cap \Xi_p \); we then show that \( \xi_{\mu, p-2, k} \) satisfies (2.5).

Put \( Z = X^2 \). When \( \phi(x) \equiv 1 \) and \( \xi \in \Xi(\mu_{p-2}) \), \( T_1(\xi) \) depends on \( \xi \) only through \( E_{\xi}[Z^{p-1}] \); all other moments are fixed. As \( E_{\xi}[Z^{p-1}] \) varies over

\[
I := \left[ \min_{\Xi(\mu_{p-2})} E_{\xi}[Z^{p-1}], \max_{\Xi(\mu_{p-2})} E_{\xi}[Z^{p-1}] \right],
\]

\( T_1(\xi) \) varies over an interval containing that given by (2.4). To establish (2.7) it then suffices to show that \( E_{\xi}[Z^{p-1}] \) varies over \( I \) as \( \xi \) varies over \( \Xi(\mu_{p-2}) \cap \Xi_p \).

By a result due to Harris (1959) (see also Rustagi 1976, Section 4.5) both endpoints of \( I \) are attained at, perhaps degenerate, members of \( \Xi(\mu_{p-2}) \cap \Xi_p \). A distribution in \( \Xi(\mu_{p-2}) \cap \Xi_p \) is described by \( p \) variables, of which \( p - 1 \) are fixed by the requirement of membership in \( \Xi(\mu_{p-2}) \). There are \( q = [(p + 1)/2] \) points of support \( 0 \leq z_1 < z_2 < \cdots < z_q \leq 1 \) (\( z_1 = 0 \) if and only if \( p \) is odd) and \( q \) corresponding probabilities \( \alpha_j = P_{\xi}(Z = z_j) \). These \( p \) variables satisfy the \( p - 1 \) equations

\[
(2.8)
\]

The distributions in \( \Xi(\mu_{p-2}) \cap \Xi_p \) may then be parameterized by the remaining variable, of which \( E_{\xi}[Z^{p-1}] \) is a continuous function with range \( I \).

To see that (2.5) holds, let \( \xi \in \Xi(\mu_{p-2}) \) be arbitrary and let \( \delta = E_{\xi_{\mu, p-2, k}}[X^{2(p-1)}] - E_{\xi}[X^{2(p-1)}] \). We can represent \( T_1(\xi_{\mu, p-2, k}) - T_1(\xi) \) as a function of \( \delta \) in the following way. Define \( e_{p \times 1} = (0, \ldots, 0, 1, 0)^T, D_{p \times p} = \text{diag}(0, \ldots, 0, 1) \). Then

\[
T_1(\xi) = (b_1(\xi_{\mu, p-2, k}) - \delta e)^T (B(\xi_{\mu, p-2, k}) - \delta D)^{-1}(b_1(\xi_{\mu, p-2, k}) - \delta e).
\]

Some algebra gives

\[
T_1(\xi_{\mu, p-2, k}) - T_1(\xi) = 2 \delta \hat{\theta}_{p-1}(\xi_{\mu, p-2, k}) - \delta^2(B^{-1}(\xi_{\mu, p-2, k}))_{p-1, p-1}
\]

\[
- \frac{\delta \hat{\theta}^2_{p-1}(\xi_{\mu, p-2, k})}{1 - \delta(B^{-1}(\xi_{\mu, p-2, k}))_{p, p}}.
\]

With \( F \) and \( \{x_j\}_{j=1}^p \) as in the proof of Theorem 2.3 and \( m := (x_1^p, \ldots, x_p^p)^T \) we have \( \hat{\theta}(\xi_{\mu, p-2, k}) = B^{-1}(\xi_{\mu, p-2, k})b_1(\xi_{\mu, p-2, k}) = F^T m \). Define

\[
P(x) = \begin{pmatrix} F & f(x) \\ m^T & x^p \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ x_1 & x_2 & \cdots & x_p & x \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_1^{p-1} & x_2^{p-1} & \cdots & x_p^{p-1} & x^{p-1} \\ x_1^p & x_2^p & \cdots & x_p^p & x^p \end{pmatrix},
\]

Then

\[
|P(x)| = x^p - \hat{\theta}(\xi_{\mu, p-2, k})f(x).
\]
The left-hand side of (2.10) is evaluated by applying a well-known formula for the determinant of Vandermonde’s matrix. This yields
\[
\prod_{j=1}^{p} (x - x_j) = x^p - \sum_{j=1}^{p} \hat{\theta}_j(\xi_{\mu_{p-2}, k}) x^{j-1}
\]
as an identity in \( x \), whence
\[
\hat{\theta}_p(\xi_{\mu_{p-2}, k}) = \sum_{j=1}^{p} x_j = 0; \quad \hat{\theta}_{p-1}(\xi_{\mu_{p-2}, k}) = \frac{1}{2} \sum_{j=1}^{p} x_j^2.
\]
Substituting this into (2.9) and using (2.7) gives
\[
0 \leq T_1(\xi_{\mu_{p-2}, k}) - T_1(\xi) = \delta \sum_{j=1}^{p} x_j^2 - \delta^2(B^{-1}(\xi_{\mu_{p-2}, k}))_{p-1, p-1},
\]
so that \( \delta \geq 0 \), as required.  

