

Robust Designs for Approximately Polynomial Regression

Abbreviated title: Robust Polynomial Designs

A version of this paper has appeared in the
Journal of Statistical Planning and Inference, 1997, 64, 369-381.

Shawn X. Liu & Douglas P. Wiens¹
Mount Royal College and University of Alberta

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.

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

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

If the fitted response is exactly correct, and if, as is assumed in this paper, the errors are additive and uncorrelated, with common variance σ^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. Suppose however

¹Shawn Liu is Instructor, Department of Mathematics, Physics and Engineering, Mount Royal College, Calgary, Alberta, Canada T3E 6K6. Douglas P. Wiens is Professor, Statistics Centre, Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1. The research of both authors is supported by the Natural Sciences and Engineering Research Council of Canada.

AMS 1991 *subject classifications*. Primary 62K05, 62F35; secondary 62J05.

Key words and phrases. Bounded bias, bounded variance, approximate polynomial regression, minimax designs.

that the ‘true’ response is in fact only approximately polynomial in that for $x \in \mathcal{S}$ and some unknown function ψ we have

$$E[Y|x] = \boldsymbol{\theta}^T \mathbf{f}(x) + \left(\frac{x-a}{b}\right)^p \psi\left(\frac{x-a}{b}\right), \quad (1.2)$$

where $\boldsymbol{\theta} := (\theta_0, \dots, \theta_{p-1})^T$ and $\mathbf{f}(x) := (1, x, x^2, \dots, x^{p-1})^T$. 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 ψ at 0 ensures that the parameter $\boldsymbol{\theta}$ 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 $\hat{\boldsymbol{\theta}}$. 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, \dots, x_n . Denote this measure by ξ and define

$$\begin{aligned} \mathbf{B}(\xi) &= n^{-1} \sum_{i=1}^n \mathbf{f}(x_i) \mathbf{f}^T(x_i) = \int_{\mathcal{S}} \mathbf{f}(x) \mathbf{f}^T(x) d\xi(x), \\ \mathbf{b}(\xi, \psi) &= n^{-1} \sum_{i=1}^n \mathbf{f}(x_i) \left(\frac{x_i-a}{b}\right)^p \psi\left(\frac{x_i-a}{b}\right) = \int_{\mathcal{S}} \mathbf{f}(x) \left(\frac{x-a}{b}\right)^p \psi\left(\frac{x-a}{b}\right) d\xi(x). \end{aligned}$$

Then the bias vector and covariance matrix are

$$\begin{aligned} E[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] &= \mathbf{B}^{-1}(\xi) \mathbf{b}(\xi, \psi) \\ COV[\hat{\boldsymbol{\theta}}] &= \frac{\sigma^2}{n} \mathbf{B}^{-1}(\xi), \end{aligned}$$

so that the mean squared error matrix $E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T]$ has determinant

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

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,

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

for a given constant k^2 .

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

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

for a given constant c^2 .

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

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

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

$$\mathbf{b}(\xi_{ab}, \psi) = \mathbf{C} \mathbf{b}(\xi_{01}, \psi), \quad \mathbf{B}(\xi_{ab}) = \mathbf{C} \mathbf{B}(\xi_{01}) \mathbf{C}^T, \quad |\mathbf{B}(\xi_{ab})| = b^{p(p-1)} |\mathbf{B}(\xi_{01})|.$$

It follows that if ξ_{01} is an optimal design on $[-1, 1]$ with respect to any of the above criteria, then ξ_{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 ξ in the class Ξ of *all* probability measures on \mathcal{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 $\mathbf{f}(\mathbf{x}) = (1, x_1, \dots, x_p)^T$, 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 of this paper.

