Robust Regression Designs When the Design Space Consists of Finitely Many Points

by

K.C. Li
Purdue University

Technical Report #81-45

Department of Statistics
Purdue University

1981

#81-45

ROBUST REGRESSION DESIGNS WHEN THE DESIGN SPACE CONSISTS OF FINITELY MANY POINTS*

Short title. ROBUST REGRESSION DESIGNS

Ker-Chau Li

Department of Statistics
Purdue University
West Lafayette, IN 47907

SUMMARY

We consider the nearly linear regression problem when the assumed first degree model is contaminated by some small constant, and the design space consists of finitely many points, symmetrically distributed on the interval $[-\frac{1}{2}, \frac{1}{2}]$. Under the usual squared loss function for the estimation of the slope and the intercept, and with the use of the least squares estimators, the problem is to find the designs which are optimal in the sense of minimizing the maximum risk among symmetric designs. The results turn out to be quite different from those obtained by Li and Notz (1981) in a setup that is similar except that the design space is the whole interval $[-\frac{1}{2}, \frac{1}{2}]$. In many cases the optimal solution has a support containing more than two points.

KEY WORDS AND PHRASES. Antisymmetric contaminations, design space, least squares estimators, robust designs, symmetric contaminations.

AMS Subject Classification. Primary 62J05, 62K05: Secondary 62G35.

*Research sponsored by the National Science Foundation under grant No. MCS-78-25301. Revision supported by NSF grant No. MCS-8200631.

ROBUST REGRESSION DESIGNS WHEN THE DESIGN SPACE CONSISTS OF FINTELY MANY POINTS

1. Introduction

Suppose we have a regression setting given by

(1.1) $$Y(x_i) = f(x_i) + e_i, \quad i = 1,\ldots,n$$

where the $\{e_i\}$ are uncorrelated random variables with mean 0 and variance σ^2. The x_i are elements of a compact subset X of a Euclidean space, and f is a real-valued function on X from a class F_0. Without robustness considerations, F_0 is typically composed of linear combinations of specified functions $f_0, f_1, \ldots, f_K$. The regression problem is concerned with making some inference about the unknown coefficients of these specified f_j and the associated optimal design problem is to choose the x_i's in an optimal manner for this inference (see Kiefer (1974) and the references given there for results on optimal regression designs). However, as first discussed by Box and Draper (1959), there are some dangers (e.g. in estimation, there may result a large bias term) inherent in a strict formulation of F_0 which ignores the possibility that the true f may only be approximated by an element of F_0; in other words, the true f may be equal to an element f_0 in the assumed model plus a contamination g, which is not in F_0. Thus, instead of (1.1), we consider

$$Y(x_i) = f_0(x_i) + g(x_i) + e_i, \quad i = 1,\ldots,n$$

where x_i's and e_i's are as before, f_0 is in F_0, and g is in another class G. A careful description of some problems in this context is given by Kiefer (1973) in the case where G is a finite dimensional space, which is a common assumption made by most authors before then.

In this paper, we consider the case that $F_0 = \{f_0 \mid f_0(x) = \alpha + \beta x;$ $\alpha, \beta \in R\}$ and $G = \{g \mid |g(x)| \leq \delta, \ \min_{\alpha,\beta} \int (g(x) - \alpha - \beta x)^2 \, d\mu(x) = \int g^2(x) \, d\mu(x)\}$, with $\delta \geq 0$ and $\mu(\cdot)$ being the uniform probability measure on $X \subset R$. When $\delta = 0$, this is exactly the ideal straight line model. To allow the true mean function to be deviated from the ideal straight line, we may set $\delta > 0$. This δ measures the amount of model-violation by the sup-norm, a quantity that is extremely simple for the designer to think of. The equality that $\min_{\alpha \ \beta} \int (g(x) - \alpha - \beta(x)^2 \, d\mu(x) = \int g^2(x) \, d\mu(x)$ is merely to make the slope and the intercept of the ideal straight line identifiable. Hence our assumptions provide a natural model-robust setting for guarding against possible departures from straight lines. Since our purpose is to find good designs, the estimates will be restricted to <u>the least squares ones</u> (denoted by $\hat{\alpha}, \hat{\beta}$). Consider the loss function $w_1^2 \, (\alpha - \hat{\alpha})^2 + w_2^2 \, (\beta - \hat{\beta})^2$ with w_1 and w_2 being specified non-negative numers. Apparently, the design minimizing the maximum risk may depend on σ^2, δ, n, w_1, w_2 and X. We shall focus on the case that $X = \{ \frac{k}{2N}, -\frac{k}{2N} \mid k = 1, 2, \ldots, N\}$ for a fixed natural number N (similar results may be obtained for the case that X consists of any $2N$ or $2N + 1$ points symmetrically distributed on $[-1/2, 1/2]$). As usual, a design will be denoted by a probability measure ξ on X. Thus the number of observations to be made at point $x = \frac{k}{2N}$ is $n \, \xi(\frac{k}{2N})$. Approximate designs for which $n \, \xi(\frac{k}{2N})$ needs not to be an integer will be considered here. In addition, since the problem is symmetric about 0, we shall seek designs among those symmetric ones; i.e., $\xi(\frac{k}{2N}) = \xi(\frac{-k}{2N})$, for $k = 1,\ldots,N$. Denoting the set of all symmetric designs by Ξ, we shall find an ξ in Ξ minimizing

$$(1.2) \qquad \max_{g\in G,\ \alpha,\beta\in R} E\{w_1^2(\alpha-\hat{\alpha})^2 + w_2^2(\beta-\hat{\beta})^2\}.$$

Put $R(g,\xi;w_1,w_2) = w_1^2(\int g(t)d\xi(t))^2 + w_2^2 \cdot \frac{\sigma^2}{n}(\int t^2 d\xi(t))^{-1} + w_2^2(\int g(t)td\xi(t)/\int t^2d\xi(t))^2$. A straightforward computation shows that $E\{w_1^2(\alpha-\hat{\alpha})^2 + w_2^2(\beta-\hat{\beta})^2\} = R(g,\xi;w_1,w_2) + w_1^2 \cdot \frac{\sigma^2}{n}$.