3. Special cases

We consider the special cases \( p = 2 \) and \( p = 3 \). For \( p \geq 4 \) the designs for constant \( \phi(x) \) may be obtained directly from Theorem 2.6; see Liu and Wiens (1994) for details.

We first rephrase Theorem 2.6 in a form more amenable to numerical work. When the conditions of part (a) of Theorem 2.6 hold, the optimal Bounded Bias design for fixed \( \mu_{p-2} \) may be characterized by Eq. (2.8) with \( \xi = \xi_{\mu_{p-2}, k} \), together with
\[
\langle E_{\xi_{\mu_{p-2}, k}}[m^2(X)] \rangle = \sum_{i=1}^{q} \alpha_i z_i^{p-2} l^2(z_i) = k^2.
\]
These equations define \( B(\xi_{\mu_{p-2}, k}) \) as a function of \( \mu_{p-2} \); this function is then to be maximized over all \( \mu_{p-2} \) for which Eqs. (2.8) and (3.1) have solutions.

3.1. \( p = 2 \)

When \( p = 2 \) (straight-line regression) we have \( \xi_D = \delta_{\pm 1} \) (mass of \( \frac{1}{2} \) at each of \( \pm 1 \)), so that \( k_D^2 = 1 \). By Corollary 2.5, (2.4) implies that \( 0 \leq k^2 \leq 1 \). It is easy to verify, by using Jensen’s inequality, that (2.5) is satisfied by \( \xi_{\mu, k} = \delta_{\pm r} \), where \( r \) is the solution to \( l(r^2) = k \). Thus, the optimal Bounded Bias design is
\[
\xi_k = \delta_{\pm r} \quad \text{where} \quad l(r^2) = \min(k, 1).
\]
Similarly, the optimal Bounded Variance design is \( \xi_c = \delta_{\pm c} \), \( 0 < c < 1 \). The minimax design is \( \xi_k^* = \delta_{\pm r^*} \), where
\[
r^* = \arg \min_{[0, 1]} \frac{1 + (n/\sigma^2)l^2(r^2)}{r^2}.
\]
3.2. $p = 3$

When $p = 3$ (quadratic regression) condition (2.3) fixes $E_x[Z] = \mu_1$. The $D$-optimal design $\xi_p$ places mass $\frac{1}{3}$ at each of 0, 1 and $-1$ so that $k^2_0 = \frac{2}{3}$. We will require the following result.

**Lemma 3.1.** If $Z$ is a non-negative random variable with finite second moment, and $l(\cdot)$ is convex on the range of $Z$, then

$$
l \left( \frac{E[Z^2]}{E[Z]} \right) \leq \frac{E[Zl(Z)]}{E[Z]}.
$$

**Proof.** Let $Z \sim G$ and define a distribution function $G_*$ by $dG_*(z) = z dG(z)/E_G[Z]$. Denote expectation with respect to $G_*$ by $E_*$. Then by Jensen’s inequality,

$$
l \left( \frac{E[Z^2]}{E[Z]} \right) = l(E_*[Z]) \leq E_*[l(Z)] = \frac{E[Zl(Z)]}{E[Z]}.
$$

Eqs. (2.8) and (3.1) suggest that

$$
\xi_{\mu_1,k} = (1 - z) \delta_0 + z \delta_{\pm \sqrt{\mu_1/k}},
$$

(3.2)

where $z \in [\mu_1, 1]$ satisfies

$$
\mu_1 l^2 \left( \frac{\mu_1}{z} \right) = k^2.
$$

(3.3)

Corollary 2.5 guarantees a solution to this equation. To see that $\xi_{\mu_1,k}$ satisfies (2.5), note that if $\xi \in \Xi(\mu_1,k)$ then by Lemma 3.1 and (3.3) we have

$$
l \left( \frac{E_x[Z^2]}{\mu_1} \right) \leq E_x[Zl(Z)] = \frac{1}{\sqrt{\mu_1}} l \left( \frac{\mu_1}{z} \right) = l \left( \frac{E_{\xi_{l^1,k}}[Z^2]}{\mu_1} \right).
$$