2 General Theory

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

$$\begin{aligned} \mathbf{b}_i(\xi) &= \int_S \mathbf{f}(x) x^p (\text{sign } x)^{i-1} \phi(x) d\xi(x), \quad i = 1, 2, \\ T_i(\xi) &= \begin{cases} \mathbf{b}_i^T(\xi) \mathbf{B}^{-1}(\xi) \mathbf{b}_i(\xi) & \text{if } |\mathbf{B}(\xi)| > 0, \\ \infty & \text{if } |\mathbf{B}(\xi)| = 0. \end{cases} \\ T(\xi) &= \begin{cases} \max_{i=1,2} T_i(\xi) & \text{if } \phi(0) = 0, \\ T_1(\xi) & \text{if } \phi(0) > 0. \end{cases} \end{aligned}$$

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 $|\mathbf{B}(\xi)| > 0$,*

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

Proof: Fix a ξ with $|\mathbf{B}(\xi)| > 0$. Denote by Ψ the set of functions ψ satisfying assumption A1). Note that the functional $A(\psi) := \mathbf{b}^T(\xi, \psi) \mathbf{B}^{-1}(\xi) \mathbf{b}(\xi, \psi)$ is quadratic, hence convex, on the convex set Ψ . Thus if $\psi \in \Psi$ is not an extremal, say $\psi = (1 - \alpha)\psi_1 + \alpha\psi_2$ 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 Ψ . But these extremals have $|\psi(x)| \equiv \phi(x)$. In order that such an extremal be continuous, we must have $\psi(x) \equiv \phi(x)$ or $\psi(x) \equiv -\phi(x)$ if $\phi(0) > 0$, while if $\phi(0) = 0$ we could also have $\psi(x) \equiv (\text{sign } x)\phi(x)$ or $\psi(x) \equiv -(\text{sign } x)\phi(x)$. The result follows upon noting that $A(\psi) = A(-\psi)$. \square

The Bounded Bias problem may now be phrased as

$$\mathbf{BB):} \text{ Find } \xi_k := \operatorname{argmax}_{\Xi} \{|\mathbf{B}(\xi)| : T(\xi) \leq k^2\}$$

and the Bounded Variance problem as

$$\mathbf{BV):} \text{ Find } \xi_c := \operatorname{argmin}_{\Xi} \{T(\xi) : |\mathbf{B}(\xi)| \geq c^2\}.$$

From C3) and (1.3) the Minimax problem is

M): Find ξ_{k^*} , where

$$k^* := \operatorname{argmin}_k \left(\frac{\sigma^2}{n} \right)^p \left(1 + \frac{n}{\sigma^2} T(\xi_k) \right) / |\mathbf{B}(\xi_k)|. \quad (2.1)$$

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 $T(\xi) \leq E_\xi[X^{2p}]$ if $|\mathbf{B}(\xi)| > 0$. Thus one can make $T(\xi)$ arbitrarily small while retaining a non-singular covariance matrix $|\mathbf{B}(\xi)|$, for instance by employing a p -point design with a sufficient amount of mass at 0. It follows that $\{|\mathbf{B}(\xi)| : T(\xi) \leq k^2\} \neq \emptyset$ for every $k > 0$. Standard arguments now show that there exists a solution ξ_k to BB) for every $k > 0$, and then the convexity of $-\log|\mathbf{B}(\xi)|$ and of $T(\xi)$ may be used to show that ξ_k lies in the subclass Ξ_S of symmetric members of Ξ , i.e. those ξ satisfying $\xi(x) = 1 - \xi((-x)^-)$ 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 ξ_k does not solve the *unconstrained* problem of maximizing $|\mathbf{B}(\xi)|$ unconditionally. Let ξ_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

