Robust Designs Based on the Infinitesimal Approach

Douglas P. WIENS and Julie ZHOU

We introduce an *infinitesimal* approach to the construction of robust designs for linear models. The resulting designs are robust against small departures from the assumed linear regression response and/or small departures from the assumption of uncorrelated errors. Subject to satisfying a robustness constraint, they minimize the determinant of the mean squared error matrix of the least squares estimator at the ideal model. The robustness constraint is quantified in terms of boundedness of the Gateaux derivative of this determinant, in the direction of a contaminating response function or autocorrelation structure. Specific examples are considered. If the aforementioned bounds are sufficiently large, then (permutations of) the classically optimal designs, which minimize variance alone at the ideal model, meet our robustness criteria. Otherwise, new designs are obtained.

KEY WORDS: Approximately linear regression response; B robustness; M robustness; Regression design; V robustness.

1. INTRODUCTION

In this article we introduce an *infinitesimal* approach to the construction of robust designs for linear models. The resulting designs are robust against small departures from the assumed linear regression response and/or small departures from the assumption of uncorrelated errors. Subject to satisfying a robustness constraint, they minimize the determinant of the mean squared error (MSE) matrix of the estimator at the ideal model.

Specifically, suppose that an experimenter is to observe a random variable Y at locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in a q-dimensional design space $\mathcal{S}$. The response $E[Y|\mathbf{x}]$ is thought to be approximately linear in the elements of a p-dimensional vector $\mathbf{z}(\mathbf{x})$ of regressors and is observed subject to additive and possibly autocorrelated errors. With $f(\mathbf{x}) = E[Y|\mathbf{x}] - \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$, the observations then satisfy

$$Y_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta} + f(\mathbf{x}_i) + \varepsilon_i, \qquad i = 1, \ldots, n. \tag{1}$$

The ε_i are zero-mean random errors with covariance matrix $\sigma^2\mathbf{P}$ for some autocorrelation matrix $\mathbf{P}$. The disturbance function f is assumed to be constrained in a manner that ensures that the regression parameters are well defined; see Section 4.

The experimenter, knowing neither f nor $\mathbf{P}$, intends to compute the classical least squares estimate $\hat{\boldsymbol{\theta}}$. He thus will incur possible errors due to bias, as well as a possible loss in efficiency. He wishes to use a design for which the size, as measured by the determinant, of the MSE matrix of $\hat{\boldsymbol{\theta}}$ remains within reasonable bounds.

Versions of this problem, with respect only to variations from the fitted response function, have been investigated by, among others, Box and Draper (1959), Huber (1975), Kiefer and Wynn (1984), Li and Notz (1982), Liu (1994), Liu and Wiens (1994), Pesotchinsky (1982), and Wiens (1991, 1992). The second type of model departure—autocorrelated errors—has been investigated by Bickel and Herzberg (1979), Bischoff (1992, 1993), Constantine (1989), Herzberg and Huda (1981), Jenkins and Chanmugam (1962), Judickii (1976), Sacks and Ylvisaker (1966, 1968, 1970), and Wiens and Zhou (1995, 1996).

A possible approach to these problems is to seek robustness against *infinitesimal* departures from the model, in the sense of bounding the maximum Gateaux derivative, evaluated at the ideal model ($f = 0, \mathbf{P} = \mathbf{I}$), of the determinant of the MSE matrix of $\hat{\boldsymbol{\theta}}$. The derivatives are taken in the directions of arbitrary contaminating functions f or autocorrelation matrices $\mathbf{P}$. The boundedness ensures that the MSE remains close to its value at the ideal model, in sufficiently small neighborhoods. Subject to boundedness, we choose the design to minimize the MSE at the ideal model, thus imposing a requirement of efficiency in addition to that of robustness.

Our designs attain their optimality through the placement of the design points and through the order in which these points are implemented. As an example of this latter factor, in a quality control framework observations at similar machine settings (x) may be positively (negatively) correlated, with the correlations decreasing in magnitude over time. It turns out that for MA(1) errors, an appropriate ordering of the design points maximizes (minimizes) the number of sign changes among successive values of $x - \bar{x}$.

The infinitesimal robustness of an estimator can be studied by means of the influence function of the estimator and by the change-of-bias function (CBF) and the change-of-variance function (CVF) (see Hampel, Ronchetti, Rousseeuw, and Stahel 1986 and Hössjer 1991). The CBF and CVF quantify the local stability of the bias and variance. Extensions of these stability notions to design theory are given in Section 2, where we also introduce corresponding optimality criteria. Explicit optimal designs are given in Sections 3, 4, and 5. Various designs are compared in Section 6, and an example from the chemical engineering literature is discussed in Section 7. Proofs are given in the Appendix.

Douglas P. Wiens is Professor and Director, Statistics Centre, Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1. Julie Zhou is Assistant Professor, Department of Mathematical Sciences, Lakehead University, Thunder Bay, Ontario, Canada P7B 5E1. The research of both authors is supported by the Natural Sciences and Engineering Research Council of Canada.

Journal of the American Statistical Association
December 1997, Vol. 92, No. 440, Theory and Methods

2. DESIGN SENSITIVITY

Let ξ be the design measure (i.e., the empirical distribution function of the design points) and define

$$\mathbf{b}_{f,\xi} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i)$$

and

$$\mathbf{B}_\xi(m) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n-m} \mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_{i+m}), & 0 \le m \le n-1, \\ \mathbf{B}_\xi^T(-m), & -(n-1) \le m < 0. \end{cases}$$

Denote the $n \times p$ model matrix with rows $\mathbf{z}^T(\mathbf{x}_i)$ by $\mathbf{Z}$. Then the determinant of the MSE matrix of $n^{1/2}\hat{\boldsymbol{\theta}}$ is

$$\mathcal{D}(f, \xi, \mathbf{P}) = \sigma^{2p}|\mathbf{B}_\xi(0)|^{-2}\left|\frac{\mathbf{Z}^T\mathbf{P}\mathbf{Z}}{n}\right| \times \left(1 + \frac{n}{\sigma^2}\mathbf{b}_{f,\xi}^T\left(\frac{\mathbf{Z}^T\mathbf{P}\mathbf{Z}}{n}\right)^{-1}\mathbf{b}_{f,\xi}\right).$$

Let $\mathcal{P}$ and $\mathcal{F}$ be convex classes of autocorrelation matrices $\mathbf{P}$ and disturbance functions f, containing $\mathbf{P}_0 = \mathbf{I}$ and $f_0 = 0$. We define the CVF for ξ at $\mathbf{P}_0$, in the direction $\mathbf{P} \in \mathcal{P}$, by

$$\mathrm{CVF}(\xi, \mathbf{P}) = \frac{\frac{d}{dt}\mathcal{D}(f_0, \xi, (1-t)\mathbf{P}_0 + t\mathbf{P})_{|t=0}}{\mathcal{D}(f_0, \xi, \mathbf{P}_0)}, \qquad (2)$$

and the CBF for ξ at f_0, in the direction $f \in \mathcal{F}$, by

$$\mathrm{CBF}(\xi, f) = \frac{\frac{1}{2}\frac{d^2}{dt^2}\mathcal{D}((1-t)f_0 + tf, \xi, \mathbf{P}_0)_{|t=0}}{\sigma^{-2}\mathcal{D}(f_0, \xi, \mathbf{P}_0)}. \qquad (3)$$

The division by $\mathcal{D}(f_0, \xi, \mathbf{P}_0) = \sigma^{2p}|\mathbf{B}_\xi(0)|^{-1}$ and $\sigma^{-2}\mathcal{D}(f_0, \xi, \mathbf{P}_0)$ is for scale and affine invariance. For the CVF, it corresponds to basing the loss on $\ln \mathcal{D}$. The use of the second derivative in (3) is motivated by the observation that the quantity being differentiated is a linear function of t^2.

