The Canadian Journal of Statistics
Vol. 29, No. 1, 2001, Pages 117–128
La revue canadienne de statistique

Restricted minimax robust designs for misspecified regression models

Giseon HEO, Byron SCHMULAND and Douglas P. WIENS

Key words and phrases: Biased regression; least squares; optimal design; ozonation; polynomial regression; second order design.

MSC 2000: Primary 62K05, 62F35; secondary 62J05.

Abstract: The authors propose and explore new regression designs. Within a particular parametric class, these designs are minimax robust against bias caused by model misspecification while attaining reasonable levels of efficiency as well. The introduction of this restricted class of designs is motivated by a desire to avoid the mathematical and numerical intractability found in the unrestricted minimax theory. Robustness is provided against a family of model departures sufficiently broad that the minimax design measures are necessarily absolutely continuous. Examples of implementation involve approximate polynomial and second order multiple regression.

Quelques plans minimax restreints robustes pour des modèles de régression mal spécifiés

Résumé : Les auteurs proposent et explorent de nouveaux plans expérimentaux pour la régression. Ces plans sont minimax par rapport à une classe paramétrique restreinte et s'avèrent à la fois robustes au biais dû à un mauvais choix de modèle et raisonnablement efficaces. L'introduction de cette classe restreinte de plans est motivée par le désir d'éviter les problèmes mathématiques et numériques liés à la théorie minimax générale. Les plans sont robustes à des familles de modèles suffisamment larges pour que les mesures des plans minimax soient absolument continues. Les exemples d'implantation concernent l'approximation polynomiale et la régression multiple du second ordre.

1. INTRODUCTION

Suppose that an experimenter fits, by least squares, a regression model

$$\mathrm{E}(Y \mid \mathbf{x}) = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta} \tag{1}$$

to data $\{(Y_i, \mathbf{x}_i)\}_{i=1}^{n}$, with the $\mathbf{x}_i$ being chosen from a q-dimensional *design space* S. The mean response is linear in p regressors $z_1(\mathbf{x}), \ldots, z_p(\mathbf{x})$, each a function of independent variables $x_1, \ldots, x_q$. The experimenter is concerned that the true model might be only approximated by (1), a more precise description being

$$\mathrm{E}(Y \mid \mathbf{x}) = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta} + f(\mathbf{x}) \tag{2}$$

for some unknown but "small" function f. In this situation, she would like to choose design points that yield estimates $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ and estimates $\widehat{Y}(\mathbf{x}) = \mathbf{z}'(\mathbf{x})\widehat{\boldsymbol{\theta}}$ of $\mathrm{E}(Y \mid \mathbf{x})$ which remain relatively efficient while suffering as little as possible from the bias engendered by the model misspecification.

Under (2), the parameter $\boldsymbol{\theta}$ is not well-defined if f is unconstrained. This concern may be obviated by transferring to $\mathbf{z}'(\mathbf{x})\boldsymbol{\theta}$ the projection of f on the regressors; we may then assume that f and $\mathbf{z}$ are orthogonal in $L^2 = L^2(S, d\mathbf{x})$. This still leaves open the possibility that $\mathrm{E}(Y \mid \mathbf{x}) = f(\mathbf{x})$ is completely unknown and orthogonal to the regressors; in order to rule out this case, we place a bound on the magnitude of f. Our model then becomes

$$Y(\mathbf{x}_i) = \mathrm{E}(Y \mid \mathbf{x}_i) + \varepsilon_i, \quad i = 1, \ldots, n$$

with the mean response given by (2) and with f an arbitrary, unknown member of

$$\mathcal{F} = \left\{ f : \int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x} = \mathbf{0},\ \int_S f^2(\mathbf{x})\, d\mathbf{x} \le \eta^2 \right\}. \tag{3}$$

We assume additive, uncorrelated random errors with common variance σ^2. The radius η of $\mathcal{F}$ is fixed. It will be seen that the designs exhibited in this article depend on η^2 and σ^2 only through $\nu = \sigma^2/(n\eta^2)$, which may be chosen by the experimenter according to her judgement of the relative importance of variance versus bias. An alternate interpretation of this parameter is that it is inversely related to the premium in terms of lost efficiency relative to the variance-minimising design that the experimenter is willing to pay for robustness against model misspecification.

Various authors—Box & Draper (1959), Stigler (1971), Andrews & Herzberg (1979), Li & Notz (1982), Pesotchinsky (1982), Sacks & Ylvisaker (1984), Dette & Wong (1996), Liu & Wiens (1997), to mention but a few—have studied such problems in this framework and others. Our approach is to seek *minimax* designs which minimise (over a class of designs) the maximum (over $\mathcal{F}$) value of a measure of the mean squared error of $\widehat{Y}$. Such designs have been constructed only for particularly well-structured problems. See Huber (1975, 1981) for the case of straight line regression ($\mathbf{z}(x) = (1, x)'$) over $S = [-1/2, 1/2]$, with extensions by Wiens (1990, 1992) to the case of multiple linear regression, $\mathbf{z}(\mathbf{x}) = (1, x_1, \ldots, x_q)'$ with S a sphere in $\mathbb{R}^q$, as well as to the partial second order model with interactions, $\mathbf{z}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)'$, $S = [-1/2, 1/2] \times [-1/2, 1/2]$. In Section 2 of this article, we review a number of these results, and outline some of the difficulties encountered in extending this approach to more involved problems. It will be seen there that even the quadratic polynomial model resists a straightforward treatment.

Motivated by these considerations, we propose in Section 3 a certain parametric class of designs from which we seek a minimax member. We argue that these *restricted minimax* designs are mathematically and numerically simpler than the unrestricted designs, while performing almost as well. This is illustrated by reconsidering the examples of Section 2 with the new designs. As well, examples are given of the restricted approach in problems not attempted with the unrestricted approach.

The family of model departures against which robustness is provided is sufficiently broad that the minimax design measures are necessarily absolutely continuous. We give two methods of approximating and implementing such designs and illustrate one in a case study undertaken in Section 4.

