ELSEVIER

Journal of Statistical Planning and Inference 55 (1996) 95–106

journal of statistical planning and inference

Minimax regression designs for approximately linear models with autocorrelated errors

Douglas P. Wiens[a,*], Julie Zhou[b]

[a]*Department of Mathematical Sciences, University of Alberta, Edmonton, Alta., Canada T6G 2G1*
[b]*Department of Mathematical Sciences, Lakehead University, Thunder Bay, Ont., Canada P7B 5E1*

Received 30 March 1994; revised 18 July 1995

Abstract

We study the construction of regression designs, when the random errors are autocorrelated. Our model of dependence assumes that the spectral density $g(\omega)$ of the error process is of the form $g(\omega) = (1 - \alpha)g_0(\omega) + \alpha g_1(\omega)$, where $g_0(\omega)$ is uniform (corresponding to uncorrelated errors), $\alpha \in [0, 1)$ is fixed, and $g_1(\omega)$ is arbitrary. We consider regression responses which are exactly, or only approximately, linear in the parameters. Our main results are that *a design which is asymptotically (minimax) optimal for uncorrelated errors retains its optimality under autocorrelation if the design points are a random sample, or a random permutation, of points from this distribution.* Our results are then a partial extension of those of Wu (*Ann. Statist.* **9** (1981), 1168–1177), on the robustness of randomized experimental designs, to the field of regression design.

AMS classification: Primary 62K05, 62F35; secondary 62J05, 62M10

Keywords: Regression designs; Autocorrelated errors; Minimax robustness; Randomization; Regression spectrum

1. Introduction

In this paper we study optimal designs for regression models under certain departures from the classical assumptions. The usual formulation of the fixed-regressors linear regression model, which we write as

$$Y_i = z^{\mathrm{T}}(x_i)\theta + \varepsilon_i, \quad i = 1, \ldots, n;\ x_i \in S \subseteq \mathbb{R}^q,\ z(x_i) \in \mathbb{R}^p,\ \theta \in \mathbb{R}^p, \tag{1.1}$$

$$E[\varepsilon] = \mathbf{0}, \qquad \mathrm{COV}[\varepsilon] = \sigma^2 I, \tag{1.2}$$

* Corresponding author. Tel.: (403) 492–2052; fax.: (403) 492–2927; e-mail: wiens@robust.stat.ualberta.ca.

0378-3758/96/$15.00 © 1996—Elsevier Science B.V. All rights reserved
SSDI 0378-3758(95)00187-5

employs the following assumptions:

(I) The regression response $E[Y|\boldsymbol{x}_i] = \boldsymbol{z}^{\mathrm{T}}(\boldsymbol{x}_i)\boldsymbol{\theta}$ is exactly correct.

(II) The errors ε_i are uncorrelated.

The optimal design problem is then to choose values $\boldsymbol{x}_i$ of the independent variables in such a way as to minimize some scalar valued function of the covariance matrix of the ordinary least squares estimate $\hat{\boldsymbol{\theta}}$.

Beginning with Box and Draper (1959), numerous attempts have been made to relax assumption (I). A possible alternative assumes only that $E[Y|\boldsymbol{x}]$ is approximately linear:

$$E[Y|\boldsymbol{x}] \approx \boldsymbol{z}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{\theta}_0$$

where $\boldsymbol{\theta}_0$ minimizes

$$\int_S \{E[Y|\boldsymbol{x}] - \boldsymbol{z}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{\theta}\}^2 \,\mathrm{d}\boldsymbol{x}.$$

Then with

$$f(\boldsymbol{x}) := n^{1/2}\{E[Y|\boldsymbol{x}] - \boldsymbol{z}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{\theta}_0\},$$

the model (1.1) becomes instead

$$Y_i = \boldsymbol{z}^{\mathrm{T}}(\boldsymbol{x}_i)\boldsymbol{\theta}_0 + n^{-1/2} f(\boldsymbol{x}_i) + \varepsilon_i, \tag{1.3}$$

$$\int_S \boldsymbol{z}(\boldsymbol{x}) f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} = \boldsymbol{0}. \tag{1.4}$$

If $f(\boldsymbol{x})$ is non-zero then the regression estimates are typically biased. In order that the bias not dominate the variance, $f(\boldsymbol{x})$ is constrained by a condition such as

$$\int_S f^2(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \leqslant \eta^2 \tag{1.5}$$

for some constant η. The optimality problem is then based on the minimization of a function of the mean squared error matrix. See Huber (1975) and Wiens (1992, 1993) for details and special cases. For other approaches to the weakening of (I), see Marcus and Sacks (1976), Sacks and Ylvisaker (1978), Pesotchinsky (1982), Li and Notz (1982), Li (1984), Notz (1989) and Liu (1994).

