

Restricted Minimax Robust Designs for Misspecified Regression Models

Abbreviated title: Restricted Robust Designs

Giseon Heo, Byron Schmuland, Douglas P. Wiens*

University of Alberta

A version of this paper was accepted January 26 2000 for publication by

The Canadian Journal of Statistics

We 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. Methods of approximating and implementing such designs are illustrated in two case studies involving approximate polynomial and second order multiple regression.

AMS 1991 SUBJECT CLASSIFICATIONS: Primary 62K05, 62F35; secondary 62J05.

KEY WORDS: Biased regression; Bioassay; Dose response; Least squares; Logistic model; Optimal design; Ozonation; Polynomial regression; Second order design.

⁰Giseon Heo, Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1. Fax.: (780) 492-6826; e-mail: heo@stat.ualberta.ca. Byron Schmuland is Associate Professor of Statistics, Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1. Tel.: (780) 492-3195; fax.: (780) 492-6826; e-mail: schmu@stat.ualberta.ca. Douglas Wiens is Professor of Statistics, Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1. Tel.: (780) 492-4406; fax.: (780) 492-6826; e-mail: doug.wiens@ualberta.ca. This research is supported by the Natural Sciences and Engineering Research Council of Canada.

1. INTRODUCTION

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

$$E(Y|\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta} \tag{1}$$

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 $z_1(\mathbf{x}), \dots, z_p(\mathbf{x})$, each a function of independent variables x_1, \dots, x_q . She is concerned that the true model might be only approximated by (1), with a more precise description being

$$E(Y|\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta} + f(\mathbf{x}) \tag{2}$$

for some unknown but “small” function f . In this situation she would like to choose design points that yield estimates $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, and estimates $\hat{Y}(\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\hat{\boldsymbol{\theta}}$ of $E(Y|\mathbf{x})$, 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 $\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$ the projection of f on the regressors; we may then assume that f and $\mathbf{z}(\cdot)$ are orthogonal in $L^2 = L^2(S, d\mathbf{x})$. This still leaves open the possibility that $E(Y|\mathbf{x}) = f(\mathbf{x})$ 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

$$Y(\mathbf{x}_i) = E(Y|\mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, n$$

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

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

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 and Draper (1959), Stigler (1971), Andrews and Herzberg (1979), Li and Notz (1982), Pesotchinsky (1982), Sacks and Ylvisaker (1984), Dette and Wong (1996), Liu and 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 $\hat{Y}(\cdot)$. Such designs have been constructed only for particularly well-structured problems. See Huber (1975, 1981) for the case of straight line regression ($\mathbf{z}(x) = (1, x)^T$) over $S = [-1/2, 1/2]$, with extensions by Wiens (1990, 1992) to the case of multiple linear regression: $\mathbf{z}(\mathbf{x}) = (1, x_1, \dots, x_q)^T$ with S a sphere in \mathbb{R}^q , as well as to the partial second order model with interactions: $\mathbf{z}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)^T$, $S = [-1/2, 1/2] \times [-1/2, 1/2]$. 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. In the case studies which are undertaken in Section 4 we illustrate two methods of approximating and implementing such designs.

2. UNRESTRICTED MINIMAX DESIGNS

An exactly implementable design will correspond to a *design measure* ξ placing mass n^{-1} at each of $\mathbf{x}_1, \dots, \mathbf{x}_n$. 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 although the true model is (2), the least squares estimator $\hat{\boldsymbol{\theta}}$ is biased. With $\mathbf{b}(f, \xi) := \int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})\xi(d\mathbf{x})$ and $\mathbf{A}_\xi := \int_S \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})\xi(d\mathbf{x})$ assumed

non-singular, the bias is $E[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta} = \mathbf{A}_\xi^{-1} \mathbf{b}(f, \xi)$ and the mean squared error matrix is

$$\begin{aligned} \text{MSE}(f, \xi) &= E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \\ &= (\sigma^2/n) \mathbf{A}_\xi^{-1} + \mathbf{A}_\xi^{-1} \mathbf{b}(f, \xi) \mathbf{b}^T(f, \xi) \mathbf{A}_\xi^{-1}. \end{aligned}$$

We consider the loss functions $\mathcal{L}_Q =$ integrated MSE of the fitted responses $\hat{Y}(\mathbf{x})$, $\mathcal{L}_D =$ 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 are, with $\mathbf{A}_0 := \int_S \mathbf{z}(\mathbf{x}) \mathbf{z}^T(\mathbf{x}) d\mathbf{x}$, given by

$$\begin{aligned} \mathcal{L}_Q(f, \xi) &= \int_S E \left[\left\{ \hat{Y}(\mathbf{x}) - E(Y|\mathbf{x}) \right\}^2 \right] d\mathbf{x} \\ &= (\sigma^2/n) \text{tr}(\mathbf{A}_\xi^{-1} \mathbf{A}_0) + \mathbf{b}^T(f, \xi) \mathbf{A}_\xi^{-1} \mathbf{A}_0 \mathbf{A}_\xi^{-1} \mathbf{b}(f, \xi) + \int_S f^2(\mathbf{x}) d\mathbf{x}, \end{aligned} \quad (4)$$

$$\mathcal{L}_D(f, \xi) = \det(\text{MSE}(f, \xi)) = (\sigma^2/n)^p \frac{1}{|\mathbf{A}_\xi|} \left(1 + \frac{n}{\sigma^2} \mathbf{b}^T(f, \xi) \mathbf{A}_\xi^{-1} \mathbf{b}(f, \xi) \right), \quad (5)$$

$$\mathcal{L}_A(f, \xi) = \text{tr}(\text{MSE}(f, \xi)) = (\sigma^2/n) \text{tr} \mathbf{A}_\xi^{-1} + \mathbf{b}^T(f, \xi) \mathbf{A}_\xi^{-2} \mathbf{b}(f, \xi). \quad (6)$$

