ROBUST DESIGNS FOR NEARLY LINEAR REGRESSION

by

K. C. Li and W. Notz[1]
University of California, Berkeley
and
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #80-3

February 1980

[1]Research supported in part by the National Science Foundation, NSF Grants MCS75-22481 A02 and MSC-7901707.

ROBUST DESIGNS FOR NEARLY LINEAR REGRESSION

by K. C. Li and W. Notz[1]

University of California, Berkeley and Purdue University

Summary. In this paper we seek designs and estimators which are optimal in some sense for multivariate linear regression on cubes and simplexes when the true regression function is unknown. More precisely, we assume the unknown true regression function is the sum of a linear part plus some contamination orthogonal to the set of all linear functions in the L_2 norm with respect to Lebesgue measure. The contamination is assumed bounded in absolute value and it is shown that the usual designs for multivariate linear regression on cubes and simplices and the usual least squares estimators minimize the supremum over all possible contaminations of the expected mean square error. Additional results for extrapolation and interpolation, among other things, are discussed. For suitable loss functions optimal designs are found to have support on the extreme points of our design space.

Key Words: Multivariate linear regression, extrapolation, optimum designs, least squares estimates, L_2 norm, cubes, simplices, extreme points.

AMS Subject Classifications: Primary - 62J05, 62K05, Secondary - 62G35.

[1]Research sponsored in part by the National Science Foundation, NSF Grants MCS75-22481 A02 and MSC-7901707.

1. Introduction.

Consider the regression design problem given by

$$y(x_i) = f(x_i) + e_i, \quad i = 1,2,\ldots,n$$

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. 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 design problem is to choose the x_i in an optimal manner for this inference. Many papers have been addressed to this problem. Box and Draper (1959) have discussed some of the dangers 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, e.g., in estimation there may result a large bias term. A careful description of some problems in this context is given by Kiefer (1973) in the case where the class of possible functions f, F, is a finite dimensional space containing F_0.

In a related direction Huber (1975) formulated a problem where $X = [-\frac{1}{2}, +\frac{1}{2}]$, F_0 = {linear functions on X}, and $F = \{f(x) = a+bx+g(x); \inf_{\alpha,\beta} \int_{-\frac{1}{2}}^{+\frac{1}{2}} (g(x)-\alpha-\beta x)^2 dx = \int_{-\frac{1}{2}}^{+\frac{1}{2}} g^2(x)dx \leq c\}$. c > 0 is a given constant. Notice that if $f \in F$ then a+bx is the best linear approximation to f in the L_2 norm with respect to Lebesgue measure on $[-\frac{1}{2}, +\frac{1}{2}]$. Huber confines himself to the use of the standard least squares estimates based on the model F_0 and finds the design which minimizes the maximum risk

$$\sup_{f \in F} E\int_{-\frac{1}{2}}^{+\frac{1}{2}} (\hat{a} + \hat{b}x - f(x))^2 dx.$$

Unfortunately this formulation leads to the restriction that the designs must be absolutely continuous with respect to Lebesgue measure, otherwise the maximum risk above is infinite. This means no implementable design can have finite maximum risk.

In a similar spirit is some work by Marcus and Sacks (1976). They take $X = [-1,+1]$, F_0 = {linear functions on X}, and $F = \{f(x) = a+bx+g(x):$ $|g(x)| \leq \phi(x)\}$. $\phi(x)$ is a given function on X with $\phi(0) = 0$. For $f \in F$ the contamination $g(x)$ may be thought of as the remainder term in a first order Taylor expansion of f. Marcus and Sacks restrict the estimators of a and b to be linear but not necessarily the standard least squares estimates based on the model F_0, and restrict designs to have finite support. They look for estimates and designs to minimize the mean square error

$$\sup_{f \in F} E(\hat{a}-a)^2 + \theta^2(\hat{b}-b)^2)$$

where $\hat{a}$ and $\hat{b}$ denote the estimates of a and b, and θ is a specified constant.

They are able to solve this problem for a number of, but not all, choices of ϕ. If $\phi(x) \geq mx$ then the unique optimal design is on the points $\{-1,0,+1\}$. If ϕ is convex there is a wide range of cases for which a design can be found on two points $\{-z, +z\}$ where z depends on ϕ and θ.

It should be noted that the condition $\phi(0) = 0$ in this formulation forces the contamination $g(x)$ to be zero at $x = 0$. This fact gives special value to the point 0 and is the reason that 0 is in the support of the unique optimal design in the case $\phi(x) \geq mx$.

In this paper some of the clever ideas of Marcus and Sacks and of Huber are modified and combined to get results in some multivariate settings. More specifically we take $X \subset R^k$ to be our design space, F_0 = {linear functions on $X\}$, $F = \{f(\underset{\sim}{x}) = \beta_0 + \underset{\sim}{\beta}'\underset{\sim}{x} + g(\underset{\sim}{x})\}$ where $\beta_0 \in R$, $\underset{\sim}{\beta} \in R^k$, $\underset{\sim}{x} \in X$, $g: X \to R$ is a measurable function with respect to lebesgue measure on X, $|g| \leq c$, $c > 0$ is some constant, and $\inf \int_X (g(\underset{\sim}{x})-b_0-\underset{\sim}{b}'\underset{\sim}{x})^2 d\underset{\sim}{x} = \int_X g^2(\underset{\sim}{x})d\underset{\sim}{x}$. In this last equation the inf is over all $b_0 \in R$ and $\underset{\sim}{b} \in R^k$. Notice for $f \in F$, $\beta_0 + \underset{\sim}{\beta}'\underset{\sim}{x}$ is the best linear approximation to f in the L_2 norm with respect to lebesgue measure on X. If estimates $\hat{\beta}_0$ and $\hat{\underset{\sim}{\beta}}$ of β_0 and $\underset{\sim}{\beta}$ are restricted to be linear (but not necessarily the standard least squares estimates) and designs are restricted to have finite support, we show that the designs which minimize the mean square error

$$\sup_{f \in F} E((\hat{\beta}_0-\beta_0)^2 + \sum_{i=1}^{k} \theta_i^2(\hat{\beta}_i-\beta_i)^2)$$

have support on the extreme points of X. Using this result we then show that the usual least squares estimates and optimal designs for multivariate linear regression on the cube or simplex minimize the mean square error.

In addition some results on extrapolition and interpolation, among other things, are noted in Section 5. For suitable loss functions, optimal designs in these settings are found to have support on the extreme points of the design space.

2. General results.

The following notation will be used in this paper. Lines underneath variables will denote column vectors and primes on vectors or matrices will denote transposes. The size of a given vector or matrix will be made clear from the context. We shall use the symbol R to denote the real line and R^k to denote k-dimensional Euclidean space.

Consider the multivariate regression problem

$$(2.1) \qquad Y(\underset{\sim}{x}_m) = \beta_0 + \underset{\sim}{\beta}'\underset{\sim}{x}_m + g(\underset{\sim}{x}_m) + e_m$$

where $m = 1,2,\ldots,n$ (n is fixed), the e_m are uncorrelated random variables with mean 0 and finite variance $\sigma^2 > 0$, $\beta_0 \in R$, $\underline{\beta} = (\beta_1,\ldots,\beta_k)' \in R^k$, $X \subset R^k$, $\underline{x}_m \in X$ for $m = 1,2,\ldots,n$, and for some fixed constant $c > 0$,

$$g \in G = \{g\colon X \to R;\ |g(\underline{x})| \le c,\ \int g(\underline{x})d\underline{x} = \int x_i g(\underline{x})d\underline{x} = 0 \text{ for } i = 1,\ldots,k,$$
$$\underline{x} = (x_1,\ldots,x_k)' \in X\}.$$

In the definition of G $d\underline{x}$ is lebesgue measure on X and all integrals are over X. In fact, all integrals that appear in this section will be assumed to be over X unless otherwise noted.

