7

Bias Constrained Minimax Robust Designs for Misspecified Regression Models

Douglas P. Wiens

University of Alberta, Edmonton, AB, Canada

Abstract: We exhibit regression designs and weights which are robust against incorrectly specified regression responses and error heteroscedasticity. The approach is to minimize the maximum integrated mean squared error of the fitted values, subject to an unbiasedness constraint. The maxima are taken over broad classes of departures from the 'ideal' model. The methods yield particularly simple treatments of otherwise intractable design problems. This point is illustrated by applying these methods in a number of examples including polynomial and wavelet regression and extrapolation. The results apply to generalized M-estimation as well as to least squares estimation. Two open problems - one concerning designing for polynomial regression and the other concerning lack of fit testing - are given.

Keywords and phrases: Extrapolation, generalized M-estimation, lack of fit, Legendre polynomials, optimal design, polynomial regression, wavelets, weighted least squares

7.1 Introduction

In this article we synthesize some recent findings concerning the interplay between the choice of design points for regression based data analyses, and the weights used in weighted least squares or generalized M-estimation. The regression model which we envisage is described in Section 7.2, and is one for which the ordinary least squares estimates are biased due to uncertainty in the specification of the response function by the experimenter. Furthermore, we allow for the possibility of error heteroscedasticity of a very general form.

The experimenter seeks robustness in the form of protection from the bias and efficiency loss engendered by model misspecification. The recommendations made here are that these robustness issues be addressed both at the design stage and at the estimation stage. We exhibit designs and regression weights which minimize the maximum loss, with the maximum evaluated over broad classes of departures from the fitted response function and from homoscedasticity, subject to a side condition of unbiasedness. The designs also leave degrees of freedom to allow for the exploration of regression responses other than the one initially fitted.

An appealing consequence of our methods is that some otherwise very intractable design problems become amenable to simple remedies. This latter point is illustrated by applying these methods in a number of examples. The case of multiple linear regression, over ellipsoidal or rectangular design spaces, is addressed in Section 7.3. Section 7.4 covers polynomial regression, a particularly difficult problem in robust design theory. Wavelet regression is treated in Section 7.5; in Section 7.6 we consider designs for the extrapolation of the regression estimates beyond the design space. In Section 7.7 we discuss post-design inference, in particular lack of fit testing. Section 7.8 treats generalized M-estimation. Two intriguing open problems are given - one in Section 7.4 and the other in Section 7.7.

7.2 General Theory

We 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 p-dimensional regressors $\mathbf{z}(\mathbf{x})$: $E[Y|\mathbf{x}] \approx \mathbf{z}^T(\mathbf{x})\theta$. The sites $\mathbf{x}_i$ are chosen from $\mathcal{S}$, a Euclidean design space with finite volume defined by $\int_{\mathcal{S}} d\mathbf{x} = \Omega^{-1}$. We define the "true" value of θ by requiring the linear approximation to be most accurate in the L^2-sense:

$$\theta := \arg\min_{\mathbf{t}} \int_{\mathcal{S}} (E[Y|\mathbf{x}] - \mathbf{z}^T(\mathbf{x})\mathbf{t})^2 d\mathbf{x}. \tag{7.1}$$

We then define $f(\mathbf{x}) = E[Y|\mathbf{x}] - \mathbf{z}^T(\mathbf{x})\theta$ and $\epsilon(\mathbf{x}) = Y(\mathbf{x}) - E[Y|\mathbf{x}]$; these definitions together with (7.1) imply that

$$\text{(i) } Y(\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\theta + f(\mathbf{x}) + \epsilon(\mathbf{x}), \text{ (ii) } \int_{\mathcal{S}} \mathbf{z}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \mathbf{0}. \tag{7.2}$$

We allow for the possibility that the variance of $\epsilon(\mathbf{x})$ is proportional to a function $g(\mathbf{x})$, with g varying over some class $\mathcal{G}$: $\mathrm{VAR}[\varepsilon(\mathbf{x})] = \sigma^2 g(\mathbf{x})$. We take the bound

$$\int_{\mathcal{S}} g^2(\mathbf{x}) d\mathbf{x} \leq \Omega^{-1}, \tag{7.3}$$

implying that $\sigma^2 = sup_{g \epsilon \mathcal{G}} \left(\int_{\mathcal{S}} \mathrm{VAR}^2[\epsilon(\mathbf{x})] \Omega d\mathbf{x} \right)^{1/2}$. Note that (7.3) allows for homoscedastic errors $\mathcal{G} = \{\mathbf{1}\}$, where $\mathbf{1}(\mathbf{x}) \equiv 1$.

We require a further condition in order that errors due to bias not swamp those due to variance. In keeping with the L^2-nature of (7.1), we assume that

$$\int_{\mathcal{S}} f^2(\mathbf{x}) d\mathbf{x} \leq \eta^2, \tag{7.4}$$

for some positive constant η^2. Neither σ^2 nor η^2 need be known to the experimenter in order for our results to be applied.

The balancing of bias and variance can be achieved in other ways than (7.4). Pesotchinsky (1982) and Li and Notz (1982) among others take

$$|f(\mathbf{x})| \leq \phi(\mathbf{x}) \tag{7.5}$$

for a known function $\phi(\mathbf{x})$ satisfying various assumptions. An apparently unavoidable consequence of this rather thin neighbourhood structure is that the resulting 'robust' designs have their mass concentrated at a small number of extreme points of the design space, and thus afford little opportunity to test the model for lack of fit or to fit alternate models. The neighbourhoods implied by (7.4) can on the other hand be viewed as somewhat broad, since for them any design with finite maximum loss must be absolutely continuous, hence must be approximated at the implementation stage. For details and discussion see Wiens (1992), where the conclusion is reached 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.*" Indeed, simulation studies carried out to compare the implementations of the continuous designs with some common competitors have consistently shown these designs to be very successful at reducing the mean squared error, against realistic departures which are sufficiently large to destroy the performance of the classical procedures yet small enough to be generally undetectable by the usual tests. See Wiens (1998, 1999) and Fang and Wiens (1999).