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

Lemma 2.1. *Suppose that $\|\mathbf{z}(\mathbf{x})\|$ is bounded in \mathbf{x} on S , and that for each $\mathbf{a} \neq \mathbf{0}$ the set $\{\mathbf{x} : \mathbf{a}^T \mathbf{z}(\mathbf{x}) = 0\}$ 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 $m(\cdot)$ satisfying $\int_S \|\mathbf{z}(\mathbf{x})\|^2 m^2(\mathbf{x}) d\mathbf{x} < \infty$.*

Theorem 2.2. *Let S and ξ be as in Lemma 2.1. Define matrices $\mathbf{K}_\xi = \int_S \mathbf{z}(\mathbf{x}) \mathbf{z}^T(\mathbf{x}) m^2(\mathbf{x}) d\mathbf{x}$, $\mathbf{H}_\xi = \mathbf{A}_\xi \mathbf{A}_0^{-1} \mathbf{A}_\xi$ 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*

$$\max_{\mathcal{F}} \mathcal{L}_Q(f, \xi) = \eta^2 \left[\nu \text{tr}(\mathbf{A}_\xi^{-1} \mathbf{A}_0) + \lambda_{\max}(\mathbf{K}_\xi \mathbf{H}_\xi^{-1}) \right], \quad (7)$$

$$\max_{\mathcal{F}} \mathcal{L}_D(f, \xi) = \eta^2 (\sigma^2/n)^{p-1} \left[\nu + \lambda_{\max}(\mathbf{G}_\xi \mathbf{A}_\xi^{-1}) \right] / |\mathbf{A}_\xi|, \quad (8)$$

$$\max_{\mathcal{F}} \mathcal{L}_A(f, \xi) = \eta^2 \left[\nu \text{tr}(\mathbf{A}_\xi^{-1}) + \lambda_{\max}(\mathbf{G}_\xi \mathbf{A}_\xi^{-2}) \right], \quad (9)$$

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 2.1. Wiens (1992) considered the approximate multiple linear regression model, with $\mathbf{x} = (x_1, \dots, x_q)^T$ varying over a q -dimensional sphere S centred at the origin, and $\mathbf{z}(\mathbf{x}) = (1, \mathbf{x}^T)^T$. 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 2.2 below were encountered. The A -optimality case is reconsidered in Example 3.1.

Example 2.2. We illustrate some of the difficulties which can be encountered in the minimax approach without further restrictions on the design density, by considering approximate quadratic regression: $\mathbf{z}(x) = (1, x, x^2)^T$, 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}_ξ are

$$\begin{aligned} H_{11} = h_0 &:= 9/4 - 30\alpha_2 + 180\alpha_2^2, & H_{13} = H_{31} = h_1 &:= 9\alpha_2/4 - 15\alpha_4 - 15\alpha_2^2 + 180\alpha_2\alpha_4, \\ H_{22} = h_2 &:= 12\alpha_2^2, & H_{33} = h_3 &:= 9\alpha_2^2/4 - 30\alpha_2\alpha_4 + 180\alpha_4^2, \end{aligned}$$

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

$$l_i(\xi) = \nu \operatorname{tr}(\mathbf{A}_\xi^{-1} \mathbf{A}_0) + \lambda_i(\xi), \quad i = 0, 1; \quad l(\xi) = \max(l_0(\xi), l_1(\xi)).$$

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

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

Step 2: Put $\xi_* = \begin{cases} \xi_0, & \text{if } l_0(\xi_0) \leq l_1(\xi_1); \\ \xi_1, & \text{otherwise.} \end{cases}$

It follows that $l(\xi_*) \leq \min(l_0(\xi_0), l_1(\xi_1))$ 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 2.1 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 $l_i(\xi)$ among all designs ξ , $i = 0, 1$.

Step 2': Put $\xi_* = \begin{cases} \xi_0, & \text{if } \lambda_0(\xi_0) \geq \lambda_1(\xi_0); \\ \xi_1, & \text{if } \lambda_1(\xi_1) \geq \lambda_0(\xi_1). \end{cases}$

In this example Step 1' may be carried out in stages, by first fixing $\alpha_0 (= 1), \alpha_2$ and α_4 . This fixes \mathbf{A}_ξ 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 $m_0(x) = (a - b/x^2 + cx^2)^+$. 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 $m_1(x) = ((a + bx^2 + cx^4) / (d + ex^2 + fx^4))^+$. 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 is taken in the next section.

Example 2.3. The situation for approximate quadratic regression is somewhat simpler if there is no intercept: $\mathbf{z}(x) = (x, x^2)^T$. In this case

$$\begin{aligned} \lambda_0(\xi) &= \frac{k_2}{12\alpha_2^2}, \quad \lambda_1(\xi) = \frac{k_4}{80\alpha_4^2}, \\ l_i(\xi) &= \nu \left(\frac{1}{12\alpha_2} + \frac{1}{80\alpha_4} \right) + \lambda_i(\xi). \end{aligned}$$

To carry out Step 1' above, we first minimise k_{2i+2} for fixed α_0, α_2 and α_4 by minimising

$$\int_S x^{2i+2} m^2(x) + 2(a - bx^2 - cx^4) m(x) dx$$

for Lagrange multipliers a, b, c . The integrand is minimised pointwise by

$$m_i(x) = \frac{1}{x^{2i}} \left(cx^2 + b - \frac{a}{x^2} \right)^+, \quad i = 0, 1,$$

with a, b, c determined by the side conditions and α_2, α_4 then determined as in Example 2.3. An equivalent and numerically simpler procedure is to set $m_i(x) = x^{-2i} (c'x^2 + b' - a'/x^2)^+ / J_i$, where $J_i := \int_S x^{-2i} (c'x^2 + b' - a'/x^2)^+ dx$, to minimise the loss at ξ_i over a', b', c' unconditionally, and to then recover the original parameters from $(a, b, c) = (a', b', c') / J_i$.