The conditions $\int g(\underline{x})d\underline{x} = \int x_i g(\underline{x})d\underline{x} = 0$ are equivalent to requiring

$$\int g^2(\underline{x})d\underline{x} = \inf_{b_0 \in R,\ \underline{b} \in R^k} \int (g(\underline{x}) - b_0 - \underline{b}'\underline{x})^2 d\underline{x}$$

which says the best linear approximation to g in the L_2 norm with respect to lebesgue measure $d\underline{x}$ on X is the function 0. This condition insures the uniqueness of the β_i in our model (2.1).

A discrete probability measure ξ on X will be called a p-exact design for p observations if $\xi(\underline{x}) = j(\underline{x})/p$, where $p > 0$ and $j(\underline{x})$ are integers, and $\underline{x} \in X$. We shall denote by Ξ_p the class of all such designs.

We also define

$$D_p = \{\text{probability measures } \xi \text{ on } X;\ \text{card}(\text{supp}\,\xi) \le p\}$$

$$D = \bigcup_{p=1}^{\infty} D_p.$$

For $\xi \in \Xi_n$, let the linear estimators of the β_i be defined by

$$\hat{\beta}_i = \int Y(\underline{x}) b_i(\underline{x}) d\xi(\underline{x}), \qquad i = 0,\ldots,k \tag{2.2}$$

where the b_i are real valued functions on X. In the discrete case where there may be more than one observation Y at a point $\underline{x}$, we interpret $Y(\underline{x})$ to be the average of the observations at $\underline{x}$.

We shall consider the weighted expected mean square error due to the design ξ and the estimators $\hat{\beta}_i$, namely

(2.3) $$\sum_{i=0}^{k} \theta_i^2 E(\hat{\beta}_i-\beta_i)^2, \quad \theta_i \geq 0, \quad i = 0,\ldots,k.$$

This mean square error can be rewritten as the sum of a "variance" term

$$\sum_{i=0}^{k} \theta_i^2 E(\hat{\beta}_i-E\hat{\beta}_i)^2$$

and a "bias" term

$$\sum_{i=0}^{k} \theta_i^2 (\beta_i-E\hat{\beta}_i)^2.$$

Using (2.2) we can write the variance term as

(2.4) $$\sum_{i=0}^{k} \theta_i^2 E(\hat{\beta}_i-E\hat{\beta}_i)^2 = (\sigma^2/n) \sum_{i=0}^{k} \theta_i^2 \int b_i^2(\underset{\sim}{x})d\xi(\underset{\sim}{x})$$

and the bias term is determined by the equations

(2.5) $$E\hat{\beta}_j - \beta_j = \sum_{i=0, i\neq j}^{k} \beta_i \int x_i b_j(\underset{\sim}{x})d\xi(\underset{\sim}{x}) + \beta_j[\int x_j b_j(\underset{\sim}{x})d\xi(\underset{\sim}{x})-1] + \int b_j(\underset{\sim}{x})g(\underset{\sim}{x})d\xi(\underset{\sim}{x})$$

for $j = 0,\ldots,k$. Let $x_0 \equiv 1$.

Define $B_j(\underset{\sim}{x}) = \theta_j b_j(\underset{\sim}{x})$. Then if the β_i are unbounded, in order for the weighted mean square error to be bounded we must have $\int x_i b_j(\underset{\sim}{x})d\xi(\underset{\sim}{x}) = \delta_{ij}$ or

(2.6) $$\int x_i B_j(\underset{\sim}{x})d\xi(\underset{\sim}{x}) = \theta_j \delta_{ij}, \; 0 \leq i, \; j \leq k$$

where δ_{ij} is the Kronecker delta. This is equivalent to saying that the linear estimators are unbiased if $g = 0$.

Let $\rho = \sigma^2/n$, $\underline{B} = (B_0,\dots,B_k)'$, and define

$$(2.7) \qquad L(\underline{B},\xi,g) = \sum_{i=0}^{k} \left(\int B_i(\underline{x})g(\underline{x})d\xi(\underline{x})\right)^2 + \rho\int\left(\sum_{i=0}^{k} B_i^2(\underline{x})\right)d\xi(\underline{x}).$$

Notice that $L(\underline{B},\xi,g)$ is equal to (2.3) with condition (2.6) imposed. Condition (2.6) and $L(\underline{B},\xi,g)$ are well defined for $\xi \in D$ and from now on we shall not restrict ξ to be an exact design for a particular p, but rather allow ξ to be a design in D.

For any $\xi \in D$ define $R(\xi) = \{\underline{B}$ satisfying (2.6) w.r.t. $\xi\}$. Our objective is to find $\xi \in D$ and $\underline{B} \in R(\xi)$ which minimize $\sup_{g\in G} L(\underline{B},\xi,g)$. A key result is the following.

THEOREM 2.1. Suppose $\xi \in D$. For all $\underline{x}^* \in X$, if $\xi(\underline{x}^*) \neq 0$ and there exist $\underline{x}_1,\underline{x}_2,\dots,\underline{x}_m \in X$ such that $\underline{x}^* = \sum_{\ell=1}^{m} \alpha_\ell \underline{x}_\ell$, where $0 < \alpha_\ell < 1$ for all ℓ and $\sum_{\ell=1}^{m} \alpha_\ell = 1$, then there exists $\xi^* \in D$ such that

(i) $\quad \text{supp}\,\xi^* = \text{supp}\,\xi \cup \{\underline{x}_1,\dots,\underline{x}_m\} - \{\underline{x}^*\}$

(ii) $\quad \min_{\underline{B}\in R(\xi^*)} \max_{g\in G} L(\underline{B},\xi^*,g) \leq \min_{\underline{B}\in R(\xi)} \max_{g\in G} L(\underline{B},\xi,g).$

Proof. Let

$$G^* = \{g\in G;\ g(\underline{x}^*) = \sum_{\ell=1}^{m} \alpha_\ell g(\underline{x}_\ell)\}$$

Since for any $g \in G$ we have $|\sum_{\ell=1}^{m} \alpha\, g(\underline{x}\ )| \leq \sum_{\ell=1}^{m} \alpha\ |g(\underline{x}\ )| \leq \sum_{\ell=1}^{m} \alpha_\ell\ c = c$, it is easy to see that for all $g \in G$ there exists $g^* \in G^*$ such that $g(\underline{x}) = g^*(\underline{x})$ if $\underline{x} \neq \underline{x}^*$.

