Minimax Robust Designs and Weights for Approximately Specified Regression Models With Heteroscedastic Errors

Douglas P. WIENS

This article addresses the problem of constructing designs for regression models in the presence of both possible heteroscedasticity and an approximately and possibly incorrectly specified response function. Working with very general models for both types of departure from the classical assumptions, I exhibit minimax designs and correspondingly optimal weights. Simulation studies and a case study accompanying the theoretical results lead to the conclusions that the robust designs yield substantial gains over some common competitors, in the presence of realistic departures that are sufficiently mild so as to be generally undetectable by common test procedures. Specifically, I exhibit solutions to the following problems: P1, for ordinary least squares, determine a design to minimize the maximum value of the integrated mean squared error (IMSE) of the fitted values, with the maximum being evaluated over both types of departure; P2, for weighted least squares, determine both weights and a design to minimize the maximum IMSE, and P3, choose weights and design points to minimize the maximum IMSE, subject to a side condition of unbiasedness. Solutions to P1 and P2 are given for multiple linear regression with no interactions and a spherical design space. For P3 the solution is given in complete generality; as an example, I consider polynomial regression. In this case the minimax design turns out to be a smoothed version of the D-optimal design, with modes coinciding exactly with the support points of this classical design.

KEY WORDS: Equileverage designs; Legendre polynomials; Multiple regression; Optimal design; Polynomial regression; Robustness; Weighted least squares.

1. INTRODUCTION

In this article I study designs and weights for regression models, with an eye to attaining robustness against two violations of the classical assumptions:

1. The response is taken to be only *approximately* linear in the regressors,

$$Y(\mathbf{x}) = E[Y|\mathbf{x}] + \varepsilon(\mathbf{x})$$

and

$$E[Y|\mathbf{x}] = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta} + f(\mathbf{x}), \tag{1}$$

for a p-dimensional vector $\mathbf{z}$ of regressors, depending on a q-dimensional vector $\mathbf{x}$ of independent variables. The response error function f represents uncertainty about the exact nature of the regression response. One estimates $\boldsymbol{\theta}$ but not f, leading to biased estimation of $E[Y|\mathbf{x}]$.

2. The random errors, although uncorrelated with mean 0, are possibly heteroscedastic,

$$\mathrm{var}[\varepsilon(\mathbf{x})] = \sigma^2 g(\mathbf{x}), \tag{2}$$

for a function g satisfying assumptions given later.

Violation 1 is commonly dealt with at the design stage, an approach initiated in the seminal work of Box and Draper (1959) and continued by Huber (1975), Pesotchinsky (1982), myself (Wiens 1992), and others. Violation 2 has most commonly been viewed purely as an estimation problem, handled by using weighted least squares with weights inversely proportional to an estimate of $g(\mathbf{x})$. This approach requires considerable structure (e.g., groups of identifiable homoscedastic observations), and even then the usual variance estimates can be inefficient and inconsistent (Carroll and Cline 1988).

When handled at the design stage, heterogeneity poses special problems in that one may have no knowledge whatsoever of the function $g(\mathbf{x})$. Wong (1992) and Wong and Cook (1993) discussed characterizations of and algorithms for constructing optimal designs for heteroscedastic regression models with *known* efficiency functions $\lambda(\mathbf{x}) = g(\mathbf{x})^{-1}$. Bayes designs assuming a known efficiency function have been constructed by Dasgupta, Mukhopadhyay, and Studden (1992). Pritchard and Bacon (1977) combined Bayesian and sequential approaches—one is to choose the next design point to maximize the mode of the posterior density of the parameters. Heteroscedasticity is dealt with by using weighted least squares with a power transformation, the power being estimated along with the regression parameters. Power transformations in this framework also have been studied by Schulz and Endrenyi (1983), who take the power to be a random variable uniformly distributed over [0, 1]. Bayesian models to determine weights in iterative weighted least squares regression were studied by Hooper (1993).

In this article I adopt a *minimax* approach to the problem of unstructured and unknown heteroscedasticity and response misspecification. As loss function I take the integrated mean squared error (IMSE) of the fitted values $\hat{Y}(\mathbf{x})$,

$$\mathrm{IMSE} = \int_S E[(\hat{Y}(\mathbf{x}) - E[Y|\mathbf{x}])^2]\,d\mathbf{x}.$$

Here and elsewhere the integration is over the *design space* $\mathcal{S}$; that is, the region to be explored by the experimenter.

I consider the following problems:

Douglas P. Wiens is Professor and Director, Statistics Centre, Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1 (E-mail: doug.wiens@ualberta.ca). This research was supported by the Natural Sciences and Engineering Research Council of Canada.

Journal of the American Statistical Association
December 1998, Vol. 93, No. 444, Theory and Methods

P1. For *ordinary* least squares (OLS), determine a design to minimize the maximum, over f and g, value of the IMSE.

P2. For *weighted* least squares (WLS), determine both weights and a design to minimize the maximum IMSE.

P3. Choose weights and design points to minimize the maximum IMSE, subject to a side condition of unbiasedness.

A precise definition of the regression models to be considered and preliminary reductions of problems P1–P3 are given in Section 2. Solutions to these problems are given in Sections 3–5. It turns out that the optimal designs are absolutely continuous with respect to Lebesgue measure; methods of approximating and implementing them are discussed. Two comparative studies and a case study are detailed in Sections 6 and 7 respectively. Conclusions are given in Section 8. All derivations are provided in the Appendix.

2. PRELIMINARIES

The minimax problems considered in this article require maximizing the loss over model response errors f and variance functions g, and minimizing the resulting maxima over the class of designs and possibly over the class of weights as well. In this section I give the solutions to the easier of these variational problems. I thereby reduce each of P1 and P2 to single minimizations over a class of densities, and P3 to a single minimization over the class of weights.

Suppose that the experimenter is to take n uncorrelated observations on a random variable Y whose mean is thought to vary in an approximately linear manner with regressors $\mathbf{z}(\mathbf{x})$: $E[Y|\mathbf{x}] \approx \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$. The sites $\mathbf{x}_i$ are chosen from $\mathcal{S}$, a design space with finite volume defined by $\int_{\mathcal{S}} d\mathbf{x} = \Omega^{-1}$. Define the "true" value of $\boldsymbol{\theta}$ by requiring the linear approximation to be most accurate in the L^2 sense:

$$\boldsymbol{\theta} := \operatorname*{argmin}_{t} \int_{\mathcal{S}} (E[Y|\mathbf{x}] - \mathbf{z}^T(\mathbf{x})\mathbf{t})^2 \, d\mathbf{x}.$$

Then define $f(\mathbf{x}) = E[Y|\mathbf{x}] - \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$ and $\varepsilon(\mathbf{x}) = Y(\mathbf{x}) - E[Y|\mathbf{x}]$, so that (1) holds. As at (2), allow for the possibility that the variance of $\varepsilon(\mathbf{x})$ varies with $\mathbf{x}$. Define $\sigma^2 = \sup_g (\int_{\mathcal{S}} \mathrm{var}^2[\varepsilon(\mathbf{x})]\Omega \, d\mathbf{x})^{1/2}$. The model is then summarized by (1), (2), and

$$\int_{\mathcal{S}} \mathbf{z}(\mathbf{x}) f(\mathbf{x}) \, d\mathbf{x} = \mathbf{0} \tag{3a}$$

$$\int_{\mathcal{S}} f^2(\mathbf{x}) \, d\mathbf{x} \le \eta^2, \tag{3b}$$

and

$$\int_{\mathcal{S}} g^2(\mathbf{x}) \, d\mathbf{x} \le \Omega^{-1}. \tag{3c}$$

The first condition of (3) is a consequence of the definitions of $\boldsymbol{\theta}$ and of f. Some condition like the second is necessary so that errors due to bias do not swamp those due to variance. The third condition follows from the definition of σ^2 and also allows for homoscedastic errors $g = \mathbf{1}$, where $\mathbf{1}(\mathbf{x}) \equiv 1$.

The optimal designs for P3 do not depend on the parameters σ^2 or η. Those for P1 and P2 depend on these values only through the quantity $\nu = \sigma^2/(n\eta^2)$, which may be interpreted as representing the relative importance to the experimenter of variance versus bias: $\nu = 0$ corresponding to a "pure bias" problem, $\nu = \infty$ giving a "pure variance" problem.

I propose estimating $\boldsymbol{\theta}$ by least squares, possibly weighted with nonnegative weights $w(\mathbf{x})$. Let ξ be the design measure; that is, the measure assigning mass n^{-1} to each of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. Define matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}$ and a vector $\mathbf{b}$ by $\mathbf{A} = \int_{\mathcal{S}} \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x}) \, d\mathbf{x}$, $\mathbf{B} = \int_{\mathcal{S}} w(\mathbf{x})\mathbf{z}(\mathbf{x}) \mathbf{z}^T(\mathbf{x}) \, d\xi(\mathbf{x})$, $\mathbf{D} = \int_{\mathcal{S}} w^2(\mathbf{x})g(\mathbf{x})\mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x}) \, d\xi(\mathbf{x})$, and $\mathbf{b} = \int_{\mathcal{S}} w(\mathbf{x})\mathbf{z}(\mathbf{x})f(\mathbf{x}) \, d\xi(\mathbf{x})$. In a more familiar regression notation these are $\mathbf{B} = n^{-1}\mathbf{Z}^T\mathbf{W}\mathbf{Z}$ and $\mathbf{D} = n^{-1}\mathbf{Z}^T\mathbf{W}\mathbf{G}\mathbf{W}\mathbf{Z}$, where $\mathbf{Z}$ is the $n \times p$ model matrix with rows $\mathbf{z}^T(\mathbf{x}_i)$ and $\mathbf{W}$ and $\mathbf{G}$ are the $n \times n$ diagonal matrices with diagonal elements $w(\mathbf{x}_i)$ and $g(\mathbf{x}_i)$. The motivation for writing these quantities as integrals with respect to ξ will become apparent later, when I broaden the class of allowable design measures to include continuous designs. Note also that although it is mathematically convenient to treat ξ as a probability distribution, I do so only in the formal sense of a nonnegative measure with a total mass of unity—there is no implication that the $\mathbf{x}_i$ are measured with error.