We find that the first inequality in Step 2' holds for all ν , so that ξ_0 is the minimax design. See Table 3 for some numerical values, and Example 3.3 for a different approach to this problem.

Example 2.4. Wiens (1990) found that for the partial second order regression model with interactions: $\mathbf{z}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)^T$ 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 2.2 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 3.4 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

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

with the β_j restricted in such a way that $m(\cdot)$ 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 2.2.

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 3.1 they generally have the limiting behaviour that one would expect, tending to the continuous uniform design as $\nu \rightarrow 0$ and to the classical, variance-minimising designs as

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

ν	Unrestricted design ¹			Restricted design ²		
	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	5233.269	-485.606	2125.479	5233.276

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

$\nu \rightarrow \infty$. As well the designs are numerically straightforward, having the same parametric form regardless of the structure of the eigenvalues which appear in Theorem 2.2. This fact has enabled us to construct the restricted designs in cases not readily amenable to an unrestricted treatment.

Theorem 3.1. *Assume that S is a compact subset of \mathbb{R}^p satisfying the conditions of Lemma 2.1 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 $\nu \rightarrow 0$ satisfies $B(\xi_0) = \inf_{\xi \in \Xi'} B(\xi)$, and (ii) any weak limit point ξ_∞ of ξ_ν as $\nu \rightarrow \infty$ satisfies $V(\xi_\infty) = \inf_{\xi \in \Xi'} V(\xi)$.*

To apply Theorem 3.1 in the case $\nu \rightarrow 0$, 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 3.1, $\inf_{\xi \in \Xi'} B(\xi) = \inf_{\xi \in \Xi} B(\xi) = \lim_{\nu \rightarrow 0} B(\xi_\nu)$. In the case $\nu \rightarrow \infty$, 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 $V(\xi'_\nu) \rightarrow V(\xi_\infty)$ by Theorem 3.1 and so $\inf_{\xi \in \Xi'} V(\xi) = \inf_{\xi \in \Xi} V(\xi) = \lim_{\nu \rightarrow \infty} V(\xi_\nu)$. The details of such constructions are straightforward in particular examples, and will not be given here.

Example 3.1. For the model of Example 2.1, (10) gives $m(\mathbf{x}) = \left(\beta_0 + \sum_{j=1}^q \beta_j x_j^2\right)^+$. With $a := \beta_0$, and $b := \beta_1 = \dots = \beta_q$ for exchangeability, this density agrees exactly with the

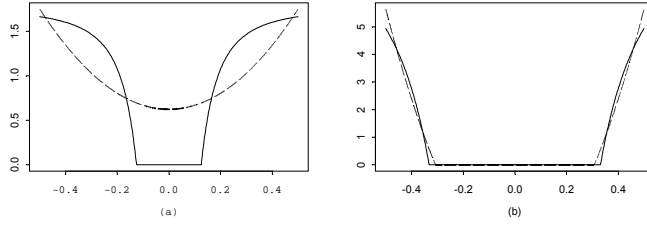


Figure 1: Unrestricted (solid lines) and restricted (broken lines) A -optimal minimax densities for the approximate straight line model. (a) $\nu = 0.445$; (b) $\nu = 10$. Explicit descriptions of the densities are in Table 1.

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 $\nu \geq .445$. For moderately large ν the loss of the restricted minimax design is only marginally greater than that of the unrestricted design. As $\nu \rightarrow \infty$ both designs approach the variance-minimising design with mass of .5 at each of $\pm 1/2$.

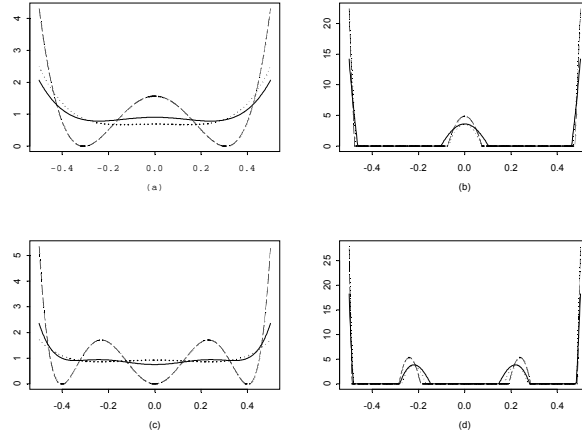


Figure 2: Q -optimal (solid lines), D -optimal (dotted lines) and A -optimal (dashed lines) minimax densities for approximate degree- q polynomial regression. (a) $q = 2$, $\nu = 1$; (b) $q = 2$, $\nu = 100$; (c) $q = 3$, $\nu = 1$; (d) $q = 3$, $\nu = 100$. Explicit descriptions of the densities are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

Example 3.2. For the polynomial model with $\mathbf{z}(x) = (1, x, \dots, x^q)^T$, Ξ' consists of those

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 $m_*(x) = \alpha(x^4 + \beta_1x^2 + \beta_2)^+$				² $m_*(x) = \alpha(x^6 + \beta_1x^4 + \beta_2x^2 + \beta_3)^+$			

designs with densities

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

See Figure 2 for plots in the quadratic and cubic cases, with values of the constants in Table 2. As noted previously by Studden (1977) for variance-minimising designs, and Wiens (1999) for bias-minimising designs, the Q - and D -optimal designs are very similar. As $\nu \rightarrow \infty$ all three tend to their variance-minimising counterparts.