2. UNRESTRICTED MINIMAX DESIGNS

An exactly implementable design will correspond to a *design measure* ξ placing mass $1/n$ at each of $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Below, we exhibit the moments of the least squares estimator under such a design. As is common in design theory, we then broaden the class of allowable measures to the class Ξ of all probability measures on S. We will find optimal designs in this class and approximate them, as necessary, prior to implementation.

When the model (1) is fitted and the true model is (2), the least squares estimator $\widehat{\boldsymbol{\theta}}$ is biased. With $\mathbf{b}(f, \xi) = \int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x})\, \xi(d\mathbf{x})$ and $\mathbf{A}_\xi = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})\, \xi(d\mathbf{x})$ assumed non-singular, the bias is $\mathrm{E}\,(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\theta} = \mathbf{A}_\xi^{-1}\mathbf{b}(f, \xi)$ and the mean squared error matrix is

$$\mathrm{MSE}(f, \xi) = \mathrm{E}\left\{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\right\} = (\sigma^2/n)\, \mathbf{A}_\xi^{-1} + \mathbf{A}_\xi^{-1}\mathbf{b}(f, \xi)\mathbf{b}'(f, \xi)\mathbf{A}_\xi^{-1}.$$

We consider the loss functions $\mathcal{L}_Q$ = integrated MSE of the fitted responses $\widehat{Y}(\mathbf{x})$, $\mathcal{L}_D$ = determinant of the MSE matrix, and $\mathcal{L}_A$ = trace of the MSE matrix. These correspond to the classical notions of Q-, D- and A-optimality, and so we adopt the same nomenclature. The term Q-optimality seems to be due to Fedorov (1972); Studden (1977) and others have used instead the term I-optimality. Explicit descriptions of these loss functions, with $\mathbf{A}_0 = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})\,d\mathbf{x}$, are given by

$$\begin{aligned}
\mathcal{L}_Q(f,\xi) &= \int_S \mathrm{E}\left[\left\{\widehat{Y}(\mathbf{x}) - \mathrm{E}(Y\,|\,\mathbf{x})\right\}^2\right] d\mathbf{x} \\
&= \left(\frac{\sigma^2}{n}\right)\mathrm{tr}\left(\mathbf{A}_\xi^{-1}\mathbf{A}_0\right) + \mathbf{b}'(f,\xi)\mathbf{A}_\xi^{-1}\mathbf{A}_0\mathbf{A}_\xi^{-1}\mathbf{b}(f,\xi) + \int_S f^2(\mathbf{x})\,d\mathbf{x}, \qquad (4)\\
\mathcal{L}_D(f,\xi) &= \det\{\mathrm{MSE}(f,\xi)\} = \left(\frac{\sigma^2}{n}\right)^p \frac{1}{|\mathbf{A}_\xi|}\left\{1 + \frac{n}{\sigma^2}\mathbf{b}'(f,\xi)\mathbf{A}_\xi^{-1}\mathbf{b}(f,\xi)\right\}, \qquad (5)\\
\mathcal{L}_A(f,\xi) &= \mathrm{tr}\{\mathrm{MSE}(f,\xi)\} = \left(\frac{\sigma^2}{n}\right)\mathrm{tr}\mathbf{A}_\xi^{-1} + \mathbf{b}'(f,\xi)\mathbf{A}_\xi^{-2}\mathbf{b}(f,\xi). \qquad (6)
\end{aligned}$$

We aim to construct designs to minimise the maximum (over $\mathcal{F}$) value of the loss. The proofs of the following results are discussed in the Appendix.

LEMMA 1. *Suppose that* $||\mathbf{z}(\mathbf{x})||$ *is bounded in* $\mathbf{x}$ *on* S *and that for each* $\mathbf{a} \neq 0$, *the set* $\{\mathbf{x} : \mathbf{a}'\mathbf{z}(\mathbf{x}) = 0\}$ *has Lebesgue measure zero. If* $\sup_{\mathcal{F}} \mathcal{L}(f,\xi)$ *is finite, then* ξ *is absolutely continuous with respect to Lebesgue measure, with a density* m *satisfying* $\int_S ||\mathbf{z}(\mathbf{x})||^2 m^2(\mathbf{x})\,d\mathbf{x} < \infty$.

THEOREM 1. *Let* S *and* ξ *be as in Lemma* 1. *Define matrices* $\mathbf{H}_\xi = \mathbf{A}_\xi\mathbf{A}_0^{-1}\mathbf{A}_\xi$, $\mathbf{K}_\xi = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})m^2(\mathbf{x})\,d\mathbf{x}$, *and* $\mathbf{G}_\xi = \mathbf{K}_\xi - \mathbf{H}_\xi$, *and denote by* $\lambda_{\max}(\mathbf{A})$ *the largest eigenvalue of a matrix* $\mathbf{A}$. *Then*

$$\begin{aligned}
\max_{\mathcal{F}} \mathcal{L}_Q(f,\xi) &= \eta^2\left\{\nu\,\mathrm{tr}\,(\mathbf{A}_\xi^{-1}\mathbf{A}_0) + \lambda_{\max}(\mathbf{K}_\xi\mathbf{H}_\xi^{-1})\right\}, \qquad (7)\\
\max_{\mathcal{F}} \mathcal{L}_D(f,\xi) &= \eta^2\left(\frac{\sigma^2}{n}\right)^{p-1}\left\{\nu + \lambda_{\max}(\mathbf{G}_\xi\mathbf{A}_\xi^{-1})\right\}\Big/|\mathbf{A}_\xi|, \qquad (8)\\
\max_{\mathcal{F}} \mathcal{L}_A(f,\xi) &= \eta^2\left\{\nu\,\mathrm{tr}\,(\mathbf{A}_\xi^{-1}) + \lambda_{\max}(\mathbf{G}_\xi\mathbf{A}_\xi^{-2})\right\}, \qquad (9)
\end{aligned}$$

and so the density $m_*(\mathbf{x})$ *of a* Q-, D- *or* A-*optimal (minimax) design* ξ_* *must minimise the right hand side of* (7), (8) *or* (9) *respectively.*

Example 1. Wiens (1992) considered the approximate multiple linear regression model, with $\mathbf{x} = (x_1, \ldots, x_q)'$ varying over a q-dimensional sphere S centred at the origin and $\mathbf{z}(\mathbf{x}) = (1, \mathbf{x}')'$. The search for minimax designs was restricted to those with symmetric, exchangeable densities $m(\mathbf{x})$. The Q- and D-optimal densities were found to be of the form $m_*(\mathbf{x}) = (a + b||\mathbf{x}||^2)^+$ for appropriate constants a and b.