The literature concerning optimal design theory under departures from (II) is somewhat more sparse. In a time series context several authors — Sacks and Ylvisaker (1966, 1968), Bickel and Herzberg (1979), Bickel et al. (1981) and Ylvisaker (1987) — have studied the problem of determining time points t_i at which to observe

$$Y(t) = \sum_j \theta_j z_j(t) + \varepsilon(t)$$

under various assumptions on the process $\{\varepsilon(t)\}$. Constantine (1989) obtained designs for a special case of (1.1) under a lag-one serial correlation model for the errors. See also Cox (1951) and Kiefer and Wynn (1981, 1983).

Here, we study designs both for (1.1) and for (1.3), with (1.2) replaced by

$$E[\boldsymbol{\varepsilon}] = \mathbf{0}, \qquad \mathrm{COV}[\boldsymbol{\varepsilon}] = \sigma^2 \boldsymbol{P}, \tag{1.6}$$

where $\mathbf{P}$ is a positive semi-definite Toeplitz matrix with unit diagonal, i.e. the autocorrelation matrix of a weakly stationary process. Thus, if

$$\rho(s) := E[\varepsilon_1 \varepsilon_{1+s}]/\sigma^2,$$

then $\boldsymbol{P}_{ij} = \rho(|i-j|)$. We assume throughout that

$$\sum_{s=-\infty}^{\infty} |\rho(s)| < \infty. \tag{1.7}$$

In Section 2 we obtain the asymptotic form of the covariance matrix of $\sqrt{n}\hat{\boldsymbol{\theta}}$, under (1.1) and (1.6). We then consider a broad class of departures from (1.2) under which

$$\boldsymbol{P} = (1-\alpha)\boldsymbol{I} + \alpha \boldsymbol{Q} \tag{1.8}$$

for an arbitrary (subject to (1.7)) autocorrelation matrix $\boldsymbol{Q}$ and fixed $\alpha \in [0, 1)$. For any scalar valued function L of covariance matrices, monotonic in that

$$\boldsymbol{V}_1 \leqslant \boldsymbol{V}_2 \text{ (w.r.t. positive definiteness)} \Rightarrow L(\boldsymbol{V}_1) \leqslant L(\boldsymbol{V}_2), \tag{1.9}$$

we consider the problem of choosing a design to minimize the maximum loss L, with the maximum evaluated subject to (1.8).

We show that this problem has the following asymptotic solution. Suppose that the corresponding optimal design problem under (1.2) has an asymptotic solution described by a particular design measure, i.e. a probability measure ξ_0 on the design space S. Then *a minimax strategy under* (1.8) *consists of randomly sampling design points from* ξ_0. Alternatively, *one may take a random permutation of design points whose empirical distribution function (e.d.f.) tends weakly to* ξ_0. The latter strategy typically amounts to designing the experiment for uncorrelated errors, and then randomizing the order of implementation.

We also show that if ξ_* is a minimax design for (1.3) and (1.2), where the maximum is evaluated over f satisfying (1.4) and (1.5), then ξ_* retains its optimality under (1.6)–(1.8) as well, if the design points are randomly sampled from ξ_* or if they constitute a random permutation whose e.d.f. tends weakly to ξ_*.

It has long been argued that the proper use of randomization in experimental design is a source of robustness against model inadequacies. Wu (1981) gave this notion a formal treatment and rigorous justification. The present work can be viewed as a partial extension to the field of regression design.

2. Derivations

We first list several assumptions which are made on the sequence of designs. Assume

(A1) $\lim_{n\to\infty}(1/n)\sum_{i=1}^{n} z(x_i) = \mathbf{0}$.

For $0 \leqslant s \leqslant n-1$ define matrices

$$\boldsymbol{B}_n(s) = \frac{1}{n}\sum_{i=1}^{n-s} z(x_i)z^{\mathrm{T}}(x_{i+s})$$

and define $\boldsymbol{B}_n(s) = \boldsymbol{B}_n^{\mathrm{T}}(-s)$ for $s < 0$. Assume

(A2) For each s, $\boldsymbol{B}_n(s)$ tends to a limit $\boldsymbol{B}(s)$ as $n \to \infty$, and $\boldsymbol{B}(0)$ is positive definite.

For the approximately linear model, i.e. if $\eta > 0$ in (1.5), we define

$$\boldsymbol{b}_n = \frac{1}{n}\sum_{i=1}^{n} z(x_i)f(x_i)$$

and assume

(A3) There exists $\lim_{n\to\infty}\boldsymbol{b}_n =: \boldsymbol{b}_f$.