We propose to estimate θ by least squares, possibly weighted with nonnegative weights $w(\mathbf{x})$. Let ξ be the design measure, i.e. the distribution placing mass n^{-1} at each of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, and denote by $K(\mathbf{x})$ the corresponding distribution function. Define matrices $\mathbf{A}, \mathbf{B}, \mathbf{D}$ and a vector $\mathbf{b}$ by

$$\mathbf{A} = \int_{\mathcal{S}} \mathbf{z}(\mathbf{x}) \mathbf{z}^T(\mathbf{x}) d\mathbf{x}, \qquad \mathbf{B} = \int_{\mathcal{S}} w(\mathbf{x}) \mathbf{z}(\mathbf{x}) \mathbf{z}^T(\mathbf{x}) \xi(d\mathbf{x}),$$

$$\mathbf{D} = \int_{\mathcal{S}} w^2(\mathbf{x}) g(\mathbf{x}) \mathbf{z}(\mathbf{x}) \mathbf{z}^T(\mathbf{x}) \xi(d\mathbf{x}), \quad \mathbf{b} = \int_{\mathcal{S}} w(\mathbf{x}) \mathbf{z}(\mathbf{x}) f(\mathbf{x}) \xi(d\mathbf{x}).$$

In a more familiar regression notation these are

$$\mathbf{B} = n^{-1} \mathbf{Z}^T \mathbf{W} \mathbf{Z} \qquad \text{and} \qquad \mathbf{D} = n^{-1} \mathbf{Z}^T \mathbf{W} \mathbf{G} \mathbf{W} \mathbf{Z},$$

where $\mathbf{Z}$ is the $n \times p$ model matrix with rows $\mathbf{z}^T(\mathbf{x}_i)$ and $\mathbf{W}, \mathbf{G}$ are the $n \times n$ diagonal matrices with diagonal elements $w(\mathbf{x}_i)$ and $g(\mathbf{x}_i)$ respectively. The motivation for writing these quantities as integrals with respect to ξ will become apparent below, where we 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 measure, we do so only in the formal sense of a non-negative 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. The mean vector and covariance matrix of the weighted least squares estimate $\hat{\theta} = \mathbf{B}^{-1} \cdot n^{-1}\mathbf{Z}^T\mathbf{W}\mathbf{Y}$ (where $\mathbf{Y}$ is the data vector) are

$$E[\hat{\theta}] - \theta = \mathbf{B}^{-1}\mathbf{b},\ COV[\hat{\theta}] = \frac{\sigma^2}{n}\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}. \tag{7.6}$$

We estimate $E[Y|\mathbf{x}]$ by $\hat{Y}(\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\hat{\theta}$, and consider the resulting integrated mean squared error : $IMSE = \int_{\mathcal{S}} E[(\hat{Y}(\mathbf{x}) - E[Y|\mathbf{x}])^2]d\mathbf{x}$. This splits into terms due solely to estimation bias, estimation variance, and model misspecification:

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

where, with $\mathbf{H} = \mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1}$, the Integrated Squared Bias (ISB) and Integrated Variance (IV) are

$$ISB(f, w, \xi) = \int_{\mathcal{S}} \left(E[\hat{Y}(\mathbf{x}) - \mathbf{z}^T(\mathbf{x})\theta]\right)^2 d\mathbf{x} = \mathbf{b}^T\mathbf{H}\mathbf{b},$$
$$IV(g, w, \xi) = \int_{\mathcal{S}} VAR[\hat{Y}(\mathbf{x})]d\mathbf{x} = \frac{\sigma^2}{n} \cdot trace(\mathbf{H}\mathbf{D}).$$

We adopt the viewpoint of *approximate* design theory, and allow as a design measure ξ *any* distribution on $\mathcal{S}$. It can be shown - a formal proof can be based on that of Lemma 1 of Wiens (1992) - that if either of $\sup_f ISB(f, w, \xi)$ or $\sup_g IV(g, w, \xi)$ is to be finite, then ξ must necessarily be absolutely continuous. The resulting designs may be approximated and implemented by placing the design points at appropriately chosen quantiles of $K(\mathbf{x})$.

We say that the pair (ξ, w) is *unbiased* if it satisfies $sup_f ISB(f, w, \xi) = 0$. Equivalently, $E[\hat{\theta}] = \theta$ for all f. The pair is *minimum variance unbiased* (MVU) if it minimizes $\sup_{f,g} IMSE(f, g, w, \xi)$ subject to being unbiased. The following theorem, which can be established as in Wiens (1998) by standard variational methods, gives a necessary and sufficient condition for unbiasedness, and minimax weights. Before stating it we require some definitions. Let $k(\mathbf{x}) = K'(\mathbf{x})$ be the design density, and define $m(\mathbf{x}) = k(\mathbf{x})w(\mathbf{x})$. Assume, without loss of generality, that the average weight $\int_{\mathcal{S}} w(\mathbf{x})\xi(d\mathbf{x})$ is 1. Then m is a density on $\mathcal{S}$ and each of $\mathbf{b}$, $\mathbf{B}$ depends on (k, w) only through m. Rather than optimize over (k, w) we may optimize over (m, w) subject to the constraint $\int_{\mathcal{S}} m(\mathbf{x})/w(\mathbf{x})d\mathbf{x} = 1$. Define also $l_m(\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\mathbf{H}\mathbf{z}(\mathbf{x})$.