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 to then replace the replicates at these sites by groups of observations at distinct but nearby sites. This observation is reinforced in the first case study of Section 4.

Example 3.3. For the quadratic model without intercept, as in Example 2.3, the designs in Ξ' have densities $m(x; a, b) = (ax^2 + bx^4)^+$. See Table 3 for some numerical values. The maximum loss values are seen to be almost identical to those of the unrestricted minimax design given in Example 2.3. The relative difference between the losses is greatest near $\nu = 0$, this being caused by the fact that, lacking a constant term, the restricted density cannot approach the (unbiased) uniform density. Examples of both the unrestricted and restricted minimax densities are shown in Figure 3.

Example 3.4. For the partial second order model considered in Example 2.4, Ξ' again contains the unrestricted minimax design. For the full second order model with $q = 2$:

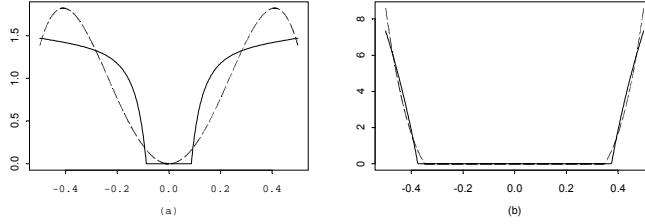


Figure 3: Unrestricted (solid lines) and restricted (broken lines) Q -optimal minimax densities for the approximate, no intercept quadratic model with regressors (x, x^2) . (a) $\nu = 0.1$; (b) $\nu = 10$. Explicit descriptions of the densities are in Table 3.

Table 3. Numerical values for the approximate, no-intercept quadratic model with regressors $(x, x^2)^T$ and the unrestricted and restricted Q -optimal minimax densities

ν	Unrestricted design ¹				Restricted design ²		
	a	b	c	$loss$	a	b	$loss$
0	0.000	1.000	0.000	1.000	22.703	-71.351	1.055
0.1	0.011	1.436	0.309	1.160	21.707	-64.716	1.184
1	0.147	2.798	2.398	2.157	12.355	-2.370	2.216
10	1.763	10.097	17.153	8.910	-32.630	267.951	8.944
100	15.449	31.572	207.962	63.262	-415.472	2027.816	63.272
1000	165.233	279.163	1795.547	563.415	-4348.981	18489.91	563.416

$${}^1m_*(x) = (cx^2 + b - a/x^2)^+ \quad {}^2m_*(x) = (ax^2 + bx^4)^+$$

$\mathbf{z}(x_1, x_2) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)^T$, the designs in Ξ' have densities

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

See Figure 4 for plots of the Q -, D - and A -optimal design densities when $\nu = 5$. Explicit expressions are:

$$Q\text{-optimality: } m_*(x_1, x_2) = 216.419 (x_1^4 + x_2^4 + 0.306x_1^2x_2^2 - 0.210(x_1^2 + x_2^2) + 0.011)^+,$$

$$D\text{-optimality: } m_*(x_1, x_2) = 369.556 (x_1^4 + x_2^4 + 0.430x_1^2x_2^2 - 0.213(x_1^2 + x_2^2) + 0.007)^+,$$

$$A\text{-optimality: } m_*(x_1, x_2) = 1.856 (x_1^4 + x_2^4 + 0.442x_1^2x_2^2 + 2.168(x_1^2 + x_2^2) + 0.149).$$

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.

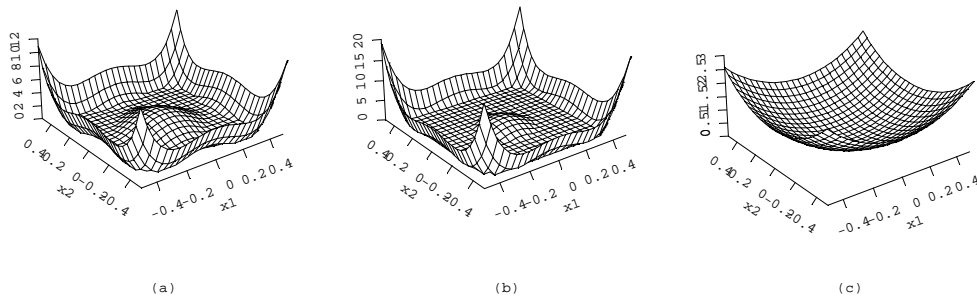


Figure 4: Restricted minimax design densities for approximate, full second order model with $\nu = 5$. (a) Q -optimal density; (b) D -optimal density; (c) A -optimal density.

4. CASE STUDIES

We discuss two case studies - one arising in the oncology literature and one from a consulting project undertaken by one of us (Heo). In the first case, a polynomial response is to be fitted. A design as in Example 3.2 is implemented by choosing the sites x_i to be n uniformly spaced quantiles of the minimax design measure. In the second case, a second order response is anticipated. A design as in Example 3.4 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, in either case, we construct discrete measures ξ_n which have the property that they converge in measure to the minimax design ξ_* as $n \rightarrow \infty$. In each case the finite sample implementation is intuitively sensible as well as robust. A balance is struck between full efficiency and robustness by placing 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. Dose response experimentation

In typical bioassays or dose response experiments, one observes the proportion p_x of subjects exhibiting a particular response as a result of exposure to, or administration of, an agent at level x . An objective is to estimate the probability $P(x)$ of the occurrence of the response. To convert this to a regression-based problem it is usual to transform to the p_x -quantile $Y = G^{-1}(p_x)$ for a suitable distribution G . For definiteness we assume G to be the logistic distribution, so that $Y = \ln(p_x/(1 - p_x))$ is the logit. The regression function