Note that (A3) imposes no restriction on the exactly linear model ($\eta = 0$), since then $\boldsymbol{b}_n = \boldsymbol{b}_f = \mathbf{0}$.

By (1.7), there exists a symmetric spectral density $g(\omega)$ on $[-\pi, \pi]$ satisfying

$$\begin{aligned} \rho(s) &= \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}s\omega} g(\omega)\,\mathrm{d}\omega = \int_{-\pi}^{\pi} \cos(s\omega)g(\omega)\,\mathrm{d}\omega, \\ g(\omega) &= (2\pi)^{-1}\sum_{s=-\infty}^{\infty} \rho(s)\mathrm{e}^{-\mathrm{i}s\omega} = (2\pi)^{-1}\sum_{s=-\infty}^{\infty} \rho(s)\cos(s\omega). \end{aligned} \tag{2.1}$$

Under a set of assumptions implied by (A1) and (A2), it is proven in Grenander and Rosenblatt (1957) that there exists a regression spectral distribution function $\boldsymbol{H}(\omega)$ ($\omega \in [-\pi, \pi]$) — a symmetric $p \times p$ matrix whose increments $\boldsymbol{H}(\omega_2) - \boldsymbol{H}(\omega_1)$ ($\omega_1 < \omega_2$) are positive semi-definite — satisfying

$$\boldsymbol{B}(s) = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}s\omega}\,\mathrm{d}\boldsymbol{H}(\omega). \tag{2.2}$$

Lemma 2.1. *Define* $\boldsymbol{C}_n = \mathrm{COV}[\sqrt{n}\hat{\boldsymbol{\theta}}]$ *and* $\boldsymbol{M}_n = \mathrm{MSE}[\sqrt{n}\hat{\boldsymbol{\theta}}]$. *Under assumptions* (A1)–(A3) *there exist* $\lim_{n\to\infty}\boldsymbol{C}_n =: \boldsymbol{C}$ *and* $\lim_{n\to\infty}\boldsymbol{M}_n =: \boldsymbol{M}$. *The limit matrices have representations*

$$\boldsymbol{C} = \sigma^2\boldsymbol{B}^{-1}(0)\left[\sum_{s=-\infty}^{\infty} \rho(s)\boldsymbol{B}(s)\right]\boldsymbol{B}^{-1}(0), \tag{2.3}$$

$$\boldsymbol{C} = 2\pi\sigma^2\boldsymbol{B}^{-1}(0)\left[\int_{-\pi}^{\pi} g(-\omega)\,\mathrm{d}\boldsymbol{H}(\omega)\right]\boldsymbol{B}^{-1}(0), \tag{2.4}$$

$$\boldsymbol{M} = \boldsymbol{C} + \boldsymbol{B}^{-1}(0)\boldsymbol{b}_f\boldsymbol{b}_f^{\mathrm{T}}\boldsymbol{B}^{-1}(0). \tag{2.5}$$

Proof. Let $\boldsymbol{Z}$ be the $n \times p$ design matrix with ith row $\boldsymbol{z}^{\mathrm{T}}(\boldsymbol{x}_i)$. Let n be large enough so that $\boldsymbol{B}_n(0) = (1/n)\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}$ is non-singular. By standard regression theory,

$$\boldsymbol{C}_n = \sigma^2 \left(\frac{1}{n}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\right)^{-1} \left(\frac{1}{n}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{Z}\right) \left(\frac{1}{n}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\right)^{-1}$$

$$= \sigma^2 \boldsymbol{B}_n^{-1}(0) \left[\sum_{|s| \leqslant (n-1)} \rho(s)\boldsymbol{B}_n(s)\right] \boldsymbol{B}_n^{-1}(0). \tag{2.6}$$

Let $\|\cdot\|$ denote the Euclidean norm. Then

$$\|\boldsymbol{B}_n(s)\| \leqslant \operatorname{tr}(\boldsymbol{B}_n(0)) \leqslant 2 \operatorname{tr} \boldsymbol{B}(0) \tag{2.7}$$

for sufficiently large n. This, the Dominated Convergence Theorem and (1.7) applied to (2.6) yield (2.3). Now substitute (2.2) into (2.3), use (1.7) and the Dominated Convergence Theorem to interchange the summation and integration, then apply (2.1) to obtain (2.4). A simple calculation using (1.4) gives

$$\boldsymbol{M}_n = \boldsymbol{C}_n + \boldsymbol{B}_n^{-1}(0)\boldsymbol{b}_n\boldsymbol{b}_n^{\mathrm{T}}\boldsymbol{B}_n^{-1}(0),$$