Hence for all ξ^* such that $\xi^*(\underline{x}^*) = 0$ we have for all $\underline{B} \in R(\xi^*)$,

$$\max_{g \in G} L(\underline{B},\xi^*,g) = \max_{g \in G^*} L(\underline{B},\xi^*,g). \tag{2.8}$$

Also for all ξ such that $\xi(\underline{x}^*) \neq 0$ we have $\max_{g \in G} L(\underline{B},\xi,g) \geq \max_{g \in G^*} L(\underline{B},\xi,g)$, since $G^* \subset G$, and hence

$$\min_{\underline{B} \in R(\xi)} \max_{g \in G} L(\underline{B},\xi,g) \geq \min_{\underline{B} \in R(\xi)} \max_{g \in G^*} L(\underline{B},\xi,g). \tag{2.9}$$

Thus is suffices to show that for any $\xi \in D$ with $\xi(\underline{x}^*) \neq 0$ there exists $\xi^* \in D$ satisfying (i) of the theorem such that

$$\min_{\underline{B} \in R(\xi^*)} \max_{g \in G^*} L(\underline{B},\xi^*,g) \leq \min_{\underline{B} \in R(\xi)} \max_{g \in G^*} L(\underline{B},\xi,g). \tag{2.10}$$

Indeed the following result is stronger and its proof yields (2.10) and hence the theorem.

LEMMA 2.1. Suppose $\xi \in D$ is such that $\xi(\underline{x}^*) \neq 0$ and there exist $\underline{x}_1,\ldots,\underline{x}_m \in X$ such that $\underline{x}^* = \sum_{\ell=1}^{m} \alpha_\ell \underline{x}_\ell$, where $0 < \alpha_\ell < 1$ for all ℓ and $\sum_{\ell=1}^{m} \alpha_\ell = 1$. Then there exists $\xi^* \in D$ satisfying (i) of Theorem 2.1 and for any $\underline{B} \in R(\xi)$ there exists $\underline{B}^* \in R(\xi^*)$ such that for all $g \in G^*$

$$L(\underline{B},\xi,g) = L(\underline{B}^*, \xi^*,g) + d$$

where d is a nonnegative constant that does not depend on g.

Proof. Define

$$\xi^*(\underline{x}) = \xi(\underline{x}) \quad \text{if} \quad \underline{x} \neq \underline{x}^*, \underline{x}_1,\ldots,\underline{x}_m$$

$$\xi^*(\underline{x}^*) = 0$$

$$\xi^*(\underline{x}_\ell) = \alpha_\ell \xi(\underline{x}^*) + \xi(\underline{x}_\ell), \quad \ell = 1,\ldots,m.$$

For $\underline{B} \in R(\xi)$ define $\underline{B}^* = (B_1^*,\ldots,B_k^*)'$ satisfying

$$\underline{B}^*(\underline{x}) = \underline{B}(\underline{x}) \quad \text{for} \quad \underline{x} \neq \underline{x}_1,\ldots,\underline{x}_m$$

$$B_j^*(\underline{x}_\ell)\xi^*(\underline{x}_\ell) = B_j(\underline{x}_\ell)\xi(\underline{x}_\ell) + \alpha_\ell B_j(\underline{x}^*)\xi(\underline{x}^*)$$

$$\text{for } \ell = 1,\ldots,m,\ j = 0,\ldots,k.$$

We must show

(2.11) $\underline{B}^* \in R(\xi^*)$

(2.12) $\int g(\underline{x})B_j^*(\underline{x})d\xi^*(\underline{x}) = \int g(\underline{x})B_j(\underline{x})d\xi(\underline{x})$ for $j = 0,\ldots,k$ and $g \in G^*$

(2.13) $\int B_j^2(\underline{x})d\xi(\underline{x}) \geq \int B_j^{*2}(\underline{x})d\xi^*(\underline{x})$ for $j = 0,\ldots,k$.

(2.11) and (2.12) follow from the following. For $g \in G^*$

(2.14) $$\sum_{\ell=1}^m g(\underline{x}_\ell)B_j^*(\underline{x}_\ell)\xi^*(\underline{x}_\ell) = \sum_{\ell=1}^m g(\underline{x}_\ell)B_j(\underline{x}_\ell)\xi(\underline{x}_\ell) + \sum_{\ell=1}^m \alpha_\ell g(\underline{x}_\ell)B_j(\underline{x}^*)\xi(\underline{x}^*)$$

$$= \sum_{\ell=1}^m g(\underline{x}_\ell)B_j(\underline{x}_\ell)\xi(\underline{x}_\ell) + g(\underline{x}^*)B_j(\underline{x}^*)\xi(\underline{x}^*)$$

(2.12) follows from (2.14). Use $g(\underline{x}_\ell)$ = i-th coordinate of $\underline{x}_\ell$ in (2.14) to get (2.11).

For (2.13) it suffices to show

(2.15) $B_j^{*2}(\underline{x}_\ell)\xi^*(\underline{x}_\ell) \leq B_j^2(\underline{x}_\ell)\xi(\underline{x}_\ell) + \alpha_\ell B_j^2(\underline{x}^*)\xi(\underline{x}^*)$ for $j = 0,\ldots,k$, $\ell = 1,\ldots,m$.

Now from the definitions of ξ^* and $\underline{B}^*$ we have

(2.16) $\xi^*(\underline{x}_\ell) = \xi(\underline{x}_\ell) + \alpha_\ell\xi(\underline{x}^*)$

(2.17) $B_j^*(\underline{x}_\ell)\xi^*(\underline{x}_\ell) = B_j(\underline{x}_\ell)\xi(\underline{x}_\ell) + \alpha_\ell B_j(\underline{x}^*)\xi(\underline{x}^*)$.

Divide (2.15), (2.16), and (2.17) by $\xi^*(\underline{x}_\ell)$ separately. Then (2.15) is simply saying $(EZ)^2 \leq EZ^2$ where Z is a random variable taking value $B_j(\underline{x}_\ell)$

with probability $\xi(\underline{x}_\ell)/\xi^*(\underline{x}_\ell)$ and value $B_j(\underline{x}^*)$ with probability $\alpha_\ell \xi(\underline{x}^*)/\xi^*(\underline{x}_\ell)$. Hence (2.15) holds and the proof of the lemma (and thus Theorem 2.1) is complete.

Theorem 2.1 implies that for finding designs $\xi \in D$ to minimize $\min_{\underline{B} \in R(\xi)} \max_{g \in G} L(\underline{B},\xi,g)$ it suffices to only consider designs ξ whose support contains only extreme points of X. We apply this to some specific situations in the next few sections.

3. Results for simplices.

In this section we apply the results of Section 2 to the case where

$$X = S_k = \{(x_1,\ldots,x_k)' \in R^k;\ \sum_{i=1}^{k} x_i = 1,\ x_i \geq 0 \text{ for all } i\}.$$

S_k is the k-1 dimensional simplex. We also assume that in the notation of Section 2, $\beta_0 = 0$, $b_0 = 0$, $B_0 = 0$, $\theta_0 = 0$, and $\theta_1 = \theta_2 = \ldots = \theta_k = 1$. Equation (2.6) becomes

$$(3.1) \qquad \int x_i B_j(\underline{x}) d\xi(\underline{x}) = \delta_{ij}, \quad 1 \leq i,j \leq k$$

and if we define for any $\xi \in D$

$$R(\xi) = \{\underline{B} = (B_1,\ldots,B_k)' \text{ satisfying (3.1) w.r.t } \xi\}$$

then we seek $\xi \in D$ and $\underline{B} \in R(\xi)$ which minimize $\sup_{g \in G} L(\underline{B},\xi,g)$ where now

$$(3.2) \qquad L(\underline{B},\xi,g) = \sum_{i=1}^{k} \left(\int B_i(\underline{x}) g(\underline{x}) d\xi(\underline{x})\right)^2 + \rho \int \left(\sum_{i=1}^{k} B_i^2(\underline{x})\right) d\xi(\underline{x}).$$

The following lemmas will be useful.