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

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

$$\begin{aligned} \tau'(0) &= - \int_{\mathcal{S}} [\mathbf{f}^T(x) \mathbf{B}^{-1}(\xi_k) \mathbf{f}(x)] d(\xi_*(x) - \xi_k(x)) \\ &= \int_{\mathcal{S}} [\mathbf{f}^T(x) \mathbf{B}^{-1}(\xi_k) \mathbf{f}(x) - \omega] d\xi_k(x) \\ &< 0, \end{aligned}$$

unless

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

But if (2.2) holds, then for *any* $\xi_* \in \Xi$ we have $\tau'(0) \geq 0$, so that the convex functional $-\log|\mathbf{B}(\xi)|$ is minimized *unconditionally* by ξ_k . Then ξ_k solves the unconstrained problem, whence $|\mathbf{B}(\xi_k)| = |\mathbf{B}(\xi_D)|$ 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 $|\mathbf{B}(\xi_\alpha)| > |\mathbf{B}(\xi_k)|$, contradicting the optimality of ξ_k . \square

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

$$\begin{aligned} m(x) &= x^p \phi(x), \quad m_i(x) = (\text{sign } x)^{i-1} m(x), \\ \hat{\boldsymbol{\theta}}_i(\xi) &= \operatorname{argmin}_{\boldsymbol{\theta}} E_\xi[\{m_i(X) - \boldsymbol{\theta}^T \mathbf{f}(X)\}^2], \\ \hat{m}_i(x; \xi) &= \hat{\boldsymbol{\theta}}_i^T(\xi) \mathbf{f}(x). \end{aligned}$$

Denote by Ξ_p the subset of Ξ_S whose members are supported on exactly p points.

Theorem 2.3 *i) If $|\mathbf{B}(\xi)| > 0$ 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 $|\mathbf{B}(\xi)| > 0$ and for $i = 1, 2$:

$$P_\xi(\hat{m}_i(X; \xi) = m_i(X)) = 1,$$

so that

$$T_1(\xi) = T_2(\xi) = T(\xi) = E_\xi[m^2(X)].$$

Proof: We calculate that $\hat{\boldsymbol{\theta}}_i(\xi) = \mathbf{B}^{-1}(\xi) \mathbf{b}_i(\xi)$, 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:

$$T_i^2(\xi) \leq E_\xi[m_i^2(X)] E_\xi[\hat{m}_i^2(X; \xi)] = E_\xi[m^2(X)] T_i(\xi).$$

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 $\mathbf{B}(\xi) = \mathbf{F} \boldsymbol{\Delta} \mathbf{F}^T$, where \mathbf{F} and $\boldsymbol{\Delta}$ are non-singular, \mathbf{F} has columns $\mathbf{f}(x_j)$, and $\boldsymbol{\Delta} = \operatorname{diag}(\xi(\{x_1\}), \dots, \xi(\{x_p\}))$. We then calculate that $\hat{m}_i(x_j; \xi) = m_i(x_j)$. \square

Corollary 2.4 $k_D^2 = E_{\xi_D}[m^2(X)]$.

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

Proof: Put $Z = X^2$. Then $m(Z) = Z^{\frac{p}{2}-1} l(Z)$ 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

$$T(\xi) \geq (E_\xi[m_1(X)])^2 = \left(E_\xi[Z^{\frac{p}{2}-1}l(Z)]\right)^2 \geq (E_\xi[Z])^{p-2}l^2(E_\xi[Z]) = \mu_1^{p-2}l^2(\mu_1),$$

where we have applied Jensen's Inequality to the convex function $z^{\frac{p}{2}-1}l(z)$. For $p = 3$ a direct calculation of $T_1(\xi)$ followed by Jensen's Inequality gives

$$T(\xi) \geq T_1(\xi) = \frac{E_\xi[Zl(Z)]^2}{\mu_1} \geq \frac{(\mu_1 l(\mu_1))^2}{\mu_1} = \mu_1 l^2(\mu_1);$$

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 Ξ_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 Ξ_p ; this conjecture remains open.