Since $w_1^2 \frac{\sigma^2}{n}$ is a constant, minimizing (1.2) amounts to minimizing

$$(1.3) \qquad \max_{g\in G,\ \alpha,\beta\in R} R(g,\xi;w_1,w_2) \quad .$$

Mathematically, it is easier to consider two classes of contaminations: $G^s = \{g \mid g\in G \text{ and } g(x) = g(-x),\ x\in X\}$ and $G^a = \{g \mid g\in G \text{ and } g(x) = -g(-x),\ x\in X\}$. Respectively in Section 2 and Section 3, we shall find the designs minimizing

$$(1.4) \qquad \max_{g\in G^s,\ \alpha,\beta\in R} R(g,\xi;w_1,w_2)$$

$$(1.5) \qquad \max_{g\in G^a,\ \alpha,\beta\in R} R(g,\xi;w_1,w_2)$$

Section 4 combines the results of Sections 2 and 3 to yield the solution of minimizing (1.3). Denote the uniform measure on $\{ \pm \frac{i}{2N}, \pm \frac{i+1}{2N}, \ldots, \pm \frac{N}{2N} \}$ by μ_i and the uniform measure on $\{ \pm \frac{i}{2N} \}$ by ζ_i. Qualitatively the robust optimal design we obtain is either a mixture of μ_1 and ζ_N, or a mixture of μ_{jo} and ζ_N, or a mixture of μ_j and ζ_{j-1} with $j \leq j^o$, where j^o is a fixed (depending only on N) integer defined by a certain cubic polynomial $\underset{\sim}{P}$ (see Section 3 for the definition). For large N, $j^o \approx (\sqrt{6} - 2)N$. Bickel and Herzberg (1979) and Bickel et al. (1981) obtained designs robust against the distributional dependence among observations. Their designs look somewhat similar to ours. The theoretical connection is however still unknown.

As an illustration of the results obtained, we list the robust designs for N = 7 in Section 5. Section 6 discusses what may happen when N tends to ∞.

Let us briefly review some relevant papers before closing this section. Huber (1975) considered essentially the same problem as ours with $X = [-1/2, 1/2]$ and took L_2 - norm to measure the amount of model-violation. However his formulation leads to the restriction that the designs must be absolutely continuous with respect to Lebesque measure, which means no implementable designs are considered. Although Huber mentioned that his designs should be approximated by finite support designs, the crucial problem about the sense and manner of approximation have still not been discussed. Marcus and Sacks (1976) considered a different class of model-violations. They took $G = \{g \mid |g(x)| \leq \phi(x)\}$ for a specified function

ϕ with $\phi(o) = 0$ The designs they considered have finite supports and they did not restrict to the least squares estimates. But the assumption that $\phi(o) = 0$ means that there is no contamination at the point $x = 0$. Pesotchinsky (1983) extended some of their results to linear regression in R^k. Li and Notz (1981) essentially used the same formulation as that of our paper except that they took X to be the entire interval [-1/2, 1/2]. The results they obtained were however rather different from what we have here. They showed that the design putting masses equally on two points 1/2 and -1/2 is optimal for any values of σ^2, δ, n, w_1, and w_2. Note that in Li and Notz, designs with infinite support were not considered. In fact there exist designs with infinte supports that are better than the two point design. Li and Notz also considered the case that $X \subset R^k$.

To avoid triviality, we assume that $N \geq 3$.

2. Symmetric contaminations

In this section, we shall find an $\xi \in \Xi$ minimizing (1.4).

First, because any symmetric g contributes no bias for the estimation of the slope, we have $R(g,\xi; w_1,w_2) = w_1^2(\int g(t)d\xi(t))^2+w_2^2 \cdot \frac{\sigma^2}{n} \cdot (\int t^2 d\xi(t))^{-1}$. Let $\xi_{(i)}$ be the i-th smallest values in $\{\xi(\frac{k}{2N}): k = 1,\ldots,N\}$ and define ξ^* by $\xi^*(\frac{k}{2N}) = \xi^*(\frac{-k}{2N}) = \xi_{(k)}$, $k = 1,\ldots,N$. It is clear that

$$(2.1) \qquad \int t^2 d\xi(t) \leq \int t^2 d\xi^*(t).$$

The least favorable contamination, say h, for ξ^* can be constructed easily. When N is even, $h(\frac{k}{2N}) = h(\frac{-k}{2N}) = \delta$ for $k \geq \frac{N}{2} + 1$, and $h(\frac{k}{2N}) = h(\frac{-k}{2N}) = -\delta$ for $k \leq \frac{N}{2}$. When h is odd, $h(\frac{k}{2N}) = h(\frac{-k}{2N}) = \delta$ for $k > \frac{N+1}{2}$, $h(\frac{k}{2N}) = h(\frac{-k}{2N}) = -\delta$ for $k < \frac{N+1}{2}$, and $h(\frac{k}{2N}) = h(-\frac{k}{2N}) = 0$ for $k = \frac{N+1}{2}$. Now taking h_ξ to be the function in $G^S_{N,\delta}$ such that if $\xi(t) = \xi_{(k)}$ then $h_\xi(t) = h(\frac{k}{2N})$, we have

$$\max_{g\in G^s} w_1^2\left(\int g(t)d\xi(t)\right)^2 \geq w_1^2\left(\int h_\xi(t)d\xi(t)\right)^2$$

$$= w_1^2\left(\int h(t)d\xi^*(t)\right)^2$$

$$= \max_{g\in G^s} w_1^2\left(\int g(t)d\xi^*(t)\right)^2.$$

This together with (2.1) shows that to minimize (1.4), we need only to minimize

(2.2) $$w_1^2\left(\int h(t)d\xi(t)\right)^2 + w_2^2\cdot\frac{\sigma^2}{n}\cdot\left(\int t^2 d\xi(t)\right)^{-1},$$

subject to

(2.3) $$0 \leq \xi\left(\tfrac{1}{2N}\right) \leq \xi\left(\tfrac{2}{2N}\right) \leq \cdots \leq \xi\left(\tfrac{1}{2}\right).$$

ξ will be said to be <u>nondecreasing</u> if (2.3) holds.