The global behavior of the designs with respect to departures in $\mathcal{P}$ and $\mathcal{F}$ will be quantified partly through the *change-of-variance sensitivity* (CVS) and the *change-of-bias sensitivity* (CBS), defined to be the suprema of CVF and CBF over $\mathcal{P}$ and $\mathcal{F}$. On representing $\mathbf{P}_{i,j}$ as $\rho(|i-j|)$ for some autocorrelation function $\rho(\cdot)$, straightforward calculations give

$$\mathrm{CVS}(\xi, \mathcal{P}) = \sup_{\mathbf{P}\in\mathcal{P}}\left\{\mathrm{trace}\left(\frac{\mathbf{Z}^T(\mathbf{P}-\mathbf{I})\mathbf{Z}}{n}\mathbf{B}_\xi^{-1}(0)\right)\right\} \qquad (4)$$

$$= \sup_{\mathbf{P}\in\mathcal{P}}\left\{\sum_{0\le|s|\le n-1} \rho(s)\mathrm{trace}(\mathbf{B}_\xi(s)\mathbf{B}_\xi^{-1}(0))\right\} \qquad (5)$$

and

$$\mathrm{CBS}(\xi, \mathcal{P}) = \sup_{f\in\mathcal{F}}\{n\mathbf{b}_{f,\xi}^T\mathbf{B}_\xi^{-1}(0)\mathbf{b}_{f,\xi}\}. \qquad (6)$$

For given α, we say that a design ξ is *V robust* if it minimizes $\mathcal{D}(f_0, \xi, \mathbf{P}_0)$; that is, maximizes $|\mathbf{B}_\xi(0)|$ subject to the constraint

$$\mathrm{CVS}(\xi, \mathcal{P}) \le \alpha, \qquad (7)$$

and is *most V robust* if α is the infimum of the CVS over a given class of designs. In Section 3 we construct V-robust and most V-robust designs for two classes $\mathcal{P}$.

For given β, we say that a design ξ is *B robust* if it maximizes $|\mathbf{B}_\xi(0)|$ subject to the constraint

$$\mathrm{CBS}(\xi, \mathcal{P}) \le \beta, \qquad (8)$$

and *most B robust* if β is the infimum of the CBS over a given class of designs. In Section 4 we note that B-robust designs coincide with the bounded bias designs of Liu (1994) and Liu and Wiens (1994). Examples are given for two classes $\mathcal{F}$.

We say that a design is *M robust* if it maximizes $|\mathbf{B}_\xi(0)|$ subject to both (7) and (8), and *most M robust* if it is both most V robust and most B robust. We consider M-robust designs in Section 5. It is not known (to us) whether most M-robust designs exist.

3. V-ROBUST DESIGNS

In this section we obtain V-robust and most V-robust designs for the classes

$$\mathcal{P}_1 = \{\mathbf{P}|\rho(s) = 0 \quad \text{for} \quad |s| \ge 2 \quad \text{and} \quad c_0 \le \rho(1) < 1 \quad \text{with} \quad c_0 > 0\}$$

and

$$\mathcal{P}_2 = \{\mathbf{P}|\rho(s) = 0 \quad \text{for} \quad |s| \ge 2 \quad \text{and} \quad -1 < \rho(1) \le -c_1 \quad \text{with} \quad c_1 > 0\}.$$

These classes correspond to MA(1) processes with positive and negative lag-1 correlations bounded away from 0. The V-robust designs presented here do not depend on the values of c_0 and c_1. We consider the multiple linear regression model—that is, $\mathbf{z}^T(\mathbf{x}) = (1, \mathbf{x}_{q\times 1}^T)$ in (1)—and restrict to the class $\Xi_{n,q}$ of n-point designs for which $\mathbf{B}_\xi(0)$ is a diagonal matrix. The latter point can of course be justified on practical grounds. Furthermore, because of affine invariance there is little loss of generality, if $\mathcal{S}$ is rotationally invariant. Then the only additional restriction imposed is the requirement that the design points be centered with respect to each coordinate axis.

We find, using (5), that

$$\mathrm{CVS}(\xi, \mathcal{P}) = \sup_{\rho(1)} 2\rho(1)\left(\frac{n-1}{n} + \sum_{j=1}^{q}\mathcal{Q}(\mathbf{x}_{(j)})\right),$$

where $\mathbf{x}_{(j)}$ is the $j+1$th column of $\mathbf{Z}$,

$$\mathcal{Q}(\mathbf{x}) := \frac{\sum_{i=1}^{n-1} x_i x_{i+1}}{\sum_{i=1}^{n} x_i^2} = \frac{\mathbf{x}^T\mathbf{Q}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}, \qquad (9)$$

and $\mathbf{Q}$ is the tridiagonal matrix with (i,j)th element $q_{ij} = \frac{1}{2}I(|i-j| = 1)$.

Theorem 1 below shows that if α is sufficiently large then (7) imposes no restriction. It is stated for $q = 1$ but the extensions to $q > 1$ are rather evident.

Theorem 1. Let $q = 1$ and $\mathcal{S} = [-1, 1]$. Define

$$\alpha_{1,n} = \begin{cases} 0, & n \text{ even} \\ \frac{2}{n(n-1)}, & n \text{ odd} \end{cases}$$

and

$$\alpha_{2,n} = \begin{cases} -4c_1 \frac{n-2}{n}, & n \text{ even} \\ -2c_1 \frac{2n^2-5n+1}{n(n-1)}, & n \text{ odd} \end{cases}$$

a. For $\alpha \geq \alpha_{1,n}$, V-robust designs for $\mathcal{P}_1$ are

$$\mathbf{x}_{(1)} = \begin{cases} \langle 1, -1, 1, -1, \ldots, 1, -1 \rangle, & n \text{ even} \\ \langle 1, -1, 1, -1, \ldots, 1, -1, 0 \rangle, & n \text{ odd}, \end{cases}$$

with $\text{CVS}(\xi, \mathcal{P}_1) = \alpha_{1,n}$.

b. For $\alpha \geq \alpha_{2,n}$, V-robust designs for $\mathcal{P}_2$ are

$$\mathbf{x}_{(1)} = \begin{cases} \langle \underbrace{1, \ldots, 1}_{n/2}, \underbrace{-1, \ldots, -1}_{n/2} \rangle, & n \text{ even} \\ \langle \underbrace{1, \ldots, 1}_{(n-1)/2}, 0, \underbrace{-1, \ldots, -1}_{(n-1)/2} \rangle, & n \text{ odd}, \end{cases}$$

with $\text{CVS}(\xi, \mathcal{P}_2) = \alpha_{2,n}$.

c. In each case, the design minimizes $\text{CVS}(\xi, \mathcal{P})$ among those designs in $\Xi_{n,1}$ that maximize $|\mathbf{B}_\xi(0)|$.

Remark 1. The designs in Theorem 1 were, for n even, given by Jenkins and Chanmugam (1962) and Constantine (1989). Jenkins and Chanmugam (1962) minimized the variance of the slope of the least squares estimator (LSE) over a subset of the set of all possible designs carried out at two levels only. The subset consists of all discrete "square wave" designs in which a run of m experiments at one level alternates with a run of m experiments at the other. Then the design problem becomes that of choosing the block size m. Constantine (1989) attempted to maximize the trace of the covariance matrix of the best linear unbiased estimator (BLUE) over $\Xi_{n,1}$. A linear approximation to $\mathbf{Z}^T\mathbf{P}^{-1}\mathbf{Z}$ was used to derive the optimal design.

To get the most V-robust designs for multiple regression, it is necessary to evaluate the extrema of $\sum_{j=1}^q \mathcal{Q}(\mathbf{x}_{(j)})$ over $\Xi_{n,q}$. This is carried out in the Appendix and yields Theorem 2. The statement of this result requires some definitions.