For sufficiently large values of ν, the A-optimal density was found to be of the form $(a - b/||\mathbf{x}||^2)^+$. See Table 1 for some numerical values when $q = 1$. For smaller values of ν, difficulties such as detailed in Example 2 below were encountered. The A-optimality case is reconsidered in Example 4.

Example 2. We illustrate some of the difficulties that can be encountered in the minimax approach without further restrictions on the design density, by considering approximate quadratic regression $\mathbf{z}(x) = (1, x, x^2)'$ over $S = [-1/2, 1/2]$. We treat Q-optimality only, the other cases being very similar. We define

$$\alpha_j = \int_S x^j m(x)\,dx, \quad k_j = \int_S x^j m^2(x)\,dx.$$

For a symmetric design ξ, the non-zero elements of $\mathbf{H}_\xi$ are

$$H_{11} = h_0 = 9/4 - 30\alpha_2 + 180\alpha_2^2,$$

$$H_{13} = H_{31} = h_1 = 9\alpha_2/4 - 15\alpha_4 - 15\alpha_2^2 + 180\alpha_2\alpha_4,$$

$$H_{22} = h_2 = 12\alpha_2^2,$$

$$H_{33} = h_3 = 9\alpha_2^2/4 - 30\alpha_2\alpha_4 + 180\alpha_4^2,$$

and the characteristic polynomial of $\mathbf{K}_\xi \mathbf{H}_\xi^{-1}$ is $|\mathbf{H}_\xi^{-1}|$ times

$$|\mathbf{K}_\xi - \lambda \mathbf{H}_\xi| = p(\lambda) = (k_2 - \lambda h_2)\left\{(k_0 - \lambda h_0)(k_4 - \lambda h_3) - (k_2 - \lambda h_1)^2\right\}.$$

There are then two candidates for the maximum eigenvalue: $\lambda_0(\xi) = k_2/h_2$, and the larger zero $\lambda_1(\xi)$ of the quadratic factor of $p(\lambda)$. Define

$$\ell_i(\xi) = \nu\,\mathrm{tr}\,(\mathbf{A}_\xi^{-1}\mathbf{A}_0) + \lambda_i(\xi), \quad i = 0, 1; \quad \ell(\xi) = \max\{\ell_0(\xi), \ell_1(\xi)\}.$$

TABLE 1. Numerical values for the approximate straight-line model; unrestricted and restricted A-optimal minimax densities.

	Unrestricted design[1]			Restricted design[2]		
ν	a	b	loss	a	b	loss
0				1	0	0
0.1				0.932	0.820	1.269
0.445	1.778	0.028	4.450	0.625	4.500	5.169
1	2.345	0.071	9.154	−0.012	12.134	9.951
10	8.815	0.969	69.263	−3.419	36.224	69.470
100	60.225	11.426	570.339	−45.250	241.806	570.394
1000	530.587	121.381	523569	−485.606	2125.479	523576

[1] $m_*(x) = (a - b/x^2)^+$, [2] $m_*(x) = (a + bx^2)^+$

There is a general prescription by which the minimax design $\xi_* = \arg\min \ell(\xi)$ may now be obtained.

Step 1: Find designs ξ_i minimising $\ell_i(\xi)$ subject to the constraint $\lambda_i(\xi) \geq \lambda_{1-i}(\xi)$, $i = 0, 1$.

Step 2: Put $\xi_* = \xi_0$ if $\ell_0(\xi_0) \leq \ell_1(\xi_1)$ and $\xi_* = \xi_1$ otherwise.

It follows that $\ell(\xi_*) \leq \min\{\ell_0(\xi_0), \ell_1(\xi_1)\}$ and that ξ_* is minimax.

The inequality constraints in Step 1 can lead to solutions so cumbersome as to be uninteresting from a practical point of view. The omitted case of Example 1 is a case in point—see Section 3.6 of Wiens (1992). Thus a more usual approach, but one not guaranteed to succeed, is:

Step $1'$: Find designs ξ_i minimising $\ell_i(\xi)$ *among all* designs ξ, $i = 0, 1$.

Step $2'$: Put $\xi_* = \xi_0$ if $\lambda_0(\xi_0) \geq \lambda_1(\xi_0)$ and $\xi_* = \xi_1$ if $\lambda_1(\xi_1) \geq \lambda_0(\xi_1)$.

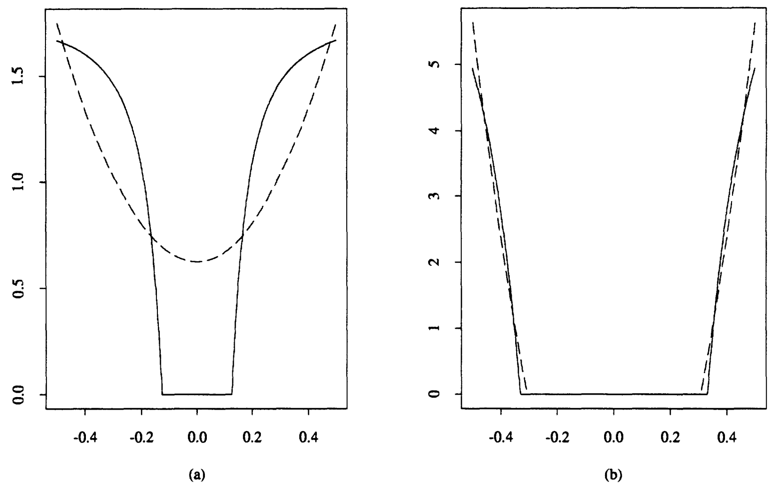

FIGURE 1: Unrestricted (solid lines) and restricted (broken lines) A-optimal minimax densities for the approximate straight line model. (a) $\nu = 0.445$; (b) $\nu = 10$. Explicit descriptions of the densities are in Table 1.

