A Method for the Determination of Optimal Model Robust Designs in the Estimation of a Linear Functional

by

M. Carl Spruill[1]
Georgia Institute of Technology
and
Purdue University

Technical Report #82-8.

Department of Statistics
Purdue University

March 1982

[1]Support was provided by NSF Grant No. MCS-8103444.

Summary. Speckman developed a minimax linear estimator robust against departures from an assumed model. We investigate some concomitant questions of optimal design. Characterization of the optimal designs is provided through theorems applicable to a broad class of specific models. Sample problems are solved to illustrate their application.

1. Introduction. In the usual regression design problem using the best linear unbiased estimator based upon N observations the variance of the estimator of $c'\theta$ may be expressed as $N^{-1}d(c,\xi)$ where ξ is the design measure. One task of the designer is the choice of the best possible design measure ξ concentrated on N or fewer points, namely the one which minimizes $d(c,\xi)$. Often it is the case that if a larger class of designs is allowed then the minimization is more easily accomplished. For example in the usual regression problem $E[Y(x)] = \theta' f(x)$, $\theta \in \mathbb{R}^k$, it is known that if the class of measures on X contains all the finitely supported measures and $f(X)$ is compact then a minimizing design must always exist in that class. Such a design is said to be optimal in the approximate theory since that design may not be feasible on N or fewer points. One may employ a method such as that given by Fedorov (1972) to pass from the approximate theory optimal design to a good design on N or fewer points.

In the model robust design problems which we consider there is a linear operator T which provides a measure of the possible contamination. The measure of the performance of the estimator is maximum mean square error and it is shown that based upon N observations the minimax linear estimator has a maximum mean square error which may be written as $N^{-1}d_{T/\sqrt{N}}(c,\xi)$ where this function of ξ is very like $d(c,\xi)$. Again one would choose the design on N or fewer points which minimizes $d_{T/\sqrt{N}}(c,\xi)$ and again minimization is more easily carried out over a larger class of designs. Obviously from the notation the minimizer of $d_{T/\sqrt{N}}(c,\xi)$ may depend upon N, the number of observations to be taken. We show that whenever the support of the minimizer contains fewer than N points the same procedure as given by Fedorov for use in the usual case continues to provide good designs on N or fewer points. Thus the wonted employment of the approximate theory works with the amendment that the number of observations be specified beforehand.

Actually the last statement requires qualification because one of the desirable characteristics of the usual design problems is that one is assured of the existence of an optimal design in an appropriately enlarged design space. Our investigations include both those with finite dimensional parameter spaces where these assurances are easily obtained in the class of finitely supported measures and those with infinite dimensional parameter spaces where those assurances are more difficult to obtain even if one is able to identify an approriate enlargement.

The contents of the paper by section are as follows. Section two is devoted to the development of some formulas for the minimax mean square error allowing observations which are second order processes. The generalization proceeds with little deviation from the lines established by Speckman (1979). In the third section theorems on optimal design are presented. These designs could be termed optimal in the approximate theory in the sense we have described. The fourth section is concerned with finding good exact designs from approximately optimal ones. In the fifth section some comparatively simple sample problems are presented. The subject of the sixth section whose contents we now summarize in some detail is an extended sample problem. In many instances more detailed arguments are given in Spruill (1982).

The problem considered is that of the extrapolation of a function to a point outside the interval on which observations may be taken. Thus we suppose that for every finite collection of points $\{x_1,\ldots,x_N\}$ in $[a,b]$ we may observe with error the values of an unknown function f and that we are to provide on the basis of the observed values $\{Y(x_1),\ldots,Y(x_N)\}$, $Y(x_j) = f(x_j) + \varepsilon_j$, $E(\varepsilon_j) \equiv 0$, $E(\varepsilon_i\varepsilon_j) = \delta_{ij}$, an estimate of the value of $f(c)$. It is assumed that on the interval