Define constants and $n \times 1$ vectors $(\mu_j, \mathbf{r}_j), 1 \leq j \leq n$, and $(\nu_j, \mathbf{s}_j), 1 \leq j \leq [(n-1)/2]$, with the μ_j and ν_j ordered from largest to smallest, by

$$\mu_j = \cos\frac{j\pi}{n+1}, \qquad (\mathbf{r}_j)_k = \sqrt{\frac{2}{n+1}} \sin\frac{kj\pi}{n+1} \tag{10}$$

and

$$\nu_j = \cos\phi_j, \qquad (\mathbf{s}_j)_k = \sqrt{\frac{2(n+1)}{n(n+2)}} \times \cot\frac{n+1}{2}\phi_j \left(1 - \frac{\cos\left(k - \frac{n+1}{2}\right)\phi_j}{\cos\frac{n+1}{2}\phi_j}\right). \tag{11}$$

Here ϕ_j is the solution, in $(2j\pi/(n+1), (2j+1)\pi/(n+1))$, to the equation

$$\tan\frac{n+1}{2}\phi - (n+1)\tan\frac{\phi}{2} = 0. \tag{12}$$

Place the $\mathbf{r}_{2j}$ and the $\mathbf{s}_j$ into a matrix $\mathbf{X}$, and define a corresponding sequence $\{\lambda_j\}$ by

$$\mathbf{X}_{n\times(n-1)} = \|\mathbf{r}_2, \mathbf{s}_1, \mathbf{r}_4, \mathbf{s}_2, \ldots, \mathbf{r}_{2[(n-1)/2]}, \mathbf{s}_{[(n-1)/2]}, \mathbf{r}_n \text{ (if } n \text{ is even)}\| \tag{13}$$

and

$$\{\lambda_j\}_{j=1}^{n-1} = \langle \mu_2, \nu_1, \mu_4, \nu_2, \ldots, \mu_{2[(n-1)/2]}, \nu_{[(n-1)/2]}, \mu_n \text{ (if } n \text{ is even)} \rangle. \tag{14}$$

Theorem 2. The most V-robust designs in $\Xi_{n,q}$ for $\mathcal{P}_k, k = 1, 2$, have model matrices $\mathbf{Z} = (\mathbf{1} \vdots \mathbf{X}_{(q;k)}\mathbf{D}_k)$, where $\mathbf{X}_{(q;k)}$ consist of the last ($k = 1$) or first ($k = 2$)q columns of $\mathbf{X}$ and $\mathbf{D}_k$ is a diagonal matrix chosen to have maximum determinant, subject to the constraint that the rows of $\mathbf{X}_{(q;k)}\mathbf{D}_k$ belong to $\mathcal{S}$. The corresponding covariance matrices of $\hat{\theta}$ at $\mathbf{P}_0 = \mathbf{I}$ are $\sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1} = \sigma^2(n^{-1} \oplus \mathbf{D}_k^{-2})$. The CVS are

$$\text{CVS}(\xi, \mathcal{P}_k) = \begin{cases} 2\left(\frac{n-1}{n} + \sum_{j=1}^q \lambda_{n-j}\right), & k = 1 \\ -2c_1\left(\frac{n-1}{n} + \sum_{j=1}^q \lambda_j\right), & k = 2. \end{cases}$$

Remarks. 2. If $\mathcal{S}$ is a q-dimensional rectangle $[-c_1, c_1] \times \cdots \times [-c_q, c_q]$, then we find that the maximizing matrix $\mathbf{D}_k$ is given by $(\mathbf{D}_k)_{jj} = c_j(\max_{1\leq i\leq n}|(\mathbf{X}_{(q;k)})_{ij}|)^{-1}$ for $j = 1, \ldots, q$. (See Fig. 1 (a)–(d) for examples with $q = 1, 2$.) If $\mathcal{S}$ is a q-dimensional sphere of radius c, then the $(\mathbf{D}_k)_{jj}$ are instead chosen to have maximum product, subject to the constraints $\sum_{j=1}^q (\mathbf{X}_{(q;k)})_{ij}^2 (\mathbf{D}_k)_{jj}^2 \leq c^2$ for $i = 1, \ldots, n$.

3. An attractive consequence of Theorem 2 is that the projection of the most V-robust design in $\Xi_{n,q}$ onto the last ($k = 1$) or first ($k = 2$)q' coordinate axes is most V-robust in $\Xi_{n,q'}$.

4. A separate but parallel development shows that if there is no constant term in the model, then $\mathbf{X}$ may be replaced by the $n \times n$ matrix consisting of the $\mathbf{r}_j$, and $\{\lambda_j\}$ by $\{\mu_j\}_{j=1}^n$, all in their natural order.

5. The empirical distribution functions of the design points of the most V-robust designs in $\Xi_{n,1}$ with $\mathcal{S} = [-1, 1]$ converge weakly, as $n \to \infty$, to the arcsine distribution with distribution function $.5 + \pi^{-1}\arcsin(x)$ and density $\pi^{-1}(1 - x^2)^{-1/2}$. An asymptotic description of the designs is then given by this limit together with the order in which the design points are to be applied. For $\mathcal{P}_1$, the points alternate in sign, with the first half of the points increasing in magnitude and the second half decreasing in magnitude. For $\mathcal{P}_2$, the first half of the points are positive and the second half negative. Within each half, the points increase and then decrease in magnitude.

It is interesting to note that the arcsine design also arises in another context—that of optimal polynomial regression

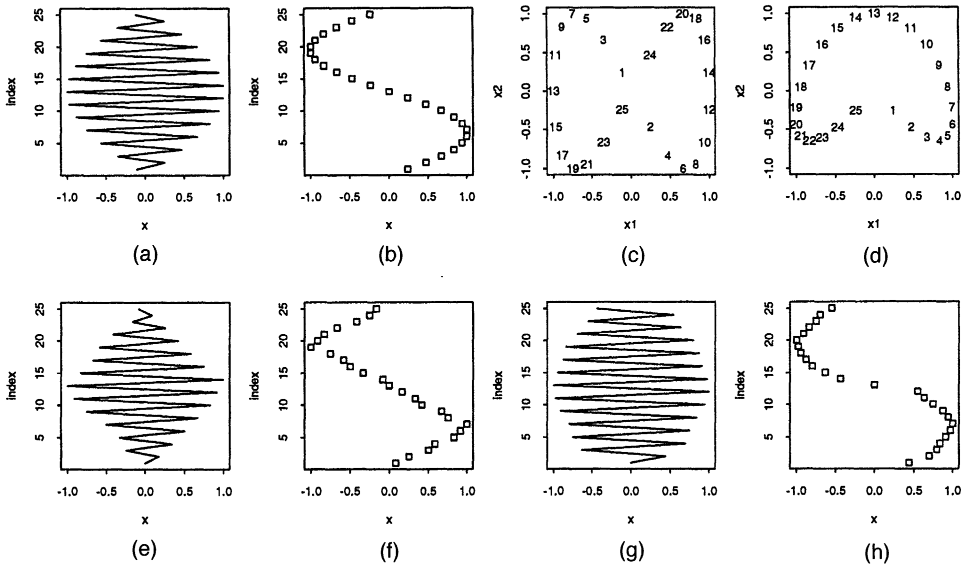

Figure 1. Most V-, Most B-, and M-Robust Designs, n = 25. (a), (b): Most V-robust designs in [−1, 1] for $\mathcal{P}_1$ and $\mathcal{P}_2$; (c), (d): most V-robust designs in [−1, 1] × [−1, 1], with the indices of design points plotted for $\mathcal{P}_1$ and $\mathcal{P}_2$; (e), (f): most B-robust designs for $\mathcal{F}_1$, ordered for $\mathcal{P}_1$ and $\mathcal{P}_2$; (g), (h): M-robust designs for $\mathcal{F}_1$, ordered for $\mathcal{P}_1$ and $\mathcal{P}_2$.

design, as the degree of the fitted polynomial tends to infinity. (See Pukelsheim 1993 for a discussion.) This suggests that the designs might be suitable for polynomial regression as well.

4. B-ROBUST DESIGNS

By virtue of (6), the B-robustness problem is that of constructing a design to maximize $|\mathbf{B}_\xi(0)|$, subject to a bound on $\sup_{f\in\mathcal{F}} \mathbf{b}_{f,\xi}^T \mathbf{B}_\xi^{-1}(0)\mathbf{b}_{f,\xi}$. Such designs then coincide with the bounded bias designs of Liu (1994) and Liu and Wiens (1997).

Example 4.1. Consider the case of multiple regression. Let $\mathcal{S}$ be a q-dimensional sphere centered at the origin, with radius r determined by the requirement that $\mathcal{S}$ have unit volume. Recall (1) and take

$$\mathcal{F}_1 = \left\{ f \,\middle|\, \int_{\mathcal{S}} \mathbf{z}(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x} = 0, \int_{\mathcal{S}} f^2(\mathbf{x})\, d\mathbf{x} \le \eta^2 \right\}.$$

The "true" parameter is defined by $\boldsymbol{\theta}_0 = \operatorname{argmin} \int_{\mathcal{S}} (E[Y|\mathbf{x}] - \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta})^2 d\mathbf{x}$. Minimax designs for this model have been considered by Wiens (1992), and by Huber (1975) when $q = 1$. Bounded bias designs were constructed by Liu (1994) when $q = 1$. Wiens (1992) showed that only absolutely continuous design measures are admissible for this problem. Thus, following Sacks and Ylvisaker (1966, 1968), who investigated optimal rates of convergence of n-point designs to absolutely continuous measures, we propose approximating the B-robust design as in the following theorem.