Assume that $\mathbf{A}$ and $\mathbf{B}$ are nonsingular and define $\mathbf{H} = \mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1}$. In this notation, the mean vector and covariance matrix of the estimate $\hat{\boldsymbol{\theta}} = (\mathbf{Z}^T\mathbf{W}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{W}\mathbf{Y} = \mathbf{B}^{-1}\int_{\mathcal{S}} w(\mathbf{x})\mathbf{z}(\mathbf{x})Y(\mathbf{x}) \, d\xi(\mathbf{x})$ are $E[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta} = \mathbf{B}^{-1}\mathbf{b}$ and $\mathrm{cov}[\hat{\boldsymbol{\theta}}] = (\sigma^2/n)\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}$. I estimate $E[Y|\mathbf{x}]$ by $\hat{Y}(\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\hat{\boldsymbol{\theta}}$ and consider the resulting IMSE. This splits into terms due solely to estimation bias, estimation variance, and model misspecification:

$$\mathrm{IMSE}(f, g, w, \xi) = \mathrm{ISB}(f, w, \xi) + \mathrm{IV}(g, w, \xi) + \int_{\mathcal{S}} f^2(\mathbf{x}) \, d\mathbf{x},$$

where the integrated squared bias (ISB) and integrated variance (IV) are

$$\mathrm{ISB}(f, w, \xi) = \int_{\mathcal{S}} (E[\hat{Y}(\mathbf{x}) - \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}])^2 \, d\mathbf{x} = \mathbf{b}^T\mathbf{H}\mathbf{b}$$

and

$$\mathrm{IV}(g, w, \xi) = \int_{\mathcal{S}} \mathrm{var}[\hat{Y}(\mathbf{x})] \, d\mathbf{x} = \frac{\sigma^2}{n} \cdot \mathrm{trace}(\mathbf{H}\mathbf{D}).$$

I adopt the viewpoint of *approximate* design theory and allow as a design measure ξ *any* distribution on $\mathcal{S}$. The optimal designs are then not discrete. It is in fact easy to see that if either of $\sup_f \mathrm{ISB}(f, w, \xi)$ or $\sup_g \mathrm{IV}(g, w, \xi)$ is to be finite, then ξ must necessarily be absolutely continuous with respect to Lebesgue measure. A formal proof can be based on that of lemma 1 of Wiens (1992).

Methods of implementing approximations to these continuous designs are discussed in Sections 6 and 7. My general approach is to divide $\mathcal{S}$ into n regions to each of which

Table 1. Constants for $m_0(x)$ of P1, Minimax for OLS, and Straight Line Regression

ν	a	b	c	γ
.01	6.67	7.33	.458	.340
.05	6.42	1.38	.465	.365
.1	5.07	.843	.467	.373
.5	3.24	.263	.472	.384
1.0	2.94	.136	.476	.406
5.0	2.49	.045	.479	.419
10	2.40	.031	.480	.427
20	2.32	.018	.480	.443
50	2.29	.015	.480	.446

ξ assigns a mass of n^{-1}, and to place one design point in each such region. This ensures that ξ is the weak limit, as $n \to \infty$, of the empirical measures of the design points. When $\mathcal{S}$ is an interval, this apportionment can be done in a natural way by placing design points at the quantiles of ξ. When $\mathcal{S}$ is a hypersphere and ξ is spherically symmetric, the well-known factorization of a spherically symmetric measure into the uniform distribution of the angles, and the univariate distribution of the norm may be exploited; see Remark 3 and Example 9. For general $\mathcal{S}$ there is considerable arbitrariness in this approach, and the derivation of implementations which are in some sense optimal is the subject of further research.

In earlier work (Wiens 1992) I discussed some alternate neighborhood structures for departures from linearity. Although the neighborhood defined by (3) is rich enough to necessitate absolutely continuous designs, alternate and evidently thinner neighborhoods that have been proposed lead to designs with mass concentrated at a small number of, generally extreme points in the design space (see, e.g., Li and Notz 1982; Marcus and Sacks 1976; Pesotchinsky 1982). Such designs allow little or no opportunity to test the fitted response, and so their robustness is somewhat questionable. My conclusion (Wiens 1992, p. 355) was that "our attitude is that an approximation to a design which is robust against more realistic alternatives is preferable to an exact solution in a neighbourhood which is unrealistically sparse."

Let $k(\mathbf{x}) = \xi'(\mathbf{x})$ be the density of ξ, and define $m(\mathbf{x}) = k(\mathbf{x})w(\mathbf{x})$. Assume, without loss of generality, that the average weight is $\int_{\mathcal{S}} w(\mathbf{x})\, d\xi(\mathbf{x}) = 1$. Then m is a density on $\mathcal{S}$ and

$$\int_{\mathcal{S}} \frac{m(\mathbf{x})}{w(\mathbf{x})}\, d\mathbf{x} = 1. \tag{4}$$

Observe that now $\mathbf{b}$ and $\mathbf{B}$, and hence $\mathrm{ISB}(f, w, \xi)$, depend on (w, ξ) only through m and that $\mathrm{IV}(f, w, \xi)$ benefits from a similar, though lesser, simplification. For this reason, rather than optimizing over $w(\cdot)$ and densities $k(\cdot)$, it is convenient to optimize over $w(\cdot)$ and densities $m(\cdot)$ subject to (4).

The max halves of the minimax solutions are given by Theorem 1.1. Before stating this, I give some definitions. Define matrices $\mathbf{C} = \int_{\mathcal{S}} m^2(\mathbf{x})\mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})\, d\mathbf{x}$ and $\mathbf{K} = \mathbf{C} - $

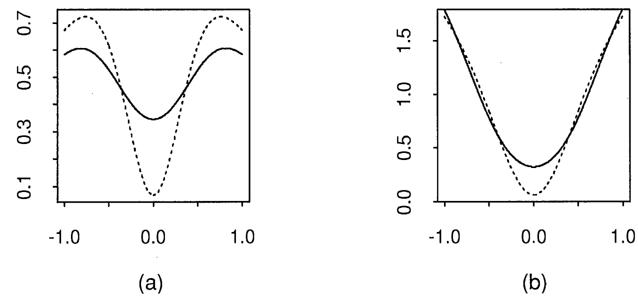


Figure 1. Minimax IMSE Designs for OLS and Straight Line Regression, as in P1. (a) Design densities and (b) least favorable variance functions; ν = .05 (——) and ν = 50 ($\cdots$).

$\mathbf{H}^{-1}$. Note that $\mathbf{K}$ is positive semidefinite:

$$\mathbf{a}^T\mathbf{K}\mathbf{a} = \int_{\mathcal{S}} \{\mathbf{a}^T[(m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}^{-1})\mathbf{z}(\mathbf{x})]\}^2\, d\mathbf{x} \ge 0 \quad \forall \mathbf{a}.$$

Let λ_m be the largest eigenvalue of $\mathbf{K}^{1/2}\mathbf{H}\mathbf{K}^{1/2}$; that is, the largest solution to $\det(\mathbf{C} - (\lambda + 1)\mathbf{B}\mathbf{A}^{-1}\mathbf{B}) = 0$. Denote by $\boldsymbol{\beta}_m$ the corresponding eigenvector, normalized so that $\|\boldsymbol{\beta}_m\| = 1$. Finally, define $l_m(\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\mathbf{H}\mathbf{z}(\mathbf{x})$.

Theorem 1.

a. Maximum ISB over all f satisfying (3a) and (3b) is

$$\max_f \mathrm{ISB}(f, w, \xi) = \eta^2\lambda_m.$$

When $m(\cdot)$ is such that $\mathbf{K}$ is nonsingular, maximum ISB is attained at $f_m(\mathbf{x}) = \eta\boldsymbol{\beta}_m^T\mathbf{K}^{-1/2}(m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}^{-1})\mathbf{z}(\mathbf{x})$.

b. Maximum IV over all g satisfying (3c) is

$$\max_g \mathrm{IV}(g, w, \xi) = \eta^2\nu\Omega^{-1/2}\left(\int_{\mathcal{S}} (w(\mathbf{x})l_m(\mathbf{x})m(\mathbf{x}))^2\, d\mathbf{x}\right)^{1/2}$$

and is attained at $g_{m,w}(\mathbf{x}) \propto w(\mathbf{x})l_m(\mathbf{x})m(\mathbf{x})$, normalized so that $\int_{\mathcal{S}} g_{m,w}^2(\mathbf{x})\, d\mathbf{x} = \Omega^{-1}$.

c. Maximum IMSE is

$$\max_{f,g} \mathrm{IMSE}(f, g, w, \xi) = \eta^2\left[1 + \lambda_m + \nu\Omega^{-1/2} \times \left(\int_{\mathcal{S}} (w(\mathbf{x})l_m(\mathbf{x})m(\mathbf{x}))^2\, d\mathbf{x}\right)^{1/2}\right].$$

Table 2. Relative Efficiencies re1 (No Contamination) and re2 (Least Favorable Contamination) of ξ_0 of Example 1, With OLS, Versus the "Optimal" and Uniform Designs

q	$re1_{opt}$	$re1_{unif}$	$re2_{unif}$
1	.719	1.079	1.010
2	.694	1.041	1.013
3	.706	1.009	1.006
4	.734	1.002	1.003
5	.762	1.000	1.004

NOTE: $re1_{unif}(q)$ = 1.000 for $q \ge 6$.