The monotonicity of $l(\cdot)$ then gives

$$
E_x[Z^2] \leq E_{\xi_{l^1,k}}[Z^2],
$$

as required. Maximizing $|B(\xi_{\mu_1,k})| = \mu_1^3 (1/z - 1)$ then yields, for $k^2 < \frac{2}{3}$, the optimal Bounded Bias design $\xi = \xi_{\mu_1,k}$, where

$$
\mu_1^* = \arg \max_{\mu_1} \left\{ \mu_1^3 \left( \frac{1}{z} - 1 \right) : \mu_1 l^2 \left( \frac{\mu_1}{z} \right) = k^2, \mu_1 \leq z \leq 1 \right\}.
$$

The optimal Bounded Variance design is, for $0 \leq c^2 \leq \frac{k}{2} = |B(\xi_D)|$, given by $\xi_c = \xi_{\mu_1,c}$, where $\alpha = \alpha(\mu_1) = \mu_1^3 / (\mu_1^3 + c^2)$ and

$$
\mu_1^{**} = \arg \min_{\mu_1} \left\{ \mu_1 l^2 \left( \frac{\mu_1}{z} \right) : 0 \leq \mu_1 \leq \alpha(\mu_1) \right\}.
$$
The minimax design is \( \xi_{L^*} = (1 - \alpha^*) \delta_0 + \alpha^* \delta_+ \sqrt{\mu_1 / \alpha^*} \), where

\[
(\alpha^*, \mu_1^*) = \arg \min_{(\alpha, \mu_1)} \left\{ \frac{\alpha(1 + (n/\sigma^2) \mu_1 I^2(\mu_1/\alpha))}{\mu_1 (1 - \alpha)} : 0 \leq \mu_1 \leq \alpha \leq 1 \right\}.
\]

4. Comparisons

4.1. Linear/quadratic discrimination

Stigler (1971) (see Studden, 1982), for extensions) considered (1.2) in the case \( p = 2, \psi(x) \equiv 1 \), and obtained designs to minimize the determinant of the covariance matrix of the estimates of the constant and linear regression coefficients, subject to a bound of \( (C \sigma^2 / n) \) \( (C \geq 4) \) on the variance of the estimate of the quadratic coefficient were a quadratic model to be fitted. We compare the resulting designs

\[
\xi_S^{(C)}(\pm 1) = \frac{1}{4} + \frac{1}{2} \sqrt{\frac{1}{4} - \frac{1}{C}}, \quad \xi_S^{(C)}(0) = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{1}{C}}
\]

for \( C = 5 \) and \( C = 10 \), with the designs of Section 3.2 with \( \psi(x) \equiv 1 \). As measures of performance we use

1. The \( D \)-efficiencies as defined in Pukelsheim and Rosenberger (1993). These are standardized efficiencies relative to the design which is optimal in a given situation. Thus, for example, the \( D \)-efficiency \( D(\xi; \theta_{(A)}) \) for \( \theta_{(A)} = (\theta_0, \theta_1)^T \), assuming the true response to be exactly linear, is the ratio, raised to the power \( \frac{1}{2} \), of the determinant of the inverse of the \( 2 \times 2 \) covariance matrix of \( \hat{\theta}_{(A)} \), using the design \( \xi \), to that using the optimal design \( \xi_0 = \delta_{\pm1} \). We consider \( D(\xi; \theta_{(A)}) \) as well as \( D(\xi; \theta_{(B)}) \) and \( D(\xi; \theta_{(B)}) \) for which the optimal design is \( \xi_1 = \frac{1}{2} \delta_0 + \frac{1}{2} \delta_{\pm1} \), and \( D(\xi; \theta_{(2)}) \) for which the optimal design is \( \xi_2 = \frac{1}{2} \delta_0 + \frac{1}{2} \delta_{\pm1} \). In the case of \( D(\xi; \theta_{(2)}) \) it is assumed that the full quadratic model is first fitted and the variance of \( \theta_2 \) obtained from the covariance matrix for \( \theta_{(B)} \).

2. The maximal bias \( T(\xi) = b_1^T(\xi) B^{-1}(\xi) b_1(\xi) \), where \( b_1(\xi) = (E_3[X^3], E_2[X^4], E_2[X^3])^T \).

3. The standardized determinant of the MSE matrix for \( D(\xi; \theta_{(B)}) \) at (1.3), i.e.

\[
\text{mse} = \frac{\sigma^2}{n} \left( \left(1 + \frac{T(\xi)}{\sigma^2/n} \right) \left( |B(\xi)| \right)^{1/3} \right);
\]

with \( \sigma^2/n = 0.01 \) and \( \sigma^2/n = 1 \).