<u>LEMMA 3.1</u>. Suppose $\xi \in D$ and $\underline{B} \in R(\xi)$. If $\sum_{i=1}^{k} B_i^2$ is not constant on supp ξ then there exists $\xi^* \in D$ with supp ξ^* = suppξ and there exists $\underline{B}^* \in R(\xi^*)$

with $\sum_{i=1}^{k} B_i^{*2}$ constant on supp ξ^* and $\inf_{g \in G} [L(\underline{B},\xi,g) - L(\underline{B}^*,\xi^*,g)] > 0$.

Proof. Let $\xi^*(\underline{x}) = \alpha(\sum_{i=1}^{k} B_i^2(\underline{x}))^{1/2}\xi(\underline{x})$, where α is the constant making ξ^* a probability measure. Let $B_i^*(\underline{x}) = B_i(\underline{x})/\alpha(\sum_{j=1}^{k} B_j^2(\underline{x}))^{1/2}$ for $i = 1,\ldots,k$ unless the denominator $\alpha(\sum_{j=1}^{k} B_j^2(\underline{x}))^{1/2}$ is 0 in which case define $B_i^*(\underline{x}) = 0$. Notice $B_i^*(\underline{x})\xi^*(\underline{x}) = B_i(\underline{x})\xi(\underline{x})$ for all $\underline{x}$ and i so that $\underline{B}^* \in R(\xi^*)$. Also notice $\sum_{i=1}^{k} B_i^{*2}(\underline{x}) = 1/\alpha^2$ is constant on supp ξ^*. Since

$$\begin{aligned}\int(\sum_{i=1}^{k} B_i^{*2}(\underline{x}))d\xi^*(\underline{x}) &= (1/\alpha^2)\int d\xi^*(\underline{x}) \\ &= [(1/\alpha)\int d\xi^*(\underline{x})]^2 \\ &= [\int(\sum_{i=1}^{k} B_i^2(\underline{x}))^{1/2} d\xi(\underline{x})]^2 \\ &\leq \int(\sum_{i=1}^{k} B_i^2(\underline{x}))d\xi(\underline{x})\end{aligned}$$

with strict inequality unless $\sum_{i=1}^{k} B_i^2(\underline{x})$ is constant on supp ξ, the lemma follows.

LEMMA 3.2. Suppose $1 \leq r < s \leq k$, where r and s are integers. If $r < s-1$ let

$$\pi_{rs}(\underline{x}) = (x_1,\ldots,x_{r-1},x_s,s_{r+1},\ldots,x_{s-1},x_r,x_{s+1},\ldots,x_k)$$

and if $r = s-1$

$$\pi_{rs}(\underline{x}) = (x_1,\ldots,x_{r-1},x_s,x_r,x_{s+1},\ldots,x_k).$$

In other words, π_{rs} interchanges the r-th and s-th coordinates of a point in R^k.

Define $\pi_{rs} \circ g = g \circ \pi_{rs}$. For any $\xi \in D$ and $\underline{B} \in R(\xi)$ there exist $\psi \in D$ and $\underline{A} \in R(\psi)$ such that

(3.3) $A_i(\underline{x}) = A_i(\pi_{rs}(\underline{x}))$ if $i \neq r,\ i \neq s$

$A_r(\underline{x}) = A_s(\pi_{rs}(\underline{x}))$

$A_s(\underline{x}) = A_r(\pi_{rs}(\underline{x}))$

$\psi(\underline{x}) = \psi(\pi_{rs}(\underline{x}))$

$\text{supp } \psi = T = \{\underline{x} \in \text{supp } \xi \text{ or } \pi_{rs}(\underline{x}) \in \text{supp } \xi\}$

and

$$\sup_{g \in G} L(\underline{A},\psi,g) \leq \sup_{g \in G} L(\underline{B},\xi,g).$$

Proof. By Lemma 3.1 if $\sum_{i=1}^{k} B_i^2$ is not constant on supp ξ we may replace $\underline{B}$ and ξ by $\xi^* \in D$, $\underline{B}^* \in R(\xi^*)$ satisfying $L(\underline{B}^*, \xi^*, g) < L(\underline{B},\xi,g)$ and such that $\sum_{i=1}^{k} B_i^{*2}$ is constant on supp ξ^* = supp ξ. So we shall assume $\sum_{i=1}^{k} B_i^2$ is constant on supp ξ. Then

$$\int(\sum_{i=1}^{k} B_i^2(\underline{x}))d\xi(\underline{x}) = [\int(\sum_{i=1}^{k} B_i^2(\underline{x}))^{\frac{1}{2}}d\xi(\underline{x})]^2.$$

Let

$\xi^0(\underline{x}) = \xi(\pi_{rs}(\underline{x}))$

$B_i^0(\underline{x}) = B_i(\pi_{rs}(\underline{x}))$ if $i \neq r,\ i \neq s$

$B_r^0(\underline{x}) = B_s(\pi_{rs}(\underline{x}))$

$B_s^0(\underline{x}) = B_r(\pi_{rs}(\underline{x}))$.

For each $g \in G$ we have $L(\underline{B},\xi,g) = L(\underline{B}^0,\xi^0,\pi_{rs}\ g)$ and so

(3.4)
$$\sup_{g \in G} L(\underline{B},\xi,g) = \sup_{g \in G} L(\underline{B}^0,\xi^0,\pi_{rs}\ g)$$
$$= \sup_{g \in G} L(\underline{B}^0,\xi^0,g)$$

since $g \in G$ if and only if $\pi_{rs} o g \in G$.

Let $u_i(\underline{x}) = B_i(\underline{x})\xi(\underline{x})$, $u_i^o(\underline{x}) = B_i^o(\underline{x})\xi^o(\underline{x})$ for $i = 1,\ldots,k$. The u_i and u_i^o are defined on T (defined in statement of the lemma). Let $\underline{u} = (u_1,\ldots,u_k)'$, $\underline{u}^o = (u_1^o,\ldots,u_k^o)'$. Using (3.2) we get

$$L(\underline{B},\xi,g) = L^o(\underline{u},g)$$

$$L(\underline{B}^o,\xi^o,g) = L^o(\underline{u}^o,g)$$

where

$$L^o(\underline{v},g) = \sum_{i=1}^{k} \Big(\sum_{\underline{x}\in T} g(\underline{x})v_i(\underline{x})\Big)^2 + \rho\Big(\sum_{\underline{x}\in T} \Big(\sum_{i=1}^{k} v_i^2(\underline{x})\Big)^{\frac{1}{2}}\Big)^2.$$

$L^o(\underline{v},g)$ is convex in $\underline{v}$ and clearly $\sup_{g\in G} L^o(\underline{v},g)$ is convex in $\underline{v}$ also. Let $w_i = (u_i+u_i^o)/2$ for $i = 1,\ldots,k$. Notice

$$w_i(\underline{x}) = w_i(\pi_{rs}(\underline{x})) \quad \text{if } i \neq r,\ i \neq s$$

$$w_r(\underline{x}) = w_s(\pi_{rs}(\underline{x}))$$

$$w_s(\underline{x}) = w_r(\pi_{rs}(\underline{x})).$$

We have $\sup_{g\in G} L^o(\underline{w},g) \leq \sup_{g\in G} L(\underline{B},\xi,g)$ by convexity and (3.4). Define

$$\psi(\underline{x}) = \alpha\Big(\sum_{i=1}^{k} w_i^2(\underline{x})\Big)^{\frac{1}{2}}$$

where α makes ψ a probability measure. Let $A_i(\underline{x}) = w_i(\underline{x})/\psi(\underline{x})$ if $\psi(\underline{x}) > 0$ and $A_i(\underline{x}) = 0$ if $\psi(\underline{x}) = 0$, for $i = 1,\ldots,k$. One can check $\underline{A} \in R(\psi)$ and $\underline{A}$, ψ are as stated in the lemma.