whence (A2) and (A3) yield (2.5). □

Suppose now that the error process $\{\varepsilon_i\}$ is the sum of two uncorrelated processes $\{\varepsilon_i'\}$ and $\{\varepsilon_i''\}$, where $\{\varepsilon_i'\}$ is white noise with variance $(1-\alpha)\sigma^2$ for $\alpha \in [0, 1)$, and $\{\varepsilon_i''\}$ is weakly stationary with variance $\alpha\sigma^2$ and absolutely summable autocorrelation function $\rho_1(s)$. Then (1.6) and (1.8) hold, with $\boldsymbol{Q}_{ij} = \rho_1(|i-j|)$. Furthermore,

$$\rho(s) = (1-\alpha)\rho_0(s) + \alpha\rho_1(s), \quad \text{where } \rho_0(s) = \begin{cases} 1, & s = 0; \\ 0, & s \neq 0. \end{cases}$$

Denote by $g_1(\omega)$ the spectral density corresponding to $\rho_1(s)$, and let $g_0(\omega) = (2\pi)^{-1}\boldsymbol{I}(|\omega| \leqslant \pi)$. Then $\rho(s)$ has spectral density $g(\omega) = (1-\alpha)g_0(\omega) + \alpha g_1(\omega)$. See Bickel and Herzberg (1979) and Samarov (1987) for other instances of this model for $\{\varepsilon_i\}$. From Lemma 2.1,

$$\boldsymbol{C} = \sigma^2\{(1-\alpha)\boldsymbol{B}^{-1}(0) + \alpha\boldsymbol{B}^{-1}(0)\left[\sum_{s=-\infty}^{\infty} \rho_1(s)\boldsymbol{B}(s)\right]\boldsymbol{B}^{-1}(0)\}. \tag{2.8}$$

A consequence of the following lemma is that $\boldsymbol{C}$ is non-singular for all $\alpha \in [0, 1)$.

Lemma 2.2. *If $\rho_1(s)$ is absolutely summable then the matrix $\sum_{s=-\infty}^{\infty} \rho_1(s)\boldsymbol{B}(s)$ is positive semi-definite.*

Proof. In the same manner that (2.2) led to (2.4) we obtain

$$\sum_{s=-\infty}^{\infty} \rho_1(s)\boldsymbol{B}(s) = 2\pi \int_{-\pi}^{\pi} g_1(-\omega)\,\mathrm{d}\boldsymbol{H}(\omega),$$

so that for any $p \times 1$ vector $\boldsymbol{a}$,

$$\boldsymbol{a}^{\mathrm{T}}\left[\sum_{s=-\infty}^{\infty} \rho_1(s)\boldsymbol{B}(s)\right]\boldsymbol{a} = 2\pi \int_{-\pi}^{\pi} g_1(-\omega)\,\mathrm{d}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{H}(\omega)\boldsymbol{a}) \geqslant 0. \qquad \square$$

Grenander and Rosenblatt (1957) gave conditions under which the LSE $\hat{\boldsymbol{\theta}}$ is fully efficient. In particular, these hold if

$$\boldsymbol{H}(\omega) = (2\pi)^{-1}\omega \boldsymbol{B}(0) \tag{2.9}$$

in which case it is obvious from (2.2) that $\boldsymbol{B}(s) = \boldsymbol{0}$ for $s \neq 0$. If the design points $\boldsymbol{x}_i$ constitute a random sample from a particular design distribution, then (2.9) holds. In this case however the effect on $\mathrm{COV}[\sqrt{n}\hat{\boldsymbol{\theta}}]$ of the sampling variation must be taken into account. It is not clear that the conclusions of Lemma 2.1 continue to hold.

Lemma 2.3 below shows that the effect of this sampling variation is asymptotically negligible. For this lemma we assume that design points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are randomly sampled from a distribution function ξ on S. Define

$$\boldsymbol{B}(\xi) = \int_S z(\boldsymbol{x})z^{\mathrm{T}}(\boldsymbol{x})\,\mathrm{d}\xi(\boldsymbol{x}), \qquad \boldsymbol{b}_f(\xi) = \int_S z(\boldsymbol{x})f(\boldsymbol{x})\,\mathrm{d}\xi(\boldsymbol{x}).$$

Replace assumptions (A1)–(A3) by

(B1) The design points $\{\boldsymbol{x}_i\}_{i=1}^n$ are randomly chosen from ξ, and are uncorrelated with $\{\varepsilon_i\}_{i=1}^n$.

(B2) $E_\xi[z(\boldsymbol{x})] = \boldsymbol{0}$.

(B3) $\boldsymbol{B}(\xi)$ is positive definite.