Let $\Xi(\boldsymbol{\mu}_{p-2})$ denote those members of Ξ_S with fixed even moments

$$E_\xi[X^{2j}] = \mu_j, \quad j = 0, \dots, p-2, \quad (\mu_0 = 1) \tag{2.3}$$

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

We say that a vector $\boldsymbol{\mu}_{p-2}$ is *admissible* for BB) if

$$\min_{\Xi(\boldsymbol{\mu}_{p-2})} T(\xi) \leq k^2 \leq \max_{\Xi(\boldsymbol{\mu}_{p-2})} T(\xi). \tag{2.4}$$

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

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

$$E_{\xi_{\boldsymbol{\mu}_{p-2}, k}}[X^{2(p-1)}] = \max_{\Xi(\boldsymbol{\mu}_{p-2}, k)} E_\xi[X^{2(p-1)}]. \tag{2.5}$$

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

$$\mu_{p-2}^* = \operatorname{argmax}\{|\mathbf{B}(\xi_{\mu_{p-2}, k})| : \mu_{p-2} \text{ is admissible for BB}\}. \quad (2.6)$$

- ii) If $k^2 < k_D^2$ 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 $k^2 \geq k_D^2$ then the optimal Bounded Bias design is ξ_D .

Proof: Only the case $k^2 < k_D^2$ requires a proof. i) Suppose that μ_{p-2} is admissible. Note that (2.3) fixes all of the elements of $\mathbf{B}(\xi)$ except that element - $E_\xi[X^{2(p-1)}]$ - in its lower right hand corner. Subject to (2.3), $|\mathbf{B}(\xi)|$ 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 $|\mathbf{B}(\xi_{\mu_{p-2}, k})|$ over admissible vectors μ_{p-2} .

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

$$T_1(\xi_{\mu_{p-2}, k}) = k^2. \quad (2.7)$$

By Theorem 2.3ii) 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 ξ 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 ξ varies over $\Xi(\mu_{p-2}) \cap \Xi_p$.

By a result due to Harris (1959) (see also Section 4.5 of Rustagi (1976)) 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 $q = \lfloor \frac{p+1}{2} \rfloor$ points of support $0 \leq z_1 < z_2 < \dots < z_q \leq 1$ ($z_1 = 0$ if and only if p is odd) and q corresponding probabilities $\alpha_i = P_\xi(Z = z_i)$. These p variables satisfy the $p - 1$ equations

$$(E_\xi[Z^j] =) \sum_{i=1}^q \alpha_i z_i^j = \mu_j, \quad j = 0, \dots, p - 2. \quad (2.8)$$

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 $\xi \in \Xi(\mu_{p-2}, k)$ 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 δ in the following way. Define $\mathbf{e}_{p \times 1} = (0, \dots, 0, 1, 0)^T$, $\mathbf{D}_{p \times p} = \operatorname{diag}(0, \dots, 0, 1)$. Then

$$T_1(\xi) = \left(\mathbf{b}_1(\xi_{\mu_{p-2}, k}) - \delta \mathbf{e} \right)^T \left(\mathbf{B}(\xi_{\mu_{p-2}, k}) - \delta \mathbf{D} \right)^{-1} \left(\mathbf{b}_1(\xi_{\mu_{p-2}, k}) - \delta \mathbf{e} \right).$$

Some algebra gives

$$\begin{aligned}
T_1(\xi\boldsymbol{\mu}_{p-2,k}) - T_1(\xi) &= 2\delta\hat{\boldsymbol{\theta}}_{p-1}(\xi\boldsymbol{\mu}_{p-2,k}) - \delta^2 \left(\mathbf{B}^{-1}(\xi\boldsymbol{\mu}_{p-2,k}) \right)_{p-1,p-1} \\
&\quad - \frac{\delta\hat{\boldsymbol{\theta}}_p^2(\xi\boldsymbol{\mu}_{p-2,k})}{1 - \delta \left(\mathbf{B}^{-1}(\xi\boldsymbol{\mu}_{p-2,k}) \right)_{p,p}}.
\end{aligned} \tag{2.9}$$