<u>LEMMA 3.3</u>. Suppose we are given design $\xi \in D$ and functions $(B_1,\ldots,B_k)' = \underline{B}$ such that $\underline{B} \in R(\xi)$. For any $g \in G$ there exists $g^* \in G$ with $|g^*(\underline{x})| = c$ on supp ξ and such that $L(\underline{B},\xi,g^*) \geq L(\underline{B},\xi,g)$.

Proof. $L(\underset{\sim}{B},\xi,g)$ is a convex quadratic function of $g(\underset{\sim}{x})$ for any $\underset{\sim}{x} \in \text{supp}\ \xi$. Thus it can be maximized by assigning $g(\underset{\sim}{x})$ its extreme values, namely $\pm c$. Let g^* be a function derived from g by redefining g at each point $\underset{\sim}{x}$ in supp ξ so as to maximize $L(\underset{\sim}{B},\xi,g)$ as a function of g and so that $|g^*| = c$ on supp ξ. Clearly the values of g^* off supp ξ can be chosen so that $g^* \in G$ (just take $g^*(\underset{\sim}{x}) = g(\underset{\sim}{x})$ for $\underset{\sim}{x} \in \text{supp}\ \xi$).

Armed with these lemmas and the results of Section 2 we can proceed to determine which $\xi \in D$ and $\underset{\sim}{B} \in R(\xi)$ minimize $\max\limits_{g \in G} L(\underset{\sim}{B},\xi,g)$.

Recall that Section 2 showed that we could restrict ourselves to designs $\xi \in D$ whose support was only on the extreme points of S_k, i.e. the k points $(1,0,\ldots,0)'$, $(0,1,0,\ldots,0)'$, $(0,0,1,0,\ldots,0)',\ldots,$ $(0,\ldots,0,1)'$. Restricting ourselves to such designs, applying lemma 3.2 and the construction in its proof for all $1 \leq r,\ s \leq k$, we see that we can further restrict ourselves to designs $\xi \in D$ having support on the extreme points of S_k and satisfying $\xi(\underset{\sim}{x}) = \xi(\pi_{rs}(\underset{\sim}{x}))$ for all extreme points $\underset{\sim}{x}$ of S_k and all $1 \leq r,\ s \leq k$. The only design $\xi^* \in D$ satisfying these restrictions is the design ξ^* taking value $1/k$ at each of the k extreme points of S_k. Thus applicat·on of the results of Section 2 and Lemma 3.2 yield that there exists $\underset{\sim}{B}^* \in R(\xi^*)$ satisfying (3.3) for all $1 \leq r,\ s \leq k$ such that

$$\max_{g \in G} L(\underset{\sim}{B}^*,\xi^*,g) \leq \min_{\underset{\sim}{B} \in R(\xi)} \max_{g \in G} L(\underset{\sim}{B},\xi,g)$$

for all $\xi \in D$.

To find $\underset{\sim}{B}^*$ notice that since $\underset{\sim}{B}^*$ must satisfy (3.3) and also the conditions for what it means to be an element of $R(\xi^*)$ (see equation (3.1)), we must have

(3.6) $$B_i(\underline{x}(i)) = k$$
$$B_i(\underline{x}(j)) = \lambda \qquad \text{for } i \neq j$$

for all $1 \leq i, j \leq k$, where $\underline{x}(i)$ is the extreme point of S_k whose i-th coordinate is 1 and whose other coordinates are all 0. Here λ is some constant real number.

For any $\underline{B} \in R(\xi^*)$ satisfying (3.6) for some λ we have by Lemma 3.3 that

(3.7) $$\max_{g \in G} L(\underline{B},\xi^*,g) = \max_{g \in G}\left[\sum_{i=1}^{k} \left(\int B_i(\underline{x})g(\underline{x})d\xi^*(\underline{x})\right)^2 + \rho\int\left(\sum_{i=1}^{k} B_i^2(\underline{x})\right)d\xi^*(\underline{x})\right]$$
$$= k[(k-1)|\lambda|c/k + c]^2 + \rho(k^2+(k-1)\lambda^2)/k.$$

The $\underline{B} \in R(\xi^*)$ satisfying (3.6) for some λ which minimizes (3.6) is easily seen to be the one with $\lambda = 0$, call it $\underline{B}^*$. We therefore have proved the following theorem.

<u>THEOREM 3.1</u>. Let $\xi^* \in D$ be the design putting mass 1/k on each of the k extreme points of S_k, i.e. the points $\underline{x}(1) = (1,0,\ldots,0)'$, $\underline{x}(2) = (0,1,0,\ldots,0)'$, $\ldots$, $\underline{x}_k = (0,\ldots,0,1)'$. Let $\underline{B}^* = (B_1^*,\ldots,B_k^*)' \in R(\xi^*)$ be such that $B_i^*(\underline{x}(j)) = k\delta_{ij}$, where δ_{ij} is the Kronecker delta. Then we have

$$\max_{g \in G} L(\underline{B}^*,\xi^*,g) = \min_{\xi \in D} \min_{\underline{B} \in R(\xi)} \max_{g \in G} L(\underline{B},\xi,g).$$

By Theorem 3.1 it follows that the design ξ^* above and the function $\underline{B}^* = (B_1^*,\ldots,B_k^*)'$ where $B_i^*(\underline{x}) = kx_i$ (x_i is the i-th coordinate of $\underline{x}$) minimize $\max_{g \in G} L(\underline{B},\xi,g)$. Notice ξ^* and the B_i^* are the "usual" optimal design and least squares estimates for multivariate linear regression on S_k.

4. <u>Results for cubes</u>.

In this section we apply the results of Section 2 to the case where $X = I^k$ = k-fold Cartesian product of the closed interval $I = [-1,+1]$. I^k is the k-dimensional cube centered at the origin with sides of length 2. All other notation is as in Section 2. We seek $\xi \in D$ and $\underline{B} \in R(\xi)$ which will minimize $\max_{g \in G} L(\underline{B},\xi,g)$. The following lemmas will prove useful.

LEMMA 4.1. Suppose $\xi \in D$ and $\underline{B} \in R(\xi)$. If $\sum_{i=0}^{k} B_i^2$ is not constant on supp ξ then there exists $\xi^* \in D$ with supp ξ_k^* = supp ξ and there exists $\underline{B}^* \in R(\xi^*)$ with $\sum_{i=0}^{k} B_i^{*2}$ constant on supp ξ^* and $\inf_{g \in G} [L(\underline{B},\xi,g) - L(\underline{B}^*,\xi^*,g)] > 0$.

Proof. The proof is similar to Lemma 3.1 and is therefore omitted.