Theorem 7.2.1 *a) The pair (ξ, w) is unbiased iff $m(\mathbf{x}) \equiv \Omega$.*
b)For fixed $m(\mathbf{x})$, maximum Integrated Variance is

$$\sup_g IV(g, w, \xi) = \frac{\sigma^2}{n}\Omega^{-1/2}\left(\int_{\mathcal{S}}(w(\mathbf{x})l_m(\mathbf{x})m(\mathbf{x}))^2 d\mathbf{x}\right)^{1/2},$$

attained at the least favourable variance function $g_{m,w}(\mathbf{x}) \propto w(\mathbf{x})l_m(\mathbf{x})m(\mathbf{x})$. Maximum IV *is minimized by weights $w_m(\mathbf{x}) \propto \left(l_m^2(\mathbf{x})m(\mathbf{x})\right)^{-1/3} I(m(\mathbf{x}) > 0)$.*
c) MVU *designs and weights (ξ_*, w_*) for heteroscedastic errors are given by*

$$k_*(\mathbf{x}) = \frac{(\mathbf{z}(\mathbf{x})^T\mathbf{A}^{-1}\mathbf{z}(\mathbf{x}))^{\frac{2}{3}}}{\int_{\mathcal{S}}(\mathbf{z}(\mathbf{x})^T\mathbf{A}^{-1}\mathbf{z}(\mathbf{x}))^{\frac{2}{3}}d\mathbf{x}}, \quad (7.7)$$

$$w_*(\mathbf{x}) = \Omega/k_*(\mathbf{x}). \quad (7.8)$$

The least favourable variances satisfy $g_*(\mathbf{x}) = w_*(\mathbf{x})^{-1/2}$. If the errors are homoscedastic ($\mathcal{G} = \{\mathbf{1}\}$) the exponents $2/3$ in (7.7) are replaced by $1/2$.

Part *c)* of Theorem 7.2.1 is an immediate consequence of parts *a)* and *b)*, since $m(\mathbf{x}) \equiv \Omega$ implies that $l_m(\mathbf{x}) \propto \mathbf{z}(\mathbf{x})^T\mathbf{A}^{-1}\mathbf{z}(\mathbf{x})$. Note that under heteroscedasticity the minimax weights $w_*(\mathbf{x})$ are equal to $g_*(\mathbf{x})^{-2}$; recall that if $g(\mathbf{x})$ is *known* then the *efficient* weights are proportional to $g(\mathbf{x})^{-1}$.

In our consideration of the special cases in the following sections, we will take mathematically convenient, canonical forms of the design spaces. This is justified in each case by the following lemma.

Lemma 7.2.2 *Suppose that the variables $\mathbf{x} \in \mathcal{S}$ are subjected to an affine transformation $\mathbf{x} \to \mathbf{M}\mathbf{x}+\mathbf{c} =: \tilde{\mathbf{x}}$ with $\mathbf{M}$ nonsingular, that $\tilde{\mathcal{S}} := \{\tilde{\mathbf{x}}|\mathbf{x} \in \mathcal{S}\}$ and that the regressors are equivariant in that $\mathbf{z}(\tilde{\mathbf{x}}) = \mathbf{P}\mathbf{z}(\mathbf{x})$ for some nonsingular matrix $\mathbf{P}$. If (ξ_*, w_*) are MVU for the design problem with regressors $\mathbf{z}(\mathbf{x})$, $\mathbf{x} \in \mathcal{S}$ then the induced design $\tilde{\xi}_*$ with distribution function $\tilde{K}_*(\tilde{\mathbf{x}}) = K_*(\mathbf{M}^{-1}(\tilde{\mathbf{x}}-\mathbf{c}))$, and weights $\tilde{w}_*(\tilde{\mathbf{x}}) = w_*(\mathbf{M}^{-1}(\tilde{\mathbf{x}} - \mathbf{c}))$, are MVU for the design problem with regressors $\mathbf{z}(\tilde{\mathbf{x}})$, $\tilde{\mathbf{x}} \in \tilde{\mathcal{S}}$.*

Lemma 7.2.2 is established by checking that (7.7) and (7.8) hold when applied to $(\tilde{K}_*, \tilde{w}_*)$.

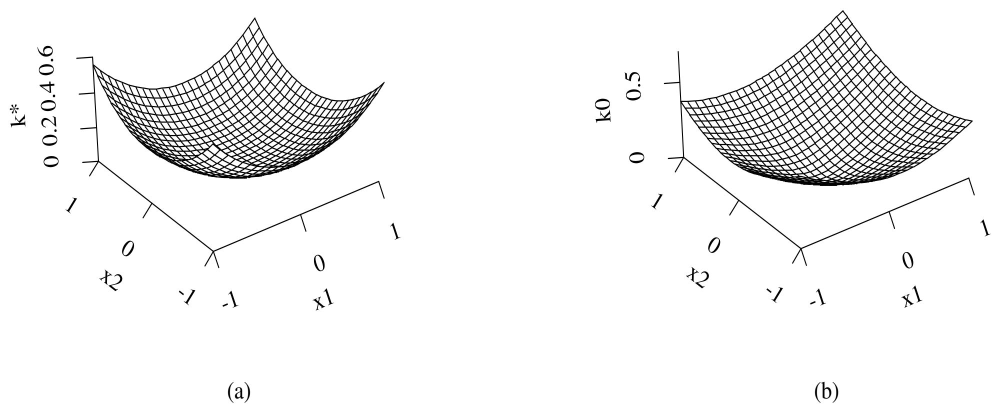

Figure 7.1: MVU design densities over the square $\mathcal{S} = [-1, 1] \times [-1, 1]$, robust against heteroscedasticity as well as response uncertainty. (a) Design for estimating a second order response. (b) Design for extrapolation of a first order fit to $\mathcal{T} = \{[-1, 2] \times [-1, 2]\} \setminus \mathcal{S}$.

7.3 Fitting a Second Order Response in Several Regressors

The following two examples illustrate the ease with which the MVU approach can be applied to otherwise quite intractable problems. Unconstrained, i.e. without the requirement of unbiasedness, minimax designs were obtained by Wiens (1990) for a greatly simplified version of the model in Section 7.3.2: $\mathbf{z}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$, $\mathcal{S} = [-1, 1] \times [-1, 1]$, $\mathcal{G} = \{\mathbf{1}\}$; even then the solution was very complex.