Theorem 3. Let $\beta > 0$ be given, and put $\beta' = \beta/(n\eta^2), \gamma_0 = r^2/(q+2)$. Define a distribution function $H_0(u), 0 \le u \le r$ by its density $h_0(u) = (qu^{q-1}/r^q)g_0(u)$ as follows:

1. If $0 < \beta' \le 4/(q(q+4))$, then

$$g_0(u) = 1 + \left(\frac{\gamma}{\gamma_0} - 1\right)\left(\frac{q+4}{4}\right)\left(\frac{u^2}{\gamma_0} - q\right),$$

where γ is determined from $\beta' = q(q+4)((\gamma/\gamma_0) - 1)^2/4$.

2. If $\beta' \ge 4/(q(q+4))$, then

$$g_0(u) = \left\{ \left[\left(\frac{u}{r}\right)^2 - b \right] \middle/ K_q(b) \right\} I(r\sqrt{b} \le u \le r),$$

where $K_q(b) = (1-b) - 2(1 - b^{(q/2)+1})/(q+2)$ and (b, γ) are determined by the equations

$$\gamma = \gamma_0 \cdot \frac{K_{q+2}(b)}{K_q(b)}, \qquad \frac{q\gamma - br^2}{r^2 K_q(b)} - 1 = \beta'.$$

Define $a = [2^{-q}n], b = n - 2^q a$. Put $u_i = H_0^{-1}(i/a), i = 1, \ldots, a$ and for each $i \le a$ let $\mathbf{t}_{i,1}, \ldots, \mathbf{t}_{i,2^q}$ be a random permutation of points equally spaced over the surface of the unit sphere in $\mathbb{R}^q$. Then the empirical distribution function ξ_n of the points $\mathbf{x}_{i,j} = \mathbf{t}_{i,j}u_i$, together with b $\mathbf{0}$'s, is an n-point design that converges weakly to a B-robust design ξ_∞ with density $\xi'_\infty(\mathbf{x}) = g_0(\|\mathbf{x}\|)$. The limiting distribution of $\|\mathbf{x}\| = U$ is $H_0(u)$, and $\mathrm{CBS}(\xi_\infty, \mathcal{F}_1) = \beta$.

Remarks. 6. The notion of points equally spaced over the surface of the unit sphere is to be interpreted in terms of the angles defined by these points. When $q = 1$, we are asserting only that the design is symmetric. When $q = 2$, one may take $\mathbf{t}_{i,j} = (\cos\psi_{i,j}, \sin\psi_{i,j})^T$, where $\psi_{i,1}, \psi_{i,2}, \psi_{i,3}, \psi_{i,4}$ is a random permutation of $\{\psi_i + (j-1)\pi/2, j = 1,2,3,4\}$ and $\psi_1, \ldots, \psi_{[n/4]}$ is a random permutation of angles equally spaced over $[0, 2\pi)$. Note that $\xi_n \in \Xi_{n,q}$.

7. The minimum value of $\mathrm{CBS}(\xi, \mathcal{F}_1)$ is $\beta = 0$. This is attained only by the uniform distribution on $\mathcal{S}$, which is then most B robust.

Example 4.2. For approximate polynomial regression $\mathbf{z}^T(x) = (1, x, x^2, \ldots, x^{p-1})$ in (1), Liu and Wiens (1997) constructed bounded bias designs for the class

$$\mathcal{F}_2 = \{f \| x^p f(x)| \le \phi(x) \,\forall\, x \in \mathcal{S} = [-1,1]\},$$

where ϕ is a given nonnegative function. Liu (1994) considered the case $p = 2$. The designs are similar to the classically optimal designs that minimize variance alone, in that they have all mass at p symmetrically placed points. For $p = 3$ and $\phi(x) \equiv 1$, the solution is

$$\xi = \begin{cases} \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{\pm(2\beta')^{1/6}} & \text{for} \quad 0 \le \beta' \le \frac{1}{2} \\ (1-\beta')\delta_0 + \beta'\delta_{\pm 1} & \text{for} \quad \frac{1}{2} \le \beta' \le \frac{2}{3} \\ \frac{1}{3}\delta_0 + \frac{2}{3}\delta_{\pm 1} & \text{for} \quad \beta' \ge \frac{2}{3}; \end{cases}$$

where $\beta' = \beta/n$ and δ_a is point mass at a. In this case, then, the concept "most B robust" leads to the clearly impractical design δ_0.

5. M-ROBUST DESIGNS

Here we outline an approach to the construction of M-robust designs for multiple regression. Let ξ^* be B robust in a class of designs. Note that the B robustness is unaffected by a permutation of the design points. Suppose then that there is a permutation of the points of support of ξ^* for which the corresponding design ξ^{**} has $\mathrm{CVS}(\xi^{**}, \mathcal{P}) \le \alpha$. Then ξ^{**} satisfies both (7) and (8). It maximizes $|\mathbf{B}_\xi(0)|$ in the class of designs satisfying (8), hence a fortiori in the smaller class of designs satisfying both (7) and (8). It is thus M robust.

Theorem 4. Let $\mathbf{z}^T(\mathbf{x}) = (1, \mathbf{x}^T_{q\times 1})$. Suppose that $\xi^* \in \Xi_{n,q}$ is B robust.

a. If $\alpha \ge 2(n-1)/n$ and there is a permutation $\langle \mathbf{x}_1^{**}, \ldots, \mathbf{x}_n^{**}\rangle$ of the support points of ξ^* for which $\mathcal{Q}(\mathbf{x}_{(j)}) \le 0$ for $j = 1, \ldots, q$, then the corresponding design ξ^{**} is M robust for $\mathcal{P}_1$.
b. If $\alpha \ge -2c_1(n-1)/n$ and there is a permutation $\langle \mathbf{x}_1^{**}, \ldots, \mathbf{x}_n^{**}\rangle$ for which $\mathcal{Q}(\mathbf{x}_{(j)}) \ge 0$ for $j = 1, \ldots, q$, then ξ^{**} is M robust for $\mathcal{P}_2$.

Example 5.1. Let $\mathcal{F}_1, \mathcal{S}$ and ξ_∞ be as in Example 4.1, with $q = 2$ and n a multiple of 4. Obtain the $\mathbf{t}_{i,j}$ by the method described in Remark 6. Then $\xi_n \in \Xi_{n,2}$. Now apply the design points by alternating between quadrants I and III for the first $n/2$ points, and then between quadrants II and IV for the remaining $n/2$ points. It is readily checked that $\mathcal{Q}(\mathbf{x}_{(j)}) < 0$ for $j = 1, 2$. The conditions of Theorem 4 are then met asymptotically, so that ξ_n is "asymptotically M robust" for $\{\mathcal{F}_1, \mathcal{P}_1\}$. Similarly, if the design points are applied quadrant by quadrant, with the first $n/4$ from quadrant I, the next $n/4$ from quadrant II, and so on, then the design is asymptotically M robust for $\{\mathcal{F}_1, \mathcal{P}_2\}$.

6. COMPARISONS

In this section we compare the relative performance of various designs, with respect to straight line regression over $\mathcal{S} = [-1, 1]$. To check design robustness, we use several other loss functions and autocorrelation structures besides those used in the derivations. The true model is given by (1) with $n = 25$, $f(x) = \eta(45/8)^{1/2}(x^2 - 1/3)$ (so that $f \in \mathcal{F}_1$ and $\int_{-1}^1 f^2(x)dx = \eta^2$), and $\boldsymbol{\varepsilon} = \sigma_0 \mathbf{P}^{1/2}\mathbf{w}$, where $\mathbf{w}$ is a vector of white noise with variance σ_w^2, $\mathbf{P}$ is an autocorrelation matrix to be specified, and $\sigma_0^2 = \mathrm{var}(\varepsilon_t/\sigma_w)$. The choice of f may be motivated by noting that: (a) this f is least favorable, in a minimax sense, for straight line regression (see Huber 1975) and (b) a quadratic disturbance represents the most common and worrisome departure from linearity in most applications.