Table 3. Constants for $m_0(x)$ of P2), Minimax for WLS, and Straight Line Regression

ν	a	b	c	γ
.01	2.33	21.8	1.11	.335
.05	1.67	6.47	1.11	.336
.1	1.61	3.62	1.11	.338
.5	1.60	1.12	1.11	.354
1.0	1.49	.890	1.11	.356
5.0	1.49	.548	1.11	.432
10	1.48	.524	1.13	.440
20	1.60	.426	1.16	.496
50	2.02	.265	1.26	.577

By Theorem 1c, problem P1 can now be phrased as follows:

P1: Find a density $m_0(\mathbf{x})$ to minimize

$$\begin{aligned}\eta^{-2} \max_{f,g} \ &\mathrm{IMSE}(f, g, w = \mathbf{1}, \xi) \\ &= 1 + \lambda_m + \nu\Omega^{-1/2} \\ &\quad \times \left(\int_{\mathcal{S}} (l_m(\mathbf{x}) m(\mathbf{x}))^2 \, d\mathbf{x} \right)^{1/2}. \end{aligned} \tag{5}$$

Then $k_0 = m_0$ is the optimal (minimax) design density for OLS estimation.

Problem P2 requires the derivation of optimal weights for fixed $m(x)$; these are given in Theorem 2. The "unbiasedness" in the statement of problem P3 refers to the requirement $\max_f \ \mathrm{ISB}(f, w, \xi) = 0$. Equivalently, $E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$ for all f. Theorem 2b states that this holds only when $m(\mathbf{x}) \equiv \Omega$, the uniform density on $\mathcal{S}$.

Theorem 2.

a. For fixed $m(\mathbf{x})$, the optimal weights minimizing $\max_g$ $\mathrm{IV}(g, w, \xi)$ subject to (4) are given by $w_m(\mathbf{x}) = \alpha_m (l_m^2(\mathbf{x}) m(\mathbf{x}))^{-1/3} I(m(\mathbf{x}) > 0)$, where $\alpha_m = \int_{\mathcal{S}} (l_m(\mathbf{x}) m^2(\mathbf{x}))^{2/3} \, d\mathbf{x}$. Then $\max_g \ \mathrm{IV}(g, w_m, \xi) = \eta^2 \nu \Omega^{-1/2} \alpha_m^{3/2}$.
b. Subject to (4), the requirement $\max_f \ \mathrm{ISB}(f, w, \xi) = 0$ holds if and only if $m(\mathbf{x}) \equiv \Omega$.

By virtue of Theorem 2a and 2b, problems P2 and P3 now become

P2: Find a density $m_0(\mathbf{x})$ minimizing

$$\eta^{-2} \max_{f,g} \ \mathrm{IMSE}(f, g, w_m, \xi) = 1 + \lambda_m + \nu\Omega^{-1/2}\alpha_m^{3/2}.$$

Then the design density $k_0(\mathbf{x}) = \alpha_{m_0}^{-1}(m_0^2(\mathbf{x}) l_{m_0}(\mathbf{x}))^{2/3}$ and the weights $w_{m_0}(\mathbf{x}) = m_0(\mathbf{x})/k_0(\mathbf{x})$ are (minimax) optimal for WLS estimation.

P3: Find weights $w_0(\mathbf{x}) \propto (\mathbf{z}^T(\mathbf{x})\mathbf{A}^{-1}\mathbf{z}(\mathbf{x}))^{-2/3}$, satisfying (4) with $m(\mathbf{x}) \equiv \Omega$. Then $w_0(\mathbf{x})$ and the design density $k_{w_0,\Omega}(\mathbf{x}) = \Omega / w_0(\mathbf{x})$ are optimal in that they minimize $\max_{f,g} \ \mathrm{IMSE}(f, g, w, \xi)$, subject to the unbiasedness condition $\max_f \ \mathrm{ISB}(f, w, \xi) = 0$.

3. MINIMAX DESIGNS FOR ORDINARY LEAST SQUARES

For P1 and P2, I consider only multiple linear regression without interactions—$\mathbf{z}(\mathbf{x}) = (1, \mathbf{x}_{q\times 1}^T)^T$—with $\mathcal{S}$ a q-dimensional ellipsoid $\{\mathbf{x} | (\mathbf{x}-\mathbf{a})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{a}) \le 1\}$ for some positive-definite matrix $\boldsymbol{\Sigma}$. Via the transformation $\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\mathbf{a}) \to \mathbf{x}$, assume that $\mathbf{a} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$, so that $\mathcal{S}$ is the hypersphere of unit radius centered at the origin. Then $\Omega^{-1} = \pi^{q/2}/\Gamma[(q/2)+1]$ and $\mathbf{A} = \Omega^{-1}(1 \oplus (q+2)^{-1}\mathbf{I}_q)$. A design density $m_0(\mathbf{x})$ appropriate for the spherical design space transforms to $m_{\mathbf{a},\boldsymbol{\Sigma}}(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-1/2} m_0(\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\mathbf{a}))$ in the original elliptical region. Under these transformations, the loss function is altered only by a constant multiple depending on $|\boldsymbol{\Sigma}|$, so that if m_0 is optimal, then this optimality is transferred to $m_{\mathbf{a},\boldsymbol{\Sigma}}$.

In particular, assume that the independent variables have been transformed to a common scale. There being no a priori reason to give preference to one quadrant over another, restrict to densities $m(\mathbf{x})$ that are invariant under permutations and sign changes $\mathbf{x} \to (\pm x_{i_1}, \ldots, \pm x_{i_n})$. Then $\mathbf{B} = 1 \oplus \gamma \mathbf{I}_q$, where $\gamma = \int_{\mathcal{S}} x_1^2 m(\mathbf{x}) \, d\mathbf{x}$.

Theorem 3 (Minimax Designs for OLS). For $0 \le u \le 1$, put $l(u; \gamma) = 1 + u^2/((q+2)\gamma^2)$ and define

$$h_0(u; \gamma) = a\nu(b + u^2)/(1 + c\nu l^2(u; \gamma)),$$

where the positive constants $a = a(\gamma), b = b(\gamma)$, and $c = c(\gamma)$ satisfy

$$\int_0^1 \frac{qu^{q-1}}{\Omega} h_0(u; \gamma) \, du = 1, \tag{6}$$

$$\int_0^1 \frac{u^{q+1}}{\Omega} h_0(u; \gamma) \, du = \gamma, \tag{7}$$

and

$$2c \left(\int_0^1 l^2(u; \gamma) q u^{q-1} h_0^2(u; \gamma) \, du \right)^{1/2} = 1. \tag{8}$$

Define

$$\gamma_0 = \operatorname*{argmin}_{\gamma \ge 0} (\Omega^{-1} a(\gamma)(b(\gamma) + q\gamma) + (4\Omega^2 c(\gamma))^{-1}). \tag{9}$$

Denote expectation with respect to the density (qu^{q-1}/Ω) $h_0(u; \gamma_0)$ by E_0. If

$$E_0[(q + 1 - l(U; \gamma_0)) h_0(U; \gamma_0)] \ge 0, \tag{10}$$

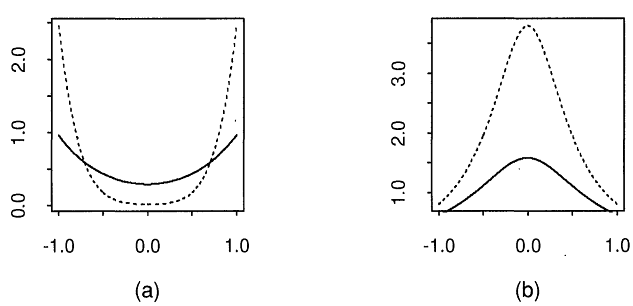


Figure 2. Minimax IMSE Design Densities (a) and Optimal Weights (b) for WLS and Straight Line Regression, as in P2; ν = .05 (——) and ν = 50 (· · ·).

Table 4. Relative Efficiencies re1 (No Contamination) and re2 (Least Favorable Contamination) of ξ_0 of Example 3, With and Without the Optimal Weights w_{m_0}, Versus the "Optimal" and Uniform Designs and OLS

	Optimal weighting of ξ_0			Constant weighting of ξ_0		
q	$re1_{opt}$	$re1_{unif}$	$re2_{unif}$	$re1_{opt}$	$re1_{unif}$	$re2_{unif}$
1	.731	1.096	1.344	.796	1.194	.915
2	.717	1.075	1.593	.775	1.162	.872
3	.731	1.044	1.788	.782	1.117	.828
4	.742	1.012	1.919	.787	1.073	.794
5	.769	1.009	1.993	.817	1.073	.750

then the minimax design ξ_0 has density $k_0(\mathbf{x}) = m_0(\mathbf{x}) = h_0(\|\mathbf{x}\|; \gamma_0)$. Minimax IMSE is

$$\max_{f,g} \text{ IMSE } (f, g, w = \mathbf{1}, \xi_0) = \eta^2\nu(\Omega^{-1}a(\gamma_0)(b(\gamma_0) + q\gamma_0) + (4\Omega^2 c(\gamma_0))^{-1}), \quad (11)$$

attained at $g_{m_0,\mathbf{1}}(\mathbf{x}) = 2cl(\|\mathbf{x}\|; \gamma_0)m_0(\mathbf{x})$ and $f_{m_0}(\mathbf{x}) \propto m_0(\mathbf{x}) - \Omega$.

Remark 1. The integrand in (6) is the density of $U = \|\mathbf{X}\|$, when $\mathbf{X}$ has the density $m_0(\mathbf{x})$. The parametric form of $m_0(\cdot)$ given in Theorem 3 is such that it minimizes (5), subject to the marginal second moment γ being fixed. Equations (7) and (8) allow (5) to be expressed as a function of γ alone; a further minimization over γ then results in (9). The symmetry of $m(\cdot)$ implies that λ_m in (5) is one of only two distinct eigenvalues $\lambda_m^{(1)}$ or $\lambda_m^{(2)}$; condition (10) ensures that λ_{m_0} is indeed the larger of $\lambda_{m_0}^{(1)}$ and $\lambda_{m_0}^{(2)}$. I have numerical evidence for, but have been unable to prove, the conjecture that (10) is vacuous in that it holds for all values of (ν, q).