For any $a \geq 0$, define $\Xi_a = \{\xi\colon \int h(t)d\xi(t) = a\}$. We first minimize (2.2) subject to (2.3) over the class Ξ_a. This is equivalent to maximizing

(2.4) $$2\sum_{k=1}^{N}\left(\tfrac{k}{2N}\right)^2\xi\left(\tfrac{k}{2N}\right),$$

subject to (2.3) and

(2.5) $$2\sum_{k=1}^{N} h\left(\tfrac{k}{2N}\right)\xi\left(\tfrac{k}{2N}\right) = a.$$

Thus by the knowledge of linear programming, we claim that the solution vector $(\xi(\frac{1}{2N}),\ldots,\xi(\frac{N}{2N}))$ cannot take more than 2 distinct nonzero values as its coordinates. Note that besides (2.5) there is another linear constraint involved:

$$(2.6) \qquad 2\sum_{k=1}^{N} \xi\left(\frac{k}{2N}\right) = 1.$$

The above claim can also be verified directly by taking $x_i = \xi(\frac{i}{2N}) - \xi(\frac{i-1}{2N})$, $i = 1,\ldots,N$, rewriting (2.2) - (2.6) in terms of x_i, and showing that at most two x_i's are nonzero.

Next, the following lemma further reduces our consideration to ξ with the form $p\zeta_N + (1-p)\mu_1$ where $0 \le p \le 1$. Recall the definition of μ_j from Section 1.

Lemma 2.1. For ξ of the form $p\mu_i + (1-p)\mu_j$ where $1 \ge p \ge 0$ and $\{i,j\} \ne \{1,N\}$, there exists a ξ' such that $\max_{g\in G^s} R(g,\xi;\, w_1,w_2) \ge \max_{g\in G^s} R(g,\xi';\, w_1,w_2)$.

The proof of this lemma will be given in the Appendix. Write $S_N = \int t^2 d\mu_1(t)$. Now, by a simple computation, we get

$$(2.7) \qquad \max_{g\in G^s} R(g,p\zeta_N+(1-p)\mu_1;\, w_1,w_2) = w_1^2p^2\delta^2 + w_2^2\cdot\frac{\sigma^2}{n}\ \left(\frac{1}{4}p + S_N(1-p)\right)^{-1}.$$

Thus minimizing (2.7) over $\{0 \le p \le 1\}$, we obtain the solution ξ minimizing (1.4). This is stated in the following:

Proposition 2.1. Suppose p solves the equation

$$(2.8) \qquad 2p\left(\left(\frac{1}{4} - S_N\right)p + S_N\right)^2 = \min\left\{\left(\frac{w_2}{w_1}\right)^2\cdot\frac{\sigma^2}{n\delta^2}\left(\frac{1}{4} - \delta_N\right),\ \frac{1}{8}\right\}.$$

Then $p\mu_N + (1-p)\mu_1$ minimizes (1.4) over $\xi \in \Xi$.

3. Antisymmetric Contaminations

In this section, we shall find the ξ minimizing (1.5). All the proofs of the lemmas will be given in the Appendix. For any $g \in G^a$ it is clear that $R(g,\xi;\, w_1,w_2) = w_2^2[\frac{\sigma^2}{n}\cdot(\int t^2 d\xi(t))^{-1} + (\int g(t)t d\xi(t)/\int t^2 d\xi(t))^2]$. Thus we may assume $w_1 = 0$ and $w_2 = 1$ without loss of generality. For convenience, write $R(g,\xi) = R(g,\xi;\, 0,1)$. We proceed to compute $\max_{g\in G^a} R(g,\xi)$. Let $\underset{\sim}{i} = (i_1,\ldots,i_N)$ be a permutation of $(1,2,\ldots,N)$. Define $h_{\underset{\sim}{i}}$ by

$$\text{(3.1)}\qquad h_{\underset{\sim}{i}}\left(\frac{i_j}{2N}\right) = \begin{cases} 1 & , \quad \text{for } j > j^* \\ (i_{j^*})^{-1}\left(-\sum_{j>j^*} i_j + \sum_{j<j^*} i_j\right), & \quad \text{for } j = j^* \\ -1 & , \quad \text{for } j < j^*, \\ -h_{\underset{\sim}{i}}\left(\frac{-i_j}{2N}\right) & \end{cases}$$

where j^* is the unique integer such that

$$\text{(3.2)}\qquad -i_{j^*} < \sum_{j>j^*} i_j - \sum_{j<j^*} i_j \le i_{j^*}.$$

Take $\Xi_{\underset{\sim}{i}} = \{\xi:\ \xi(\frac{i_1}{2N}) \le \xi(\frac{i_2}{2N}) \le \cdots \le \xi(\frac{i_N}{2N})\}$.

<u>Lemma 3.1</u>. If $\xi \in \Xi_{\underset{\sim}{i}}$, then

$$\text{(3.3)}\qquad \max_{g\in G^a} R(g,\xi) = \frac{\sigma^2}{n}\cdot\left(\int t^2 d\xi(t)\right)^{-1} + \delta^2\left(\int h_{\underset{\sim}{i}}(t)\,t\,d\xi(t) \Big/ \int t^2 d\xi(t)\right)^2.$$

Unlike the case of Section 2, the nondecreasing ξ^* (defined in Section 2) does not always improve ξ. This makes the problem of minimizing (1.5) more difficult to solve. But some techniques used before will still be useful. More precisely, for any $a \ge 0$, take $\Xi_{\underset{\sim}{i},a} = \{\xi:\ \xi \in \Xi_{\underset{\sim}{i}}$ and $\int h_{\underset{\sim}{i}}(t)\,t\,d\xi(t) = a\}$. First we minimize (3.3) subject to $\xi \in \Xi_{\underset{\sim}{i},a}$. This is equivalent to maximizing $\int t^2 d\xi(t) = 2\sum_{k=1}^{N} (\frac{k}{2N})^2 \xi(\frac{k}{2N})$, subject to the conditions that $2\sum_{k=1}^{N} h_{\underset{\sim}{i}}(\frac{k}{2N})\cdot\frac{k}{2N}\cdot\xi(\frac{k}{2N}) = a$, $2\sum_{k=1}^{N} \xi(\frac{k}{2N}) = 1$, and $0 \le \xi(\frac{i_1}{2N}) \le \cdots \le \xi(\frac{i_N}{2N})$.