(B4) The eigenvalues of $\boldsymbol{B}_n(0)$ are bounded above, and away from zero, as $n \to \infty$.

(B5) $f(\boldsymbol{x})$ is bounded on S.

Note that by (B1), (B2) and the Strong Law of Large Numbers,

$$\boldsymbol{B}_n(0) \to \boldsymbol{B}(\xi) \quad \text{(a.s.)}, \tag{2.10}$$

$$\boldsymbol{B}_n(s) \to \boldsymbol{0} \quad \text{(a.s.) for } s \neq 0, \tag{2.11}$$

$$\boldsymbol{b}_n \to \boldsymbol{b}_f(\xi) \quad \text{(a.s.)}. \tag{2.12}$$

Then (B3) and (2.10) imply that (B4) holds with probability one as $n \to \infty$.

Lemma 2.3. *Assume* (B1)–(B5). *Let* $\boldsymbol{C}_n$ *and* $\boldsymbol{M}_n$ *be as in Lemma* 2.1. *Then the conclusions of Lemma* 2.1 *hold, with* $\boldsymbol{B}(0) = \boldsymbol{B}(\xi)$, $\boldsymbol{b}_f = \boldsymbol{b}_f(\xi)$ *and* $\boldsymbol{H}(\omega)$ *as in* (2.9); *i.e.*

$$\boldsymbol{C}_n \to \boldsymbol{C}_\xi := \sigma^2\boldsymbol{B}^{-1}(\xi), \tag{2.13}$$

$$\boldsymbol{M}_n \to \boldsymbol{M}_{f,\xi} := \boldsymbol{C}_\xi + \boldsymbol{B}^{-1}(\xi)\boldsymbol{b}_f(\xi)\boldsymbol{b}_f^{\mathrm{T}}(\xi)\boldsymbol{B}^{-1}(\xi). \tag{2.14}$$

Proof. Define random matrices $\boldsymbol{D}_n$ and random vectors $\boldsymbol{d}_n$ by

$$\boldsymbol{D}_n = \boldsymbol{B}_n^{-1}(0)\left[\sum_{|s| \leqslant (n-1)} \rho(s)\boldsymbol{B}_n(s)\right]\boldsymbol{B}_n^{-1}(0), \qquad \boldsymbol{d}_n = \boldsymbol{B}_n^{-1}(0)\boldsymbol{b}_n.$$

By conditioning on $\{x_i\}_{i=1}^n$ and using the orthogonality asserted in (B1) we obtain

$$\boldsymbol{C}_n \doteq \sigma^2 E[\boldsymbol{D}_n], \qquad \boldsymbol{M}_n = \boldsymbol{C}_n + E[\boldsymbol{d}_n \boldsymbol{d}_n^{\mathrm{T}}]. \tag{2.15}$$

We claim

(i) $\boldsymbol{D}_n \to \boldsymbol{B}^{-1}(\xi)$ (a.s.),

(ii) $\boldsymbol{d}_n \to \boldsymbol{B}^{-1}(\xi)\boldsymbol{b}_f(\xi)$ (a.s.),

(iii) $\|\boldsymbol{D}_n\|, \|\boldsymbol{d}_n\|$ are bounded.

By (iii) we may take limits inside the expectation signs in (2.15); applying (i) and (ii) will then yield (2.13) and (2.14).

For (i), let $\delta > 0$ be arbitrary and let N be large enough that $\sum_{s \geqslant N} |\rho(s)| < \delta$. Denote by $\lambda_{1,n}$ and $\lambda_{p,n}$ the largest and smallest eigenvalues of $\boldsymbol{B}_n(0)$. Then for $n > N$ we have, using (2.7),

$$\begin{aligned} \|\boldsymbol{D}_n - \boldsymbol{B}_n^{-1}(0)\| &\leqslant 2 \sum_{1 \leqslant s \leqslant N} |\rho(s)| \, \|\boldsymbol{B}_n^{-1}(0)\boldsymbol{B}_n(s)\boldsymbol{B}_n^{-1}(0)\| \\ &\quad + 2 \sum_{N < s < n} |\rho(s)| \, \|\boldsymbol{B}_n^{-1}(0)\|^2 \operatorname{tr}(\boldsymbol{B}_n(0)). \end{aligned} \tag{2.16}$$

The first sum above tends to 0 (a.s.) by (2.10) and (2.11). The second is bounded by

$$2p^3(\lambda_{1,n}/\lambda_{p,n}^2) \sum_{N < s < n} |\rho(s)| < K\delta$$

for some constant K independent of n, by (B4). Letting first n, then N tend to ∞ in (2.16) gives (i).