In this example, Step $1'$ may be carried out in stages, by first fixing $\alpha_0 (= 1)$, α_2 and α_4. This fixes $\mathbf{A}_\xi$ as well, so that only $\lambda_i(\xi)$ need be minimised, subject to the three side conditions. These are standard variational problems. For $i = 0$, the solution is $m_0(x) = \left(a - b/x^2 + cx^2\right)^+$. The Lagrange multipliers a, b, c are functions of α_2 and α_4 defined through the side conditions, and α_2, α_4 are then varied to minimise the loss for a given value of ν. Similarly for $i = 1$, the solution is of the form $m_1(x) = \left((a + bx^2 + cx^4) \,/\, (d + ex^2 + fx^4)\right)^+$; see Heo (1998) for details.

We find, unfortunately, that both inequalities in Step $2'$ fail. One can then either carry out Steps 1 and 2 above—a quite unappealing proposition—or seek more tractable solutions within a restricted class of designs. The latter tack in taken in the next section.

We remark that if this quadratic model has no intercept so that $\mathbf{z}(x) = (x, x^2)'$, then both Steps $1'$ and $2'$ can be carried out successfully—see Heo, Schmuland & Wiens (1999).

Example 3. Wiens (1990) found that for the partial second order regression model with interactions $\mathbf{z}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)'$ and a square design space S centred at $\mathbf{0}$, the Q-optimal symmetric, exchangeable design density was of the form $m_*(\mathbf{x}) = (a + b(x_1^2 + x_2^2) + cx_1^2x_2^2)^+$. For this model, the eigenvalues in Theorem 1 have a quite simple structure, since the relevant matrices are diagonal. For the full second order model, this is no longer the case; due to the ensuing computational difficulties, this model was not considered. We obtain designs for the full model in Example 6 below.

3. RESTRICTED MINIMAX DESIGNS

Assume that S is symmetric about $\mathbf{0}$ and invariant under permutations of the coordinate axes. The symmetry can often be arranged through an affine transformation of the independent variables, in which case there is no loss of generality. Invariance under permutations of the axes is a natural requirement when there is no *a priori* reason to prefer one coordinate over another. For the approximate regression model defined by (2) and (3), we propose to search for minimax designs within the class Ξ' of measures with densities of the form

$$m(\mathbf{x}) = \left(\sum_j \beta_j z_j(x_1^2, \ldots, x_q^2)\right)^+, \qquad (10)$$

with the β_j restricted in such a way that m is exchangeable. The squaring of the independent variables ensures the symmetry of $m(\mathbf{x})$. The optimal design in Ξ' is obtained by choosing the β_j to minimise the appropriate maximum loss function in Theorem 1, subject to the constraint that m be a density on S.

As is seen in the examples below, these *restricted minimax* designs perform almost as well as the unrestricted designs in those cases in which the latter have been constructed. By Theorem 2, they generally have the limiting behaviour that one would expect, tending to the continuous uniform design as $\nu \to 0$ and to the classical, variance-minimising designs as $\nu \to \infty$. Also the designs are numerically straightforward, having the same parametric form regardless of the structure of the eigenvalues which appear in Theorem 1. This fact has enabled us to construct the restricted designs in cases that are not readily amenable to an unrestricted treatment.

THEOREM 2. *Assume that S is a compact subset of $\mathbb{R}^p$ satisfying the condition of Lemma* 1 *and that* $\mathbf{z}(\mathbf{x})$ *is continuous in* $\mathbf{x}$ *on S. Then for each $\nu > 0$, there is a minimax design measure ξ_ν in Ξ'. Express each maximum loss* (7)–(9) *as ν times "variance" plus "bias:"* $\sup_{f \in \mathcal{F}} \mathcal{L}(f, \xi) = \nu V(\xi) + B(\xi)$. *Then* (i) *any weak limit point ξ_0 of ξ_ν as $\nu \to 0$ satisfies* $B(\xi_0) = \inf_{\xi \in \Xi'} B(\xi)$, *and* (ii) *any weak limit point ξ_∞ of ξ_ν as $\nu \to \infty$ satisfies* $V(\xi_\infty) = \inf_{\xi \in \Xi'} V(\xi)$.

To apply Theorem 2 in the case $\nu \to 0$, suppose that "1" is an element of $\mathbf{z}(\mathbf{x})$, i.e., that the model contains an intercept. Then the continuous uniform design ξ_0 is a member of Ξ'. By Theorem 2b of Wiens (1998), this is the unique minimiser of $B(\xi)$ in Ξ and by Theorem 2, $\inf_{\xi \in \Xi'} B(\xi) = \inf_{\xi \in \Xi} B(\xi) = \lim_{\nu \to 0} B(\xi_\nu)$. In the case $\nu \to \infty$, suppose that the minimiser ξ_∞ of $V(\xi)$ in Ξ is unique and is such that we can construct a sequence of designs $\xi'_\nu \in \Xi'$ tending weakly to ξ_∞. Then $V(\xi'_\nu) \to V(\xi_\infty)$ by Theorem 2 and so $\inf_{\xi \in \Xi'} V(\xi) = \inf_{\xi \in \Xi} V(\xi) = \lim_{\nu \to \infty} V(\xi_\nu)$. The details of such constructions are straightforward in particular examples and will not be given here.

Example 4. For the model of Example 1, (10) gives $m(\mathbf{x}) = \left(\beta_0 + \sum_{j=1}^q \beta_j x_j^2\right)^+$. With $a = \beta_0$, and $b = \beta_1 = \cdots = \beta_q$ for exchangeability, this density agrees exactly with the form of the Q- and D-optimal densities $m_*(\mathbf{x})$. See Table 1 and Figure 1 for a comparison of the unrestricted and restricted A-optimal design densities when $q = 1$. In the unrestricted case, the design is available only for $\nu \geq 0.445$. For moderately large ν, the loss of the restricted minimax design is only marginally greater than that of the unrestricted design. As $\nu \to \infty$, both designs approach the variance-minimising design with mass of 0.5 at each of $\pm 1/2$.

Example 5. For the polynomial model with $\mathbf{z}(x) = (1, x, \ldots, x^q)'$, Ξ' consists of those designs with densities

$$m(x; a, \mathbf{b}) = \left(a + \sum_{j=1}^{q} b_j x^{2j}\right)^+.$$

From the plots in Figure 2, one sees that a rough guide to implementation is to locate the $q+1$ sites at which these classical designs place all of their mass, and then to replace the replicates at these sites by groups of observations at distinct but nearby sites. This observation is reinforced in the examples of Section 4.