Denoting σ_w^{-2} times the MSE matrix of $n^{1/2}\hat{\boldsymbol{\theta}}$ by $\mathbf{C}$, we find that

$$\mathbf{C} = \sigma_0^2 \begin{pmatrix} 1 & 0 \\ 0 & \tau_2^{-1} \end{pmatrix} \frac{\mathbf{Z}^T\mathbf{P}\mathbf{Z}}{n} \begin{pmatrix} 1 & 0 \\ 0 & \tau_2^{-1} \end{pmatrix} + \frac{45}{8}\nu \begin{pmatrix} \tau_2 - \frac{1}{3} \\ \frac{\tau_3}{\tau_2} \end{pmatrix} \left(\tau_2 - \frac{1}{3}, \frac{\tau_3}{\tau_2}\right),$$

where $\mathbf{Z} = (\mathbf{1} \,\vdots\, \mathbf{x})$ is the model matrix, $\tau_k = \Sigma_{i=1}^n x_i^k/n$ is the kth moment of the design, and $\nu = n\eta^2/\sigma_w^2$. The value of ν may be viewed as reflecting the relative importance of bias versus variance in the mind of the experimenter.

We consider the cases $\nu = 0$, for which the fitted model is exactly correct, and $\nu = 1$, in which case $(\int f^2)^{1/2}$ is of the same magnitude as a standard error. We take $\mathbf{P} = \mathbf{P}_j(\rho)$ to be the autocorrelation matrix of one of the following error processes $\{\varepsilon_t\}$ with lag-one autocorrelation ρ: (a) $j = 1$: MA(1) with $\rho \ge 0$; (b) $j = 2$: MA(1) with $\rho \le 0$; (c) $j = 3$: AR(1) with $\rho \ge 0$; or (d) $j = 4$: AR(1) with $\rho \le 0$. For $j = 1, 2$ we have $\sigma_0^2 = 1 + \theta^2$, where $\theta \in [-1, 1]$ satisfies $\rho = -\theta/(1+\theta^2)$. For $j = 3, 4$, we have $\sigma_0^2 = (1-\rho^2)^{-1}$.

The following designs are considered:

- vrob.pos, vrob.neg: the V-robust designs of Theorem 1, for $\mathcal{P}_1$ and $\mathcal{P}_2$.
- mostvrob.pos, mostvrob.neg: the most V-robust designs of Theorem 2, for $\mathcal{P}_1$ and $\mathcal{P}_2$ (see Fig. 1a and 1b).
- mostbrob.pos, mostbrob.neg: the most B-robust designs for $\mathcal{F}_1$, approximated as in Theorem 3 and then ordered so as to be robust for $\mathcal{P}_1$ and $\mathcal{P}_2$. Thus the unordered design points are equally spaced: $\pm i/[n/2], i = 1, \ldots, [n/2]$, and 0 if n is odd. These points are then ordered in the same manner as those in mostvrob.pos and mostvrob.neg (see Fig. 1e and 1f). Note that these designs are also M robust for $(\mathcal{P}_k, \mathcal{F}_1), k = 1, 2$, for sufficiently large α.
- uniform: n monotonically increasing points equally spaced over $[-1, 1]$. Apart from the ordering, this design then coincides with the two mostbrob designs. It is included primarily to illustrate the gains to be realized by an ordering that is appropriate for the sign of ρ.
- mrob.pos, mrob.neg: M-robust designs as in Section 5, ordered so as to be robust for $\mathcal{P}_1$ and $\mathcal{P}_2$. The unordered design points are obtained as in Theorem 3, with $\beta' = 4/5$, the boundary between the two

cases. The calculations then give that these points are $\pm(i/[n/2])^{1/3}, i = 1, \ldots, [n/2]$, and 0 if n is odd (see Fig. 1g and 1h).

For each $\mathbf{P}_j, j = 1, \ldots, 4$ we have computed (a) det $= |\mathbf{C}|$ evaluated at $\mathbf{P}(\rho)$; (b) trace $=$ trace($\mathbf{C}$) evaluated at $\mathbf{P}(\rho)$; and (c) IMSE $=$ trace($\mathbf{CA}$) evaluated at $\mathbf{P}(\rho)$, where $\mathbf{A} = \int_{-1}^{1}(1,x)^T(1,x)dx = \text{diag}(2, 2/3)$. Apart from an additive term η^2, IMSE is the integrated mean squared error $\int_{-1}^{1} E(\hat{y}(x) - E[Y|x])^2 dx$.

Figure 2 provides representative plots of the loss against ρ for each design, appropriate for the sign of ρ. We have omitted the plots for the uniform design, because for it the loss for $\rho > 0$ becomes so great as to obscure the differences between the more robust designs. The poor performance of the uniform design derives largely from the large variance of the slope estimate (see Table 1) and, when $\rho > 0$, the inappropriate ordering of the design points.

We have included only the plots for $\nu = 1$ and loss $=$ det or trace, because those in the excepted cases tell much the same story. The ranking of the designs turns out to be independent of the loss function used and largely independent of $|\rho|$. When $\nu = 0$, the V-robust designs are of course most efficient, followed by the M-robust designs. Note, however, that the V-robust designs with mass concentrated at ± 1 allow little opportunity to test the appropriateness of the linear model, and then only when n is odd. When $\nu = 1$, the three loss functions are each minimized by either the most V-robust or the M-robust designs. The performances of the most V-robust versus the M-robust designs are almost indistinguishable, as are their shapes (see Fig. 1).

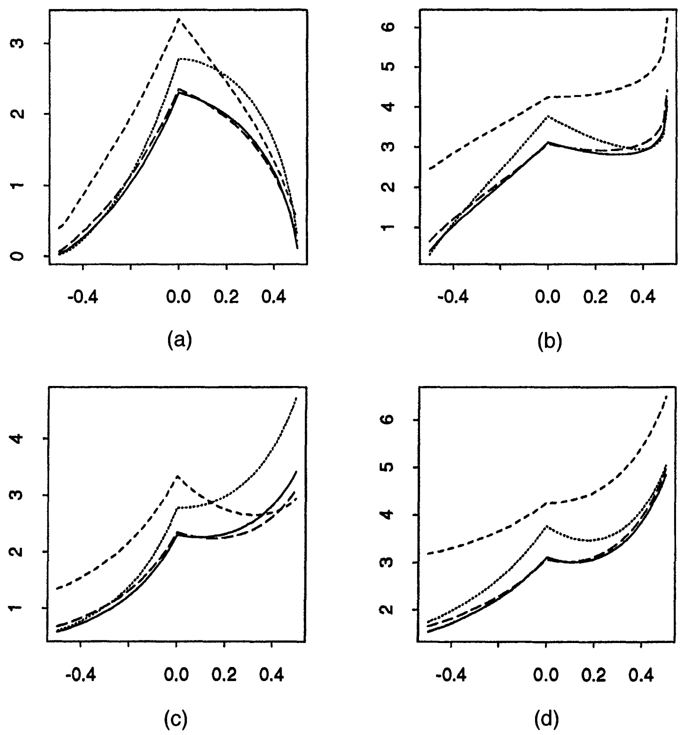

Figure 2. Comparative Losses; ν = 1. Autocorrelation models and loss functions are (a) MA(1), det; (b) MA(1), trace; (c) AR(1), det; (d) AR(1), trace. – – – mrob; ······ mostbrob; —— mostvrob; - - - vrob.

Table 1. Bias, Variance, and Power Measures Assuming Uncorrelated Errors

Design	τ_2 − 1/3	1/τ_2	Signal-to-noise ratio	Power
vrob	.627	1.042	.216	.136
mostvrob	.188	1.917	.680	.213
mostbrob	.028	2.769	.584	.197
mrob	.281	1.629	.422	.171

NOTE: For the uniform design, all values coincide with those for the mostbrob designs when the errors are uncorrelated.