Remark 2. For the numerical work, (6)–(8) were first solved for a, γ, and c, for fixed b. Then (11) was minimized over b and (10) was verified. See Table 1 for some representative values of the constants and Figure 1 for plots of the minimax densities and least favorable variance functions in the case $q = 1$ (straight line regression). It is interesting to note that the modes of the design densities are not at the extremes of the design space, even for large values of ν. (Recall that $\nu = \infty$ corresponds to a "pure variance" problem.) For smaller ν the design becomes more uniform, with the modes remaining in the "shoulders" of the design space.

Remark 3. To implement these designs, one may use the fact that if $\mathbf{X}$ has a spherically symmetric density $k(\|\mathbf{x}\|)$ on $\|\mathbf{x}\| \leq 1$, then $\mathbf{X}/\|\mathbf{X}\|$ is uniformly distributed over the surface of the unit sphere, independently of $U = \|\mathbf{X}\|$, which has density $h(u) = qu^{q-1}\Omega^{-1}k(u)$ on [0, 1]. Denote the corresponding distribution function of U by H_U. A possible implementation then consists of choosing $a_{n,q}$ design points uniformly distributed over each of the annuli $\|\mathbf{x}\| = H_U^{-1}(i/[n/a_{n,q}]), i = 1, \ldots, [n/a_{n,q}]$, and $n - a_{n,q}[n/a_{n,q}]$ points at $\mathbf{0}$. See Example 9, where $a_{17,2} = 3$.

Example 1. When the fitted model $E[Y|\mathbf{x}] = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$ is correct and the variances are homogeneous, the relative efficiency of ξ_0 relative to another design ξ_1, also symmetric with identical marginals and with second moment γ_1, is

$$\begin{aligned} \text{re1}(q) &= \frac{\text{IMSE}(f = \mathbf{0}, g = \mathbf{1}, w = \mathbf{1}, \xi_1)}{\text{IMSE}((f = \mathbf{0}, g = \mathbf{1}, w = \mathbf{1}, \xi_0)} \\ &= \frac{\text{IV}(\mathbf{1}, \mathbf{1}, \xi_1)}{\text{IV}(\mathbf{1}, \mathbf{1}, \xi_0)} = \frac{\Omega^{-1}\left(1 + \frac{q}{(q+2)\gamma_1}\right)\left(\frac{\sigma^2}{n}\right)}{\Omega^{-1}\left(1 + \frac{q}{(q+2)\gamma_0}\right)\left(\frac{\sigma^2}{n}\right)}. \end{aligned} \quad (12)$$

Table 2 gives some representative values of $\text{re1}(q)$ for ξ_1 the optimal design in this situation—all mass at $\|\mathbf{x}\| = 1$ and $\gamma_1 = q^{-1}$—and for ξ_1 the uniform design, with constant density Ω and $\gamma_1 = (q+2)^{-1}$. Also given are values of

$$\text{re2}(q) = \frac{\eta^{-2}\max_{f,g} \text{ IMSE}(f, g, \mathbf{1}, \xi_1)}{\eta^{-2}\max_{f,g} \text{ IMSE}(f, g, w, \xi_0)}, \quad (13)$$

with $w = \mathbf{1}$. The denominator of (13) is given in (11). For ξ_1, the aforementioned "optimal" design, the numerator is ∞; for the uniform design, it is obtained from Theorem 1c and is $1 + \nu\Omega^{-1}((q^3 + 6q^2 + 13q + 4)/(q + 4))^{1/2}$. I have used $\nu = \Omega$, so that the bound $(\int_S f^2(\mathbf{x})\, d\mathbf{x}/\int_S d\mathbf{x})^{1/2} \leq \sigma/\sqrt{n}$ is of the same order as a standard error.

Example 2. A design minimizing $\max_g \text{ IV}(g, \mathbf{1}, \xi)$ alone may be obtained by letting $\nu \to \infty$ in Theorem 3. One then obtains $m_0(\mathbf{x}) = h_0(\|\mathbf{x}\|; \gamma_0)$, where $h_0(u; \gamma) = d(b + u^2)/l^2(u; \gamma), d = d(\gamma)$ and $b = b(\gamma)$ satisfy (6) and (7), and $\gamma_0 = \text{argmin}_\gamma d(\gamma)(b(\gamma) + q\gamma)$. Minimax IV is $\max_g \text{IV}(g, \mathbf{1}, \xi_0) = (d(\gamma_0)(b(\gamma_0) + q\gamma_0)/\Omega^3)^{1/2}\sigma^2/n$, attained by $g_{m_0,\mathbf{1}}(\mathbf{x}) \propto (b(\gamma_0) + \|\mathbf{x}\|^2)/l(\|\mathbf{x}\|; \gamma_0)$. See Figure 1 for close approximations ("$\nu = 50$") to m_0 and $g_{m_0,\mathbf{1}}$ when $q = 1$, in which case $\gamma_0 = .464, d(\gamma_0) = 4.64, b(\gamma_0) = .004$, and $\max_g \text{ IV}(g, \mathbf{1}, \xi_0) = 4.17\sigma^2/n$.

4. MINIMAX DESIGNS AND WEIGHTS FOR WEIGHTED LEAST SQUARES

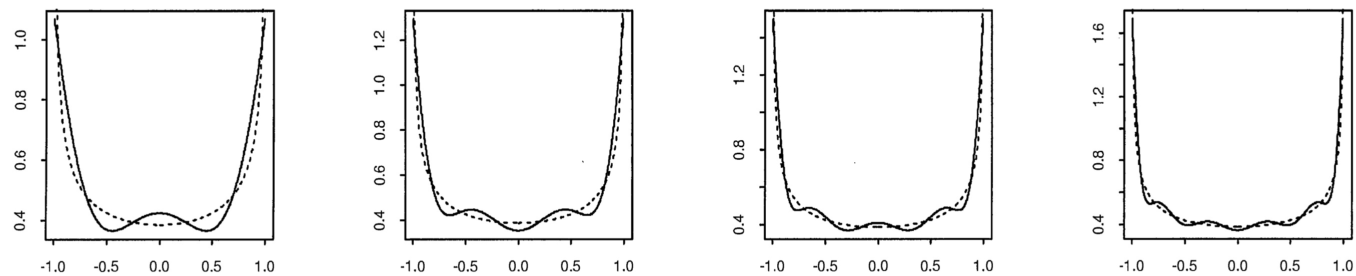
I take the same multiple linear regression model and spherical design space as in Section 3.

Theorem 4 (Minimax Designs and Weights for WLS). Let $l(u; \gamma)$ be as in Theorem 3. Define $h_0(u; \gamma)$ to be the positive root of

$$h_0(u; \gamma) + c\nu l^{2/3}(u; \gamma)h_0^{1/3}(u; \gamma) - a\nu(b + u^2) \equiv 0;$$

Table 5. Relative Efficiencies re1 (No Contamination) and re2 (Least Favorable Contamination) of $\xi_{w_0,\Omega}$ of Example 5, With and Without the Optimal Weights w_0, Versus the "Optimal" and Uniform Designs and OLS

	Optimal weighting of ξ_0			Constant weighting of ξ_0		
q	$re1_{opt}$	$re1_{unif}$	$re2_{unif}$	$re1_{opt}$	$re1_{unif}$	$re2_{unif}$
1	.696	1.044	1.087	.745	1.118	.993
2	.692	1.037	1.075	.732	1.098	1.002
3	.720	1.028	1.059	.751	1.073	1.004
4	.749	1.021	1.046	.774	1.056	1.003
5	.774	1.016	1.037	.795	1.043	1.002



(a) (b) (c) (d)

Figure 3. Minimum IMSE, Subject to max ISB = 0, Design Densities for qth Degree Polynomial WLS Regression, as in P3. (a) q = 2, (b) q = 3, (c) q = 4, (d) q = 5. · · ·, the limiting density, as q → ∞.

that is, $h_0^{1/3}(u;\gamma) = y^{1/3}(u) - c\nu l^{2/3}(u;\gamma)y^{-1/3}(u)/3$, where

$$y(u) = \frac{\nu}{2}\left[a(b+u^2) + \sqrt{\frac{4}{27}\,c^3\nu l^2(u;\gamma) + (a(b+u^2))^2}\right].$$

The positive constants $a = a(\gamma), b = b(\gamma)$, and $c = c(\gamma)$ satisfy (6), (7), and

$$c = \left(\int_0^1 (qu^{q-1}/\Omega)l^{2/3}(u;\gamma)h_0^{4/3}(u;\gamma)\,du\right)^{1/2}. \quad (14)$$

Define

$$\gamma_0 = \operatorname*{argmin}_{\gamma\geq 0} a(\gamma)(b(\gamma)+q\gamma).$$

If (10) holds, then the minimax design ξ_0 and weights w_{m_0} have $m_0(\mathbf{x}) = h_0(\|\mathbf{x}\|;\gamma_0)$. Minimax IMSE is

$$\max_{f,g}\ \mathrm{IMSE}(f,g,w_{m_0},\xi_0) = \eta^2\nu\Omega^{-1}a(\gamma_0)(b(\gamma_0)+q\gamma_0), \quad (15)$$

attained at $g_{m_0,w_{m_0}}(\mathbf{x}) = (k_0(\mathbf{x})/\Omega)^{1/2}$ and $f_{m_0}(\mathbf{x}) \propto m_0(\mathbf{x}) - \Omega$. The minimax design density k_0 and minimax weights w_{m_0} are given by

$$k_0(\mathbf{x}) = c^{-2}l^{2/3}(\|\mathbf{x}\|;\gamma_0)m_0^{4/3}(\mathbf{x})$$

and

$$w_{m_0}(\mathbf{x}) = m_0(\mathbf{x})/k_0(\mathbf{x}).$$

Table 3 gives values of the constants and Figure 2 plots the minimax densities and weights, both for the case $q = 1$.