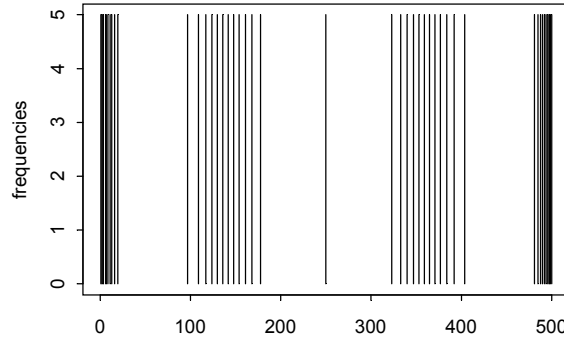


Figure 5: Implementation of the D -optimal design of §4.1; approximate cubic response.

$E(Y|x) = E(G^{-1}(p_x))$ is then approximated by $G^{-1}(P(x))$. Since $P(x)$ is unknown a further approximation - $E(Y|x) \approx \mathbf{z}^T(x)\boldsymbol{\theta}$ - is often made, where $\mathbf{z}^T(x)\boldsymbol{\theta}$ is a polynomial in x , typically of low degree. The model described by (2) and (3) is then appropriate. Of course $\text{VAR}(Y|x)$ may vary with x as well, due to the nature of the data as proportions and to the transformation. As in Wiens (1998) this can also be incorporated at the design stage, although we have not done so in this example.

Hoel and Jennrich (1979) describe an experiment cited by Guess, Crump and Peto (1977) in which 235 experimental animals are exposed to varying doses of a carcinogen. They consider the construction of designs for extrapolation in this situation, assuming that $E(Y|x)$ is a cubic polynomial. The range of x is $[1, 500]$ and their design for minimum variance estimation of $P(1/2)$ places $n_x = (63, 125, 35, 12)$ observations at $x = (1, 82.6, 342, 500)$ respectively. Wong and Lachenbruch (1996) also consider cubic estimation in dose response experimentation; the variance minimising D -optimal design exhibited by them and transformed to $[1, 500]$ places an equal number of observations at each of $x = (1, 138.9, 362.1, 500)$. Note that neither of these designs allows for the estimation of polynomial responses of degree higher than three.

We have computed an implementation of the restricted minimax D -optimal design for approximate cubic regression, as in Example 3.2 with $\nu = 10$. The density is $m_*(x) = 5633.710(x^6 - 0.3128x^4 + 0.0239x^2 - .0002)^+$ on $[-1/2, 1/2]$. Since experimentation of this type requires replication, we chose to place $n_x = 5$ observation at each of $N = 47$ sites x_i , with x_i being the $((i - 1)/(N - 1))$ -quantile of the minimax design measure ξ_* . Thus with

$N_0 := (N - 1)/2$ we found points $\{x_i\}_{i=1}^{N_0-1} \subset (0, 1/2)$ satisfying $\int_{x_{i-1}}^{x_i} m_*(x)dx = (2N_0)^{-1}$, with $x_0 := 0$. The design is then $\{\pm x_i\}_{i=1}^{N_0-1}$ together with $\{0, \pm 1/2\}$. When transformed to $[1, 500]$ this gave the design of Figure 5. As anticipated in Example 3.2, this minimax robust design can be viewed as being obtained from the variance-minimising design by breaking up the four groups of replicates into clusters of replicates. The averages of the clusters are $\{4.28, 139.76, 361.24, 496.72\}$. Smaller values of ν lead to a more uniform spread to the clusters; larger values to closer agreement with the variance-minimising design.

4.2. Wastewater ozonation

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 and 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 3.4.

Both factors were linearly transformed to the range $[-1/2, 1/2]$. The Q -optimal design ξ_* , with $\nu = 5$ as in Figure 4(a), was then implemented as follows to yield $n = 48$ design points. We chose $n_0 := n/8$ points (x_1, x_2) in $\{0 \leq x_1 \leq x_2 \leq 1/2\}$ 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 $e_{2j,2k} := \sum_{i=1}^{n_0} (x_{1i}^{2j} x_{2i}^{2k} + x_{1i}^{2k} x_{2i}^{2j}) / (2n_0)$ matched up as closely as possible with the theoretical moments $E_{\xi_*}[X_1^{2j} X_2^{2k}]$ obtained from ξ_* . We did this for the $J(J + 3)/2$ choices (k, j) with $k = 0, \dots, j$ and $j = 1, \dots, J$, 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

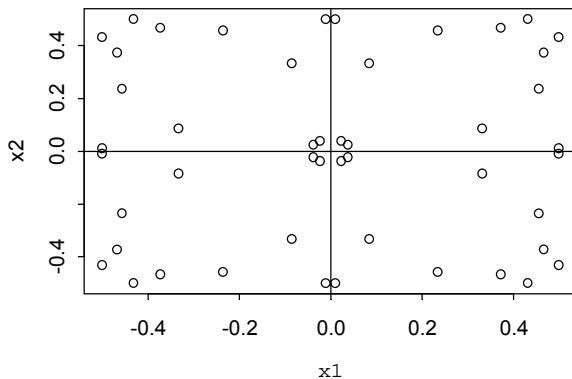


Figure 6: Implementation of the Q -optimal design of Figure 4(a) for approximate second order regression; $n = 48$.