Thus as in Section 2 we obtain that the solution vector $(\xi(\frac{1}{2N}),\ldots,\xi(\frac{n}{2N}))$ can not take more than 2 distinct nonzero values as its coordinates. But this class of design measures is still too large. We need the following lemma to focus the search of the solutions to a smaller class.

Lemma 3.2. The design measure ξ minimizing (1.5) over $\xi \in \Xi$ is nondecreasing and takes at most 2 distinct nonzero values as its probability masses.

Write $\underset{\sim}{h} = h_{\underset{\sim}{i}}$ for $\underset{\sim}{i} = (1,2,\ldots,N)$. By Lemmas 3.1 and 3.2, it suffices to minimize

$$\frac{\sigma^2}{n}\cdot\left(\int t^2 d\xi(t)\right)^{-1} + \delta^2\left(\int \underset{\sim}{h}(t)\, t\, d\xi(t)/\int t^2 d\xi(t)\right)^2 \tag{3.4}$$

over the class

$$\Xi^0 = \{\xi = p\mu_i + (1-p)\mu_j \mid 0 \le p \le 1 \text{ and } 1 \le i < j \le N\}. \tag{3.5}$$

The problem now looks relatively simple. But to directly carry out the minimization (i.e., to express (3.4) in terms of p,i and j and then to take derivatives etc.) is still not easy because of the complexity of $\underset{\sim}{h}$. We find that the following lemma helps simplifying the matter greatly.

Take η_j to be the uniform design measure on $\{\pm \frac{j}{2N}, \pm \frac{j+1}{2N}, \ldots, \pm \frac{N-1}{2N}\}$; denote $\underset{\sim}{P}(x) = x^3+3(N-1)x^2-6(N^2+3N-2)x+N(N+1)(2N+1)$; let j^0 be the largest integer such that $\underset{\sim}{P}(j^0) > 0$ and $1 \le j^0 \le N$.

Lemma 3.3. For any j such that $1 < j < N$, there exists a pair of nonnegative numbers (r_j, s_j) such that $r_j + s_j = 1$,

$$\int t\cdot\underset{\sim}{h}(t)\, d\eta_j(t) = \int t\cdot\underset{\sim}{h}(t)\, d(r_j\, \zeta_{j-1}+s_j\zeta_N)(t), \tag{3.6}$$

$$\int t^2 d\eta_j(t) \le \int t^2 d(r_j\, \zeta_{j-1}+s_j\zeta_N)(t) \quad , \text{ for } j > j^0 \tag{3.7}$$

and

$$\int t^2 d\eta_j(t) \ge \int t^2 d(r_j\, \zeta_{j-1}+s_j\zeta_N)(t) \quad , \text{ for } j \le j^0. \tag{3.8}$$

Take $\Xi^1 = \{\xi = p\zeta_N+(1-p)\mu_{j^0} : 0 \le p \le 1\}$ and $\Xi^2 = \{\xi = p\mu_j+(1-p)\zeta_{j-1} : j = 2,\ldots,j^0$ and $0 \le 1-p \le p\cdot(N-j+1)^{-1}\}$. Clearly, $\Xi^1 \cup \Xi^2$ is a subclass of Ξ^0.

Using Lemma 3.3, we now demonstrate that the solution of minimizing (3.4) over Ξ^0 can be found in $\Xi^1 \cup \Xi^2$.

For any $\xi = p\mu_i + (1-p)\mu_j$ in Ξ^0, consider the following six different cases:

(i) $j = N$ and $i = j^0$.

(ii) $j = N$ and $i > j^0$.

(iii) $j = N$ and $i < j^0$.

(iv) $j^0 < j < N$.

(v) $j \leq j^0$ and $i = j-1$

(vi) $j \leq j^0$ and $i < j-1$.

Cases (i) and (v) lead to the desired "$\xi \in \Xi^1 \cup \Xi^2$". For case (ii), write $\xi = p\cdot(N-i)(N-i+1)^{-1}\cdot\eta_i+[p\cdot(N-i+1)^{-1}+(1-p)]\zeta_N$. Construct ξ' $= p\cdot(N-i)(N-i+1)^{-1}\cdot((1-\varepsilon)\eta_i+\varepsilon(r_i\zeta_{i-1}+s_i\zeta_N))+[p\cdot(N-i+1)^{-1}+(1-p)]\zeta_N$ where $\varepsilon > 0$, and r_i and s_i are defined in Lemma 3.3. Clearly, for a suitably chosen ε, ξ' is of the form $p'\mu_{i-1}+(1-p')\zeta_N$. Also, by (3.6) and (3.7) of Lemma 3.3 and (3.4), ξ' is at least as good as ξ. Repeating the above argument several times, we end up with an ξ of the case (i). This settles Case (ii). For Case (iii) the argument is similar. But instead of moving some masses from $\{\pm \frac{i}{2N}, \ldots, \pm \frac{N-1}{2N}\}$ to $\{\pm \frac{i-1}{2N}, \pm \frac{N}{2N}\}$, we now use (3.6) and (3.8) of Lemma 3.3 and move some masses from $\{\pm \frac{i}{2N}, \pm \frac{N}{2N}\}$ to $\{\pm \frac{i+1}{2N}, \ldots, \pm \frac{N-1}{2N}\}$. Repeating this argument several times, we may end up with the Case (i) or the Case (v) as desired. For the Case (vi), moving some masses from $\{\pm \frac{i}{2N}\}$ to $\{\pm \frac{j-1}{2N}\}$, we can reduce the bias $\int \underset{\sim}{h}(t)t\,d\xi(t)$ (here note that $h(t) < 0$ for $|t| < j^0/2N$) and increase the design variance $\int t^2 d\xi(t)$. For the Case (iv), using argument similar to that in Case (ii), we may obtain an ξ' of the form $p'\mu_i + (1-p')\zeta_N$ or $p\mu_i+(1-p)[p'\mu_{j^o}+(1-p')\zeta_N]$. The former case belongs to Cases (i) $\sim$ (iii); for the latter case, the arguments similar to those for the Case (vi) and the Case (iii) will then lead to the desired result.

We now restrict our attention to $\Xi^1 \cup \Xi^2$. The design variances $\int t^2 d\xi(t)$ in this class run from S_N to $\frac{1}{4}$ and the correspondence between designs and design variances is one to one. Moreover, as the design variance increases the bias $\int h(t)\, t\, d\xi(t)$ decreases. Thus each design in this class should be optimal for some (σ,δ). To actually find the optimal design for a given (σ,δ), we need to compute (3.4) for $\xi \in \Xi^1 \cup \Xi^2$. This is done in the following.