7.3.1 $\mathcal{S}$ an ellipsoid

Let $\mathbf{x} = (x_1, ..., x_q)^T$. Suppose that the experimenter anticipates fitting a full second order model, so that $\mathbf{z}(\mathbf{x})$ contains the elements $1, x_i, x_i^2, x_i x_j$ $(1 \leq i < j \leq q)$, and that $\mathcal{S}$ is a q-dimensional ellipsoid. By virtue of Lemma 7.2.2 we may assume that $\mathcal{S} = \{\mathbf{x} | \; ||\mathbf{x}|| \leq 1\}$. We then find that the MVU design has density

$$k_*(\mathbf{x}) \propto \left(1 + \frac{||\mathbf{x}||^2}{\mu_1} + \left(\sum_{i=1}^{q} x_i^4\right)\left(\frac{1}{\mu_2} - \frac{1}{2\mu_3}\right) + ||\mathbf{x}||^4 \left(\frac{1}{2\mu_3} - \frac{\mu_1}{\mu_2(\mu_2 + q\mu_1)}\right)\right)^{\frac{2}{3}}, \tag{7.9}$$

where

$$\begin{aligned}\mu_1 &= \Omega\int_{\mathcal{S}} x_1^2 d\mathbf{x} = \frac{1}{q+2},\\ \mu_2 &= \Omega\int_{\mathcal{S}} x_1^4 d\mathbf{x} = \frac{3}{(q+2)(q+4)},\\ \mu_3 &= \Omega\int_{\mathcal{S}} x_1^2 x_2^2 d\mathbf{x} = \frac{1}{(q+2)(q+4)}.\end{aligned}$$

7.3.2 $\mathcal{S}$ a q-dimensional rectangle

If $\mathbf{x}$ is as in Section 7.3.1 but $\mathcal{S}$ is a q-dimensional rectangle, which we may take as $\{\mathbf{x}|\ |x_i| \le 1,\ i = 1, ..., q\}$ by appealing to Lemma 7.2.2, then $k_*(\mathbf{x})$ is as in (7.9) with $\mu_1 = 1/3, \mu_2 = 1/5, \mu_3 = 1/9$. See Figure 7.3(a) for a plot of $k_*(\mathbf{x})$ with $q = 2$.

7.4 Fitting a Polynomial Response

The problem of designing for a polynomial fit, when it is assumed that the fitted polynomial form is exactly correct, has a long and rich history - see Pukelsheim (1993). Considerations of robustness against incorrectly specified response functions have entered the literature relatively recently. Stigler (1971) proposed a robustness criterion of designing for a polynomial fit of particular degree, whilst maintaining a lower bound on the efficiency if the true response is a higher order polynomial of fixed maximum degree. He gave particular solutions in the case of a linear fit with a quadratic alternative. Studden (1982) characterized this problem in terms of canonical moments and gave solutions for a linear fit against an alternative of arbitrary maximum degree. More recent approaches have focussed on minimizing a scalar-valued function of the covariance matrix of the estimates of the coefficients of the fitted polynomial, subject to a bound on the bias under alternative models. One can also seek to minimize bias while bounding variance. See Montepiedra and Fedorov (1997) for applications of this approach with polynomial alternatives and Liu and Wiens (1997) for arbitrary alternatives in neighbourhoods defined by (7.5).

Unconstrained minimax solutions for models given by (7.2) and (7.4), with $\mathbf{z}(x) = (1, x, x^2, \ldots x^q)^T$, have been notoriously elusive. It is a straightforward matter to maximize the loss over f - see Theorem 1 of Wiens (1992). However, even in the case $q = 2$ the problem of finding a design to minimize the resulting maximum loss remains unsolved. For $q = 1$ the solution was obtained by Huber (1975). Rychlik (1987) addressed concerns over the required continuity, hence lack of implementability, of Huber's (1975) design by proposing that one

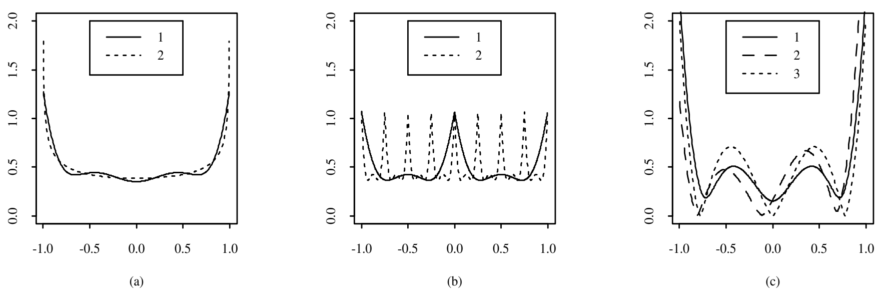

Figure 7.2: MVU design densities on $[-1,1]$, robust against heteroscedasticity as well as response uncertainty. (a) Polynomial regression. 1: $k_*(x, q=3)$; 2: $k_*(x,\infty)$. (b) Wavelet regression. 1: $k_{3,0}(x)$; 2: $k_{3,2}(x)$. (c) Extrapolation of a cubic fit. 1: $k_0(x, t_0 = 1.5)$, symmetric extrapolation; 2: $k_0(x, t_1 = 1.5)$, asymmetric extrapolation; 3: $k_0(x,\infty)$.

require f to be a polynomial of fixed maximum degree, but to otherwise satisfy (7.2) and (7.4). For IMSE loss he showed that, for large values of η, any design is minimax as long as its moments agree with those of Huber's design, up to a certain order. Thus he was able to construct discrete minimax designs. Heo (1998) extended these results to determinant loss.

In contrast, the MVU designs and weights are easily evaluated, for polynomial fits of any degree. By Lemma 7.2.2 we may take $\mathcal{S} = [-1,1]$. For q^{th} degree polynomial regression we write the optimal design density k_* as $k_*(x;q)$. This density turns out to be most conveniently expressed in terms of orthogonal polynomials, and this leads to an interesting connection to the classical D-optimal design ξ_D, i.e. the discrete measure minimizing the determinant of the covariance matrix of the ordinary least squares estimate $\hat{\theta}$. In the following result we denote by $P_q(x)$ the q^{th} degree Legendre polynomial on $\mathcal{S}$, normalized by $\int_{-1}^{1} P_q^2(x)dx = (q+.5)^{-1}$.