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 $\sum_{j,k} (e_{2j,2k} - E_{\xi_*}[X_1^{2j} X_2^{2k}])^2$, yielding the implementation shown in Figure 6 with

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

5. SUMMARY

We have presented new, parametric classes of regression designs. Within several such classes we have isolated members which 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 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 which are optimal with respect to generalisations of the common Q -, D - and A -optimality criteria. Two implementation methods have been illustrated in the case studies. 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 2.1: The assumption on S implies that \mathbf{A}_0 is invertible. Lemma 1 of Wiens (1992) then states that if $\sup_{\mathcal{F}} \mathcal{L}(f, \xi)$ is finite, ξ is absolutely continuous with respect to Lebesgue measure. Let $m(\mathbf{x})d\mathbf{x} = \xi(d\mathbf{x})$ represent its density.

Suppose that $\int_S \|\mathbf{z}(\mathbf{x})\|^2 m(\mathbf{x})^2 d\mathbf{x} = \infty$. Choose an index j so that $\int_S z_j(\mathbf{x})^2 m^2(\mathbf{x}) d\mathbf{x} = \infty$ and let $f_n(\mathbf{x}) = m(\mathbf{x})z_j(\mathbf{x})1_{\{m \leq n\}}(\mathbf{x})$. This is a bounded function and $\|f_n\|_{L^2} \rightarrow \infty$ as $n \rightarrow \infty$. Define $g_n(\mathbf{x}) = f_n(\mathbf{x}) - c_n(\mathbf{x})$, where $c_n(\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\mathbf{A}_0^{-1} \int_S f_n(\mathbf{y})\mathbf{z}(\mathbf{y})d\mathbf{y}$. Note that $\|\int_S f_n(\mathbf{y})\mathbf{z}(\mathbf{y})d\mathbf{y}\| \leq \int_S \|\mathbf{z}(\mathbf{y})\|^2 m(\mathbf{y})d\mathbf{y} \leq \sup_S \|\mathbf{z}(\mathbf{x})\|^2$, and so the functions c_n are uniformly bounded and $|\|f_n\|_{L^2} - \|g_n\|_{L^2}| \leq \|c_n\|_{L^2}$. Now let $h_n(\mathbf{x}) = \eta g_n(\mathbf{x}) / \|g_n\|_{L^2}$ so that $h_n \in \mathcal{F}$. The j^{th} coordinate of $\mathbf{b}(h_n, \xi)$ is

$$\begin{aligned} \int_S h_n(\mathbf{x})z_j(\mathbf{x})m(\mathbf{x})d\mathbf{x} &= (\eta / \|g_n\|_{L^2}) \left\{ \int_S m^2(\mathbf{x})z_j^2(\mathbf{x})1_{\{m \leq n\}}(\mathbf{x})d\mathbf{x} - \int_S m(\mathbf{x})z_j(\mathbf{x})c_n(\mathbf{x})d\mathbf{x} \right\} \\ &= \eta \left\{ \frac{\|f_n\|_{L^2}^2 - d_n}{\|f_n\|_{L^2} - e_n} \right\}, \end{aligned}$$

where (d_n) and (e_n) are bounded sequences of numbers. Since $\|f_n\|_{L^2} \rightarrow \infty$, we see that $\|\mathbf{b}(h_n, \xi)\| \rightarrow \infty$ as $n \rightarrow \infty$. This implies that $\sup_{\mathcal{F}} \mathcal{L}(f, \xi) = \infty$ which gives a contradiction.

Proof of Theorem 2.2: Note that

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

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

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

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

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

The maxima of the three quadratic forms over β are $\lambda_{\max}(\mathbf{I} + \mathbf{G}_\xi^{1/2} \mathbf{H}_\xi^{-1} \mathbf{G}_\xi^{1/2}) = \lambda_{\max}(\mathbf{K}_\xi \mathbf{H}_\xi^{-1})$, $\lambda_{\max}(\mathbf{G}_\xi^{1/2} \mathbf{A}_\xi^{-1} \mathbf{G}_\xi^{1/2}) = \lambda_{\max}(\mathbf{G}_\xi \mathbf{A}_\xi^{-1})$ 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}_ξ is non-singular, the inclusion (A.1) is proven as Theorem 1 of Wiens (1992). If \mathbf{G}_ξ is singular (as at the continuous uniform design) we proceed as follows. Take any design ξ_1 for which the corresponding matrix \mathbf{G}_{ξ_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

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

for sufficiently small $t > 0$. We have

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

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

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

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

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

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

Proof of Theorem 3.1: We begin with some preliminary calculations. First note that $\|\mathbf{A}_\xi\| := \lambda_{\max}(\mathbf{A}_\xi)$ is uniformly bounded in ξ :

$$\begin{aligned}\|\mathbf{A}_\xi\| &= \sup_{\|\beta\|=1} \int_S (\beta^T \mathbf{z}(\mathbf{x}))^2 \xi(d\mathbf{x}) \leq \int_S \sup_{\|\beta\|=1} (\beta^T \mathbf{z}(\mathbf{x}))^2 \xi(d\mathbf{x}) \\ &= \int_S \|\mathbf{z}(\mathbf{x})\|^2 \xi(d\mathbf{x}) \leq \sup_{\mathbf{x} \in S} \|\mathbf{z}(\mathbf{x})\|^2 =: M < \infty.\end{aligned}$$