Let $V_j = (\sum_{k\geq j}^{N} k^2)/4N^2(N-j+1)$ and $B_j = j(j-1)/4N(N-j+1)$. For $\xi \in \Xi^1$,

$$\max_{g\in G^a} R(g,\xi) = \frac{\sigma^2}{n}\, x + \delta^2 \cdot (1-4V_{j^o})^{-2} \cdot [2(1-2B_{j^o})-x(2V_{j^o}-B_{j^o})]^2,$$ where x is defined as $(\int t^2 d\xi(t))^{-1} = (\frac{1}{4}p+(1-p)V_{j^o})^{-1}$. For $\xi \in \Xi^2$, $\max_{g\in G^a} R(g,\xi) =$

$$\frac{\sigma^2}{n} \cdot x+\delta^2 \cdot (4V_j-(j-1)^2N^{-2})^{-2} \cdot [2(2B_j+ \frac{j-1}{N})-x \cdot \frac{j-1}{N} \cdot (2V_j+(j-1)N^{-1} \cdot B_j)]^2,$$

where x is defined to be $(\int t^2 d\xi(t))^{-1} = [(\frac{j-1}{2N})^2 p+V_j(1-p)]^{-1}$. These two functions are quadratic in x and can be rewritten as

$$\underset{\sim}{Q}(x) = \delta^2F_j^2H_j^2x^2 + (\frac{\sigma^2}{n} -2\delta^2F_j^2H_jG_j)x+\delta^2F_j^2G_j^2,$$

for $x \in [V_j^{-1}, V_{j-1}^{-1}]$, $j = 2,\ldots,j^o,\infty$, where

$F_j = [4V_j-(\frac{j-1}{N})^2]^{-1}$, $G_j = 4B_j+2(j-1)/N$, $H_j = (j-1)N^{-1}(2V_j+(j-1)N^{-1}B_j)$ for $j = 2,\ldots,j^o$; and $F_\infty = (1-4V_{j^o})^{-1}$, $G_\infty = 2(1-2B_{j^o})$, $H_\infty = 2V_{j^o}-B_{j^o}$, $V_\infty = \frac{1}{4}$, and $V_{\infty-1} = V_{j^o}$.

Lemma 3.4. $\underset{\sim}{Q}(x)$ is convex for $x \in [V_\infty^{-1}, V_1^{-1}]$.

Let $T_j = (G_j - H_j V_j^{-1}) F_j^2 H_j$ and $R_j = (G_j - H_j V_{j-1}^{-1}) F_j^2 H_j$ for $j = 2, \ldots, j^o, \infty$. By standard techniques and Lemma 3.4, the minimizer of $\underset{\sim}{Q}(x)$ over $x \in [V_\infty^{-1}, V_1^{-1}] = [4, S_N^{-1}]$ (hence the ξ minimizing (1.5)), can be derived directly. This is stated in the following

Proposition 3.1. The design ξ which minimizes (1.5) is described below:

(i) If $R_j \leq \frac{\sigma^2}{2n\delta^2} \leq T_j$ for some j such that $2 \leq j \leq j^o$, then

$$\xi = \left(\frac{V_j - \alpha}{V_j - (j-1)^2 N^{-2}}\right) \zeta_{j-1} + \left(\frac{\alpha - (j-1)^2 N^{-2}}{V_j - (j-1)^2 N^{-2}}\right) \mu_j,$$

where $\alpha = 2\delta^2 F_j^2 H_j^2 / (2\delta^2 F_j^2 G_j H_j - \sigma^2/n)$;

(ii) If $T_{j-1} \leq \sigma^2/2n\delta^2 \leq R_j$, for some j such that $2 \leq j \leq j^o$, then

$\xi = \mu_{j-1}$;

(iii) If $T_{j^o} \leq \sigma^2/2n\delta^2 \leq R_\infty$, then

$$\xi = \mu_{j^o};$$

(iv) If $R_\infty \leq \sigma^2/2n\delta^2 \leq T_\infty$, then

$$\xi = (\alpha - V_{j^o})\left(\frac{1}{4} - V_{j^o}\right)^{-1} \zeta_N + \left(\frac{1}{4} - \alpha\right)\left(\frac{1}{4} - V_{j^o}\right)^{-1} \mu_{j^o}$$

where $\alpha = 2\delta^2 F_\infty H_\infty / (2\delta^2 F_\infty^2 H_\infty G_\infty - \sigma^2/n)$,

(v) If $\frac{\sigma^2}{2n\delta^2} \geq T_\infty$, then

$$\xi = \zeta_N.$$

4. Robust designs

In this section we shall find the ξ minimizing (1.3).

First, suppose $w_1 = 0$. It is easy to see that $\max_{g\in G} R(g,\xi;\ 0,w_1) = \max_{g\in G^a} R(g,\xi;\ 0,w_1)$ because any g can be written a sum of a symmetric function g_1 and an antisymmetric function g_2 and the bias $\int g(t)td\xi(t) = \int g_2(t)td\xi(t)$. Thus Proposition 3.1 is applicable.

Theorem 4.1. When $w_1 = 0$, the ξ minimizing (1.3) is equal to the ξ of Proposition 3.1.

Next, we turn to the case $w_1 \neq 0$. The designs of the form $p\mu_N + (1-p)\mu_1$ are reasonable candidates because of Proposition 2.1. In fact, $R(g,\ p\mu_N + (1-p)\mu_1;\ w_1,w_2) = \frac{1}{4}\ p^2(g(\frac{1}{2})+g(-\frac{1}{2}))^2 \cdot w_1^2 + \frac{\sigma^2}{n} w_2^2 \cdot (\frac{1}{4}\ p+S_N(1-p))^{-1} + \frac{1}{16} \cdot w_2{}^2 \cdot (g(\frac{1}{2})-g(-\frac{1}{2}))^2 p^2(\frac{1}{4}\ p+S_N(1-p))^{-2}$. The maximum over $g \in G$ occurs when $g(\frac{1}{2}) = g(-\frac{1}{2}) = \delta$ or $g(\frac{1}{2}) = -g(-\frac{1}{2}) = \delta$. The former occurs when

$$w_2/w_1 \leq 2(\tfrac{1}{4}\ p+(1-p)S_N). \tag{4.1}$$

This is the case where the symmetric contamination dominates the antisymmetric one; i.e., $\max_{g\in G} R(g,p\zeta_N+(1-p)\mu_1;\ w_1,w_2) = \max_{g\in G^S} R(g,p\zeta_N+(1-p)\mu_1)$. Now, we obtain the following

Theorem 4.2. If $\frac{\sigma^2}{n\delta^2} \geq \frac{1}{2}(\frac{w_2}{2w_1} - S_N)(\frac{1}{4} - S_N)^{-2}$ and $\frac{w_2}{w_1} \leq \frac{1}{2}$, then $\xi = p\zeta_N+(1-p)\mu_1$ minimizes (1.3), where p solves the equation (2.8).