LEMMA 4.2. Suppose q is an integer satisfying $1 \leq q \leq k$. Let

$$\lambda_q(\underline{x}) = (x_1,\ldots,x_{q-1},-x_q,x_{q+1},\ldots,x_k)'.$$

Define $\lambda_q \circ g = g \circ \lambda_q$. For any $\xi \in D$ and $\underline{B} \in R(\xi)$ there exists $\psi \in D$ and $\underline{A} \in R(\psi)$ such that

$$(4.1) \quad A_i(\underline{x}) = A_i(\lambda_q(\underline{x})) \quad \text{if} \quad i \neq q$$

$$A_q(\underline{x}) = -A_q(\lambda_q(\underline{x}))$$

$$\psi(\underline{x}) = \psi(\lambda_q(\underline{x}))$$

$$\text{supp } \psi = T = \{\underline{x} \in I^k;\ \underline{x} \in \text{supp } \xi \text{ or } \lambda_q(\underline{x}) \in \text{supp } \xi\}$$

and

$$\sup_{g \in G} L(\underline{A},\psi,g) \leq \sup_{g \in G} L(\underline{B},\xi,g).$$

Proof. The proof is similar to Lemma 3.2 with obvious modifications and is therefore omitted.

LEMMA 4.3. Suppose we are given design $\xi \in D$ and $\underline{B} \in R(\xi)$. For any $g \in G$ there exists $g^* \in G$ with $|g^*(\underline{x})| = c$ on supp ξ and such that $L(\underline{B},\xi,g^*) \geq L(\underline{B},\xi,g)$.

Proof. The proof is similar to Lemma 3.3 and is therefore omitted.

Equipped with these lemmas and the results of Section 2 we can proceed to find the $\xi \in D$ and $\underline{B} \in R(\xi)$ which minimize $\max_{g \in G} L(\underline{B},\xi,g)$.

Recall that the results of Section 2 allow us to restrict ourselves to designs $\xi \in D$ whose support is only on the extreme points of I^k, i.e. on the 2^k corners of the cube. Restricting ourselves to such design, applying Lemma 4.2 and the construction involved in its proof for all $1 \leq q \leq k$, we find that we can restrict ourselves to designs $\xi \in D$ having support on the 2^k corners of I^k and satisfying $\xi(\underline{x}) = \xi(\lambda_q(\underline{x}))$ for all $1 \leq q \leq k$ and all $\underline{x}$ which are extreme points of I^k. The only design satisfying these restrictions is easily seen to be the design $\xi^* \in D$ taking value $1/2^k$ at each corner of I^k. Thus application of the results of Section 2 and Lemma 4.2 yield that there exists $\underline{B}^* \in R(\xi^*)$ satisfying (4.1) for all $1 \leq q \leq k$ such that

$$\max_{g \in G} L(\underline{B}^*,\xi^*,g) \leq \min_{\underline{B} \in R(\xi)} \max_{g \in G} L(\underline{B},\xi,g)$$

for all $\xi \in D$.

To find $\underline{B}^*$ notice $\underline{B}^*$ must satisfy (4.1) and the conditions for what it means to be in $R(\xi^*)$ (see equation (2.6)). It is not difficult to verify that this means

(4.2) $B_i^*(\underline{x}) = \theta_i$ if the i-th coordinate of $\underline{x}$ is +1

$B_i^*(\underline{x}) = -\theta_i$ if the i-th coordinate of $\underline{x}$ is -1.

$B_0^*(\underline{x}) = \theta_0$.

Therefore we have the following result.

THEOREM 4.1. Let $\xi^* \in D$ be the design putting mass $1/2^k$ on each of the 2^k corners of the cube I^k. Let $\underline{B}^* = (B_0^*, B_1^*, \ldots, B_k^*)' \in R(\xi^*)$, where the B_i^* are as in equation (4.2). Then we have

$$\max_{g \in G} L(\underline{B}^*, \xi^*, g) = \min_{\xi \in D} \min_{\underline{B} \in R(\xi)} \max_{g \in G} L(\underline{B}, \xi, g).$$

As in section 3 ξ^* and $\underline{B}^*$ (or more precisely the estimates b_i^* arising from the B_i^*) in Theorem 4.1 are just the "usual" optimal design and least squares estimates for multivariate linear regression on I^k.

Let $\xi^* \in D$, $\underline{B}^* \in R(\xi^*)$ be as in Theorem 4.1. D, $\underline{B}^* \in R(\xi^*)$ be as in Theorem 4.1. Define

$$G(c) = \{g: I^k \to R;\ |g(\underline{x})| = c \text{ for all } \underline{x} \in I^k\}.$$

It is easy to verify, using Lemma 4.3, that

$$\max_{g \in G} L(\underline{B}, \xi, g) = \max_{g \in G(c)} L(\underline{B}, \xi, g)$$

for any $\xi \in D$, $\underline{B} \in R(\xi)$.

Now we restrict to the case $\theta_o = 1 \geq \theta_1 = \ldots = \theta_k = \theta > 0$. The minimax mean square error $\sup_{g \in G} L(\underline{B}^*, \xi^*, g)$ can be written as

$$\sup_{g \in G} L(\underline{B}^*, \xi^*, g) = \sup_{g \in G(c)} [(\int g(\underline{x}) d\xi^*(\underline{x}))^2 + \theta^2 \sum_{i=1}^{k} (\int x_i g(\underline{x}) d\xi^*(\underline{x}))^2]$$

$$+ \rho \int (1 + \theta^2 \sum_{i=1}^{k} x_i^2) d\xi^*(\underline{x}).$$

Let $\underline{x}_j = (x_{1j}, \ldots, x_{kj})' \in I^k$ for $j = 1, \ldots, m$ be any set of m distinct points in I^k. Let ξ_0 be the probability measure putting mass 1/m on each of these m points. Define

(4.3)
$$X(\xi_0) = \begin{bmatrix} 1 & \cdots & 1 \\ \\ \underset{-}{x}_1 & \cdots & \underset{-}{x}_m \end{bmatrix}$$

$$g(\xi_0) = (g(\underset{-}{x}_1),\ldots,g(\underset{-}{x}_m))'.$$

Notice $X(\xi_0)$ is a $(k+1) \times m$ matrix and $g(\xi_0)$ is a $m \times 1$ column vector. If we use J_m to denote the $m \times m$ matrix all of whose entries are 1, we get

$$\begin{aligned}
\sup_{g \in G} L(\underset{-}{B}^*,\xi_0,g) &= \sup_{g \in G(c)} [(\int g(\underset{-}{x})d\xi_0(\underset{-}{x}))^2 + \theta^2 \sum_{i=1}^{k} (\int x_i g(\underset{-}{x})d\xi_0(\underset{-}{x}))^2] \\
&\quad + \rho\int(1 + \theta^2 \sum_{i=1}^{k} x_i^2)d\xi_0(\underset{-}{x}) \\
&= \sup_{g \in G(c)} [(1-\theta^2)(\int g(\underset{-}{x})d\xi_0(\underset{-}{x}))^2 + \theta^2(\int g(\underset{-}{x})d\xi_0(\underset{-}{x}))^2 \\
&\quad + \theta^2 \sum_{i=1}^{k} (\int x_i g(\underset{-}{x})d\xi_0(\underset{-}{x}))^2] + \rho(1-\theta^2) \\
&\quad + \theta^2\rho \int(1+\sum_{i=1}^{k} x_i^2)d\xi_0(\underset{-}{x}) \\
&= \sup_{g \in G(c)} [(1-\theta^2)g'(\xi_0)J_m g(\xi_0) \\
&\quad + \theta^2 g'(\xi_0)X'(\xi_0)X(\xi_0)g(\xi_0)]/m^2 \\
&\quad + (1-\theta^2)\rho + (\theta^2\rho \operatorname{tr} X(\xi_0)X'(\xi_0))/m \\
&\geq \sup_{g \in G(c)} L(\underset{-}{B}^*,\xi^*,g) \\
&= \sup_{g \in G} L(\underset{-}{B}^*,\xi^*,g) \\
&= c^2 + \rho + k\theta^2\rho.
\end{aligned}$$