With \mathbf{F} and $\{x_j\}_{j=1}^p$ as in the proof of Theorem 2.3 and $\mathbf{m} := (x_1^p, \dots, x_p^p)^T$ we have $\hat{\boldsymbol{\theta}}(\xi\boldsymbol{\mu}_{p-2,k}) = \mathbf{B}^{-1}(\xi\boldsymbol{\mu}_{p-2,k})\mathbf{b}_1(\xi\boldsymbol{\mu}_{p-2,k}) = \mathbf{F}^{-T}\mathbf{m}$. Define

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

Then

$$\frac{|\mathbf{P}(x)|}{|\mathbf{F}|} = x^p - \hat{\boldsymbol{\theta}}(\xi\boldsymbol{\mu}_{p-2,k})\mathbf{f}(x). \tag{2.10}$$

The left side of (2.10) is evaluated by applying a well known formula for the determinant of Vandermonde's matrix. This yields

$$\prod_{j=1}^p (x - x_j) = x^p - \sum_{j=1}^p \hat{\theta}_j(\xi\boldsymbol{\mu}_{p-2,k})x^{j-1}$$

as an identity in x , whence

$$\hat{\theta}_p(\xi\boldsymbol{\mu}_{p-2,k}) = \sum_{j=1}^p x_j = 0; \quad \hat{\theta}_{p-1}(\xi\boldsymbol{\mu}_{p-2,k}) = \frac{1}{2} \sum_{j=1}^p x_j^2.$$

Substituting this into (2.9) and using (2.7) gives

$$0 \leq T_1(\xi\boldsymbol{\mu}_{p-2,k}) - T_1(\xi) = \delta \sum_{j=1}^p x_j^2 - \delta^2 \left(\mathbf{B}^{-1}(\xi\boldsymbol{\mu}_{p-2,k}) \right)_{p-1,p-1},$$

so that $\delta \geq 0$, as required. \square

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 $\boldsymbol{\mu}_{p-2}$ may be characterized by equations (2.8) with $\xi = \xi\boldsymbol{\mu}_{p-2,k}$, together with

$$(E_{\xi\boldsymbol{\mu}_{p-2,k}}[m^2(X)] =) \sum_{i=1}^q \alpha_i z_i^{p-2} I^2(z_i) = k^2. \tag{3.1}$$

These equations define $|\mathbf{B}(\xi_{\boldsymbol{\mu}_{p-2,k}})|$ as a function of $\boldsymbol{\mu}_{p-2}$; this function is then to be maximized over all $\boldsymbol{\mu}_{p-2}$ for which equations (2.8) and (3.1) have solutions.

3.1 $p = 2$.

When $p = 2$ (straight line regression) we have $\xi_{\mathbf{D}} = \delta_{\pm 1}$ (mass of $\frac{1}{2}$ at each of ± 1), so that $k_{\mathbf{D}}^2 = 1$. 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 $\xi_{\boldsymbol{\mu}_0,k} = \delta_{\pm r}$, where r is the solution to $l(r^2) = k$. Thus the optimal Bounded Bias design is

$$\xi_k = \delta_{\pm r}, \text{ where } l(r^2) = \min(k, 1).$$

Similarly the optimal Bounded Variance design is $\xi_c = \delta_{\pm c}$, $0 < c \leq 1$. The Minimax design is $\xi_{k^*} = \delta_{\pm r^*}$, where

$$r^* = \operatorname{argmin}_{[0,1]} \frac{1 + \frac{n}{\sigma^2} l^2(r^2)}{r^2}.$$

3.2 $p = 3$.