See Figure 2 for plots in the quadratic and cubic cases with values of the constants in Table 2. As noted previously for variance-minimising designs (Studden 1977) and for mse-minimising designs (Wiens 2000), the Q- and D-optimal designs are very similar. As $\nu \to \infty$, all three tend to their variance-minimising counterparts.

Example 6. For the partial second order model as considered in Example 3, Ξ' again contains the unrestricted minimax design. For the full second order model with $q = 2$, $\mathbf{z}(x_1, x_2) =$

$(1, x_1, x_2, x_1x_2, x_1^2, x_2^2)'$, the designs in Ξ' have densities

$$m(x_1, x_2; a, b, c, d) = \left(a + b(x_1^2 + x_2^2) + cx_1^2x_2^2 + d(x_1^4 + x_2^4)\right)^+ .$$

See Figure 3 for plots of the Q-, D- and A-optimal design densities when $\nu = 5$.

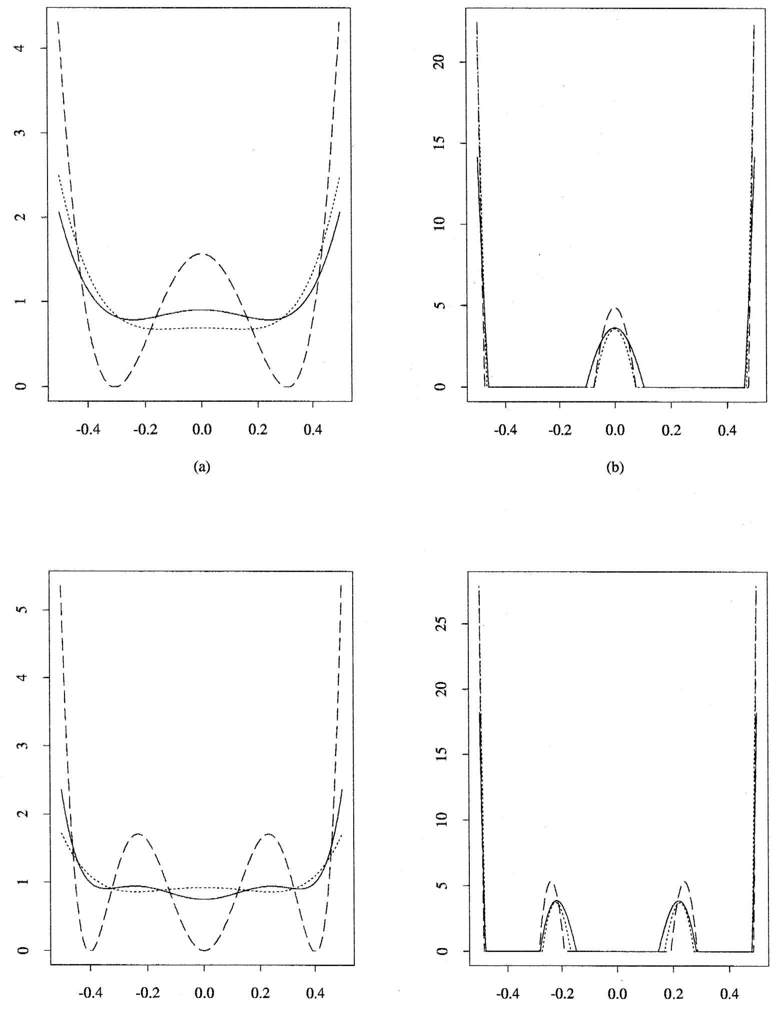

FIGURE 2: Q-optimal (solid lines), D-optimal (dotted lines) and A-optimal (dashed lines) minimax densities for approximate degree-q polynomial regression. (a) $q = 2, \nu = 1$; (b) $q = 2, \nu = 100$; (c) $q = 3, \nu = 1$; (d) $q = 3, \nu = 100$. Explicit descriptions of the densities are in Table 2.

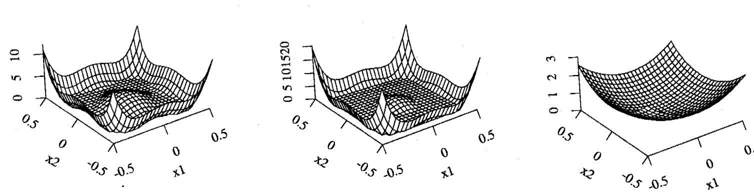

FIGURE 3: Restricted minimax design densities for approximate, full second order model with $\nu = 5$. (a) Q-optimal density; (b) D-optimal density; (c) A-optimal density.

TABLE 2: Numerical values for the approximate quadratic and cubic models; restricted minimax densities.

	Quadratic model[1]				Cubic model[2]			
	α	β_1	β_2		α	β_1	β_2	β_3
				$\nu=1$				
Q	34.845	−0.117	0.026		375.733	−0.265	0.021	0.002
D	35.095	−0.044	0.020		202.398	−0.102	0.008	0.003
A	178.081	−0.188	0.009		2753.817	−0.323	0.026	0.000
				$\nu=100$				
Q	1606.184	−0.224	0.002		25589.67	−0.332	0.025	0.000
D	2984.049	−0.225	0.001		39087.81	−0.334	0.026	0.000
A	3904.564	−0.232	0.001		59911.30	−0.355	0.031	−0.001

[1] $m_*(x) = \alpha(x^4 + \beta_1 x^2 + \beta_2)^+$, [2] $m_*(x) = \alpha(x^6 + \beta_1 x^4 + \beta_2 x^2 + \beta_3)^+$

Explicit expressions, obtained by expressing the appropriate loss functions in terms of a, b, c and d and then minimising the loss numerically over these constants, are

Q-optimality:

$$m_*(x_1, x_2) = 216.419\,(x_1^4 + x_2^4 + 0.306x_1^2x_2^2 - 0.210(x_1^2 + x_2^2) + 0.011)^+\,,$$

D-optimality:

$$m_*(x_1, x_2) = 369.556\,(x_1^4 + x_2^4 + 0.430x_1^2x_2^2 - 0.213(x_1^2 + x_2^2) + 0.007)^+\,,$$

A-optimality:

$$m_*(x_1, x_2) = 1.856\,(x_1^4 + x_2^4 + 0.442x_1^2x_2^2 + 2.168(x_1^2 + x_2^2) + 0.149)\,.$$

All three designs have concentrations of mass at the boundary of the square, in particular at $(\pm 1/2, \pm 1/2)$ and to a lesser extent at $(\pm 1/2, 0)$ and $(0, \pm 1/2)$. Substantial mass is however placed all along the boundary and, in the Q- and D-cases, near the origin. The designs can roughly be described as smoothed versions of central composite designs.