If $k+1$ is such that a $(k+1) \times (k+1)$ Hadamard matrix X exists (in standard form so that the first row and column are all $+1$) then any exact design ψ

on k+1 points whose support is such that $X(\psi) = X$, where $X(\psi)$ is as in (4.3), satisfies

$$\sup_{g\in G} L(\underline{B}^*,\psi, g) = \sup_{g\in G(c)} [(c-\theta^2)g'(\psi)J_{k+1}g(\psi) + \theta^2 g'(\psi)X'(\psi)X(\psi)g(\psi)]/(k+1)^2 + (1-\theta^2)\rho+(\theta^2\rho \text{ tr } X(\psi)X'(\psi))/(k+1).$$

Recalling that since $X(\psi)$ is a Hadamard matrix $X(\psi)X'(\psi) = X'(\psi)X(\psi) = (k+1)\,\mathrm{diag}(1,\ldots,1)$ = the diagonal matrix all of whose diagonal entries are k+1, we have that

$$\begin{aligned}\sup_{g\in G} L(\underline{B}^*,\psi,g) &= \sup_{g\in G(c)} [(1-\theta^2)g'(\psi)J_{k+1}g(\psi) + \theta^2(k+1)g'(\psi)g(\psi)]/(k+1)^2 \\ &\quad + (1-\theta^2)\rho + \theta^2\rho(k+1)^2/(k+1) \\ &= [(1-\theta^2)c^2(k+1)^2 + \theta^2(k+1)^2c^2]/(k+1)^2 \\ &\quad + (1-\theta^2)\rho + \theta^2\rho(k+1) \\ &= c^2 + \rho + k\theta^2\rho\end{aligned}$$

where we have used the fact that if $g \in G(c)$ then $|g(\underline{x})| = c$ for all $\underline{x} \in I^k$.

We see that ψ gives the same minimax value as ξ^*. Since $\underline{B}^* \in R(\psi)$ we have:

THEOREM 4.2. Suppose $\theta_0 = 1 \geq \theta_1 = \ldots = \theta_k = \theta > 0$. Suppose k+1 is such that a $(k+1) \times (k+1)$ Hadamard matrix exists. Let $\underline{B}^* = (B_0^*,B_1^*,\ldots,B_k^*)'$ where $B_0^*(\underline{x}) = 1$, $B_i^*(\underline{x}) = \theta x_i$ for $i = 1,\ldots,k$. Let ψ be an exact design supported on k+1 points in I^k such that $X(\psi)$, as defined in (4.3), is a Hadamard matrix in standard form. Then we have

$$\max_{g \in G} L(\underline{B}^*,\psi,g) = \min_{\xi \in D} \min_{\underline{B} \in R(\xi)} \max_{g \in G} L(\underline{B},\xi,g)$$

where M is as in Theorem 3.1.

Theorem 4.2 allows one to reduce the support of a minimax design in special cases.

REMARK. Suppose A is a Lebesgue measurable set in I^k, A is invariant under coordinate reflections, and $(1,\ldots,1)' \in A$. Then the above arguments work when we restrict $\underline{x} \in A$ and Theorems 3.1 and 3.2 again hold.

5. Additional results.

The results of Section 2 can be applied to other situations also. Using the notation of Section 2 we get the following when $\theta_0 = \theta_1 = \ldots = \theta_k = 1$.

Case 1. Global performance; interpolation (integrated error).

Let w be a nonnegative measure on X. With the same model as in equation (2.1) let

$$\hat{Y}(\underline{x}) = \hat{\beta}_0 + \hat{\underline{\beta}}'\underline{x}$$

be an estimate of $EY(\underline{x})$. $\hat{\beta}_0$ and $\hat{\underline{\beta}} = (\hat{\beta}_1,\ldots,\hat{\beta}_k)'$ are as in (2.2). Assume our loss function is

$$\int(\beta_0 + \underline{\beta}'\underline{x} + g(\underline{x}) - \hat{Y}(\underline{x}))^2 w(d\underline{x}).$$

For $\underline{z} = (z_1,\ldots,z_k)' \in X$, $\xi \in D$, $\underline{B} \in R(\xi)$ define

$$L_{\underline{z}}(\underline{B},\xi,g) = \rho\int(B_0(\underline{x}) + \sum_{i=1}^{k} z_i B_i(\underline{x}))^2 d\xi(\underline{x})$$
$$+ (g(\underline{z}) - \int(B_0(\underline{x}) + \sum_{i=1}^{k} z_i B_i(\underline{x}))g(\underline{x})d\xi(\underline{x}))^2.$$

Then the corresponding risk will be

$\int L_{\underline{z}}(\underline{B},\xi,g)w(d\underline{z})$.

The following is analogous to Theorem 2.1.

THEOREM 5.1. Suppose $\xi \in D$. For all $x^* \in X$ if $\xi(\underline{x}^*) \neq 0$, $w(\underline{x}^*) = 0$, and there exists $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_m \in X$ such that $\underline{x}^* = \sum_{\ell=1}^{m} \alpha_\ell \underline{x}_\ell$, where $0 < \alpha_\ell < 1$ for all ℓ and $\sum_{\ell=1}^{m} \alpha_\ell = 1$, then there exists $\xi^* \in D$ such that

(i) $\xi^*(\underline{x}^*) = 0$

(ii) $\text{supp } \xi^* = \text{supp } \xi \cup \{\underline{x}_1,\ldots,\underline{x}_m\} - \{\underline{x}^*\}$

(iii) $\min_{\underline{B}\in R(\xi^*)} \max_{g\in G} \int L_{\underline{z}}(\underline{B},\xi^*,g)w(d\underline{z}) \leq \min_{\underline{B}\in R(\xi)} \max_{g\in G} \int L_{\underline{z}}(\underline{B},\xi,g)w(d\underline{z})$.

Proof. Following the proof given for theorem 2.1 it is straightforward to show that the following counterparts to equations (2.8), (2.9), and (2.10) are true.

First, for all ξ^* such that $\xi^*(\underline{x}^*) = 0$ we have for all $\underline{B} \in R(\xi)$, since $w(\underline{x}^*) = 0$,

$$(5.1)\quad \max_{g\in G} \int L_{\underline{z}}(\underline{B},\xi^*,g)w(d\underline{z}) = \max_{g\in G^*} \int L_{\underline{z}}(\underline{B},\xi^*,g)w(d\underline{z})$$

where G^* here is the counterpart of the G^* defined in the proof of Theorem 2.1.

Second, for all ξ such that $\xi(\underline{x}^*) \neq 0$ we have

$$(5.2)\quad \min_{\underline{B}\in R(\xi)} \max_{g\in G} \int L_{\underline{z}}(\underline{B},\xi,g)w(d\underline{z}) \geq \min_{\underline{B}\in R(\xi)} \max_{g\in G^*} \int L_{\underline{z}}(\underline{B},\xi,g)w(d\underline{z})$$

Finally, it suffices to show that for any $\xi \in D$ with $\xi(\underline{x}^*) \neq 0$ there exists $\xi^* \in D$ satisfying (i) and (ii) of the theorem such that

$$(5.3)\quad \min_{\underline{B}\in R(\xi^*)} \max_{g\in G^*} \int L_{\underline{z}}(\underline{B},\xi^*,g)w(d\underline{z}) \leq \min_{\underline{B}\in R(\xi)} \max_{g\in G^*} \int L_{\underline{z}}(\underline{B},\xi,g)w(d\underline{z}).$$

The following lemma, analogous to Lemma 2.1, proves (5.3) and hence the theorem.