Proof. By Proposition 2.1 and the above argument, if p solves (2.8) and (4.1) holds, then $\xi = p\zeta_N + (1-p)\mu_1$ minimizes (1.3). On the other hand, it is straightforward to see that (2.8) and (4.1) imply the sufficient conditions in this theorem. $\square$

In a similar spirit, we shall derive the conditions under which the antisymmetric contaminations dominate the symmetric ones. This is suggested by the following lemma whose proof was given in Li (1981).

Lemma 4.1. For any $\xi \in \Xi^1 \cup \Xi^2$,

$$\max_{g \in G} R(g,\xi;\, w_1,w_2) = \max_{g \in G^s \cup G^a} R(g,\xi;\, w_1,w_2).$$

It remains to actually compute $\max_{g \in G^s} R(g,\xi;\, w_1,w_2)$ and $\max_{g \in G^a} R(g,\xi;\, w_1,w_2)$ for $\xi \in \Xi^1 \cup \Xi^2$. This involves only straightforward calculation. It turns out that the equality $\max_{g \in G} R(g,\xi;\, w_1,w_2) = \max_{g \in G^a} R(g,\xi;\, w_1,w_2)$ holds

(i) for $\xi \in \Xi^1$ (i.e., $\xi = p\zeta_N + (1-p)\mu_{j^o}$), if

$$(4.2)\quad \frac{w_2}{w_1} \geq 2(N-j^o+1)^{-1} \cdot [j-1+p(N-2j^o+2)] \cdot F_\infty H_\infty \cdot \left(\frac{\sigma^2}{n\delta^2}\right)^{-1},$$

and

(ii) for $\xi \in \Xi^2$ (i.e., $\xi = p\zeta_{j-1} + (1-p)\mu_j$), if

$$(4.3)\quad \frac{w_2}{w_1} \geq 2(N-j+1)^{-1} (j-1-pN) \cdot F_j H_j \cdot \left(\frac{\sigma^2}{n\delta^2}\right)^{-1}.$$

Now, by Proposition 3.1, we have

Theorem 4.3. If the choices of $\frac{\sigma^2}{n\delta^2}$ and $\frac{w_2}{w_1}$ are such that the assumptions of Proposition 3.1 hold and either (4.2) or (4.3) is satisfied, then the design ξ in Proposition 3.1 minimizes (1.3).

5. An illustration

In the following example, we take $N = 7$.

First, consider the case that $0 \leq \frac{w_2}{w_1} \leq \frac{1}{2}$. Theorem 4.2 is applicable here, with $S_N = S_7 = 0.1020$. The design space X is $\{\pm \frac{1}{14}, \pm \frac{2}{14}, \ldots, \pm \frac{1}{2}\}$; ζ_i puts masses equally on $\frac{i}{14}$ and $-\frac{i}{14}$; μ_i puts masses uniformly on $\{\pm \frac{i}{14}, \pm \frac{i+1}{14}, \ldots, \pm \frac{1}{2}\}$. Denote $\frac{\sigma^2}{n\delta^2} \equiv \lambda$ and $\frac{w_2}{w_1} \equiv \omega$. The following table lists the solutions according to different values of λ and ω.

Table 1

Range of $\frac{w_2}{w_1}$ $(\equiv \omega)$	Range of $\frac{\sigma^2}{n\delta^2}$ $(\equiv\lambda)$	Optimal designs	p
$0.2041 \sim 0.5000$	$11.42\omega-2.331 \sim 0.8448\cdot\omega^{-2}$	$p\zeta_7+(1-p)\mu_1$	$p(0.1480p+0.1020)^2=0.07398\omega^2\lambda$
$0.0000 \sim 0.2041$	$0.000 \sim 0.8448\cdot\omega^{-2}$		
$0.0000 \sim 0.5000$	$0.8448\cdot\omega^{-2} \sim +\infty$	ζ_7	——

Note that $p(0.1480p + 0.1020)^2$ is an increasing function of p. Thus we see that for a fixed ω, as λ increases, p also increases. This simply says that if the sampling variances tend to dominate the model violations then we tend to use the classical two-points design ζ_7. On the other hand, if the amount of the model violation turns out more serious than the sampling variances then we tend to use the uniform design μ_1. The above table illustrates how to actually achieve optimality by a suitable mixture of ζ_7 and μ_1.

Next, we turn to the case that $\frac{w_2}{w_1} \geq \frac{1}{2}$. Theorem 4.3 is applicable now. The value of j^o here turns out to be 3. Table 2 provides the optimal designs found by this theorem.