The performance measures are given in Table 1 for the designs described above and for the Bounded Bias designs \( \xi_{bb}^{(5/12)} = \frac{7}{12} \delta_0 + \frac{5}{12} \delta_{\pm1} \), for which \( k^2 = \frac{5}{12} \), and \( \xi_{bb}^{(1/4)} = \frac{1}{2} \delta_0 + \frac{1}{2} \delta_{\pm(1/2)^n} \), for which \( k^2 = \frac{1}{4} \). We note that \( \xi_1 \) and \( \xi_2 \) are also Bounded Bias designs for \( k^2 = \frac{3}{4} \) and \( k^2 = \frac{1}{2} \), respectively, and are the Minimax designs as \( \sigma^2/n \) tends to \( \infty \) and 0, respectively. They are also special cases of Bounded Variance designs.
Table 1
Designs for linear and quadratic model discrimination

<table>
<thead>
<tr>
<th>Design</th>
<th>Efficiencies</th>
<th>max. bias</th>
<th>mse</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\theta_2$</td>
<td>$\theta_{(B)}$</td>
<td>$\theta_{(A)}$</td>
</tr>
<tr>
<td>$\xi_0$</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$\xi_1$</td>
<td>0.89</td>
<td>0.94</td>
<td>0.71</td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>0.97</td>
<td>0.88</td>
<td>0.65</td>
</tr>
<tr>
<td>$\xi_3$</td>
<td>0.63</td>
<td>0.84</td>
<td>0.94</td>
</tr>
<tr>
<td>$\xi_4$</td>
<td>0.63</td>
<td>0.75</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Table 2
Designs for quadratic and cubic model discrimination

<table>
<thead>
<tr>
<th>Design</th>
<th>Efficiencies</th>
<th>max. bias</th>
<th>mse</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\theta_3$</td>
<td>$\theta_{(B)}$</td>
<td>$\theta_{(A)}$</td>
</tr>
<tr>
<td>PR Section 2.1</td>
<td>1</td>
<td>0.93</td>
<td>0.75</td>
</tr>
<tr>
<td>PR Section 2.2</td>
<td>0.85</td>
<td>1</td>
<td>0.87</td>
</tr>
<tr>
<td>PR Section 2.3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PR Section 3.1</td>
<td>0.72</td>
<td>0.94</td>
<td>0.84</td>
</tr>
<tr>
<td>PR Section 3.2</td>
<td>0.42</td>
<td>0.89</td>
<td>0.94</td>
</tr>
<tr>
<td>PR Section 4.1</td>
<td>0.66</td>
<td>0.98</td>
<td>0.91</td>
</tr>
<tr>
<td>PR Section 4.2</td>
<td>0.64</td>
<td>0.96</td>
<td>0.90</td>
</tr>
<tr>
<td>PR Section 4.3</td>
<td>1.00</td>
<td>0.94</td>
<td>0.75</td>
</tr>
<tr>
<td>PR Section 5.1</td>
<td>0.50</td>
<td>0.93</td>
<td>0.94</td>
</tr>
<tr>
<td>PR Section 5.2</td>
<td>0.50</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>$\xi_{bb}$</td>
<td>0.97</td>
<td>0.92</td>
<td>0.74</td>
</tr>
<tr>
<td>$\xi_{bc}$</td>
<td>0.96</td>
<td>0.87</td>
<td>0.70</td>
</tr>
<tr>
<td>$\xi_{m}$</td>
<td>0.97</td>
<td>0.95</td>
<td>0.77</td>
</tr>
</tbody>
</table>

4.2 Quadratic/cubic discrimination

Pukelsheim and Rosenberger (1993) reviewed and extended the literature on designs intended to give efficient estimates in a quadratic model, yet guard against a cubic response. Table 2 gives some comparative measures. The relative efficiencies are those for estimation of the cubic coefficient $\theta_3$, for the full $4 \times 1$ parameter vector $\theta_{(B)}$ and for the vector $\theta_{(A)}$ of parameters in the quadratic model. The ten designs in Pukelsheim and Rosenberger (1993) are referred to by the relevant section number, e.g. “PR Section 2.1”. These are compared with the Bounded Bias design $\xi_{bb}(\pm 1) = 0.1495$, $\xi_{bb}(\pm 0.444) = 0.3505$ for which $k^2 = 0.3$, with the Bounded Variance design $\xi_{bc}(\pm 1) = 0.1290$, $\xi_{bc}(\pm 0.443) = 0.3710$ for which $c^2 = 0.003$, and with the
design \( \xi_m(\pm 1) = 0.1703, \xi_m(\pm 0.445) = 0.3297 \) which is minimax when \( \sigma^2 = 0.01n \). Note the similarity in the performances of these three designs and those of PR Section 2.1, optimal for estimation of \( \theta_3 \), and PR Section 4.3, which minimizes a mixture of the loss measures for \( \theta_3 \) and \( \theta_{(4)} \).

Acknowledgements

This paper has benefitted from the comments of a referee. The research of both authors is supported by the Natural Sciences and Engineering Research Council of Canada.

References