$[a,c]$, $c > b$, f is essentially a polynomial whose degree does not exceed $m - 1$. Allowance is made for the possibility that $f^{(m)}$ is non-zero by choosing a linear estimator to minimize the maximum mean square error over f for which $f^{(m-1)}$ is absolutely continuous and $\left(\int_a^c (f^{(m)}(s))^2 ds\right)^{\frac{1}{2}} \leq \varepsilon$. The maximum mean square error of the resulting estimator may be found using Speckman's methods and depends of course on the selection of the points $\{x_1,\ldots,x_N\}$. We prove that the optimal locations of these points are at $a \leq x_1^* < x_2^* < \ldots < x_m^* \leq b$ and that the observations should be taken in the proportions (see theorems 6.2 and 6.3) $p_i^* = \dfrac{|\phi_{x_i^*}(c)|}{\sum_{j=1}^{m} |\phi_{x_j^*}(c)|}$ at x_i^*, where $\phi_{x_j^*}(y) = \dfrac{\prod_{j\neq i} (y-x_j^*)}{\prod_{j\neq i} (x_i^*-x_j^*)}$ are the Lagrange interpolation polynomials to the points x_i^*. The resulting estimator is the same as the usual one $\hat{f}(c)$, where $\hat{f}$ is the unique polynomial of degree $m - 1$ passing through the points $\{(x_j^*,\bar{y}(x_j^*))\}_{j=1}^{m}$ and $\bar{y}(x_j^*)$ is the average of the observations at x_j^*.

The points $\{x_1^*,\ldots,x_m^*\}$ are gotten from the solution of the following complementary problem P_η for $\eta = (N\varepsilon^2)^{-1}$.

P_η: For $\eta > 0$ and $a < b < c$ fixed minimize

$$\rho(f) = \sup_{a\leq x\leq b} |f(x)|^2 + \eta \int_a^c (f^{(m)}(x))^2$$

over all f in the Sobolev space $W_m^2[a,c]$ such that $f(c) = 1$.

We prove that if $\eta>0$ then a solution f_0 exists, is unique, and equioscillates on $[a,b]$. That is, setting $||f||_\infty = \sup_{a\leq x\leq b} |f(x)|$ here and throughout the remainder, there are m points $a \leq x_1^* < \ldots < x_m^* \leq b$ such that $f_0(x_j^*) = (-1)^{j-m}||f_0||_\infty$ for $j=1,2,\ldots,m$. We also provide a formula which, given $x_1^*,\ldots,x_m^*$, completely determines the solution. Thus f_0 is shown to be a polynomial spline of degree 2m - 1 on each of the subintervals $[x_i^*, x_{i+1}^*]$, $i=0,\ldots,m$, where $x_0^* = a$ and $x_{m+1}^* = c$. The knots are located at points $x_1^*,\ldots,x_m^*$ where the solution f_0 attains its extreme values on [a,b]. At the left end of the interval, one can have $a < x_1$ in which case $f_0(x) = (-1)^{1-m}||f_0||_\infty$ $x \in [a,x_1]$. Even though this is the case the location of the points $\{x_j^*\}_{j=1}^m$ is unique. Thus the optimal design is unique. It is readily observed that the optimal proportions for the robust extrapolation problem correspond to those found by Hoel and Levine (1964) for unbiased extrapolation except that the locations $\{x_i^*\}$ may differ. The estimators are the same and Hoel and Levine's results as well as their extension by Karlin and Studden (1966b) may be subsumed under ours. We do not present a method of determining the locations $\{x_i^*\}_{i=1}^m$ in general. We can show by an example in the case m = 2 that the locations can differ. In the example they differ, depending upon N and ε, with the locations a and b given by the Hoel-Levine results. If $a < b < c$ then one of the two points of support is always b. The left hand endpoint however is located in the interior of (a,b) when either the contamination ε or the sample size N is sufficiently large. In studying the model-robust extrapolation of a mean function Huber (1975) derived similar characterizations when the interval of observation is a half-line and the contamination is $\sup_{0 \le x < +\infty} |f^{(m)}(x)| < \varepsilon$. We are able to make direct comparisons when m = 2. The model under our consideration leads to a location of right-most support point to the right of Huber's by a factor of roughly $(3/2)^{1/3} - 1$.