Table 6. Relative Efficiencies re1 (No Contamination) and re2 (Least Favorable Contamination) of $\xi_{w_0,\Omega}$ of Example 7, With and Without the Optimal Weights w_0, Versus the D-Optimal and Uniform Designs and OLS

	Optimal weighting of ξ_0			Constant weighting of ξ_0		
q	$re1_D$	$re1_{unif}$	$re2_{unif}$	$re1_D$	$re1_{unif}$	$re2_{unif}$
1	.696	1.044	1.087	.745	1.118	.993
2	.848	1.060	1.157	.902	1.127	1.057
3	.915	1.067	1.211	.964	1.124	1.115
4	.952	1.071	1.255	.996	1.121	1.163
5	.976	1.074	1.290	1.015	1.117	1.204
10	1.028	1.078	1.406	1.054	1.106	1.341

NOTE: With optimal weighting of ξ_0, $\lim_{q\to\infty} re2_{unif} = \infty$ and $\lim_{q\to\infty} re1_D = \lim_{q\to\infty} re1_{unif} = 1.084$.

A general prescription seems to be that for large values of ν (variance dominant) the design places most of its mass near the extremes of the design space, but the more extreme design points receive relatively little weight in the regression. For smaller values of ν (bias dominant), the design tends to become more uniform, with extreme design points still being somewhat downweighted.

Example 3. Efficiencies of (ξ_0, w_{m_0}) under exact linearity and homoscedasticity, relative to other designs and OLS, may be computed as in Example 1, but with the denominator of $\mathrm{re1}(q)$ using

$$\mathrm{IV}\ (g = \mathbf{1}, w_0, \xi_0) = c^2\Omega^{-1}E_0[l^{1/3}(\|\mathbf{X}\|;\gamma_0)h_0^{-1/3}(\|\mathbf{X}\|;\gamma_0)]\sigma^2/n,$$

where E_0 denotes expectation with respect to $m_0(\mathbf{x})$. When ξ_0 is used with OLS, this denominator is as in (12), but with γ_0 replaced by $\gamma_* = \int_0^1 x_1^2 k_0(\mathbf{x};\gamma_0)\,d\mathbf{x}$. Table 4 gives some numerical values of re1 and re2, with ξ_0 used both with the minimax weights and with OLS. The numerator of re2 is as in Example 1. The denominator for WLS is given in (15); that for OLS is obtained from Theorem 1c with $w = \mathbf{1}$ and $m = k_0$. I again take $\nu = \Omega$.

Example 4. Letting $\nu \to \infty$ in Theorem 4 gives a design and weights minimizing $\max_g\ \mathrm{IV}(g,w,\xi)$. Then $m_0(\mathbf{x}) = h_0(\|\mathbf{x}\|;\gamma_0)$, where $h_0(u;\gamma) = d(b+u^2)^3/l^2(u;\gamma), d = d(\gamma)$ and $b = b(\gamma)$ satisfy (6) and (7), and $\gamma_0 = \mathrm{argmin}_\gamma d(b+q\gamma)^3$. Minimax IV is $\max_g \mathrm{IV}(g,w_{m_0},\xi_0) = (d(\gamma_0)(b(\gamma_0)+q\gamma_0)^3/\Omega^3)^{1/2}\sigma^2/n$. The minimax weights are $w_{m_0}(\mathbf{x}) = (b(\gamma_0)+q\gamma_0)/(b(\gamma_0)+\|\mathbf{x}\|^2)$, and the minimax design has density $k_0(\mathbf{x}) = d(\gamma_0)(b(\gamma_0)+\|\mathbf{x}\|^2)^4/((b(\gamma_0)+q\gamma_0)l^2(\|\mathbf{x}\|;\gamma_0))$. Figure 2 gives close approximations ($\nu = 50$) to k_0 and w_{m_0} when $q = 1$, in which case $\gamma_0 = .700, d(\gamma_0) = 4.81, b(\gamma_0) = .117$, and $\max_g \mathrm{IV}(g,w_{m_0},\xi_0) = 4.58\sigma^2/n$.

5. MINIMUM VARIANCE UNBIASED DESIGNS FOR WEIGHTED LEAST SQUARES

Problem P3 can be solved in considerably more generality than seems feasible with P1 and P2. In this section I do not restrict the class of competing designs in any way, and do not impose any special structure on the design space.

Table 7. ISB, IV and IMSE for the Designs of Example 8; Heteroscedastic Errors and Incorrect Fitted Response

	OLS			WLS		
Design	ISB	IV	IMSE	ISB	IV	IMSE
Minimax	.017	.237 (.217)	.254	.001	.225 (.231)	.225
D-optimal	.194	.195 (.200)	.389	.194	.195 (.200)	.389
Uniform	.003	.269 (.232)	.272	.004	.246 (.232)	.250

NOTE: Values in parentheses are those of IV under homoscedasticity.

Theorem 5 (Minimax Designs and Weights for Unbiased WLS). Subject to the unbiasedness condition $\max_f \text{ ISB}(f,w,\xi) = 0$, the maximum loss $\max_{f,g} \text{IMSE}(f,g,w,\xi)$ is minimized by the design $\xi_{w_0,\Omega}$ with density

$$k_{u_0,\Omega}(\mathbf{x}) = \frac{(\mathbf{z}(\mathbf{x})^T\mathbf{A}^{-1}\mathbf{z}(\mathbf{x}))^{2/3}}{\int_S(\mathbf{z}(\mathbf{x})^T\mathbf{A}^{-1}\mathbf{z}(\mathbf{x}))^{2/3}\,d\mathbf{x}}$$

and weights $w_0(\mathbf{x}) = \Omega/k_{w_0,\Omega}(\mathbf{x})$. Minimax IMSE, attained at $g_0(\mathbf{x}) = (\mathbf{k}_{w_0,\Omega}(\mathbf{x})/\Omega)^{1/2}$, is

$$\max_{f,g} \text{IMSE}(f,g,w_0,\xi_{w_0,\Omega}) = \eta^2\left[1+\nu\Omega^{-1/2}\left(\int_S(\mathbf{z}(\mathbf{x})^T\mathbf{A}^{-1}\mathbf{z}(\mathbf{x}))^{2/3}\,d\mathbf{x}\right)^{3/2}\right].$$

Note that the minimax weights $w_0(\mathbf{x})$ are equal to $g_0(\mathbf{x})^{-2}$; if $g(\mathbf{x})$ is known, then the efficient weights are proportional to $g(\mathbf{x})^{-1}$.

Example 5. For the multiple linear regression model of Sections 3 and 4, I obtain $k_{w_0,\Omega} \propto (1+(q+2)\|\mathbf{x}\|^2)^{2/3}$. Table 5 presents relative efficiencies as q varies, with $\nu = \Omega$. With $q = 2$, this design is evaluated numerically in Example 9 and then transformed and implemented in Section 7.

Example 6. For multiple linear regression without interactions—$\mathbf{z}(\mathbf{x}) = (1,x_1,\ldots,x_q)^T$—on $S = [-1,1]^q$, the optimal design density is $k_{w_0,\Omega}(\mathbf{x}) \propto (1+3\|\mathbf{x}\|^2)^{2/3}$. If the independent variables have been linearly transformed to lie in this hypercube, then optimal designs in terms of the original units may be obtained as in the discussion at the beginning of Section 3.

Example 7. For $\mathbf{z}(x) = (1,x,x^2,\ldots,x^q)^T$ (degree-q polynomial regression) on $\mathcal{S} = [-1,1]$, the optimal design densities $k_{w_0,\Omega}(x;q)$ are as follows, for $q = 2,3,4,5$:

$$k_{w_0,\Omega}(x;2) = .425(1-2x^2+5x^4)^{2/3},$$

$$k_{w_0,\Omega}(x;3) = .081(9+45x^2-165x^4+175x^6)^{2/3};$$

$$k_{w_0,\Omega}(x;4) = .095(9-36x^2+294x^4-644x^6+441x^8)^{2/3},$$

and

$$k_{w_0,\Omega}(x;5) = .043(25+175x^2-1{,}750x^4+6{,}510x^6 - 9{,}555x^8+4{,}851x^{10})^{2/3}.$$

There is an interesting connection between $k_{w_0,\Omega}(x;q)$ and the Legendre polynomials, and hence to the classical D-optimal design ξ_D; that is, the $(q+1)$-point measure minimizing the determinant of the covariance matrix of $\hat{\boldsymbol{\theta}}$.

Lemma 1. Denote by $P_m(x)$ the mth degree Legendre polynomial on $[-1,1]$, normalized by $\int_{-1}^1 P_m^2(x)\,dx = (m+.5)^{-1}$. Define a density on $[-1,1]$ by $h_q(x) = (q+1)^{-1}\mathbf{z}^T(x)\mathbf{A}^{-1}\mathbf{z}(x)$. Then $k_{w_0,\Omega}(x;q) \propto h_q(x)^{2/3}$ and

$$h_q(x) = .5(P_q(x)P'_{q+1}(x) - P'_q(x)P_{q+1}(x)), \quad (16)$$

$$\lim_{q\to\infty} k_{w_0,\Omega}(x;q) = \frac{(1-x^2)^{-1/3}}{2^{1/3}\beta(\frac{2}{3},\frac{2}{3})}. \quad (17)$$

It can be shown that the local maxima of $h_q(x)$, and hence those of $k_{w_0,\Omega}(x;q)$, are the 0s of $(1-x^2)P'_q(x)$. These are precisely the points of support of the D-optimal design ξ_D. In this sense $k_{w_0,\Omega}(\cdot;q)$ is a smoothed version of ξ_D, which has the limiting density $(1-x^2)^{-1/2}/\pi = \lim_{q\to\infty} h_q(x)$. This should of course not be taken as an endorsement of using ξ_D in these situations—having only as many sites as parameters, ξ_D affords no opportunity to even test the adequacy of the fitted response.