Theorem 7.4.1 *Define a density on* $[-1,1]$ *by* $h_q(x) = (q+1)^{-1}\mathbf{z}^T(x)\mathbf{A}^{-1}\mathbf{z}(x)$. *Then the design density* $k_*(x;q)$*, MVU for polynomial regression with heteroscedastic errors, satisfies*

$$k_*(x;q) \propto h_q(x)^{\frac{2}{3}} = .5(P_q(x)P'_{q+1}(x) - P'_q(x)P_{q+1}(x))^{\frac{2}{3}}.$$

See Wiens (1998) for a proof. The exponents $2/3$ are replaced by $1/2$ for homoscedastic errors.

It can be shown that the local maxima of $h_q(x)$, hence those of $k_*(x;q)$, are the zeros of $(1-x^2)P_q^{'}(x)$. But these are precisely the points of support of ξ_D. In this sense, $k_*(\cdot;q)$ is a smoothed version of ξ_D, which has the limiting density $(1-x^2)^{-1/2}/\pi = \lim_{q\to\infty} h_q(x)$. The limiting MVU density is

$$k_*(x;\infty) = \frac{(1-x^2)^{-\frac{1}{3}}}{2^{\frac{1}{3}}\beta(\frac{2}{3},\frac{2}{3})}.$$

See Figure 7.4(a) for plots of $k_*(x;3)$ and $k_*(x;\infty)$.

Placing designs points at the modes of $k_*(x;q)$ would recover the classically optimal but non-robust design ξ_D. We recommend placing design points at the n quantiles $x_i = K_*^{-1}\left(\frac{i-1}{n-1};q\right)$, $i=1,...,n$. Of course replication at a smaller number of locations is also an option, and is the subject of current research.

7.5 Wavelet Regression

The flexibility of wavelets in function representation methods has in recent years stimulated interest in wavelet approximations of regression response functions for the analysis of experimental data - see e.g. Antoniadis, Gregoire and McKeague (1994) and Benedetto and Frazier (1994). A primary attraction of wavelets in regression is their ability to approximate response functions which lack the smoothness properties of, e.g., polynomial responses. For instance if one assumes only that the response function is square integrable, with no further smoothness assumptions, then a *Haar* wavelet approximation is suitable. When the response is smoother, but not necessarily continuous, then a *multiwavelet* (Alpert 1992) approximation may be more appropriate. The class of multiwavelets contains the Haar wavelets.

Related design questions have arisen concurrently. For models in which the response function can be represented exactly as a linear combination of Haar wavelets see Herzberg and Traves (1994) with extensions by Xie (1998). Both unconstrained and constrained minimax design problems for wavelet approximations of the response function have been studied by Oyet (1997, 1998). Here we summarize the constrained solutions; details may be found in Oyet (1997) and Oyet and Wiens (1999).

Take $\mathcal{S}=[0,1)$ and assume that the regression response $E[Y|x]$ is in the space $L^2(\mathcal{S})$ of square integrable functions, so that it may be approximated arbitrarily closely by linear combinations of multiwavelets. Denote by $\mathbf{z}_{N,m}(x)$ the $N\cdot 2^{m+1}\times 1$ vector consisting of the wavelets $\{\phi_l(x),\ {}_N\nu_l^{-j,k}(x) \mid j = 0,...,m,\ k=0,...,2^j-1,\ l=0,...,N-1\}$ in some order. The elements of $\mathbf{z}_{N,m}$ form an orthogonal basis for $L^2(\mathcal{S})$ as $m\to\infty$. When $N=1$ they coincide with the Haar wavelets; in general they are described as follows. Recall from Section 7.4 that P_l denotes a normalized Legendre polynomial. In this notation $\phi_l(x) = \sqrt{2l+1}P_l(2x-1)I_{[0,1)}(x)$; also ${}_N\nu_l^{-j,k}(x) = 2^{j/2}{}_N\nu_l(\{2^jx\})I([2^jx]=k)$, where

$\{x\} = x - [x]$ denotes the *fractional* part of x. The *primary wavelets* ${}_N\nu_l$ can in turn be developed recursively; examples are ${}_1\nu_0(x) = I_{[0,1/2)}(x) - I_{[1/2,1)}(x)$ and

$$\begin{aligned} {}_2\nu_0(x) &= \sqrt{3}(4|x-1/2|-1)\cdot I_{[0,1)}(x), \\ {}_2\nu_1(x) &= 2(1-3|x-1/2|)\cdot(I_{[0,1/2)}(x) - I_{[1/2,1)}(x)). \end{aligned}$$

By virtue of the orthogonality of $\mathbf{z}_{N,m}(x)$, Theorem 7.2.1c) can be reduced to a particularly simple form in this case.

Theorem 7.5.1 *(Oyet and Wiens 1999) For the multiwavelet approximation the MVU design density $k_{N,m}(x)$ for homoscedastic errors is*

$$\kappa_N \cdot \left[\phi_{N-1}(\{2^{m+1}x\})\phi'_N(\{2^{m+1}x\}) - \phi'_{N-1}(\{2^{m+1}x\})\phi_N(\{2^{m+1}x\})\right]^{\frac{1}{2}},$$

where the normalizing constant is

$$\kappa_N = \left(\int_0^1 \left[\phi_{N-1}(x)\phi'_N(x) - \phi'_{N-1}(x)\phi_N(x)\right]^{\frac{1}{2}} dx\right)^{-1}.$$

The exponents $1/2$ are replaced by $2/3$ for heteroscedastic errors. The limiting density (for homoscedasticity) is

$$k_{\infty,m}(x) = \frac{\{2^{m+1}x\}^{-1/4}(1-\{2^{m+1}x\})^{-1/4}}{\beta(\frac{3}{4},\frac{3}{4})}.$$

A comparison with Theorem 7.4.1 reveals that $k_{N,m}(x)$ is a scaled and dilated copy of $k_*(x;N-1)$, extended periodically over $\mathcal{S}$ with period $length(\mathcal{S})/2^{m+1}$. Some particular cases are $k_{2,m}(x) = 2.51\cdot[(\{2^{m+1}x\}-1/2)^2+1/12]^{1/2}$ and $k_{3,m}(x) = 8.00\cdot[((\{2^{m+1}x\}-1/2)^2-1/20)^2+1/100]^{1/2}$. See Figure 7.4(b) for plots of $k_{3,0}(x)$ and $k_{3,2}(x)$ for heteroscedasticity, scaled to $x \in \mathcal{S} = [-1,1)$ for purposes of comparison with the densities in (a) and (c).