When $p = 3$ (quadratic regression) condition (2.3) fixes $E_{\xi}[Z] = \mu_1$. The D -optimal design $\xi_{\mathbf{D}}$ places mass $\frac{1}{3}$ at each of 0, 1 and -1 so that $k_{\mathbf{D}}^2 = \frac{2}{3}$. 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 $dG_*(z) = zdG(z)/E_G[Z]$. Denote expectation with respect to G_* by E_* . Then by Jensen's Inequality,

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

□

Equations (2.8) and (3.1) suggest that

$$\xi_{\boldsymbol{\mu}_1,k} = (1 - \alpha)\delta_0 + \alpha\delta_{\pm\sqrt{\frac{\mu_1}{\alpha}}}, \quad (3.2)$$

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

$$\mu_1 l^2\left(\frac{\mu_1}{\alpha}\right) = k^2. \quad (3.3)$$

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

$$l\left(\frac{E_\xi[Z^2]}{\mu_1}\right) \leq \frac{E_\xi[Zl(Z)]}{\mu_1} = \sqrt{\frac{T_1(\xi)}{\mu_1}} \leq \frac{k}{\sqrt{\mu_1}} = l\left(\frac{\mu_1}{\alpha}\right) = l\left(\frac{E_{\xi_{\mu_1, k}}[Z^2]}{\mu_1}\right).$$

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

$$E_\xi[Z^2] \leq E_{\xi_{\mu_1, k}}[Z^2],$$

as required. Maximizing $|\mathbf{B}(\xi_{\mu_1, k})| = \mu_1^3 \left(\frac{1}{\alpha} - 1\right)$ then yields, for $k^2 < \frac{2}{3}$, the optimal Bounded Bias design $\xi_k = \xi_{\mu_1^*, k}$, where

$$\mu_1^* = \operatorname{argmax}_{\mu_1} \left\{ \mu_1^3 \left(\frac{1}{\alpha} - 1\right) : \mu_1 l^2\left(\frac{\mu_1}{\alpha}\right) = k^2, \mu_1 \leq \alpha \leq 1 \right\}.$$

The optimal Bounded Variance design is, for $0 \leq c^2 \leq \frac{4}{27} = |\mathbf{B}(\xi_D)|$, given by $\xi_c = \xi_{\mu_1^{**}, c}$, where $\alpha = \alpha(\mu_1) = \frac{\mu_1^3}{\mu_1^3 + c^2}$ and

$$\mu_1^{**} = \operatorname{argmin}_{\mu_1} \left\{ \mu_1 l^2\left(\frac{\mu_1}{\alpha}\right) : 0 \leq \mu_1 \leq \alpha(\mu_1) \right\}.$$

The Minimax design is $\xi_{k^*} = (1 - \alpha^*)\delta_0 + \alpha^* \delta_{\pm\sqrt{\frac{\mu_1^*}{\alpha^*}}}$, where

$$(\alpha^*, \mu_1^*) = \operatorname{argmin}_{(\alpha, \mu_1)} \left\{ \frac{\alpha(1 + \frac{n}{\sigma^2} \mu_1 l^2(\frac{\mu_1}{\alpha}))}{\mu_1^3(1 - \alpha)} : 0 \leq \mu_1 \leq \alpha \leq 1 \right\}.$$

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 $\frac{C\sigma^2}{n}$ ($C \geq 4$) 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 $\phi(x) \equiv 1$. 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 $D(\xi; \boldsymbol{\theta}_{(A)})$ for $\boldsymbol{\theta}_{(A)} = (\theta_0, \theta_1)^T$, 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 2×2 covariance matrix of $\hat{\boldsymbol{\theta}}_{(A)}$, using the design ξ , to that using the optimal design $\xi_0 = \delta_{\pm 1}$. We consider $D(\xi; \boldsymbol{\theta}_{(A)})$ as well as $D(\xi; \boldsymbol{\theta}_{(B)})$ ($\boldsymbol{\theta}_{(B)} = (\theta_0, \theta_1, \theta_2)^T$) for which the optimal design is $\xi_1 = \frac{1}{3}\delta_0 + \frac{2}{3}\delta_{\pm 1}$, and $D(\xi; \boldsymbol{\theta}_2)$ for which the optimal design is $\xi_2 = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{\pm 1}$. In the case of $D(\xi; \boldsymbol{\theta}_2)$ it is assumed that the full quadratic model is first fitted and the variance of $\hat{\boldsymbol{\theta}}_2$ obtained from the covariance matrix for $\hat{\boldsymbol{\theta}}_{(B)}$.
2. The maximal bias $T(\xi) = \mathbf{b}_1^T(\xi)\mathbf{B}^{-1}(\xi)\mathbf{b}_1(\xi)$, where $\mathbf{b}_1(\xi) = (E_\xi[X^3], E_\xi[X^4], E_\xi[X^5])^T$.
3. The standardized determinant of the MSE matrix for $D(\xi; \boldsymbol{\theta}_{(B)})$ as at (1.3), i.e. $mse = \frac{\sigma^2}{n}((1 + \frac{T(\xi)}{\sigma^2/n})/(|\mathbf{B}(\xi)|))^{1/3}$; with $\frac{\sigma^2}{n} = .01$ and $\frac{\sigma^2}{n} = 1$.