Claim (ii) above is an immediate consequence of (2.10) and (2.12), using (B3). For (iii), note that

$$\|\boldsymbol{D}_n\| \leqslant \|\boldsymbol{B}_n^{-1}(0)\| \sum_{s=-\infty}^{\infty} |\rho(s)| \leqslant \{p^{1/2}\lambda_{p,n}\} \sum_{s=-\infty}^{\infty} |\rho(s)|,$$

which is bounded by (B4). Also

$$\|\boldsymbol{d}_n\| \leqslant \|\boldsymbol{B}_n^{-1}(0)\| (\operatorname{tr} \boldsymbol{B}_n(0))^{1/2} \left(\frac{1}{n} \sum_{i=1}^{n} f^2(x_i) \right)^{1/2},$$

which is bounded as above, using also (B5). □

Remark. Lemma 2.3 continues to hold with (B1) replaced by

(B1′) The design points are a random permutation $\{x_{\pi(i)}\}_{i=1}^n$ of $\{x_i\}_{i=1}^n$, where the e.d.f. ξ_n of the $\{x_i\}$ tends weakly to ξ.

To see this, note that the proof of Lemma 2.3 goes through essentially unchanged under (B1′), with (2.11) replaced by

$$E_\pi[\boldsymbol{B}_n(s)] \to \boldsymbol{0} \quad \text{for } s \neq 0.$$

This follows from

$$E_\pi[\boldsymbol{B}_n(s)] = \frac{1}{n}\sum_{i=1}^{n-s} E_\pi[z_{\pi(i)}z_{\pi(i+s)}^{\mathrm{T}}]$$

$$= \frac{n-s}{n-1} E_{\xi_n}[z(x)]E_{\xi_n}[z^{\mathrm{T}}(x)] - \frac{n-s}{n(n-1)}\boldsymbol{B}_n(0),$$

where the second equality is obtained by first conditioning on $\pi(i)$.

Now consider the problem of choosing a design measure ξ to minimize the supremum — over all absolutely summable $\rho_1(s)$ — value of $L(\boldsymbol{C}) = L(\boldsymbol{C}; \rho_1, \xi)$, where L is monotonic in the sense of (1.9). Suppose that ξ_0 minimizes $L(\boldsymbol{C}; \rho_0, \xi) = L(\sigma^2\boldsymbol{B}^{-1}(\xi))$, i.e. is an optimal design under (1.2). We then have

Theorem 2.4. *An asymptotically minimax design in the exactly linear model* (1.1) *is attained by randomly sampling design points from* ξ_0, *if this design measure and the sampling mechanism satisfy* (B1) *or* (B1′) *and* (B2)–(B4).

Proof. For any design measure ξ, we have

$$\sup L(\boldsymbol{C}; \rho_1, \xi) \geqslant L(\boldsymbol{C}; \rho_0, \xi) = L(\sigma^2\boldsymbol{B}^{-1}(\xi)) \geqslant L(\sigma^2\boldsymbol{B}^{-1}(\xi_0)) = \sup L(\boldsymbol{C}; \rho_1, \xi_0),$$

where the last equality follows from the randomization. □

For the approximately linear model (1.3), denote the loss $L(\boldsymbol{M})$ by $L(\boldsymbol{M}; \rho_1, f; \xi)$ and consider the problem of minimizing the supremum, over all absolutely summable $\rho_1(s)$ and all f satisfying (1.4) and (1.5), of $L(\boldsymbol{M}; \rho_1, f; \xi)$. Suppose that ξ_* minimizes the supremum, over f, of $L(\boldsymbol{M}; \rho_0, f; \xi) = L(\boldsymbol{M}_{f,\xi})$, i.e. is a minimax design for the approximately linear model under (1.2).

Theorem 2.5. *An asymptotically minimax design in the approximately linear model* (1.3) *is attained by randomly sampling design points from the design measure* ξ_*, *provided that this design measure and the sampling mechanism satisfy* (B1) *or* (B1′) *and* (B2)–(B4), *and that* $\sup_f L(\boldsymbol{M}; \rho_0, f; \xi_*)$ *is attained at a bounded* f.

The proof of Theorem 2.5 is very similar to that of Theorem 2.4 and so is omitted Note that (B5) is required to hold only when sampling from ξ_* — only the least favourable f need be bounded. This is typically satisfied — see the examples in Huber (1975) and Wiens (1992, 1993).