Table 2

Range of $\frac{\sigma^2}{n\delta^2}$ $(\equiv\lambda)$	$\frac{w_2}{w_1} \geq$ this value	Optimal designs	Values of p
.0000 ∿.01514	$\lambda^{-1}[.01749(.1108-\lambda)^{-1}-.1583]$	$p\mu_2+(1-p)\zeta_1$	$.09972(.1108-\lambda)^{-1}-.04511$
.01514 ∿.03528	1.655	μ_2	———
.03528 ∿.1090	$\lambda^{-1}[.2564(.5544-\lambda)^{-1}-.4356]$	$p\mu_3+(1-p)\zeta_2$	$.5229(.5544-\lambda)^{-1}-.1739$
.1090 ∿.3224	1.286	μ_3	———
.3224 ∿2.072	$\lambda^{-1}[2.975(4.221-\lambda)^{-1}-.3486]$	$p\zeta_7+(1-p)\mu_3$	$4.784(4.221-\lambda)^{-1}-1.227$
2.072 ∿ + ∞	.5000	ζ_7	———

6. Asymptotics

Let ξ_∞ denote the measure to which the ξ minimizing (1.3) tends, as $N \to \infty$. Denote the uniform probability measure on $\{t: x \leq |t| \leq \frac{1}{2}\}$ by μ_x. First it is clear that as $N \to \infty$ and $j/2N \to x$, we have $S_N \to \int_{1/2}^{1/2} t^2 dt = \frac{1}{12}$, $\mu_j \to \mu_x$, $N^{-3}\underline{p}_N(j) \to 2(2x-1)(2x^2+4x-1)$, $\frac{j^0}{2N} \to (\sqrt{6}-2)/2$, $V_j \to \frac{1}{12}(1+2x+4x^2)$ (for $j \leq j^0$), $B_j \to x^2(1-2x)^{-1}$ (for $j \leq j^0$), and $T_j - R_j \to 0$.

Let ζ be the design which puts masses equally on points $\frac{1}{2}$ and $-\frac{1}{2}$. Then the following is an asymptotic version of Theorem 4.2.

Theorem 6.1. If $\frac{\sigma^2}{n\delta^2} \geq 9 \cdot \frac{w_2}{w_1} - \frac{3}{2}$ and $\frac{w_2}{w_1} \leq \frac{1}{2}$, then $\xi_\infty = p\zeta + (1-p)\mu_0$, where p is a nonnegative number solving

$$p(2p+1)^2 = \min\{12\cdot(\frac{w_2}{w_1})^2 \cdot \frac{\sigma^2}{n\delta^2}, 9\}.$$

Similarly, we obtain the following asymptotic version of Theorem 4.3.

Theorem 6.2. Let $\lambda = \frac{\sigma^2}{n\delta^2}$. There are three cases:

(i) If $\lambda \geq \frac{2\sqrt{6}}{3}$ and $\frac{w_2}{w_1} \geq \frac{1}{2}$, then $\xi_\infty = \zeta$.

(ii) If $\frac{2\sqrt{6}}{9} \leq \lambda \leq \frac{2\sqrt{6}}{3}$ and $\frac{w_2}{w_1} \geq \frac{1}{3} \cdot [2+(\sqrt{6}-2)p]\cdot\lambda^{-1}$, where

$p = \frac{2(\sqrt{6}+2)}{2(\sqrt{6}+2)-3\lambda} - \frac{\sqrt{6}}{2}$, then $\xi_\infty = p\zeta + (1-p)\mu_{(\sqrt{6}-2)/2}$.

(iii) If $\lambda \leq \frac{2\sqrt{6}}{9}$ and $\frac{w_2}{w_1} \geq \frac{1+2x+4x^2}{6x}$, where x solves the equation

$$\lambda = \frac{24x^3(1+4x^3)}{(1-2x)^3(1+4x)(1+2x+4x^2)},$$

then $\xi_\infty = \mu_x$.

Acknowledgments. I am indebted to my thesis advisor Professor Jack Kiefer for the guidance and many helpful discussions on this work. I also thank Professor Jerome Sacks whose valuable suggestions lead to the present shorter proof of Lemma 3.2.

Appendix

Proof of Lemma 2.1. Consider the case that N is even only. (The case that N is odd can be established similarly).

(i) Suppose that $j \geq i \geq \frac{N}{2}+1$. Then it is clear that $R(h,\xi;\, w_1,w_2) \geq R(h,\zeta_N;\, w_1,w_2)$.

(ii) Suppose that $j \geq i > 1$ and $i \leq \frac{N}{2}$. Choose $\varepsilon > 0$ small enough so that $\xi' = (1-p)\mu_j+(p-\varepsilon)\mu_i+ \varepsilon\cdot(\frac{N}{2})\cdot(N-i+1)^{-1}\zeta_N+ \varepsilon\cdot(\frac{N}{2}-i+1)(N-i+1)^{-1}\zeta_{i-1}$ is non-decreasing. Now it is clear that $R(h,\xi;\, w_1,w_2) \geq R(h,\xi';\, w_1,w_2)$.

(iii) Suppose that $N > j \geq \frac{N}{2}+1$ and $i = 1$. Then $R(h,\xi;\, w_1,w_2) \geq R(h,p\mu_1+(1-p)\zeta_N;\, w_1,w_2)$.

(iv) Suppose that $\frac{N}{2} \geq j > 1$ and $i = 1$. This is similar to (ii). □

Proof of Lemma 3.1. First, we see that if $g \in G^a$ is such that $g(\frac{k'}{2N}) \leq g(\frac{k''}{2N})$ for some k' and k'' with $\xi(\frac{k'}{2N}) > \xi(\frac{k''}{2N})$, then there exists a $g^* \in G^a$ such that $g^*(\frac{k'}{2N}) \geq g^*(\frac{k''}{2N})$ and $\int tg(t)d\xi(t) \leq \int tg^*(t)\, d\xi(t)$. This, together with the fact that if g is an extreme point of G^a then $\#\{k \| g(\frac{k}{2N})| \neq \delta,\, k > 0\} \leq 1$, proves the desired result. □

Proof of Lemma 3.2. We need only to consider the class of ξ such that $\xi(\frac{\cdot}{2N})$ takes at most two distinct values. Let $\underset{\sim}{i} = (i_1,\ldots,i_N)$ be the permutation of $(1,2,\ldots,N)$ such that $\xi(i_j/2N) \leq \xi(i_k/2N)$ for any $j \leq k$ and if $\xi(i_j/2N) = \xi(i_k/2N)$ and $j < k$, then $i_j < i_k$. Suppose ξ is not nondecreasing. Let j be the largest number such that $i_j \neq j$. It is clear that $\xi(i_j/2N) > \xi(j/2N)$. Recall j^* from (3.1).