4. IMPLEMENTATIONS AND CASE STUDY

The implementation of a continuous measure as a discrete design involves some arbitrariness. In this section, we discuss two possible methods and illustrate one with a case study. First consider the case of a single independent variable x. Here one may choose the sites x_i to be n uniformly spaced quantiles of the minimax design measure, i.e., $x_i = \xi_*^{-1}\{(i-1)/(n-1)\}$, where for notational convenience ξ_* is identified with its distribution function. One can also force replications through an obvious modification of this technique; see Heo, Schmuland & Wiens (1999) for an example in an oncology setting. In the experiment described there, x represents the level of a carcinogen, and one is to estimate the probability $\mathrm{P}(x)$ of a particular response at level x. Replication is desirable in order to estimate the probabilities $\mathrm{P}(x_i)$ by sample proportions. The logits are then approximated by a cubic polynomial in x. The results obtained in Heo, Schmuland & Wiens (1999) reinforce the observation made in Example 5, in that the design may be obtained from the variance-minimising design by breaking up its four large groups of replicates into clusters of replicates at nearby sites.

The following case study arose from a consulting project undertaken by the first author. In it, a second order response is anticipated. A design as in Example 6 is implemented by choosing the sites in such a manner that the empirical moments, up to a certain order which is $O(n)$, match those obtained from the minimax density. Thus, as in the method of the previous paragraph, we construct a discrete measure ξ_n which has the property that it converges in measure to the minimax design ξ_* as $n \to \infty$.

In each of these cases, the finite sample implementation is intuitively sensible as well as robust. A balance is struck between full efficiency and robustness as we place observations at varied locations near the sites at which the variance-minimising designs place all of their mass. This 'within-site' variation permits the fitting and exploration of alternate models.

4.1. Case study.

Prairie farmers in Alberta have traditionally stocked dugouts with trout for recreational purposes. Some are now attempting commercial fish culturing indoors, year-round. Because of limited water supplies, attempts are being made to recycle waste water for this purpose. Most solids in wastewater from trout-rearing facilities settle readily, but a suspension of fine "particulate" material remains. Several studies have shown that fine particulate adversely affects fish health and productivity. The wastewater engineering research team at the Alberta Environmental Centre conducted a bench-scale experiment to determine the amount of total suspended solid (TSS) remaining after applying ozone (O_3) at application rates ranging from 0 to 2 mg/L (see Heo & James 1995). Because ozonation is to be used for disinfection and the associated capital cost is high, the team wanted to determine an optimal O_3 rate, minimising the worst cost. Another factor which is important in the removal of suspended solids is the gas to liquid ratio, denoted GL. Uncertainties about the exact nature of the relationship between TSS, O_3 and GL led to the assumption of an approximate second order model as in Example 6.

Both factors were linearly transformed to the range $[-1/2, 1/2]$. The Q-optimal design ξ_*, with $\nu = 5$ as in Figure 3(a), was then implemented as follows to yield $n = 48$ design points. We chose $n_0 = n/8$ points (x_1, x_2) in $\{0 \le x_1 \le x_2 \le 1/2\}$ and then obtained the remaining $7n_0$ sites by symmetry and exchangeability. The n_0 points $\{x_{1i}, x_{2i}\}_{i=1}^{n_0}$ were chosen such that the moments

$$e_{2j,2k} = \sum_{i=1}^{n_0} \left(x_{1i}^{2j} x_{2i}^{2k} + x_{1i}^{2k} x_{2i}^{2j} \right) / (2n_0)$$

matched up as closely as possible with the theoretical moments $\mathrm{E}_{\xi_*}(X_1^{2j} X_2^{2k})$ obtained from ξ_*. We did this for the $J(J+3)/2$ choices (k, j) with $k = 0, \ldots, j$ and $j = 1, \ldots, J$, with J being the smallest integer for which $J(J+3)/2$ exceeds the number, $2n_0$, of coordinates to be chosen. Thus $n = 48$, $n_0 = 6$ yielded $J = 4$ and 14 even order moments to be matched up. Of course, all moments with at least one odd order are zero, and the 14 moments obtained by exchanging j and k will be matched as well. The matching was done by numerical minimisation of

$$\sum_{j,k} \left\{ e_{2j,2k} - \mathrm{E}_{\xi_*} \left(X_1^{2j} X_2^{2k} \right) \right\}^2,$$

yielding the implementation shown in Figure 4 with

$$\begin{aligned} \{x_{1i}, x_{2i}\}_{i=1}^{n_0} &= \{(0.011, 0.500), (0.023, 0.038), (0.085, 0.332), \\ &\qquad (0.235, 0.456), (0.373, 0.466), (0.432, 0.500)\}. \end{aligned}$$

5. SUMMARY

We have presented new, parametric classes of regression designs. Within several such classes, we have isolated members that are minimax robust against a broad class of departures from the assumed linear (in the regressors) model. In those cases in which minimax members of a broader, infinite-dimensional, class of designs have already been obtained, it has been seen that they often