At first glance, the shapes of the loss functions in Figure 2a—decreasing in $|\rho|$—and the similar (though less extreme) behaviors exhibited in the other three plots may seem counterintuitive. That these are in fact features of appropriate orderings of the design points is made plausible by analyzing the dominant term in $\mathbf{C}$. For MA(1) errors and any symmetric design, this is

$$\sigma_0^2 \frac{\mathbf{Z}^T\mathbf{P}\mathbf{Z}}{n}$$

$$= \left(\frac{2}{1+\sqrt{1-4\rho^2}}\right)(\text{diag}(1+2\rho, (1+2\rho)\tau_2 - \rho\delta)$$

$$+ O(n^{-1})),$$

where $\delta = \sum_{i=1}^{n-1}(x_i - x_{i+1})^2/n$. From this, one obtains that the determinant of this matrix tends to 0 as $|\rho| \to 1/2$, as do the trace and IMSE as $\rho \to -1/2$, if $\delta = 4\tau_2$ when $\rho > 0$ and $\delta = 0$ when $\rho < 0$. In each case we ignore the $O(n^{-1})$ terms. These equalities are not quite attainable, but they are upper and lower bounds. The robust designs considered here order the design points in such a way as to maximize or minimize δ for fixed τ_2, subject to the other constraints placed on them, according to whether ρ is positive or negative.

Table 1 gives values, for each design pair (.pos, .neg), of the following:

a. $\tau_2 - 1/3$, the dominant term in the bias when $\nu \neq 0$
b. $1/\tau_2$, determining the variance of the slope estimate
c. The signal-to-noise ratio $\lambda^2 = (45/8)\nu(\tau_4 - \tau_2^2)$ evaluated at $\nu = 1$. As used by Huber (1975), this is $E[Z]^2/\text{var}[Z]$, where $Z = \sum_{i=1}^{n} Y_i(f(x_i) - \overline{f_0})$ (with $\overline{f_0} := \sum_{i=1}^{n} f(x_i)/n$) is the test statistic of the most powerful (Neyman–Pearson) test, assuming iid normal errors, of the null hypothesis that $E[Y|x]$ is linear, versus the alternative $f(x)$ given earlier.
d. The power of the level $\alpha = .1$ Neyman–Pearson test, when $\nu = 1$. For purposes of comparison, note that the best design for such a test (when the errors are iid), placing one-half of the observations at $x = 0$ and one-quarter at each of $x = \pm 1$, has $\lambda^2 = 1.41$ and a power of .325. However, this design has a power of 0 against a response function linear at $-1, 0$, and 1, regardless of the behavior at other points.

The "Power" column of Table 1 indicates that the nonlinear disturbance $f(x)$ is so slight as to be very difficult to detect; nonetheless, of the designs considered here, the most powerful are the most V-robust designs.

Table 2. Simulated Size and Power of One-Sided 5% Durbin–Watson Tests

Response/error structure	Design: vrob	mostvrob	mostbrob	mrob	uniform
Independent	.051 (2)	.052 (2)	.051 (2)	.051 (2)	.018 (1)
ν = 0 MA(1), ρ = .2	.206 (4)	.209 (4)	.208 (4)	.208 (4)	.106 (3)
MA(1), ρ = −.2	.203 (4)	.205 (4)	.203 (4)	.204 (4)	.204 (4)
Independent	.052 (2)	.071 (2)	.068 (2)	.062 (2)	.028 (2)
ν = 1 MA(1), ρ = .2	.210 (4)	.248 (4)	.241 (4)	.232 (4)	.130 (3)
MA(1), ρ = −.2	.201 (4)	.171 (3)	.175 (3)	.183 (4)	.172 (3)

NOTE: The size and power refer to the proportion of "conclusive" rejections only. Standard errors in the third decimal place are in parentheses. The AR(1) values were very similar to those for MA(1) models and have been omitted.

Table 2 gives simulated values of the power of the one-sided Durbin–Watson test for serial correlation, with a nominal level of 5%. Here 12,000 sets of residuals were simulated for $\nu = 0$ and $\nu = 1$ and each of five correlation structures represented by their autocorrelation matrices $\mathbf{I}, \mathbf{P}_1(.2), \mathbf{P}_2(-.2), \mathbf{P}_3(.2)$ and $\mathbf{P}_4(-.2)$. The hypotheses and designs used were those appropriate for the sign of ρ.

The simulations indicate that the uniform design fares exceedingly poorly against positive autocorrelation. Among the more robust designs, when $\nu = 0$, the size and power of the test do not depend on the choice of design. When $\nu = 1$, there is essentially no change in the performance of the V-robust design. However, the power under the most V-robust, most B-robust, and M-robust designs then tends to be larger at $|\rho|$ than at $-|\rho|$. To a lesser extent, the frequency of type I errors increases when $\nu = 1$, for all but the V-robust designs.

In general, the powers of the tests considered here are rather low in all cases, illustrating (once again) the point that barely detectable departures from the assumed model can have severe effects on the precision and accuracy of estimates. We are concerned here solely with the amelioration of such effects through the use of an appropriate design. (For further discussion of the effects of dependence and of the attainment of robustness against such effects through suitable estimation procedures, see Field and Wiens 1994, Hampel et al. 1986, chap. 8, and Samarov 1987, and references cited therein.)

7. EXAMPLE

Werther, Hartge, and Rensner (1990) reviewed measurements of fluid-dynamic properties for gas-solid beds. They discussed several techniques for the determination of solids concentrations, velocities, and mass flows in various gas-solid fluidized beds—low-velocity bubbling fluidized beds, turbulent beds, circulating fluidized beds, and so on. One technique for determining the integral concentration C_v of solids in a low-velocity bubbling fluidized bed is pressure-drop measurements. Because the solid particles are held in a state of suspension by the upward-flowing gas, the pressure drop of the gas through the bed is equal to the weight of the solids per unit area of the bed. If $p = p(h)$ is the pressure over the height h of the fluidized bed, then the relationship between the pressure change Δp and the height change Δh is approximately linear within the bed; that is, $\Delta p \approx \theta \Delta h$, where $\theta = \rho_s C_v g$ for known constants ρ_s (the density of the solid) and g (the acceleration due to gravity).

Werther et al. (1990) discussed the estimation of C_v through fitting the aforementioned relationship, with additive error, after making a small number of measurements (p_i, h_i). With $\Delta p_i = p_i - p_{i+1}$ and $x_i = h_{i+1} - h_i$, the statistical model is

$$\Delta p_i = \theta x_i + \varepsilon_i, \qquad i = 1, \dots, n-1. \tag{15}$$

If the errors in the measurement of p_i can be assumed to be stationary and uncorrelated, then $\{\varepsilon_i\}$ in (15) is an MA(1) process with negative correlation $\rho(1) = -.5$. Other sources of variation—in particular the reported carry-over effect between periodic flushings of the pressure gauge—render this model somewhat approximate. It seems quite safe, however, to assume that $\rho(1)$ is negative. The simulation studies in Section 6 have shown that the most V-robust designs are quite robust against misspecifications of the error structure. We thus take $\{x_i\}$ to be the most V-robust design for $\mathcal{P}_2$ and straight-line regression through the origin; that is, $x_i \propto \sin(i\pi/n)$ (see Remark 4 of Sec. 3). Then with $h_1 = h$ and $h_n = H$, we obtain, for $i = 1, \dots, n-1$,

$$\begin{aligned} h_{i+1} &= h + (H-h) \cdot \sum_{j=1}^{i} x_j \Big/ \sum_{j=1}^{n-1} x_j \\ &= h + (H-h) \cdot \frac{\sin \frac{(i+1)\pi}{2n} \sin \frac{i\pi}{2n}}{\sin \frac{(n-1)\pi}{2n}}. \end{aligned}$$

We have simulated data from this model for a low-velocity bubbling fluidized bed, using input values obtained in collaboration with I. Nirdosh, Department of Chemical Engineering, Lakehead University. The solids are sand particles with $\rho_s = 2,200$ kg/m^3, and the height of the bed is .4 m. We used 60,000 sets of MA(2) errors, with $\rho(1) = -.36$, $\rho(2) = -.08$, with each of two designs: the most V-robust, with $\{h_i\} = \{.020, .068, .152, .248, .332, .380\}$, and the uniform, with $\{h_i\} = \{.020, .092, .164, .236, .308, .380\}$. The resulting estimates of the bias and standard error of $\hat{\theta}$ are $(-.002, 3.63)$ for the V-robust design and $(-.010, 4.04)$ for the uniform design. We obtained similar values with other MA(2) models.