The performance measures are given in Table 1 for the designs described above and for the Bounded Bias designs $\xi_{bb}^{(5/12)} = \frac{7}{12}\delta_0 + \frac{5}{12}\delta_{\pm 1}$, for which $k^2 = \frac{5}{12}$, and $\xi_{bb}^{(1/4)} = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{\pm(1/2)^{1/6}}$, for which $k^2 = \frac{1}{4}$. We note that ξ_1 and ξ_2 are also Bounded Bias designs for $k^2 = \frac{2}{3}$ and $k^2 = \frac{1}{2}$ respectively, and are the Minimax designs as $\frac{\sigma^2}{n}$ tends to ∞ and 0 respectively. They are also special cases of Bounded Variance designs.

Table 1. Designs for Linear and Quadratic Model Discrimination

Design	Efficiencies			max. bias	mse	
	θ_2	$\boldsymbol{\theta}_{(B)}$	$\boldsymbol{\theta}_{(A)}$		$\frac{\sigma^2}{n} = .01$	$\frac{\sigma^2}{n} = 1$
ξ_0	0	0	1	∞	∞	∞
ξ_1	.89	1	.82	.67	.077	2.24
ξ_2	1	.94	.71	.50	.074	2.29
$\xi_S^{(5)}$.80	.99	.85	.72	.080	2.28
$\xi_S^{(10)}$.40	.84	.94	.89	.100	2.77
$\xi_{bb}^{(5/12)}$.97	.88	.65	.42	.075	2.41
$\xi_{bb}^{(1/4)}$.63	.75	.63	.25	.075	2.71

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 θ_3 , for the full 4×1 parameter vector $\boldsymbol{\theta}_{(B)}$ and for the vector $\boldsymbol{\theta}_{(A)}$ 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 §2.1”. These are compared with the Bounded Bias design $\xi_{bb}(\pm 1) = .1495$, $\xi_{bb}(\pm .444) = .3505$ for which $k^2 = .3$, with the

Bounded Variance design $\xi_{bv}(\pm 1) = .1290$, $\xi_{bv}(\pm .443) = .3710$ for which $c^2 = .003$, and with the design $\xi_m(\pm 1) = .1703$, $\xi_m(\pm .445) = .3297$ which is minimax when $\sigma^2 = .01n$. Note the similarity in the performances of these three designs and those of PR §2.1, optimal for estimation of θ_3 , and PR §4.3, which minimizes a mixture of the loss measures for θ_3 and $\theta_{(A)}$.