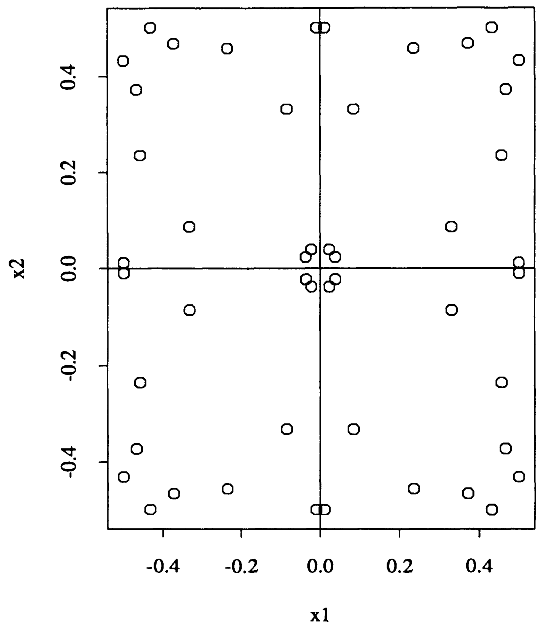

FIGURE 4: Implementation of the Q-optimal design of Figure 3(a) for approximate second order regression; $n = 48$.

coincide with the minimax members of the restricted classes of designs studied here. When they do not, it is typically the case that the new designs are mathematically and numerically simpler than those previously obtained, or sought but not obtained due to their extreme complexity. Examples have been given of polynomial and second order designs that are optimal with respect to generalisations of the common Q-, D- and A-optimality criteria. Two implementation methods have been discussed. The resulting designs are intuitively sensible as well as robust, and roughly correspond to breaking up the replicates in the classical, variance-minimising designs into clusters of observations at nearby sites.

APPENDIX: DERIVATIONS

Proof of Lemma 1. See Heo, Schmuland & Wiens (1999).

Proof of Theorem 1. Note that

$$\mathbf{G}_\xi = \int_S \left[\left\{m(\mathbf{x})\mathbf{I} - \mathbf{A}_\xi \mathbf{A}_0^{-1}\right\} \mathbf{z}(\mathbf{x})\right] \left[\left\{m(\mathbf{x})\mathbf{I} - \mathbf{A}_\xi \mathbf{A}_0^{-1}\right\} \mathbf{z}(\mathbf{x})\right]' \, d\mathbf{x},$$

so that $\mathbf{G}_\xi$ is positive semi-definite. We will prove that

$$\eta \mathbf{G}_\xi^{1/2}(S_p) \subseteq \{\mathbf{b}(f,\xi) : f \in \mathcal{F}\} \subseteq \eta \mathbf{G}_\xi^{1/2}(B_p), \tag{A.1}$$

where $S_p = \{\mathbf{b} : ||\mathbf{b}|| = 1\}$ and $B_p = \{\mathbf{b} : ||\mathbf{b}|| \leq 1\}$ are the unit sphere and the unit ball in $\mathbb{R}^p$, respectively. Using (4)–(6), this gives

$$\begin{aligned}
\sup_{f\in\mathcal{F}} \mathcal{L}_Q(f,\xi) &= (\sigma^2/n)\,\mathrm{tr}\left(\mathbf{A}_\xi^{-1}\mathbf{A}_0\right) + \eta^2 \sup_{||\beta||=1} \beta'\left(\mathbf{I} + \mathbf{G}_\xi^{1/2}\mathbf{H}_\xi^{-1}\mathbf{G}_\xi^{1/2}\right)\beta, \\
\sup_{f\in\mathcal{F}} \mathcal{L}_D(f,\xi) &= (\sigma^2/n)^p \frac{1}{|\mathbf{A}_\xi|}\left\{1 + \frac{1}{\nu} \sup_{||\beta||=1} \beta'\left(\mathbf{G}_\xi^{1/2}\mathbf{A}_\xi^{-1}\mathbf{G}_\xi^{1/2}\right)\beta\right\}, \\
\sup_{f\in\mathcal{F}} \mathcal{L}_A(f,\xi) &= (\sigma^2/n)\,\mathrm{tr}(\mathbf{A}_\xi^{-1}) + \eta^2 \sup_{||\beta||=1} \beta'\left(\mathbf{G}_\xi^{1/2}\mathbf{A}_\xi^{-2}\mathbf{G}_\xi^{1/2}\right)\beta.
\end{aligned}$$

The maxima of the three quadratic forms over β are

$$\begin{aligned}\lambda_{\max}(\mathbf{I}+\mathbf{G}_\xi^{1/2}\mathbf{H}_\xi^{-1}\mathbf{G}_\xi^{1/2}) &= \lambda_{\max}(\mathbf{K}_\xi\mathbf{H}_\xi^{-1}),\\ \lambda_{\max}(\mathbf{G}_\xi^{1/2}\mathbf{A}_\xi^{-1}\mathbf{G}_\xi^{1/2}) &= \lambda_{\max}(\mathbf{G}_\xi\mathbf{A}_\xi^{-1})\end{aligned}$$

and

$$\lambda_{\max}(\mathbf{G}_\xi^{1/2}\mathbf{A}_\xi^{-2}\mathbf{G}_\xi^{1/2}) = \lambda_{\max}(\mathbf{G}_\xi\mathbf{A}_\xi^{-2})$$

respectively, yielding (7)–(9).

If $\mathbf{G}_\xi$ is non-singular, the inclusion (A.1) is proven as Theorem 1 of Wiens (1992). If $\mathbf{G}_\xi$ is singular (as at the continuous uniform design), we proceed as follows. Take any design ξ_1 for which the corresponding matrix $\mathbf{G}_{\xi_1}$ is invertible. Put $\xi_t = (1-t)\xi + t\xi_1$ and let $p(t) = |\mathbf{G}(\xi_t)|$. Then $p(t)$ is a polynomial in $t \in [0,1]$ with $p(0) = 0$ and $p(1) > 0$, so that $p(t)$ is non-constant and non-negative on $[0,1]$. Thus $p(t) > 0$ for all sufficiently small $t > 0$.