7.6 Extrapolation Designs

Spruill (1984) and Dette and Wong (1996) constructed extrapolation designs for polynomial regression, robust against various misspecifications of the degree of the polynomial. Draper and Herzberg (1973) extended the methods of Box and Draper (1959) to extrapolation under response uncertainty. In their approach one estimates a first order model but designs with the possibility of a second order model in mind; the goal is extrapolation to one fixed point outside of the spherical design space. Huber (1975) obtained designs for extrapolation of a response, assumed to have a bounded derivative of a certain order but to be otherwise arbitrary, to one point outside of the design interval. These results were corrected and extended by Huang and Studden (1988).

If the goal is extrapolation to a *range* of values outside of $\mathcal{S}$, robust against broad classes of alternate models, then the conclusions of Theorem 7.2.1 apply with only minor changes and yield easily implemented procedures. Suppose that one is to extrapolate the estimates of the regression response to a region $\mathcal{T}$ disjoint from $\mathcal{S}$, with (7.2)(i) holding on $\mathcal{T} \cup \mathcal{S}$. In the expressions of Section 7.2 replace the range $\mathcal{S}$ of the integrals defining IMSE, ISB and IV by $\mathcal{T}$, obtaining in this way integrated mean squared *prediction* error, etc. Redefine $\mathbf{H}$ to be $\mathbf{B}^{-1}\mathbf{A}_{\mathcal{T}}\mathbf{B}^{-1}$, where $\mathbf{A}_{\mathcal{T}} = \int_{\mathcal{T}} \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})d\mathbf{x}$. Then Theorem 7.2.1a),b) apply to the minimization of the resulting IMPSE, subject to $\sup IPB = 0$. The minimax design density under heteroscedasticity is (Fang and Wiens 1999)

$$k_0(\mathbf{x}) = \frac{(\mathbf{z}(\mathbf{x})^T\mathbf{A}^{-1}\mathbf{A}_{\mathcal{T}}\mathbf{A}^{-1}\mathbf{z}(\mathbf{x}))^{\frac{2}{3}}}{\int_{\mathcal{S}}(\mathbf{z}(\mathbf{x})^T\mathbf{A}^{-1}\mathbf{A}_{\mathcal{T}}\mathbf{A}^{-1}\mathbf{z}(\mathbf{x}))^{\frac{2}{3}}d\mathbf{x}} \tag{7.10}$$

(the exponents are 1/2 for homoscedastic errors) and correspondingly optimal weights are $w_0(\mathbf{x}) = \Omega/k_0(\mathbf{x})$.

7.6.1 Extrapolation of a polynomial fit

For q^{th} degree polynomial regression we find that

$$k_0(x) \propto \left(\sum_{i,j} \alpha_{ij} P_i(x) P_j(x)\right)^{2/3}, \qquad x \in \mathcal{S} = [-1, 1],$$

where $\alpha_{ij} = (i + .5)(j + .5)\int_{\mathcal{T}} P_i(x)P_j(x)dx$ for $0 \le i, j \le q$. For example, for quadratic regression and a symmetric extrapolation region, i.e. $\mathcal{T} = [-t_0,\ t_0]\backslash S$, we find that $k_0(x) = k_0(x; t_0)$ is given by

$$k_0(x;t_0) \propto [5t_0^3(t_0+1)(3x^2-1)^2 - t_0(t_0+1)(5x^4-22x^2+5) + 4(1-2x^2+5x^4)]^{\frac{2}{3}},$$

(an even function of x) while for one-sided extrapolation ($\mathcal{T} = [1,\ t_1]$) we find that

$$\begin{aligned} k_0(x;t_1) \propto [&5t_1^4(3x^2-1)^2 + 5t_1^3(3x-1)(x+1)(3x^2-1) - t_1^2(5x^4 - 30x^3 - 22x^2 \\ &+10x+5) - t_1(x+1)(5x^3 - 15x^2 - 7x + 5) + 2(10x^4 + 5x^3 - 4x^2 + x + 2)]^{\frac{2}{3}}. \end{aligned}$$

For both types of extrapolation region and for arbitrary but fixed q, $k_0(x;\infty) \propto \left(P_q^2(x)\right)^{2/3}$. See Figure 7.4(c) for plots of $k_0(x;1.5)$ in both the symmetric and asymmetric cases, and of $k_0(x;\infty)$.

7.6.2 Extrapolation of a first order response in several variables

$\mathcal{S}$ an ellipsoid

Suppose that $\mathcal{S}$ is as in Section 7.3.1 and has been transformed to a sphere of unit radius, that $\mathcal{T}$ is, after this transformation, the annulus $\{\mathbf{x}|1 < \ ||\mathbf{x}|| \leq t_2\}$ and that $\mathbf{z}(\mathbf{x}) = (1, x_1, ..., x_q)^T$. Evaluating (7.10) gives the optimal unbiased extrapolation design density

$$k_0(\mathbf{x}, t_2) \propto \left\{1 + (q+2)\frac{t_2^{q+2} - 1}{t_2^q - 1}\|\mathbf{x}\|^2\right\}^{2/3}.$$

$\mathcal{S}$ a q-dimensional rectangle

Suppose that again $\mathbf{z}(\mathbf{x}) = (1, x_1, ..., x_q)^T$ but that $\mathcal{S}$ is the q-dimensional cube $[-1, 1]^q$, and that the extrapolation region is the possibly asymmetric perimeter $\mathcal{T} = [-t_3, t_4]^q \backslash \mathcal{S}$, where $t_3, t_4 \geq 1$. One of t_3, t_4 may be unity, for one-sided extrapolation. We find that

$$k_0(\mathbf{x}, \mathbf{t}) \propto \left\{\left(1 + 3\mu_1 \sum_{i=1}^{q} x_i\right)^2 + 9\left(\mu_2 - \frac{1}{3\mu_3^q}\right)\|\mathbf{x}\|^2 - \frac{1}{\mu_3^q}\right\}^{2/3}, \qquad (7.11)$$

where $\mu_1 = (t_4 - t_3)/2$, $\mu_2 = (t_4 + t_3)^2/12$ and $\mu_3 = (t_4 + t_3)/2$. For symmetric extrapolation $t_3 = t_4$ and $\mu_1 = 0$.