Figure 3 presents plots of $k_{w_0,\Omega}(x;q), q = 2,3,4,5$ together with the limiting density. Table 6 gives efficiencies relative to the D-optimal and uniform designs, with $\nu = \Omega = .5$. For WLS, the numerator of rel_D may be determined by using the results of section 9.5 of Pukelsheim (1993). With $\mathbf{B}$ denoting the relevant moment matrix, it is

$$\text{IV}(g=\mathbf{1},w=\mathbf{1},\xi_D) = \frac{\sigma^2}{n}\cdot\text{tr}\mathbf{A}\mathbf{B}^{-1} = \frac{\sigma^2}{n}\cdot\frac{4q(q+1)}{2q+1}.$$

The other required quantities may be obtained from Theorems 1 and 1 and Lemma 2.

6. COMPARISONS

Example 8 (Quadratic Regression). Consider a regression model as at (1), (2), and (3) with $\mathbf{z}(x) = (1,x,x^2)^T, -1 \le x \le 1$, normally distributed errors with $\sigma^2 = 1$, and sample size $n = 24$. As response error $f(x)$, take the cubic Legendre polynomial with the normalization

Table 8. Simulated Size and Power of Test of Homoscedasticity in Example 8

	α = 0		α = 2	
Design	η^2 = 0	η^2 = 1/12	η^2 = 0	η^2 = 1/12
Minimax	.047 (1)	.041 (1)	.212 (3)	.196 (3)
D-optimal	.047 (1)	.046 (1)	.229 (3)	.229 (3)
Uniform	.042 (1)	.039 (1)	.204 (3)	.208 (3)

NOTE: Standard errors in the third decimal place are in parentheses.

Table 9. Simulated Power of Test That θ_3 = 0 in Example 8

	OLS; η^2 = 2/n		WLS; η^2 = 2/n	
Design	α = 0	α = 2	α = 0	α = 2
Minimax	.186 (2)	.196 (3)	.190 (2)	.211 (3)
Uniform	.186 (2)	.243 (3)	.188 (2)	.192 (2)

NOTE: Standard errors in the third decimal place are in parentheses.

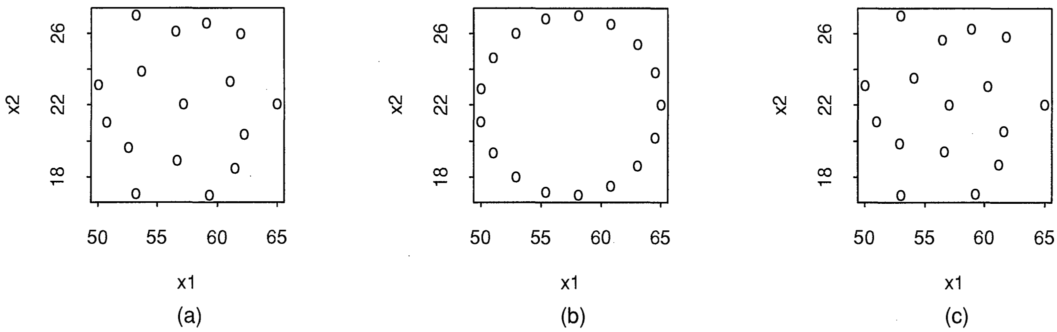


Figure 4. Designs of Example 9 Transformed for Stackloss Data. (a) P3, (b) C, (c) U.

$\int_S f^2(x)\,dx = \eta^2 = 1/12$, so that $\nu = \Omega$. The true response is then cubic, with cubic coefficient $\theta_3 = 1.25\sqrt{7/6}$. The variance function is $g(x) \propto (1+x^2)^\alpha$ with $\alpha = 2$, normed so that $\int_S g^2(x)\,dx = \Omega^{-1}$. Three designs are compared: the minimax design of Example 7, the D-optimal design with eight observations at each of 0 and ± 1, and the uniform design. For the two continuous designs observations are made at $\xi^{-1}((i-1)/(n-1)), i = 1, \ldots, n$.

Values of ISB, IV, and IMSE are given in Table 7 for both OLS and WLS fits. As benchmarks, the values of IV under homoscedasticity ($\alpha = 0$) are also included. In the exact quadratic model ($\eta = 0$), the ISB is of course 0 in all cases. For the uniform design, the weights are $w(x) = 1/g(x)$; that is, optimal for *known* $g(x)$. Even under such conditions favorable to the uniform design, it is by these measures outperformed by the minimax design, as is the D-optimal design. Note that estimates obtained from the D-optimal design, with no more sites than parameters, are necessarily unaffected by the use of WLS.

Are these model inadequacies likely to be detected? To answer this question, I have obtained the powers of two tests, each at a nominal level of 5%:

1. To test H: $\alpha = 0$ versus K: $\alpha \neq 0$, apply the scores test of Cook and Weisberg (1983), after preliminary OLS fits using the three designs. In terms of $g_i' = \partial g(x_i)/\partial\alpha_{|\alpha=0} = \log(1+x_i^2)$ (ignore the additive constant) and $\overline{g'} = \sum_i g_i'/n$, the test statistic is

$$T = \frac{[\sum_i (g_i' - \overline{g'})(e_i^2/\hat\sigma^2)]^2}{2\sum_i (g_i' - \overline{g'})^2}. \tag{18}$$

Here $e_1, \ldots, e_n$ are the OLS residuals and $\hat\sigma^2 = \sum_i e_i^2/n$. The values of T are compared to those of the χ_1^2 distribution.

2. To test a quadratic versus cubic response, (i.e., H: $\theta_3 = 0$ vs. K: $\theta_3 \neq 0$), use the usual normal-theory F test after fitting a cubic response. This excludes the D-optimal design from consideration. Both OLS and WLS fits are compared for the minimax and uniform designs.

The powers, based on 25,000 simulations, are given in Tables 8 and 9. The same 600,000 N(0, 1) values were used in preparing each column of the tables. The weights used with the uniform design were again optimal for known $g(x)$.

For the three designs the powers were comparable, and low. This is somewhat alarming in view of the fact that in each case the parametric form of the alternative was correctly specified. A lesson to be (re-)learned is that barely detectable model deviations can have a significant effect on the accuracy and precision of the estimates. I remark that for smaller values of $n (n \leq 20)$ or $\alpha (\alpha \leq 1)$, the performance of the D-optimal design in the test for heteroscedasticity was disastrous, with the powers falling to or significantly below the size of the test. The performance of this design also deteriorated rapidly under an asymmetric variance function. That of the minimax and uniform designs was quite stable under asymmetry.

Example 9 (Multiple Regression With Two Regressors). Take the multiple linear regression model, with $q = 2$, of Sections 3 and 4 and Example 5 and compare the corresponding designs (P1, P2, P3) with the uniform (U) and classically optimal (C) designs. Use $f(\mathbf{x}) = \sqrt{12/n}(\|\mathbf{x}\|^2 - .5)$ and $n = 17$, so that $\eta^2 = \pi/n$ and $\nu = \Omega$, and $g(\mathbf{x}) \propto (1+\|\mathbf{x}\|^2)^\alpha$ with $\alpha = 2$ and $\int_S g^2(\mathbf{x})\,d\mathbf{x} = \Omega^{-1}$.

Design C consists of n points equally spaced over the boundary $\|\mathbf{x}\| = 1$. It is classically optimal in the sense of minimizing IMSE$(\mathbf{0}, \mathbf{1}, \mathbf{1}, \xi)$; that is, it minimizes the IV (as well as the trace and the determinant of the covariance matrix of the regression estimates) under the assumptions

Table 10. ISB, IV, and IMSE for the Designs of Example 9; Heteroscedastic Errors and Incorrect Fitted Response

	OLS			WLS		
Design	ISB	IV	IMSE	ISB	IV	IMSE
P1	.003	.606 (.531)	.608	.018	.541 (.584)	.558
P2	.047	.582 (.471)	.630	.000	.532 (.518)	.532
P3	.019	.591 (.496)	.610	.001	.537 (.535)	.538
U	.002	.612 (.534)	.613	.031	.541 (.534)	.572
C	.554	.594 (.370)	1.148	.554	.594 (.370)	1.148

NOTE: Values in parentheses are those of IV under homoscedasticity.

Table 11. Simulated Size and Power of Test of Homoscedasticity in Example 9

	$\alpha = 0$		$\alpha = 2$	
Design	$\eta^2 = 0$	$\eta^2 = \pi/n$	$\eta^2 = 0$	$\eta^2 = \pi/n$
P1	.051 (1)	.055 (1)	.050 (1)	.042 (1)
P2	.060 (2)	.085 (2)	.013 (1)	.013 (1)
P3	.056 (1)	.069 (2)	.030 (1)	.023 (1)
U	.051 (1)	.052 (1)	.052 (1)	.043 (1)

NOTE: Standard errors in the third decimal place are in parentheses.