To prove the right-hand inclusion in (A.1), let $f \in \mathcal{F}$ and pick $\mathbf{b}_t \in B_p$ so that

$$\eta\mathbf{G}_{\xi_t}^{1/2}\mathbf{b}_t = \mathbf{b}(f,\xi_t) \tag{A.2}$$

for sufficiently small $t > 0$. We have

$$\|\mathbf{b}(f,\xi_t) - \mathbf{b}(f,\xi)\| \le t\left\{\int_S f^2(\mathbf{x})\,d\mathbf{x}\right\}^{1/2}\left\{\int_S \|\mathbf{z}(\mathbf{x})\|^2 (m_1 - m)^2(\mathbf{x})\,d\mathbf{x}\right\}^{1/2}, \tag{A.3}$$

so $\mathbf{b}(f,\xi_t) \to \mathbf{b}(f,\xi)$as $t \to 0$. Similarly $\mathbf{G}_{\xi_t} \to \mathbf{G}_\xi$ and hence $\mathbf{G}_{\xi_t}^{1/2} \to \mathbf{G}_\xi^{1/2}$, as the mapping $\mathbf{G} \mapsto \mathbf{G}^{1/2}$ is continuous on the space of symmetric positive semi-definite matrices. Then $\mathbf{G}_{\xi_t}^{1/2} \to \mathbf{G}_\xi^{1/2}$ uniformly on the compact set B_p. Choose a subsequence $t_n \to 0$ and $\mathbf{b} \in B_p$ so that $\mathbf{b}_{t_n} \to \mathbf{b}$ and let $t_n \to 0$ in (A.2) above to obtain $\eta\mathbf{G}_\xi^{1/2}\mathbf{b} = \mathbf{b}(f,\xi)$.

For the left-hand inclusion in (A.1), fix $\mathbf{s} \in S_p$ and pick $f_t \in \mathcal{F}$ so that

$$\eta\mathbf{G}_{\xi_t}^{1/2}\mathbf{s} = \mathbf{b}(f_t,\xi_t) \tag{A.4}$$

for sufficiently small $t > 0$. As before, the left-hand side converges to $\eta\mathbf{G}_\xi^{1/2}\mathbf{s}$ as $t \to 0$. Since $\mathcal{F}$ is weakly compact in L^2, we can choose a subsequence $t_n \to 0$ and $f \in \mathcal{F}$ so that $f_{t_n} \to f$ weakly in L^2. Then

$$\|\mathbf{b}(f,\xi) - \mathbf{b}(f_{t_n},\xi_{t_n})\| \le \|\mathbf{b}(f,\xi) - \mathbf{b}(f_{t_n},\xi)\| + \|\mathbf{b}(f_{t_n},\xi) - \mathbf{b}(f_{t_n},\xi_{t_n})\|.$$

The first term goes to zero by weak convergence, and the second term goes to zero by (A.3). Letting $t_n \to 0$ in (A.4), we obtain $\eta\mathbf{G}_\xi^{1/2}\mathbf{s} = \mathbf{b}(f,\xi)$.

Proof of Theorem 2. See Heo, Schmuland & Wiens (1999).

ACKNOWLEDGEMENTS

The helpful comments of the Editor, an Associate Editor and a referee are gratefully acknowledged.

REFERENCES

D. A. Andrews & A. M. Herzberg (1979). The robustness and optimality of response surface designs. *Journal of Statistical Planning and Inference*, 3, 249–257.

G. E. P. Box & N. R. Draper (1959). A basis for the selection of a response surface design. *Journal of the American Statistical Association*, 54, 622–654.

H. Dette & W. K. Wong (1996). Robust optimal extrapolation designs. *Biometrika*, 83, 667–680.

V. V. Fedorov (1972). *Theory of Optimal Experiments*. Academic Press, New York.

G. Heo (1998). *Optimal Designs for Approximately Polynomial Regression Models*. Unpublished doctoral dissertation, University of Alberta, Edmonton, Alberta, Canada.

G. Heo & W. James (1995). *The Effects of Ozone Application Rate and Gas: Liquid Ratio on Removing Fine Suspended Particles in a Bubble Column Reactor*. Technical report, Alberta Environmental Centre.

G. Heo, B. Schmuland & D. P. Wiens (1999). *Restricted Minimax Robust Designs for Misspecified Regression Models*. Technical report 99.04, Statistics Centre, Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada: http://www.stat.ualberta.ca.

P. J. Huber (1975). Robustness and designs. In *A Survey of Statistical Design and Linear Models: Proceedings of an International Symposium on Statistical Design and Linear Models, Colorado State University, Fort Collins, March 19–23, 1973* (J. N. Srivastava, ed.), North Holland, Amsterdam, pp. 287–303.

P. J. Huber (1981). *Robust Statistics*. Wiley, New York.

K. C. Li & W. Notz (1982). Robust designs for nearly linear regression. *Journal of Statistical Planning and Inference*, 6, 135–151.

S. X. Liu & D. P. Wiens (1997). Robust designs for approximately polynomial regression. *Journal of Statistical Planning and Inference*, 64, 369–381.

L. Pesotchinsky (1982). Optimal robust designs: Linear regression in $\mathbf{R}^k$. *The Annals of Statistics*, 10, 511–525.

J. Sacks & D. Ylvisaker (1984). Some model robust designs in regression. *The Annals of Statistics*, 12, 1324–1348.

S. M. Stigler (1971). Optimal experimental designs for polynomial regression. *Journal of the American Statistical Association*, 66, 311–318.

W. J. Studden (1977). Optimal designs for integrated variance in polynomial regression. *Statistical Decision Theory and Related Topics II: Proceedings of a Symposium, Purdue University, Lafayette, Indiana, May 17–19, 1976* (S. S. Gupta & D. S. Moore, eds.), Academic Press, New York, pp. 411–420.

D. P. Wiens (1990). Robust, minimax designs for multiple linear regression. *Linear Algebra and its Applications, Second Special Issue on Linear Algebra and Statistics*, 127, 327–340.

D. P. Wiens (1992). Minimax designs for approximately linear regression. *Journal of Statistical Planning and Inference*, 31, 353–371.

D. P. Wiens (1998). Minimax robust designs and weights for approximately specified regression models with heteroscedastic errors. *Journal of the American Statistical Association*, 93, 1440–1450.

D. P. Wiens (2000). Robust weights and designs for biased regression models: Least squares and generalized M-estimation. *Journal of Statistical Planning and Inference*, to appear.

Received 28 October 1999
Accepted 25 January 2000

Giseon HEO: heo@stat.ualberta.ca
Byron SCHMULAND: schmu@stat.ualberta.ca
Douglas P. WIENS: wiens@stat.ualberta.ca
Dept. of Mathematical Sciences, University of Alberta
Edmonton, Alberta, Canada T6G 2G1