Table 2. Designs for Quadratic and Cubic Model Discrimination

Design	Efficiencies			max. bias	mse	
	θ_3	$\theta_{(B)}$	$\theta_{(A)}$		$\frac{\sigma^2}{n} = .01$	$\frac{\sigma^2}{n} = 1$
PR §2.1	1	.93	.75	.34	.097	4.30
PR §2.2	.85	1	.87	.50	.100	4.14
PR §2.3	0	0	1	∞	∞	∞
PR §3.1	.72	.94	.84	.40	.101	4.34
PR §3.2	.42	.89	.94	.59	.117	4.72
PR §4.1	.66	.98	.91	.57	.105	4.28
PR §4.2	.64	.96	.90	.55	.107	4.35
PR §4.3	1.00	.94	.75	.34	.097	4.29
PR §5.1	.50	.93	.94	.60	.112	4.51
PR §5.2	.50	.92	.93	.58	.112	4.55
ξ_{bb}	.97	.92	.74	.30	.096	4.36
ξ_{bv}	.96	.87	.70	.26	.097	4.53
ξ_m	.97	.95	.77	.34	.096	4.25

Acknowledgements

This paper has benefitted from the comments of a referee.

References

- Box, G.E.P. and Draper, N.R. (1959), "A Basis for the Selection of a Response Surface Design," *Journal of the American Statistical Association*, 54, 622-654.
- Harris, B. (1959), "Determining Bounds on Integrals with Applications to Cataloging Problems," *Annals of Mathematical Statistics*, 30, 521-548.
- Huber, P.J. (1975), "Robustness and Designs," in: *A Survey of Statistical Design and Linear Models*, ed. J.N. Srivastava, Amsterdam: North Holland, pp. 287-303.
- Huber, P.J. (1981), *Robust Statistics*, New York: Wiley.
- Li, K.C. and Notz, W. (1982), "Robust Designs for Nearly Linear Regression," *Journal of Statistical Planning and Inference*, 6, 135-151.

- Li, K.C. (1984), “Robust Regression Designs When the Design Space Consists of Finitely Many Points,” *Annals of Statistics*, 12, 269-282.
- Liu, S.X. (1994), “Some Topics in Robust Estimation and Experimental Design,” unpublished Ph.D. dissertation, University of Calgary, Department of Mathematics and Statistics.
- Liu, S.X. and Wiens, D.P. (1994), “Robust Designs for Approximately Polynomial Regression,” Technical Report 94.03, University of Alberta, Statistics Centre, Department of Mathematical Sciences. http://www.stat.ualberta.ca/people/schmu/stats_center/reports.html
- Marcus, M.B. and Sacks, J. (1976), “Robust Designs for Regression Problems,” in: *Statistical Theory and Related Topics II*, ed. S.S. Gupta and D.S. Moore, New York: Academic Press, pp. 245-268.
- Pesotchinsky, L. (1982), “Optimal Robust Designs: Linear Regression in R^k ,” *Annals of Statistics*, 10, 511-525.
- Pukelsheim, F. (1993), *Optimal Design of Experiments*, Toronto: John Wiley.
- Pukelsheim, F. and Rosenberger, J.L. (1993), “Experimental Designs for Model Discrimination,” *Journal of the American Statistical Association*, 88, 642-649.
- Rustagi, J.S. (1976). *Variational Methods in Statistics*, New York: Academic Press.
- Sacks, J. and Ylvisaker, D. (1978), “Linear Estimation for Approximately Linear Models,” *Annals of Statistics*, 6, 1122-1137.
- Stigler, S. (1971), “Optimal Experimental Design for Polynomial Regression,” *Journal of the American Statistical Association*, 66, 311-318.
- Studden, W.J. (1982), “Some Robust-Type D -Optimal Designs in Polynomial Regression,” *Journal of the American Statistical Association*, 77, 916-921.
- Wiens, D.P. (1990), “Robust Minimax Designs for Multiple Linear Regression,” *Linear Algebra Appl.* 127, 327-340.
- (1992), “Minimax Designs for Approximately Linear Regression,” *Journal of Statistical Planning and Inference*, 31, 353-371.
- (1993), “Designs for Approximately Linear Regression: Maximizing the Minimum Coverage Probability of Confidence Ellipsoids,” *Canadian Journal of Statistics*, 21, 59-70.