8. CONCLUSIONS

A fruitful approach to designing for correlated errors has been to seek an ordering of the support points, of a design optimal for uncorrelated errors, that minimizes the variance under the model of correlation. This approach was taken by Jenkins and Chanmugan (1962), resulting in the designs of Theorem 1, found to be V robust. As in Table 1, these designs tend to result in small variability of the regression estimates when the fitted model is exactly correct, but large biases otherwise. Another tack has been to control the bias under departures from linearity (Box and Draper 1959, Huber 1975, Wiens 1991), resulting in the uniform designs found by us to be most B robust (see Table 1). Sacks and Ylvisaker (1968) also noted optimality properties of uniform designs for correlated errors.

The present work shows that under departures from linearity that are almost undetectable, especially also in the presence of departures from independence, the most V-robust and M-robust designs used here yield estimates with significantly smaller MSE than those resulting from the competing designs. The favorable properties of these designs appear to be quite insensitive to the loss function used and also to the underlying autocorrelation model. The most V-robust designs have surprisingly high power with respect to tests for model departures in the direction of nonlinear response functions.

Given that statisticians routinely fit linear models by ordinary least squares in the hope that the departures from linearity/independence are sufficiently small that these assumptions will not lead them seriously astray, we recommend the use of most V-robust or M-robust designs. They are simple to compute and afford the necessary protection in such situations.

APPENDIX: PROOFS

Proof of Theorem 1

It is easily verified that each of the four designs maximizes $|\mathbf{B}_\xi(0)| = \|\mathbf{x}\|^2/n$ in $\Xi_{n,1}$ unconditionally, and has the stated value of $CVS(\xi, \mathcal{P})$, and hence is V robust. For part c, it must be verified that $CVS(\xi, \mathcal{P}_k) \geq \alpha_{k,n}$ for each k, if $\|\mathbf{x}\|^2$ is a maximum. This is straightforward on noting that (a) maximality of $\|\mathbf{x}\|^2$ forces $\mathbf{x}$ to have the (unordered) elements in the statement of the theorem, and (b) for such $\mathbf{x}$, the denominator of $\mathcal{Q}(\mathbf{x})$ is fixed and the numerator is minimized (for $\mathcal{P}_1$) or maximized (for $\mathcal{P}_2$) by the indicated orderings.

The proof of Theorem 2 requires some preliminary results.

Lemma A1. The matrix $\mathbf{Q}$ has characteristic roots (ch.roots) μ_j and corresponding orthonormal characteristic vectors (ch.vecs) $\mathbf{r}_j$, given by (10). The ch.vecs $\mathbf{r}_j$ are orthogonal to $\mathbf{1} = (1, \ldots, 1)^T$ iff j is even.

Proof. Both statements may be verified by direct calculations. The first was derived by Grenander and Szegö (1958, pp. 67–68).

Lemma A2. With definitions as at (9)–(14), we have that $\mathbf{X}^T\mathbf{X} = \mathbf{I}_{n-1}$ and

$$\min_{\Xi_{n,q}} \sum_{j=1}^{q} \mathcal{Q}(\mathbf{x}_{(j)}) = \sum_{j=1}^{q} \lambda_{n-j}, \qquad \max_{\Xi_{n,q}} \sum_{j=1}^{q} \mathcal{Q}(\mathbf{x}_{(j)}) = \sum_{j=1}^{q} \lambda_j.$$

These extrema are attained at arbitrary nonzero multiples of the first q and last q columns of $\mathbf{X}$.

Proof. Note that

$$\max_{\Xi_{n,q}} \sum_{j=1}^{q} \mathcal{Q}(\mathbf{x}_{(j)}) = \max\{\text{trace } \mathbf{QXX}^T | \mathbf{X}^T\mathbf{X} = \mathbf{I}_q, \mathbf{X}^T\mathbf{1} = \mathbf{0}\},$$

where $\mathbf{X}$ has columns $\{\mathbf{x}_{(j)}\}_{i=1}^q$. If $\mathbf{J}_{n\times n-1}$ satisfies $\mathbf{J}^T\mathbf{J} = \mathbf{I}_{n-1}$ and $\mathbf{JJ}^T = \mathbf{I}_n - (1/n)\mathbf{11}^T$, then the conditions on $\mathbf{X}$ are equivalent to "$\mathbf{X} = \mathbf{JH}$ for some $\mathbf{H}_{n-1\times q}$ with $\mathbf{H}^T\mathbf{H} = \mathbf{I}_q$". With $\mathbf{R} = \mathbf{J}^T\mathbf{QJ}$, the desired maximum is thus $\max\{\text{trace } \mathbf{RHH}^T | \mathbf{H}^T\mathbf{H} = \mathbf{I}_q\}$. Theorem 1.10.2 of Srivastava and Khatri (1979) states that this maximum is $\sum_{j=1}^q \lambda_j$, and similarly the minimum is $\sum_{j=1}^q \lambda_{n-j}$, where $\lambda_1 \geq \cdots \geq \lambda_{n-1}$ are the ch.roots of $\mathbf{R}$. These extrema are attained if $\mathbf{H}$ consists of the q corresponding orthonormalized ch.vecs of $\mathbf{R}$, and then $\mathbf{x}_{(j)}$ is (any nonzero multiple of) $\mathbf{Jh}_{(j)}$.

The ch.vecs of $\mathbf{R}$ with roots $\lambda \neq 0$ are of the form $\mathbf{h} = \mathbf{J}^T\mathbf{z}$, where $\mathbf{z}$ is a ch.vec of $\tilde{\mathbf{R}} := \mathbf{JJ}^T\mathbf{Q}$ with root λ. There is an extraneous ch.vec of $\tilde{\mathbf{R}}$ with root 0, which is useless to us. If n is odd, then $\mathbf{Q}$ has a ch.vec $\mathbf{r}_{(n+1)/2}$ with root $\lambda = 0$; this provides an additional ch.vec of $\tilde{\mathbf{R}}$.

The equations $\tilde{\mathbf{R}}\mathbf{z} = \lambda\mathbf{z}$ may be written as

$$(\mathbf{Q} - \lambda\mathbf{I})\mathbf{z} = c\mathbf{1}, \qquad c = \mathbf{1}^T\mathbf{Qz}/n. \tag{A.1}$$

Premultiplying by $\mathbf{1}^T$ gives $\lambda\mathbf{1}^T\mathbf{z} = 0$, so that if $\lambda \neq 0$, we have $\mathbf{z} = \mathbf{JJ}^T\mathbf{z} = \mathbf{Jh}$. Thus the set $\mathcal{X}$ of unordered vectors $\mathbf{x}_{(j)}$ consists of the ch.vecs $\mathbf{z}$ of $\tilde{\mathbf{R}}$ corresponding to nonzero ch.roots, plus possibly a vector arising from $\mathbf{r}_{(n+1)/2}$.

Case 1. If $c = 0$ in (A.1), then $\mathbf{z}$ is a ch.vec of $\mathbf{Q}$. By Lemma A1, there are $[n/2]$ such vectors that are orthogonal to $\mathbf{1}$. These include $\mathbf{r}_{(n+1)/2}$ iff $n + 1$ is odd and $(n + 1)/2$ is even. Thus this case contributes the vectors $\mathbf{r}_{2j}, j = 1, \ldots, [n/2]$ to $\mathcal{X}$, with corresponding roots μ_{2j}.

Case 2. Let $c \neq 0, \lambda \neq 0$ and assume that λ is not a ch.root of $\mathbf{Q}$. Then the first equation in (A.1) gives $\mathbf{z} = c(\mathbf{Q} - \lambda\mathbf{I})^{-1}\mathbf{1}$, and the second yields

$$\mathbf{1}^T(\mathbf{Q} - \lambda\mathbf{I})^{-1}\mathbf{1} = 0. \tag{A.2}$$

Writing (A.2) as $\sum_{j=1}^{[(n+1)/2]} (\mathbf{1}^T\mathbf{r}_{2j-1})^2/(\mu_{2j-1} - \lambda) = 0$ shows that there are $[(n-1)/2]$ solutions $\lambda = \nu_j$ which when ordered satisfy $\mu_{2j-1} > \nu_j > \mu_{2j+1}$. The remaining elements of $\mathcal{X}$ are then (multiples of) the vectors $\mathbf{s}_j = (\mathbf{Q} - \nu_j\mathbf{I})^{-1}\mathbf{1}$.