This also shows that $\|\mathbf{H}_\xi\| = \|\mathbf{A}_\xi \mathbf{A}_0^{-1} \mathbf{A}_\xi\| \leq \|\mathbf{A}_\xi\|^2 \|\mathbf{A}_0^{-1}\| \leq M^2 \|\mathbf{A}_0^{-1}\|$ is uniformly bounded in ξ . Consider the bias term B_Q in \mathcal{L}_Q ; the others are very similar. The calculation above implies that $\lambda_{\min}(\mathbf{A}_\xi^{-1} \mathbf{A}_0 \mathbf{A}_\xi^{-1}) = \lambda_{\max}^{-1}(\mathbf{A}_\xi \mathbf{A}_0^{-1} \mathbf{A}_\xi) \geq M^{-2} \|\mathbf{A}_0^{-1}\|^{-1} := \varepsilon > 0$. Therefore

$$B_Q(\xi) = \sup_{\|\beta\|=1} \beta^T \mathbf{G}_\xi^{1/2} \mathbf{A}_\xi^{-1} \mathbf{A}_0 \mathbf{A}_\xi^{-1} \mathbf{G}_\xi^{1/2} \beta \geq \varepsilon \sup_{\|\beta\|=1} \left\| \mathbf{G}_\xi^{1/2} \beta \right\|^2 = \varepsilon \|\mathbf{G}_\xi\|.$$

This gives us

$$\int_S m^2(\mathbf{x}) \|\mathbf{z}(\mathbf{x})\|^2 d\mathbf{x} = \text{tr}(\mathbf{K}_\xi) = \text{tr}(\mathbf{G}_\xi) + \text{tr}(\mathbf{H}_\xi) \leq c(B(\xi) + 1),$$

for some constant c .

Now fix $\nu > 0$, and let $\xi_n \in \Xi'$ satisfy $\lim_{n \rightarrow \infty} \sup_{f \in \mathcal{F}} \mathcal{L}(f, \xi_n) = \inf_{\xi \in \Xi'} \sup_{f \in \mathcal{F}} \mathcal{L}(f, \xi)$. In particular, $\sup_n B(\xi_n) < \infty$, so that the corresponding densities $m_n(\mathbf{x}) = (\beta_n^T \mathbf{z}(x_1^2, \dots, x_q^2))^+$ are bounded in $L^2(S; \|\mathbf{z}(\mathbf{x})\|^2 d\mathbf{x})$. By taking subsequences we may assume that $\beta_n / \|\beta_n\|$ converges to $\mathbf{s} \in \mathbb{R}^p$ with $\|\mathbf{s}\| = 1$, that m_n converges weakly to m_ν in $L^2(S; \|\mathbf{z}(\mathbf{x})\|^2 d\mathbf{x})$, and that ξ_n converges weakly to some probability measure ξ_ν . For any of the loss functions \mathcal{L}_Q , \mathcal{L}_D , or \mathcal{L}_A , it is not hard to show that the map $\xi \mapsto \sup_{f \in \mathcal{F}} \mathcal{L}(f, \xi)$ is lower semicontinuous from Ξ into $[0, \infty]$. Therefore $\sup_{f \in \mathcal{F}} \mathcal{L}(f, \xi_\nu) = \inf_{\xi \in \Xi'} \sup_{f \in \mathcal{F}} \mathcal{L}(f, \xi)$; that is, ξ_ν is minimax. It remains to show that $\xi_\nu \in \Xi'$.

We begin by showing that $\sup_n \|\beta_n\| < \infty$. Suppose not and write $\beta_n = \|\beta_n\| (\mathbf{s} + \mathbf{e}_n)$ where $\mathbf{e}_n \rightarrow 0$. Then $\beta_n^T \mathbf{z} = \|\beta_n\| (\mathbf{s}^T \mathbf{z} + \mathbf{e}_n^T \mathbf{z})$ diverges to ∞ if $\mathbf{s}^T \mathbf{z} > 0$, and diverges to $-\infty$ if $\mathbf{s}^T \mathbf{z} < 0$. Therefore we see that $m_n(\mathbf{x}) \rightarrow \infty$ on the set $\{\mathbf{x} : \mathbf{s}^T \mathbf{z}(x_1^2, \dots, x_q^2) > 0\}$ and $m_n(\mathbf{x}) \rightarrow 0$ on the set $\{\mathbf{x} : \mathbf{s}^T \mathbf{z}(x_1^2, \dots, x_q^2) < 0\}$. The conditions in Lemma 2.1 show that the set $\{\mathbf{x} : \mathbf{s}^T \mathbf{z}(x_1^2, \dots, x_q^2) = 0\}$ has Lebesgue measure zero. The divergence $m_n(\mathbf{x}) \rightarrow \infty$ contradicts the weak L^2 convergence of m_n unless $\{\mathbf{x} : \mathbf{s}^T \mathbf{z}(x_1^2, \dots, x_q^2) > 0\}$ also has

Lebesgue measure zero. Thus we conclude that $m_\nu = 0$ Lebesgue almost surely. For every continuous function h with closed support in the set $\{\mathbf{x} : \|\mathbf{z}(\mathbf{x})\| > 0\}$ this gives

$$\begin{aligned} \int_S h(\mathbf{x}) \xi_n(d\mathbf{x}) &= \int_S \frac{h(\mathbf{x})}{\|\mathbf{z}(\mathbf{x})\|^2} m_n(\mathbf{x}) \|\mathbf{z}(\mathbf{x})\|^2 d\mathbf{x} \\ &\rightarrow \int_S \frac{h(\mathbf{x})}{\|\mathbf{z}(\mathbf{x})\|^2} m_\nu(\mathbf{x}) \|\mathbf{z}(\mathbf{x})\|^2 d\mathbf{x} = 0 = \int_S h(\mathbf{x}) \xi_\nu(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

Since $\{\mathbf{x} : \|\mathbf{z}(\mathbf{x})\| > 0\}$ has full Lebesgue measure, we conclude that ξ_ν is singular with respect to Lebesgue measure, contradicting Lemma 2.1. Therefore we find that $\sup_n \|\beta_n\| < \infty$ must hold.