Table 12. Simulated Power of Test for Second-Order Response in Example 9

	OLS; $\eta^2 = \pi/n$		WLS; $\eta^2 = \pi/n$	
Design	$\alpha = 0$	$\alpha = 2$	$\alpha = 0$	$\alpha = 2$
P1	.075 (2)	.146 (2)	.070 (2)	.107 (2)
P2	.073 (2)	.094 (2)	.108 (2)	.102 (2)
P3	.074 (2)	.117 (2)	.091 (2)	.109 (2)
U	.077 (2)	.152 (2)	.081 (2)	.101 (2)

NOTE: Standard errors in the third decimal place are in parentheses.

that the fitted response is exactly correct and the errors are homoscedastic. To implement the four continuous designs, apply Remark 3 with $a_{n,q} = 3$. Thus first $[n/a_{n,q}] = 5$ values, $u_i = H_U^{-1}(i/5), i = 1, \ldots, 5$ were obtained. Then $a_{n,q}$ design points

$$\mathbf{x}_{ij} = u_i \cdot \begin{pmatrix} \cos\left(\phi_i + \frac{2\pi(j-1)}{a_{n,q}}\right) \\ \sin\left(\phi_i + \frac{2\pi(j-1)}{a_{n,q}}\right) \end{pmatrix} \qquad j = 1, \ldots, a_{n,q}$$

were taken, equally spaced over $\|\mathbf{x}\| = u_i$, for each value of i. The angles ϕ_i were a random permutation of $\{2\pi k/(a_{n,q}[n/a_{n,q}])\colon k = 1, \ldots, [n/a_{n,q}]\}$ and hence were equally spaced over $[0, 2\pi/a_{n,q})$. There were two design points at $\mathbf{0}$; see Figure 4.

When WLS was used, the weights used for P1 were the same as those for P2, obtained from Theorem 4. Those used for the uniform design were again optimal for known $g(x)$.

Tables 10–12 give the same performance measures as in Example 8, with Tables 11 and 12 again based on 25,000 N(0, 1) simulations. Note that C cannot be used in the test of heteroscedasticity, because for it the denominator of (18) vanishes. In Table 12 the test is the F test for the presence of second-order effects; that is, of H: $\theta_{11} = \theta_{22} = \theta_{12} = 0$.

The findings are much as in Example 8. Designs P2 and P3 in particular result in substantial decreases in IMSE, even though the departures from the fitted model would generally not be picked up by the standard tests. Note also, from Table 10, the good performance of these designs under homoscedasticity when WLS is used unnecessarily. The power of the test to detect heteroscedasticity was very low in all cases.

7. CASE STUDY

As an illustration, I transform the designs of Example 9 so as to adapt them to the Stackloss data of Brownlee (1965). The original dataset includes three independent variables: X_1, air flow; X_2, water temperature; and X_3, nitric acid concentration. The response Y measures ingoing ammonia lost as unabsorbed nitric acid. Daniel and Wood (1980) found that 4 of the 21 data points—1, 3, 4 and 21—correspond to transitional states and should be omitted from the dataset, and also that X_3 can be dropped from the model. Their analysis also indicates the presence of a quadratic effect in X_1. Ryan (1997) found some evidence of heteroscedasticity and, following Staudte and Sheather (1990), constructed an *equileverage* design for fitting the first-order model,

Table 13. Coefficient Estimates for Stackloss Data

Design	θ_0	θ_1	θ_2
P1	−37.53 (.550)	.693 (.009)	.575 (.012)
P2	−37.56 (.461)	.694 (.007)	.575 (.010)
P3	−37.53 (.500)	.694 (.008)	.575 (.011)
C	−37.75 (.370)	.700 (.006)	.574 (.008)
U	−37.54 (.564)	.693 (.009)	.575 (.012)

NOTE: SE in parentheses.

$$E[Y|\mathbf{X}] = \theta_0 + \theta_1 X_1 + \theta_2 X_2;$$
$$50 \le X_1 \le 65, \qquad 17 \le X_2 \le 27, \quad (19)$$

to the 17 points remaining after the deletion of those mentioned earlier. An equileverage design is one for which the "hat" matrix has a constant diagonal; that constructed by Ryan (1997) coincides exactly with design C of Example 9.

I applied linear transformations to the designs P1, P2, P3, C, and U so that they spanned the ranges in (19) exactly. I then simulated heteroscedastic data in the following manner. I first fitted (19) to the 17 real data points by OLS, obtaining an estimated regression function,

$$\hat{Y}^{(1)}(\mathbf{x}) = -42.00 + .776X_1 + .569X_2. \qquad (20)$$

I also fitted a second-order model, obtaining

$$\hat{Y}^{(2)}(\mathbf{x}) = -13.41 - .048X_1 + .281X_2 + .007X_1^2 + .011X_2^2 - .004X_1X_2. \quad (21)$$

For a single vector ε of iid N(0, 1) variates, and for each of the five designs, I used the following data:

$$y_i = \hat{Y}^{(2)}(\mathbf{x}_i) + \frac{1}{4}|\hat{Y}^{(2)}(\mathbf{x}_i) - \hat{Y}^{(1)}(\mathbf{x}_i)|\varepsilon_i; \qquad i = 1, \ldots, 17.$$

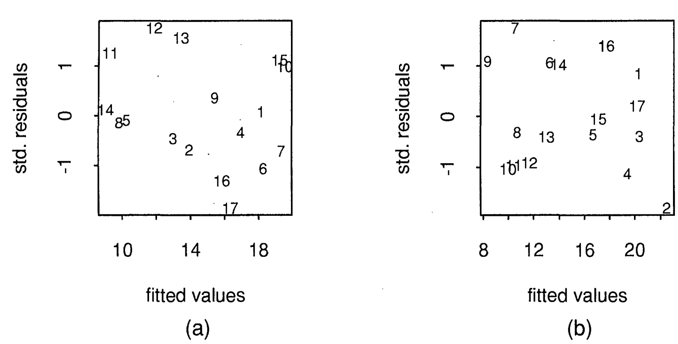


Figure 5. Studentized Residuals Versus Fitted Values by Observation Number for Stackloss Data. Designs (a) P3, (b) C, (c) U.

The simulated regression function is then that in (21), whereas the error standard deviation at a point $\mathbf{x}$ is one-fourth the difference between (21) and (20). With each design I fitted the first-order model (19) using WLS and optimal weights with P2 and P3, OLS otherwise.

Figures 4 and 5 give plots of designs P3, U, and C and the corresponding (studentized) residual plots. Those for P1 and P2 tell very much the same story as P3 and so are omitted. Coefficient estimates and standard errors are in Table 13. These estimates are quite similar for the five designs. However, the residual plots for P3 and U show evidence of both response nonlinearity and heteroscedasticity—evidence apparently not exhibited by the residuals from design C.

8. SUMMARY AND CONCLUSIONS

I have presented optimal robust designs and correspondingly optimal weights for a variety of estimation methods. Those for P1 (minimax designs for OLS) and P2 (minimax designs and weights for WLS) are at present limited to multiple regression without interactions, whereas those for P3 (minimum variance unbiased designs and weights for WLS) can be used in almost complete generality.

In Examples 5–7 I have given explicit expressions for designs applicable to a number of common situations. The designs afford considerable protection from some common and realistic model departures, and I recommend their routine use with at least a preliminary OLS fit. Any ensuing indication of heteroscedasticity should call for a WLS analysis. Given the difficulty in detecting these deviations from the model, and given the good performance of the robust designs with WLS even when the true model is homoscedastic, one should consider a WLS analysis from the start.

APPENDIX: DERIVATIONS

To avoid trivialities, and to ensure the nonsingularity of a number of relevant matrices, I assume that the design space satisfies the following condition: for each $\mathbf{a} \neq \mathbf{0}$, the set $\{\mathbf{x}\colon \mathbf{a}^T\mathbf{z}(\mathbf{x}) = 0\}$ has Lebesgue measure 0.

Proof of Theorem 1

Part a is proven as theorem 1 of Wiens (1992). Part b is the Cauchy–Schwarz inequality applied to

$$\mathrm{IV}(g, w, \xi) = (\sigma^2/n) \int_S w(\mathbf{x}) g(\mathbf{x}) m(\mathbf{x}) l_m(\mathbf{x})\, d\mathbf{x}.$$

Part c follows from Parts a and b.

Proof of Theorem 2

a. Note that w_m satisfies (4). Let $w_1(\mathbf{x})$ be any other nonnegative function satisfying (4). For $t \in [0,1]$, put $w_t = (1-t)w_m + tw_1$ and define

$$\phi(t) = \int_S (w_t(\mathbf{x}) l_m(\mathbf{x}) m(\mathbf{x}))^2\, d\mathbf{x} + 2\alpha_m^3 \int_S \frac{m(\mathbf{x})}{w_t(\mathbf{x})}\, d\mathbf{x}.$$

The function ϕ is convex in t. Then

$$\phi'(0) = 2\int_S \left(w_m(\mathbf{x}) l_m^2(\mathbf{x}) m^2(\mathbf{x}) - \alpha_m^3 \frac{m(\mathbf{x})}{w_m^2(\mathbf{x})} \right) \times (w_1(\mathbf{x}) - w_m(\mathbf{x}))\, d\mathbf{x} = 0,$$

because the integrand vanishes identically by the definition of w_m. Thus $\phi(0) \leq \phi(1)$; that is, $\int_S (w_m(\mathbf{x}) l_m(\mathbf{x}) m(\mathbf{x}))^2\, d\mathbf{x} \leq \int_S (w_1(\mathbf{x}) l_m(\mathbf{x}) m(\mathbf{x}))^2\, d\mathbf{x}$.