LEMMA 5.1. Suppose $\xi \in D$ is such that $\xi(\underline{x}^*) \neq 0$ and there exist $\underline{x}_1,\ldots,\underline{x}_m \in X$ such that $\underline{x}^* = \sum_{\ell=1}^{m} \alpha_\ell \underline{x}_\ell$, where $\alpha_\ell > 0$ for all ℓ and $\sum_{\ell=1}^{m} \alpha_\ell = 1$. Then there exists $\xi^* \in D$ satisfying (i) and (ii) of Theorem 5.1 and for any $\underline{B} \in R(\xi)$ there exists $\underline{B}^* \in R(\xi^*)$ such that for all $g \in G^*$

$$L_{\underline{z}}(\underline{B},\xi,g) = L_{\underline{z}}(\underline{B}^*,\xi^*,g) + d(\underline{z})$$

for all $\underline{z} \in X$. Here $d(\underline{z})$ is a nonnegative real number that does not depend on g but does depend on $\underline{z}$.

Proof. Analogous to Lemma 2.1.

Notice that when supp(w) is just a single point we are dealing with interpolation to a point.

Case 2. Global performance (maximal error).

Using the same notation as in Case 1 we get the following.

THEOREM 5.2. Suppose $\xi \in D$. For all $\underline{x}^*$ if $\xi(\underline{x}^*) \neq 0$ and there exists $\underline{x}_1,\ldots,\underline{x}_m \in X$ such that $\underline{x}^* = \sum_{\ell=1}^{m} \alpha_\ell \underline{x}_\ell$ where $\alpha_\ell > 0$ for all ℓ and $\sum_{\ell=1}^{m} \alpha_\ell = 1$, then there exists $\xi^* \in D$ such that

(i) $\xi^*(\underline{x}^*) = 0$

(ii) $\text{supp}\ \xi^* = \text{supp}\ \xi \cup \{\underline{x}_1,\ldots,\underline{x}_m\} - \{\underline{x}^*\}$

(iii) $\min_{\underline{B}\in R(\xi^*)} \max_{g\in G} \max_{\underline{z}\in X} L_{\underline{z}}(\underline{B},\xi^*,g) \leq \min_{\underline{B}\in R(\xi)} \max_{g\in G} \max_{\underline{z}\in X} L_{\underline{z}}(\underline{B},\xi^*,g)$.

Proof. The proof is almost the same as for Theorem 5.1, but we must take care of the point $\underline{x}^*$ carefully. The following steps are easily proved.

Step 1: For all $\xi^* \in D$ such that $\xi^*(\underline{x}^*) = 0$ we have

$$\max_{g \in G} \max_{\underline{z} \neq \underline{x}^*} L_{\underline{z}}(\underline{B},\xi^*,g) = \max_{g \in G^*} \max_{\underline{z} \neq \underline{x}^*} L_{\underline{z}}(\underline{B},\xi^*,g).$$

Step 2: For all $\xi \in D$ such that $\xi(\underline{x}^*) \neq 0$ we have

$$\min_{\underline{B} \in R(\xi)} \max_{g \in G} \max_{\underline{z} \in X} L_{\underline{z}}(\underline{B},\xi,g) \geq \min_{\underline{B} \in R(\xi)} \max_{g \in G^*} \max_{\underline{z} \neq \underline{x}^*} L_{\underline{z}}(\underline{B},\xi,g).$$

Step 3: There exists $\xi^* \in D$ satisfying (i) and (ii) such that

$$\min_{\underline{B} \in R(\xi^*)} \max_{g \in G^*} \max_{\underline{z} \neq \underline{x}^*} L_{\underline{z}}(\underline{B},\xi^*,g) \leq \min_{\underline{B} \in R(\xi)} \max_{g \in G^*} \max_{\underline{z} \neq \underline{x}^*} L_{\underline{z}}(\underline{B},\xi,g).$$

In order to complete the proof we have to show

$$\max_{g \in G^*} \max_{\underline{z} \neq \underline{x}^*} L_{\underline{z}}(\underline{B},\xi^*,g) = \max_{g \in G^*} \max_{\underline{z} \in X} L_{\underline{z}}(\underline{B},\xi^*,g)$$

which follows from a straightforward continuity argument since ξ^* is a discrete measure so we can assume g to be continuous without loss of generality.

Case 3. Extrapolation.

Let w be a nonnegative measure on $X^* - X$ where $X^* \subset R^k$ and our model is "nearly" linear, in the sense of equation (2.1), on X^*. A result analogous to Theorem 5.1 holds in this case (here "$w(\underline{x}^*) = 0$" holds always).

Case 4. For a loss function of the form

$$\theta_0(\hat{\beta}_0-\beta_0)^2 + (\hat{\underline{\beta}}-\underline{\beta})'A(\hat{\underline{\beta}}-\underline{\beta})$$

where the notation is similar to Section 2 with $\hat{\underline{\beta}}' = (\hat{\beta}_1,\ldots,\hat{\beta}_k)$ being linear estimates of $\underline{\beta}' = (\beta_1,\ldots,\beta_k)$, and A in nonnegative definite, we can prove a result analogous to Theorem 2.1. In fact, after a suitable linear transformation (which preserves extreme points) we can reduce to the case of Section 2.

Case 5. For a bilinear model

$$\beta_0 + \Sigma\beta_i x_i + \Sigma\beta_{ij} x_i x_j + g(\underline{x})$$

the method of Section 2 also works after necessary modifications (e.g. the analogue of Theorem 2.1 can be proved with the restriction that $\underline{x}_1, \ldots, \underline{x}_m$, $\underline{x}^*$ all lie on a line parallel to one of the coordinate lines. We can thus conclude the optimal design will have support on the boundary of X only).

Case 6. If $X \subset R$ and the set of contaminations G is

$$G = \{g: |g(x)| < c|x|\}$$

where $c > 0$ is a constant, then the method of Section 2 can be applied to give an alternative proof to some of the results in Marcus and Sacks (1976). Moreover, analogous results for $X \subset R^k$ can be easily established (an optimal design will have support belonging to the set $\{0\} \cup$ {extreme points of X}.

6. General Remark. Although we have used the approximate design approach to our problems the results should not be interpreted as an asymptotical conclusion. In fact, if the sample size is very large we should use a more restrictive contamination model (e.g. all g in G uniformly satisfy a certain degree of smoothness). The results would be different and could be expected to be close to those of Huber. If the sample size is not large, though, we can feel comfortable with our model.

7. Acknowledgements. Both authors wish to express their sincere thanks to Professor Jack Kiefer for his helpful suggestions and discussions.

REFERENCES

[1] 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.

[2] Huber, P. (1975). Robustness and designs. *A Survey of Statistical Design and Linear Models*, North Holland, Amsterdam, 287-303.

[3] Kiefer, J. (1973). Optimal designs for fitting biased multiresponse surfaces. *Multivariate Analysis III*, Academic Press, New York, 245-268.

[4] 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.