See Figure 7.3(b) for a plot of $k_0(\mathbf{x})$ of (7.11) with $t_3 = 1,\ t_4 = 2$.

7.7 Lack of Fit Testing

If weighted least squares is used for reasons other than to address heteroscedasticity, then degrees of freedom are lost in estimating the error variance. The following result is Theorem 3.2 of Wiens (1999).

Lemma 7.7.1 *Suppose that the data obey the linear model $\mathbf{Y} = \mathbf{Z}\theta + \varepsilon$, where $\mathbf{Z}_{n \times p}$ has rank p and the elements of ε are uncorrelated random errors with mean* 0 *and variance σ^2. The weighted least squares estimate $\hat{\theta} = \left(\mathbf{Z}^T\mathbf{W}\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{W}\mathbf{Y}$ has mean θ and covariance matrix $\sigma^2\mathbf{C}$, where*

$$\mathbf{C} = \left(\mathbf{Z}^T\mathbf{W}\mathbf{Z}\right)^{-1}\left(\mathbf{Z}^T\mathbf{W}^2\mathbf{Z}\right)\left(\mathbf{Z}^T\mathbf{W}\mathbf{Z}\right)^{-1}$$

and $\mathbf{W}$ is the diagonal matrix of regression weights. Let $\mathbf{P}_\mathbf{V}$ be the projector onto the column space of $\mathbf{V} := (\mathbf{Z} \vdots \mathbf{W}\mathbf{Z})$ and denote the rank of $\mathbf{P}_\mathbf{V}$ by r.

Then $S^2 = ||(\mathbf{I}-\mathbf{P_V})\mathbf{Y}||^2/(n-r)$ *is an unbiased estimate of* σ^2. *The vector* $(\mathbf{I}-\mathbf{P_V})\mathbf{Y}$ *is uncorrelated with* $\hat{\theta}$. *If the errors are normally distributed then* $(n-r)S^2 \sim \sigma^2\chi^2_{n-r}$, *independently of* $\hat{\theta} \sim \mathbf{N}(\theta, \sigma^2\mathbf{C})$.

The projector $\mathbf{P_V}$ will typically have rank $r = 2p$ when the weights are non-constant, so that p degrees of freedom are lost in the estimate of σ^2, relative to ordinary least squares. Note that S^2 is easily obtained as the mean square of the residuals in a regression of $\mathbf{Y}$ on the columns of $\mathbf{V}$. Inferences on linear functions of θ are then carried out in the usual way, with the required change in the degrees of freedom.

As an example, consider the standard test for lack of fit, based upon groups of replicated observations and assuming the errors to be homoscedastic. The experimenter takes n_i observations Y_{ij} at each of locations $\mathbf{x}_i$, $i = 1, ..., c$. He computes the Pure Error estimate of σ^2: $S^2_{PE} = \sum(n_i - 1)S_i^2/(n-c)$, where S_i^2 is the sample variance of $\{Y_{i,1}, ..., Y_{i,n_i}\}$, and computes also the regression estimate S^2 of Lemma 7.7.1. Then the test of the hypothesis that $f \equiv 0$ in (7.2) consists of rejecting for large values of

$$T = \frac{(n-r)S^2 - (n-c)S^2_{PE}}{(c-r)\,S^2_{PE}}.$$

Under the null hypothesis both estimates of σ^2 are unbiased. Under (7.2) S^2_{PE} is unbiased but S^2 is positively biased, with bias $E[S^2 - \sigma^2] = ||(\mathbf{I}-\mathbf{P_V})\mathbf{f}||^2/(n-r)$. Here $\mathbf{f} = (f(\mathbf{x}_1), ..., f(\mathbf{x}_1), ..., f(\mathbf{x}_c), ..., f(\mathbf{x}_c))^T$ with $f(\mathbf{x}_i)$ appearing n_i times. If the errors are normal then T has a non-central F^{c-r}_{n-c} distribution (central under the null hypothesis). In this case the power of the test is an increasing function of the non-centrality parameter, which is in any event an intuitive measure of the quality of the procedure. This non-centrality parameter is $\lambda^2 = n\mathcal{B}(f, w, \xi)/\sigma^2$, where $\mathcal{B}(f, w, \xi) = n^{-1}||(\mathbf{I}-\mathbf{P_V})\mathbf{f}||^2$. A somewhat more informative expression is determined as follows. Let ξ be the design measure placing mass n_i/n at $\mathbf{x}_i$ and define r-vectors $\mathbf{q}_w(\mathbf{x})$ and $\mathbf{d}_{f,w,\xi}$, and an $r \times r$ matrix $\mathbf{Q}$ by

$$\begin{aligned}
\mathbf{q}_w(\mathbf{x}) &= \left(\mathbf{z}^T(\mathbf{x}), \mathbf{z}^T(\mathbf{x})w(\mathbf{x})\right)^T, \\
\mathbf{d}_{f,w,\xi} &= \int_{\mathcal{S}} \mathbf{q}_w(\mathbf{x})f(\mathbf{x})\xi\,(d\mathbf{x}), \\
\mathbf{Q}_{w,\xi} &= \int_{\mathcal{S}} \mathbf{q}_w(\mathbf{x})\mathbf{q}_w^T(\mathbf{x})\xi\,(d\mathbf{x}).
\end{aligned}$$

Assume that $\mathbf{Q}_{w,\xi}$ has full rank. Then

$$\begin{aligned}
\mathcal{B}(f, w, \xi) &= \int_{\mathcal{S}} f^2(\mathbf{x})\xi\,(d\mathbf{x}) - \mathbf{d}^T_{f,w,\xi}\mathbf{Q}^{-1}_{w,\xi}\mathbf{d}_{f,w,\xi} \\
&= \int_{\mathcal{S}} \left(f(\mathbf{x}) - \mathbf{q}_w^T(\mathbf{x})\mathbf{Q}^{-1}_{w,\xi}\mathbf{d}_{f,w,\xi}\right)^2 \xi\,(d\mathbf{x}).
\end{aligned}$$

From this last expression we see that $\mathcal{B}(f, w, \xi)$ is the squared $L^2(\xi)$-distance from f to the closest linear combination of the elements of $\mathbf{q}_w(\mathbf{x})$.