(i) Suppose $\xi(i_{j^*}/2N) = \xi(i_j/2N)$. Then $h_{\underset{\sim}{i}}(j/2N) = -\delta < 0$. Let k be the smallest positive integer such that $\xi(k/2N) = \xi(i_j/2N)$. Then $k < j$ and $k \cdot h_{\underset{\sim}{i}}(k/2N) \geq j \cdot h_{\underset{\sim}{i}}(j/2N)$. Now construct ξ' from ξ by removing a little bit of masses from the points $\{\pm k/2N\}$ to $\{\pm j/2N\}$ so that ξ' is still in $\Xi_{\underset{\sim}{i}}$. It is clear that $\int th_{\underset{\sim}{i}}(t)d\xi'(t) \leq \int t\, h_{\underset{\sim}{i}}(t)d\xi(t)$ and $\int t^2 d\xi'(t) > \int t^2 d\xi(t)$. Thus by (3.3) we see that ξ' improves ξ.

(ii) Suppose $\xi(i_{j^*}/2N) = \xi(j/2N)$. In this case, $i_j \neq N$ and $h_{\underset{\sim}{i}}(i_j/2N) = \delta > 0$. First suppose $i_j = 1$. Choose $\lambda > 0$ small enough so that $\xi' = \xi - \lambda\zeta_1 + \lambda\mu_1$ is still a probability measure in $\Xi_{\underset{\sim}{i}}$. Then, it is clear that $\int t^2 d\xi'(t) > \int t^2 d\xi(t)$ and $\int h_{\underset{\sim}{i}}(t)td\xi'(t) \leq \int h_{\underset{\sim}{i}}(t)td\xi(t)$. Thus ξ' improves ξ. Next, suppose $i_j > 1$. Let s be a nonnegative number such that $s \leq 1$ and $\int |t| d\zeta_{i_j}(t) =$

$\int|t|d(s\zeta_1+(1-s)\zeta_N)(t)$. Choose λ small enough so that $\xi' = \xi-\lambda\zeta_{i_j} + \lambda\cdot(s\zeta_1+(1-s)\zeta_N)$ is a probability measure and $h_{\underset{\sim}{i}}$ is still least favorable for ξ'. It is then clear that ξ' improves ξ.

Proof of Lemma 3.3. Let $\tilde{j}$ be the j^* in (3.1) when $\underset{\sim}{i} = (1,2,\ldots,N)$. Some computation leads to $j^0 \leq \tilde{j}$. Consider the following two cases.

(i) $j \geq \tilde{j}$. Take $r_j = (N+1-j)/2(N-\underset{\sim}{h}(\frac{j}{2N})(j-1))$ and $s_j = 1-r_j$. By the fact that $\underset{\sim}{h}(\frac{j}{2N}) \leq 1$, it can be verified, after straightforward computation, that (3.6) and (3.7) hold.

(ii) $j < \tilde{j}$. Take $r_j = ((N-j)N- \frac{1}{2}(j-1)j)/(N+j-1)(N-j)$ and $s_j = 1-r_j$. Then $1 \geq r_j \geq 0$ and (3.6) holds. Now, some computations lead to

$$\int t^2 d\eta_j(t) - \int t^2 d(r_j\zeta_{j-1}+s_j\zeta_N)(t) = \underset{\sim}{p}_N(j)/24N^2(N-j).$$

Analyzing the cubic polynomial $\underset{\sim}{p}_N(x)$ carefully, we obtain (3.7) and (3.8). $\square$

Proof of Lemma 3.4. Since $\underset{\sim}{Q}$ is continuous, it suffices to show that $\lim_{x\downarrow V_{j-1}^{-1}} \underset{\sim}{Q}'(x) \geq \lim_{x\uparrow V_{j-1}^{-1}} \underset{\sim}{Q}'(x)$. This is equivalent to showing that

$(G_{j-1} - \frac{H_{j-1}}{V_{j-1}})F_{j-1}^2H_{j-1} \leq (G_j - \frac{H_j}{V_{j-1}})F_j^2H_j$. Since $F_j^2(G_j - \frac{H_j}{V_{j-1}})^2 = F_{j-1}^2(G_{j-1} - \frac{H_{j-1}}{V_{j-1}})^2$ (because of continuity of $\underset{\sim}{Q}(x)$ at $x = V_{j-1}^{-1}$), we need only to verify that $F_jH_j \geq F_{j-1}H_{j-1}$. This can be done by expressing F_jH_j explicitly in terms of N and j. It can also be shown that $F_\infty H_\infty \geq F_{j^0}H_{j^0}$. Thus the proof is complete. $\square$

References

BICKEL, P. J. and HERZBERG, A. M. (1979). Robustness of design against autocorrelation in time. I: Asymptotic theory, optimality for location and linear regression. Ann. Statist. 7, 77-95.

BICKEL, P. J., HERZBERG, A. M., and SCHILLING, M. F. (1981). Robustness of design against autocorrelation in time. II: Optimality, theoretical and numerical results for the first-order autoregressive process. J. Amer. Statist. Assoc. 76, 870-877.

BOX, G.E.P. and DRAPER, N. R. (1959). A basis for the selection of a response surface design. J. Amer. Statist. Assoc. 54, 622-654.

CHENG, C. S. and LI, K. C. (1982). A minimax approach to sample surveys. Submitted to Ann. Statist.

GODAMBE, V. P. (1982). Estimation in survey sampling: robustness and optimality. J. Amer. Statist. Assoc. 77, 393-403.

HUBER, P. J. (1975). Robustness and designs. A Survey of Statistical Design and Linear Models, North Holland, Amsterdam, 287-303.

KIEFER, J. (1973). Optimal designs for fitting biased multi-response surfaces. Multivariate Analysis III, Academic Press, New York, 245-268.

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). Ann. Statist. 2, 847-879.

LI, K. C. and NOTZ, W. (1981). Robust designs for nearly linear regression. J. Statist. Plann. Infer. 6, 135-151.

LI, K. C. (1981). Robust regression designs when the design space consists of finitely many points. Mimeograph Series #81-45, Department of Statistics, Purdue University.

LI, K. C. (1983). Minimaxity for some randomized designs - some general results. Ann. Statist. 11, 225-239.

MARCUS, M. B. and SACKS, J. (1976). Robust designs for regression problems. Statistical Decision Theory and Related Topics II, Academic Press, New York, 245-268.

PESOTCHINSKY, L. (1982). Optimal robust designs: linear regression in R^k. Ann. Statist. 10, 511-525.

WU, C. F. (1981). On the robustness and efficiency of some randomized designs. Ann. Statist. 9, 1168-1177.