Of the n equations given by $(\mathbf{Q} - \lambda\mathbf{I})\mathbf{s} = \mathbf{1}, n - 2$ are of the form

$$\left(v_{k+1} - \frac{1}{1-\lambda}\right) = 2\lambda\left(v_k - \frac{1}{1-\lambda}\right) - \left(v_{k-1} - \frac{1}{1-\lambda}\right), \qquad k = 2, \ldots, n-1.$$

This recursion, when solved and combined with the remaining two equations and with (A.2), yields

$$v_k = \left(1 - \frac{\cos\left(k - \frac{n+1}{2}\right)\phi}{\cos\frac{n+1}{2}\phi}\right) \bigg/ 2\sin^2\frac{\phi}{2},$$

where $\phi = \cos^{-1}\lambda$ satisfies (12). From this, we calculate $\sum_{k=1}^n v_k^2 = n(n+1)(n+2)/(2\sin^2\phi)$, whence normalizing $\mathbf{s}$ to have unit length gives (11). It then remains only to establish that the terms in (14) are in decreasing order. Because both μ_{2j} and ν_j are in (μ_{2j+1}, μ_{2j-1}), we require $\mu_{2j} > \nu_j$. This follows from the observation that the function on the left of (12) is strictly increasing where it is nonnegative and is negative at $\cos^{-1}\mu_{2j}$.

Proof of Theorem 2

If $\xi \in \Xi_{n,q}$ has model matrix $\mathbf{Z} = (\mathbf{1} \vdots \mathbf{X}_0)$, then for $\mathcal{P}_1$ we are to maximize $|\mathbf{Z}^T\mathbf{Z}| = n|\mathbf{X}_0^T\mathbf{X}_0|$, subject to $\sum_{j=1}^q \mathcal{Q}(\mathbf{x}_{(j)})$ being a minimum. But by Lemma A2, any $\mathbf{X}_0$ whose columns minimize $\sum_{j=1}^{q} \mathcal{Q}(\mathbf{x}_{(j)})$ must be of the form $\mathbf{X}_{(q;1)}\mathbf{D}_1$ for a diagonal matrix $\mathbf{D}_1$, and then $|\mathbf{X}_0^T\mathbf{X}_0| = |\mathbf{D}_1|^2$ is maximized by maximizing $|\mathbf{D}_1|$. The proof for $\mathcal{P}_2$ is entirely analogous.

Proof of Theorem 3

A convexity argument allows a reduction to designs with spherically symmetric densities. The maximization of $\mathbf{b}_{f,\xi}^T\mathbf{B}_{\xi}^{-1}(0)\mathbf{b}_{f,\xi}$ over $\mathcal{F}$ can be carried out as in theorem 1 of Wiens (1992); the methods of section 3 of that paper then show that the B-robust design is the solution to the problem of maximizing $E_H[U^2]$ subject to a bound on $E_H[g(U)]$, where $h(u) = (qu^{q-1}/r^q)g(u)$ is the density and H is the distribution function, under ξ, of $U := \|\mathbf{x}\|$. To solve this problem, we first fix $\gamma = E_H[U^2]/q$. With the aid of Lagrange multipliers, we find that g is of the form $g(u) = \mu(u^2+\delta)^+, \mu > 0, u \in [0,r]$, where $\mu(\gamma)$ and $\delta(\gamma)$ are chosen to make h a proper density and to attain the required bound on $E_H[g(U)]$. Then $E_H[U^2]$ is maximized over γ, yielding $H_0(u)$. The final statements of the theorem follow from the observation that by virtue of spherical symmetry, $\mathbf{T} = \mathbf{x}/\|\mathbf{x}\|$ is uniformly distributed over the surface of the unit sphere, independently of U.

Proof of Theorem 4

Under the stated conditions, in each case we have $\mathrm{CVS}(\xi^{**}, \mathcal{P}) = \sup_{\rho(1)} 2\rho(1)[(n-1)/n + \sum_{j=1}^{q} \mathcal{Q}(\mathbf{x}_{(j)})] \leq \alpha$.

[Received March 1995. Revised February 1997.]

REFERENCES

Bickel, P. J., and Herzberg, A. M. (1979), "Robustness of Design Against Autocorrelation in Time I: Asymptotic Theory, Optimality for Location and Linear Regression," *The Annals of Statistics*, 7, 77–95.

Bischoff, W. (1992), "On Exact D-Optimal Designs for Regression Models With Correlated Observations," *Annals of the Institute of Statistical Mathematics*, 44, 229–238.

——— (1993), "On D-Optimal Designs for Linear Models Under Correlated Observations With an Application to a Linear Model With Multiple Response," *Journal of Statistical Planning and Inference*, 37, 69–80.

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.

Constantine, G. M. (1989), "Robust Designs for Serially Correlated Observations," *Biometrika*, 76, 245–251.

Field, C. A., and Wiens, D. P. (1994), "One-Step M Estimators in the Linear Model, With Dependent Errors," *Canadian Journal of Statistics*, 22, 219–231.

Grenander, U., and Szegö, G. (1958), *Toeplitz Forms and Their Application*, Berkeley: University of California Press.

Hampel, F. R., Ronchetti, E., Rousseeuw, R. J., and Stahel, W. (1986), *Robust Statistics: The Approach Based on Influence Functions*, Toronto: Wiley.

Herzberg, A. M., and Huda, S. (1981), "A Comparison of Equally Spaced Designs With Different Correlation Structures in One and More Dimensions," *Canadian Journal of Statistics*, 9, 203–208.

Hössjer, O. (1991), "The Change-of-Variance Function for Dependent Data," *Probability Theory and Related Fields*, 90, 447–467.

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.

Jenkins, G. M., and Chanmugam, J. (1962), "The Estimation of Slope When the Errors are Autocorrelated," *Journal of the Royal Statistical Society*, Ser. B, 24, 199–214.

Judickii, M. I. (1976), "On Designing Experiments to Estimate the Parameters in Regression Models With Correlated Errors in the Observations," *Theory of Probability and Mathematical Statistics*, 11, 173–180.

Kiefer, J., and Wynn, H. P. (1984), "Optimum and Minimax Exact Treatment Designs for One-Dimensional Autoregressive Error Process," *The Annals of Statistics*, 12, 414–450.

Li, K. C., and Notz, W. (1982), "Robust Designs for Nearly Linear Regression," *Journal of Statistical Planning and Inference*, 6, 135–151.

Liu, S. X. (1994), "Some Topics in Robust Estimation and Experimental Design," unpublished Ph.D. dissertation, University of Calgary, Dept. of Mathematics and Statistics.

Liu, S. X., and Wiens, D. P. (in press), "Robust Designs for Approximately Polynomial Regression," *Journal of Statistical Planning and Inference*, 64.

Pesotchinsky, L. (1982), "Optimal Robust Designs: Linear Regression in R^k," *The Annals of Statistics*, 10, 511–525.

Pukelsheim, F. (1993), *Optimal Design of Experiments*, Toronto: Wiley.

Sacks, J., and Ylvisaker, D. (1966), "Designs for Regression Problems With Correlated Errors," *The Annals of Statistics*, 37, 66–89.

——— (1968), "Designs for Regression Problems with Correlated Errors; Many Parameters," *The Annals of Statistics*, 39, 49–69.

——— (1970), "Designs for Regression Problems With Correlated Errors, III," *The Annals of Statistics*, 41, 2057–2074.

Samarov, A. M. (1987), "Robust Spectral Regression," *The Annals of Statistics*, 15, 99–111.

Srivastava, M. S., and Khatri, C. G. (1979), *An Introduction to Multivariate Statistics*, New York: North-Holland.

Tang, D. (1993), "Minimax Regression Designs Under Uniform Departure Models," *The Annals of Statistics*, 21, 434–446.

Werther, J., Hartge, E.-U., and Rensner, D. (1990), "Messtechniken Fuer Gas/Feststoff-Wirbelschichtreaktoren," *Chemie-Ingenieur-Tecknik*, 62, 605–613.

Wiens, D. P. (1991), "Designs for Approximately Linear Regression: Two Optimality Properties of Uniform Designs," *Statistics and Probability Letters*, 12, 217–221.

——— (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference*, 31, 353–371.

Wiens, D. P., and Zhou, J. (1995), "Minimax Designs for Approximately Linear Models With AR(1) Errors," Technical Report 95.01, University of Alberta, Statistics Centre, Dept. of Mathematical Sciences.

——— (1996), "Minimax Regression Designs for Approximately Linear Models With Autocorrelated Errors," *Journal of Statistical Planning and Inference*, 55, 95–106.