Remark. Assumptions (A1) and (B2) each preclude fitting a response with an intercept. For an intercept model however, one may write the design matrix in partitioned form as $[\mathbf{1}|\boldsymbol{Z}]$, and the parameter vector as $(\theta_0, \boldsymbol{\theta}_1^{\mathrm{T}})$. Then if this $\boldsymbol{Z}$ satisfies the above

assumptions our optimality results continue to hold for the estimation of θ_1, since the columns of $\boldsymbol{Z}$ are orthogonal to $\mathbf{1}$.

3. Examples

Example 1. We consider the approximately linear regression model with $z^{\mathrm{T}}(x) = (1, x)$ and $\boldsymbol{\theta}_0 = (\theta_0, \theta_1)^{\mathrm{T}}$ in (1.3). Without loss of generality, the design space S is taken as $[-0.5, 0.5]$. The optimality criterion is minimax MSE of $\hat{\theta}_1$:

$$\min_{\xi} \max_{f,\rho_1} E(\hat{\theta}_1 - \theta_1)^2.$$

From Huber (1975), the minimax design for the approximately linear model with uncorrelated errors has density function

$$m_0(x) = \frac{1}{(1-2a)^2}\left(1 - \frac{a^2}{x^2}\right)^+, \quad 0 \leqslant a < 1, \quad -0.5 \leqslant x \leqslant 0.5,$$

where a depends on the ratio σ^2/η^2. When $(\sigma^2/\eta^2) \to 0$, $a \to 0$ and $m_0(x) \to 1$. Then from Theorem 2.5, an asymptotically minimax design for the correlated error model is attained by randomly sampling design points from $m_0(x)$. The design points may be chosen as follows. Let M_0 be the distribution function corresponding to m_0. Select n points $M_0^{-1}((i-0.5)/n)$, $i = 1, \ldots, n$, whose empirical distribution function tends weakly to M_0. Now take a random permutation of these points. A set of design points chosen in such a manner for $n = 16$ when $a = 0$ is $\langle 0.218, 0.094, -0.281, -0.156, -0.406, 0.469, -0.094, 0.344, 0.031, -0.344, -0.031, 0.219, 0.406, -0.219, 0.156, -0.469\rangle$.

Example 2. We consider the approximately linear regression model with $z^{\mathrm{T}}(x) = (1, x_1, x_2)$ and $\boldsymbol{\theta}_0 = (\theta_0, \boldsymbol{\theta}_1)^{\mathrm{T}}$ in (1.3). The design space S is the sphere of unit area: $\{x \mid \|x\| \leqslant r = 1/\sqrt{\pi}\}$. The loss function is the determinant of the MSE matrix of $\hat{\boldsymbol{\theta}}_1$, so that the optimality criterion is

$$\min_{\xi} \max_{f,\rho_1} \det(E[(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1)^{\mathrm{T}}]).$$

From Wiens (1992), the minimax design for uncorrelated errors is the spherically symmetric design ξ_* in which $\|x\|$ has density function

$$g_0(u) = 2\pi a u\left(1 - \frac{b}{\pi u^2}\right)^+, \quad a > 0, \ 0 \leqslant b < 1, \ 0 \leqslant u \leqslant r,$$

where a and b depend on the ratio $\sigma^2/n\eta^2$. By spherical symmetry, $x/\|x\|$ is uniformly distributed over the boundary of the unit circle, independently of $\|x\|$. Then from Theorem 2.5, an asymptotically minimax design for the correlated error model is

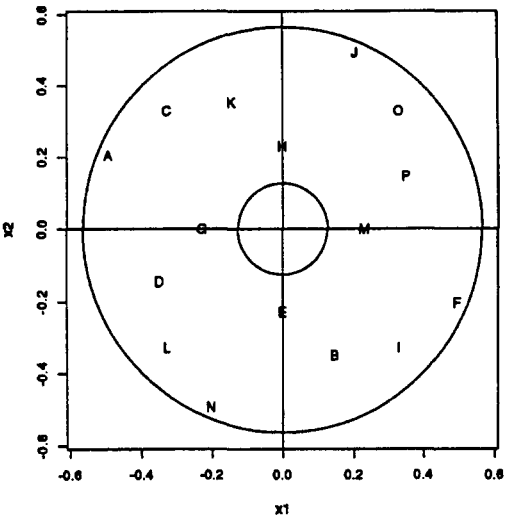

Fig. 1. Design points for Example 2. Order of implementation is alphabetical.

attained by randomly sampling design points from ξ_*. As in Example 1, we take a random permutation of design points whose empirical distribution function tends weakly to ξ_*. This may be done as follows.