Hence, by taking a further subsequence, we can assume that $\beta_n \rightarrow \beta \in \mathbb{R}^p$. But then the sequence m_n converges uniformly to the function $m_\nu(\mathbf{x}) = (\beta^T \mathbf{z}(x_1^2, \dots, x_q^2))^+$, which must be the density of ξ_ν . This proves that $\xi_\nu \in \Xi'$.

Since ξ_ν is minimax, for any $\xi \in \Xi'$ we have $B(\xi_\nu) \leq \nu V(\xi_\nu) + B(\xi_\nu) \leq \nu V(\xi) + B(\xi)$. Taking the limit as $\nu \rightarrow 0$, and then the infimum over $\xi \in \Xi'$ shows that $\lim_{\nu \rightarrow 0} B(\xi_\nu) = \inf_{\xi \in \Xi'} B(\xi)$. For any sequence $\nu_n \rightarrow 0$ let $\xi_0 \in \Xi$ so that $\xi_{\nu_n} \rightarrow \xi_0$ weakly. Since $\sup_n B(\xi_{\nu_n}) < \infty$, the argument above shows that we may assume that $\xi_0 \in \Xi'$ and that $m_n \rightarrow m_0$ uniformly. It is then easy to see that $B(\xi_{\nu_n}) \rightarrow B(\xi_0)$, so that $B(\xi_0) = \inf_{\xi \in \Xi'} B(\xi)$.

Similar arguments show that $\lim_{\nu \rightarrow \infty} V(\xi_\nu) = \inf_{\xi \in \Xi'} V(\xi)$. For any sequence $\nu_n \rightarrow \infty$ let $\xi_\infty \in \Xi$ so that $\xi_{\nu_n} \rightarrow \xi_\infty$ weakly. Since $\mathbf{x} \mapsto \mathbf{z}(\mathbf{x})$ is continuous, the map $\xi \mapsto \mathbf{A}_\xi$ is also continuous. Therefore the map $\xi \mapsto V(\xi)$ is continuous for any of our loss functions, since we have $V_Q(\xi) = \text{tr}(\mathbf{A}_\xi^{-1} \mathbf{A}_0)$, $V_D(\xi) = |\mathbf{A}_\xi^{-1}|$, and $V_A(\xi) = \text{tr}(\mathbf{A}_\xi^{-1})$. Then we have $V(\xi_\infty) = \lim_n V(\xi_{\nu_n}) = \inf_{\xi \in \Xi'} V(\xi)$, so ξ_∞ is variance minimising in Ξ' . If the variance minimising measure is unique, we can even conclude that $\lim_{\nu \rightarrow \infty} \xi_\nu = \xi_\infty$.

REFERENCES

- Andrews, D., and Herzberg, A. (1979), "The Robustness and Optimality of Response Surface Designs," *Journal of Statistical Planning and Inference*, 3, 249-257.
- 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.
- Dette, H., and Wong, W.K. (1996), "Robust Optimal Extrapolation Designs," *Biometrika*, 83, 667-680.

- Fedorov, V.V. (1972), *Theory of Optimal Experiments*, New York: Academic Press.
- Guess, H., Crump, K. and Peto, R. (1977), "Uncertainty Estimates for Low Dose Extrapolation of Animal Carcinogenicity Data," *Cancer Research*, 37, 3475-3483.
- Heo, G. (1998), "Optimal Designs for Approximately Polynomial Regression Models," unpublished Ph.D. dissertation, University of Alberta, Dept. of Mathematical Sciences.
- Heo, G. and James, W. (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.
- Hoel, P.G., and Jennrich, R.I. (1979), "Optimal Designs for Dose Response Experiments in Cancer Research," *Biometrika*, 66, 307-316.
- Huber, P.J. (1975), "Robustness and Designs," in: *A Survey of Statistical Design and Linear Models*, ed. J.N. Srivastava, Amsterdam: North Holland, pp. 287-303.
- Huber, P.J. (1981), *Robust Statistics*, New York: Wiley.
- Li, K.C., and Notz, W. (1982), "Robust Designs for Nearly Linear Regression," *Journal of Statistical Planning and Inference*, 6, 135-151.
- Liu, S.X., and Wiens, D.P. (1997), "Robust Designs for Approximately Polynomial Regression," *Journal of Statistical Planning and Inference*, 64, 369-381.
- Pesotchinsky, L. (1982), "Optimal Robust Designs: Linear Regression in R^k ," *The Annals of Statistics*, 10, 511-525.
- Sacks, J., and Ylvisaker, D. (1984), "Some Model Robust Designs in Regression," *The Annals of Statistics*, 12, 1324-1348.
- Stigler, S. (1971), "Optimal Experimental Designs for Polynomial Regression," *Journal of the American Statistical Association*, 66, 311-318.
- Studden, W.J. (1977), "Optimal Designs for Integrated Variance in Polynomial Regression," *Statistical Decision Theory and Related Topics II*, ed. Gupta, S.S. and Moore, D.S. New York: Academic Press, pp. 411-420.

- Wiens, D.P. (1990), "Robust, Minimax Designs for Multiple Linear Regression," *Linear Algebra and Its Applications, Second Special Issue on Linear Algebra and Statistics*; 127, 327-340.
- Wiens, D.P. (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference*, 31, 353-371.
- Wiens, D.P. (1998), "Minimax Robust Designs and Weights for Approximately Specified Regression Models With Heteroscedastic Errors," *Journal of the American Statistical Association*, 93, 1440-1450.
- Wiens, D.P. (1999), "Robust Weights and Designs for Biased Regression Models: Least Squares and Generalized M-estimation," *Journal of Statistical Planning and Inference*, to appear.
- Wong, W.K., and Lachenbruch, P.A. (1996), "Tutorial in Biostatistics: Designing Studies for Dose Response," *Statistics in Medicine* , 15, 343-359.