Wiens (1991) showed that for ordinary least squares procedures, i.e. $w \equiv 1$ and $\mathbf{q}_w(\mathbf{x}) = \mathbf{z}(\mathbf{x})$, the continuous uniform design with density $k(\mathbf{x}) \equiv \Omega$ has the property of maximizing the minimum value of λ^2, hence of the power of the test under the normality assumption, as f ranges over the class

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

The inequality in the definition of $\mathcal{F}^+$ serves to separate the null and alternate hypotheses. A related and open problem in the current context is to choose maximin weights w_ξ for a fixed design ξ, i.e. $w_\xi = \arg\min_w \min_{f \in \mathcal{F}^+} \mathcal{B}(f, w, \xi)$.

7.8 Generalized M-Estimation

For a Mallows-type generalized (or 'Bounded Influence') M-estimate defined by

$$\hat{\theta}_{GM} = argmin_\theta \int_S \rho \left(\frac{Y(\mathbf{x}) - \mathbf{z}^T(\mathbf{x})\theta}{\sigma} \right) w(\mathbf{x}) \xi(d\mathbf{x})$$

the asymptotic bias of $\sqrt{n}\ \hat{\theta}_{GM}$ is $\mathbf{B}^{-1}\mathbf{b}$ and, with $\psi = \rho'$, the asymptotic covariance matrix is $\nu \mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}$ where $\nu = \sigma^2 E[\psi^2(\epsilon/\sigma)]/E[\psi'(\epsilon/\sigma)]^2$. See Hampel *et al.* (1986) and Wiens (1996) for background material and details of the asymptotics, respectively. A comparison with (7.6) shows that the optimality properties of the MVU designs and weights, derived under the assumption that the estimation is to be carried out by weighted least squares, in fact hold for the case of generalized M-estimation as well.

Acknowledgments. This research was carried out with the support of the Natural Sciences and Engineering Research Council of Canada. We are also grateful for the careful review by the referee.

References

1. Alpert, B. K. (1992). Wavelets and Other Bases for Fast Numerical Linear Algebra, In *Wavelets: A Tutorial in Theory and Applications* (Ed., C. K. Chui), pp. 181–216, Boston: Academic Press.

2. Antoniadis, A., Gregoire, G. and McKeague, I. W. (1994). Wavelet methods for curve estimation, *Journal of the American Statistical Association*, **89**, 1340–1352.

3. Benedetto, J. J. and Frazier, M. W. (eds.) (1994). *Wavelets: Mathematics and Applications*, Boca Raton, Fla.: CRC Press.

4. 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.

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

6. Draper, N. R. and Herzberg, A. (1973). Some designs for extrapolation outside a sphere, *Journal of the Royal Statistical Society*, *Series B*, **35**, 268–276.

7. Fang Z. and Wiens, D. P. (1999). Robust extrapolation designs and weights for biased regression models with heteroscedastic errors, *The Canadian Journal of Statistics* (to appear).

8. Hampel, F. R., Ronchetti, E., Rousseeuw, R. J. and Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions*, New York: John Wiley & Sons.

9. Heo, G. (1998). Optimal Designs for Approximately Polynomial Regression Models, *Ph.D. Dissertation*, University of Alberta, Department of Mathematical Sciences.

10. Herzberg, A. M. and Traves, W. N. (1994). An optimal experimental design for the Haar regression model, *Canadian Journal of Statistics*, **22**, 357–364.

11. Huang, M. N. L. and Studden, W. J. (1988). Model robust extrapolation designs, *Journal of Statistical Planning and Inference*, **18**, 1–24.

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

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

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

15. Montepiedra, G. and Fedorov, V. V. (1997). Minimum bias designs with constraints, *Journal of Statistical Planning and Inference*, **63**, 97–111.

16. Oyet, A. J. (1997). Robust Designs for Wavelet Approximations, *Ph.D. Dissertation*, University of Alberta, Department of Mathematical Sciences.

17. Oyet, A. J. (1998). Wavelets and designs for heteroscedastic error regression models, *Technical Report*, Memorial University of Newfoundland, Department of Mathematics and Statistics.

18. Oyet, A. J. and Wiens, D. P. (1999). Robust designs for wavelet approximations of regression models, revisions requested by the *Journal of Nonparametric Statistics*.

19. Pesotchinsky, L. (1982). Optimal robust designs: Linear regression in R^k, *Annals of Statistics*, **10**, 511–525.

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

21. Rychlik, T. (1987). Robust Experimental Designs: A Comment on Huber's Result, *Zastosowania Matematyki Applicationes Mathematicae*, **19**, 93–107.

22. Spruill, M. C. (1984). Optimal designs for minimax extrapolation, *Journal Multivariate Analysis*, **15**, 52–62.

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

24. Studden, W. J. (1982). Some robust-type D-optimal designs in polynomial regression, *Journal of the American Statistical Association*, **77**, 916–921.

25. 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.

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

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

28. Wiens, D. P. (1996). Asymptotics of Generalized M-Estimation of Regression and Scale With Fixed Carriers, in an Approximately Linear Model, *Statistics and Probability Letters*, **30**, 271–285.

29. 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.

30. 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).

31. Xie, M.-Y. (1998). Some Optimalities of Uniform Designs and Projection Uniform Designs Under Multi-Factor Models, *Ph.D. Dissertation*, Hong Kong Baptist University, Department of Mathematics.