Define a probably distribution by

$$p_0 = 1 - \frac{[\sqrt{n}]^2}{n}, \qquad p_i = \frac{[\sqrt{n}]}{n}, \quad i = 1, 2, \ldots, [\sqrt{n}].$$

Define

$$a_{-1} = 0, \qquad a_i = G_0^{-1}\left(\sum_{j=0}^{i} p_j\right), \quad i = 0, 1, \ldots, [\sqrt{n}],$$

where G_0 is the distribution function corresponding to the optimal density function g_0. Divide the design space $S = \{\boldsymbol{x} \mid \|\boldsymbol{x}\| \leqslant r\}$ into $[\sqrt{n}] + 1$ annuli $A_0, A_1, \ldots, A_{[\sqrt{n}]}$ with

$$A_i = \{\boldsymbol{x} \mid \|\boldsymbol{x}\| \in (a_{i-1}, a_i]\},$$

so that

$$P_{G_0}(\boldsymbol{x} \in A_i) = p_i.$$

In each A_i we select np_i points equally spaced over $\|\boldsymbol{x}\| = (a_{i-1} + a_i)/2$. We then have $n - [\sqrt{n}]^2$ points in A_0 and $[\sqrt{n}]$ points in each A_i for $i = 1, \ldots, [\sqrt{n}]$. It is easy to verify that the empirical distribution of the design points tends weakly to ξ_*.

Fig. 1 gives a random permutation of $n = 16$ points chosen in this manner when $\sigma^2/n\eta^2 = 0.01$, $g_0(u) = 7.8477(u - 0.0159/u)^+$, $0.126 \leqslant u \leqslant 0.564$.

Remark. The implementation of the design of Example 2 requires that the model be transformed in such a way that a spherical design space becomes appropriate. In practice, a transformation to a square design space may be less problematical. For robust minimax designs in the approximately linear model with $z^T(x) = (1, x_1, x_2, x_1x_2)$ and $S = [-0.5, 0.5] \times [-0.5, 0.5]$, see Wiens (1990).

Acknowledgements

Research of both the authors was supported by the Natural Sciences and Engineering Research Council of Canada.

References

Bickel, P.J. and A.M. Herzberg (1979). Robustness of design against autocorrelation in time I: asymptotic theory, optimality for location and linear regression. *Ann. Statist.* **7**, 77–95.

Bickel, P.J., A.M. Herzberg and M.F. Schilling (1981). Robustness of design against autocorrelation in time II: optimality, theoretical and numerical results for the first-order autoregressive process. *J. Amer. Statist. Assoc.* **76**, 870–877.

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

Constantine, G.M. (1989). Robust designs for serially correlated observations. *Biometrika* **76**, 245–251.

Cox, D.R. (1951). Some systematic experimental designs. *Biometrika* **38**, 312–323.

Grenander, U. and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series.* Wiley, New York.

Huber, P.J. (1975). Robustness and designs. In: J.N. Srivastava, Ed., *A Survey of Statistical Design and Linear Models.* North-Holland, Amsterdam, 287–303.

Kiefer, J. and H. Wynn (1981). Optimum balanced block and latin square designs for correlated observations. *Ann. Statist.* **9**, 737–757.

Kiefer, J. and H. Wynn (1983). Autocorrelation-robust design of experiments. In: T. Leonard and C.-F. Wu, Ed., *Scientific Inference, Data Analysis and Robustness.* Academic Press, New York, 279–299.

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

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

Liu, S.X. (1994). Some topics in robust estimation and experimental design, Ph.D. thesis, University of Calgary.

Marcus, M.B. and J. Sacks (1976). Robust designs for regression problems. In: S.S. Gupta and D.S. Moore, Eds., *Statistical Decision Theory and Related Topics II.* Academic Press, New York, 245–268.

Notz, W. (1989). Optimal designs for regression models with possible bias. *J. Statist. Plann. Inference* **22**, 43–54.

Pesotchinsky, L. (1982). Optimal robust designs: linear regression in $\mathbb{R}^k$. *Ann. Statist.* **10**, 511–525.

Sacks, J. and D. Ylvisaker (1966). Designs for regression problems with correlated errors. *Ann. Math. Statist.* **37**, 66–89.

Sacks, J. and D. Ylvisaker (1968). Designs for regression problems with correlated errors; many parameters. *Ann. Math. Statist.* **39**, 49–69.

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

Samarov, A.M. (1987). Robust spectral regression. *Ann. Statist.* **15**, 99–111.

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

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

Wiens, D.P. (1993). Designs for approximately linear regression: maximizing the minimum coverage probability of confidence ellipsoids. *Canad J. Statist.* **21**, 59–70.

Wu, C.-F. (1981). On the robustness and efficiency of some randomized designs. *Ann. Statist.* **9**, 1168–1177.

Ylvisaker, D. (1987). Prediction and design. *Ann. Statist.* **15**, 1–19.