b. The sufficiency is immediate from (3a). For the necessity, denote by $\mathcal{F}$ the set of functions f satisfying (3a) and (3b), and write $\mathbf{b}(= \int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x}) m(\mathbf{x})\, d\mathbf{x})$ as $\mathbf{b}_f$. Suppose that $\mathbf{b}_f = \mathbf{0}$ for all $f \in \mathcal{F}$. To see that this requires $m(\mathbf{x})$ to be uniform, first let f be arbitrary but with $\int_S f^2(\mathbf{x})\, d\mathbf{x} < \infty$, $\int_S \|\mathbf{z}(\mathbf{x})\| f(\mathbf{x})\, d\mathbf{x} < \infty$. Let $\tilde{f}$ be proportional to $f(\mathbf{x}) - \mathbf{z}(\mathbf{x})^T \mathbf{A}^{-1} \int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x}$, normed so that $\int_S \tilde{f}^2(\mathbf{x})\, d\mathbf{x} = \eta^2$. Then $\int_S \mathbf{z}(\mathbf{x}) \tilde{f}(\mathbf{x})\, d\mathbf{x} = \mathbf{0}$, so that $\tilde{f} \in \mathcal{F}$, and hence $\mathbf{b}_{\tilde{f}}$, which is proportional to $\int_S (m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}^{-1}) \mathbf{z}(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x}$, must vanish. Because f is arbitrary, $(m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}^{-1})\mathbf{z}(\mathbf{x}) = \mathbf{0}$ a.e. $\mathbf{x} \in \mathcal{S}$. Thus $m(\mathbf{x})$ is an eigenvalue of $\mathbf{B}\mathbf{A}^{-1}$ if $\mathbf{z}(\mathbf{x}) \neq \mathbf{0}$, so that on $\mathcal{S}_0 := \{\mathbf{x} \in \mathcal{S} : \mathbf{z}(\mathbf{x}) \neq \mathbf{0}\}, m(\mathbf{x})$ can assume at most p distinct values. Decompose $\mathcal{S}_0$ as $\mathcal{S}_0 = \cup_{i=1}^s \mathcal{S}_i$, with $s \leq p$ and $m(\mathbf{x}) \equiv \alpha_i$ on $\mathcal{S}_i$. For any $\mathcal{S}_i$ with positive Lebesgue measure, the relationship $(\alpha_i \mathbf{I} - \mathbf{B}\mathbf{A}^{-1})\mathbf{z}(\mathbf{x}) \equiv \mathbf{0}$, together with assumption A), forces $\alpha_i \mathbf{I} = \mathbf{B}\mathbf{A}^{-1}$, so that at most one set $\mathcal{S}_i$ can have positive measure. Thus $m(\mathbf{x})$ is almost everywhere constant on $\mathcal{S}_0$, and hence on $\mathcal{S}$ itself because, again by A), $\mathcal{S} \setminus \mathcal{S}_0$ is of measure 0.

Proof of Theorem 3

Using the symmetry of $m(\mathbf{x})$, one finds that

$$1 + \lambda_m = \Omega^{-1} \max\left(\int_S m^2(\mathbf{x})\, d\mathbf{x}, \int_S \frac{x_1^2 m^2(\mathbf{x})}{(q+2)\gamma^2}\, d\mathbf{x} \right).$$

Under the stated conditions, $m_0(\mathbf{x})$ minimizes

$$\Phi(m) := \Omega^{-1} \int_S m^2(\mathbf{x})\, d\mathbf{x} + \nu\Omega^{-1/2} \times \left(\int_S (l_m(\mathbf{x}) m(\mathbf{x}))^2\, d\mathbf{x} \right)^{1/2},$$

and $\Omega^{-1} \int_S m_0^2(\mathbf{x})\, d\mathbf{x} = 1 + \lambda_{m_0}$.

First, fix γ so that $l_m(\mathbf{x}) = \Omega^{-1} l(\|\mathbf{x}\|; \gamma)$ no longer depends on m and $\Phi(m)$ is a convex functional of m. Observe that if $\mathbf{Q}$ is any orthogonal matrix and the density $m_{\mathbf{Q}}(\mathbf{x})$ is defined by $m_{\mathbf{Q}}(\mathbf{x}) = m(\mathbf{Q}\mathbf{x})$, then $\Phi(m_{\mathbf{Q}}(\mathbf{x})) = \Phi(m)$. Thus $\Phi(.5m + .5m_{\mathbf{Q}}) \leq \Phi(m)$, and it follows that to minimize Φ one need only consider densities satisfying $m = m_{\mathbf{Q}}$ for all $\mathbf{Q}$. Such densities are spherically symmetric: $m(\mathbf{x}) = h(\|\mathbf{x}\|)$ for some nonnegative function h satisfying $\int_0^1 (qu^{q-1}/\Omega) h(u)\, du = 1$. The integrand here is the density of $U = \|\mathbf{X}\|$. Constancy of γ is (7) with h_0 replaced by h.

Let $h_1(u)$ be any nonnegative function satisfying these constraints. Put $h_t(u) = (1-t)h_0(u;\gamma) + th_1(u)$ for $t \in [0,1]$ and define

$$\begin{aligned} \phi(t) &= \Phi(m_t) - 2\Omega^{-1} a\nu(b + q\gamma) \\ &= \Omega^{-2} \int_0^1 qu^{q-1} h_t^2(u)\, du + \nu\Omega^{-2} \end{aligned}$$

$$\times \left(\int_0^1 l^2(u;\gamma) q u^{q-1} h_t^2(u)\, du \right)^{1/2}$$
$$- 2\Omega^{-2} \int_0^1 a\nu(b+u^2) q u^{q-1} h_t(u)\, du. \qquad (A.1)$$

The multipliers have been arranged so that $\phi'(0) = 0$; this together with the convexity of $\phi(t)$ yields $\phi(0) \le \phi(1)$, so that $h_0(\|\mathbf{x}\|;\gamma)$ minimizes $\Phi(m)$ for fixed γ.

Writing the definition of $h_0(u;\gamma)$ in the form $(1 + c\nu l^2(u;\gamma))h_0(u;\gamma) = a\nu(b+u^2)$, then taking expectations and applying (6)–(8) gives $E[h_0(U;\gamma)] + (4\Omega c^2)^{-1} = a\nu(b+c\gamma)$, so that, from (A.1), $\Phi(h_0(\|\cdot\|;\gamma) = \Omega^{-1}a\nu(b+q\gamma) + \nu(4c\Omega^2)^{-1}$, which is minimized by γ_0.

Thus m_0 minimizes $\Phi(m)$ unconditionally. If (10) holds, then $\Omega^{-1}\int_S m_0^2(\mathbf{x})\,d\mathbf{x} = 1 + \lambda_{m_0}$, and it follows that m_0 is minimax. The remaining statements of the theorem follow from Theorem 1.

Proof of Theorem 4

This is very similar to that of Theorem 3 and thus is omitted.

Proof of Theorem 5

This is immediate from Theorem 1b and Theorem 2.

Proof of Lemma 2

Define $\mathbf{p}(x) = (P_0(x), \ldots, P_q(x))^T$, and let $\mathbf{P}_{q+1\times q+1}$ be the matrix of coefficients of the Legendre polynomials, defined through $\mathbf{p}(x) = \mathbf{P}\mathbf{z}(x)$. Then with

$$\mathbf{D} = \operatorname{diag}(2, \ldots, (i+.5)^{-1}, \ldots, (q+.5)^{-1})),$$

one has

$$\mathbf{D} = \int_{-1}^1 \mathbf{p}(x)\mathbf{p}^T(x)\,dx = \mathbf{P}\int_{-1}^1 \mathbf{z}(x)\mathbf{p}^T(x)\,dx = \mathbf{PAP}^T$$

so that $\mathbf{A} = \mathbf{P}^{-1}\mathbf{D}\mathbf{P}^{-1^T}$. Then calculate that

$$h_q(x) = (q+1)^{-1}\sum_{i=0}^{q}(i+.5)P_i^2(x),$$

and formula 8.915.1 of Gradshteyn and Ryzhik (1980) gives (16). A standard asymptotic expansion for Legendre polynomials—formula 8.965 of Gradshteyn and Ryzhik (1980)—yields (17).

[Received July 1996. Revised March 1998.]

REFERENCES

Box, G. E. P., and Draper, N. R. (1959), "A Basis for the Selection of a Response Surface Design," *Journal of the American Statistical Association*, 54, 622–654.

Brownlee, K. A. (1965), *Statistical Theory and Methodology in Science and Engineering*, New York: Wiley.

Carroll, R. J., and Cline, D. B. H. (1988), "An Asymptotic Theory for Weighted Least Squares With Weights Estimated by Replication," *Biometrika*, 75, 35–43.

Cook, R. D., and Weisberg, S. (1983), "Diagnostics for Heteroscedasticity in Regression," *Biometrika*, 70, 1–10.

Daniel, C., and Wood, F. S. (1980), *Fitting Equations to Data*, New York: Wiley.

Dasgupta, A., Mukhopadhyay, S., and Studden, W. J. (1992), "Compromise Designs in Heteroscedastic Models," *Journal of Statistical Inference and Planning*, 32, 363–384.

Gradshteyn, I. S., and Ryzhik, I. M. (1980), *Table of Integrals, Series, and Products*, New York: Academic Press.

Hooper, P. M. (1993), "Iterative Weighted Least Squares Estimation in Heteroscedastic Linear Models," *Journal of the American Statistical Association*, 88, 179–184.

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.

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

Marcus, M. B., and Sacks, J. (1976), "Robust Designs for Regression Problems," in *Statistical Theory and Related Topics II*, eds. S. S. Gupta and D. S. Moore, New York: Academic Press, pp. 245–268.

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

Pritchard, D. J., and Bacon, D. W. (1977), "Accounting for Heteroscedasticity in Experimental Design," *Technometrics*, 19, 109–115.

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

Ryan, T. P. (1997), *Modern Regression Methods*, New York: Wiley.

Schulz, M., and Endrenyi, L. (1983), "Design of Experiments for Estimating Parameters With Unknown Heterogeneity of the Error Variance," in *Proceedings of the Statistical Computing Section, American Statistical Association*, pp. 177–181.

Staudte, R. G., and Sheather, S. J. (1990), *Robust Estimation and Testing*, New York: Wiley.

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

Wong, W.-K. (1992), "A Unified Approach to the Construction of Minimax Designs," *Biometrika*, 79, 611–619.

Wong, W.-K., and Cook, R. D. (1993), "Heteroscedastic G-Optimal Designs," *Journal of the Royal Statistical Society*, Ser. B, 55, 871–880.