It is not surprising that the similar models lead to quite similar optimal designs although the proof is laborious. Some recompense is to be found in the aesthetically pleasing manner in which the solutions to the purely mathematical problems P_n and the purely statistical optimal design problems entwine and by their convolution lead to information not readily apparent by the consideration of either one separately. While it is not entirely clear that the problems P_n inherently merit solution it has been shown in Spruill (1981) that solutions in the case m = 2 are useful in the solution of a class of non-standard control theory problems. Nonetheless the generality of the method suggested by the theorems commends itself. For example, if instead of the estimation of f(c) one is to estimate $f'(x_0)$ for some $x_0 \in (a,b)$, then the optimal design and estimator may be found in the following way.

i) Find $f_0 \in W_m^2$ to minimize $\rho(f)$ under the restriction $f'(x_0) = 1$.

ii) Find ξ_0 whose support is contained in $\{x: |f_0(x)| = ||f_0||_\infty\}$ and which satisfies for all $f \in W_m^2[a,c]$,

$$\int_a^b f_0(x)f(x)d\xi_0(x) + \frac{1}{N\varepsilon^2} \int_a^c f_0{}^{(m)}(x)f^{(m)}(x)dx = \rho(f_0)f'(x_0).$$

iii) Estimate $f'(x_0)$ by

$$\hat{f}'(x_0) = \frac{1}{N\rho(f_0)} \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_j(x_i)f_0(x_i).$$

There are many others who treat optimal robust designs. See, for example, Huber (1975), Kiefer (1980), Li (1981), Li and Notz (1980), Marcus and Sacks (1976), Notz (1980), and Pesotchinsky (1980). None of the above-mentioned have assumptions which are exactly the same as ours.

2. Preliminaries. Let X be an arbitrary set which is to be thought of as a set indexing all the experiments available to the experimenter. We assume throughout that for each finite subset $\{x_j\}_{j=1}^{k}$, $0 < k < \infty$, of

distinct elements x_j from X and each collection $\{n_j\}_{j=1}^{k}$ of natural numbers $n_j > 0$, $j=1,\dots,k$ an experiment can be run. We further assume that the experiment results in the $N = \sum_{j=1}^{k} n_j$ uncorrelated second order stochastic processes $\{Y(x_1,1,t),\dots,Y(x_1,n_1,t),\dots,Y(x_k,n_k,t): t \in T\}$ where T is an arbitrary set. We think of T as the time parameter set although it could be space or some other parameter. For example, if the Y's are Banach space valued then one could have T as the dual space and Y(x,t) the evaluation of Y(x) at $t \in T$.

We will notationally suppress the underlying probability space $(\Omega,\mathcal{G},P)$ upon which we assume all the processes to be defined. Thus it will be assumed that Y(x,t) are all measurable functions from $(\Omega,\mathcal{G})$ into $(\mathbb{R},\mathcal{B})$ where $\mathcal{B}$ is the usual class of Borel subsets in $\mathbb{R}^1$.

The stochastic process Y(x,t) is the sum of a mean function $m_x\theta(t)$ and a zero mean error process N(t) whose covariance function K(s,t) is assumed to be known. The parameter θ is known to be a member of the separable Hilbert space Θ although its exact value is unknown. The mappings m_x are each known bounded linear mappings from Θ into the reproducing kernel Hilbert space H(K) generated by K, the value of m_x at θ in Θ being the function of $t \in T$ given by $m_x\theta$.

Our ultimate objective is to robustly estimate the value of (τ,θ), where $\tau \in \Theta$ is fixed, with a linear estimator employing the best possible choice of x's. We begin by fixing the set of observation points $x_1,\dots,x_k$ and numbers of observations $n_1,\dots,n_k$ at each point. Then the space of linear estimators which we employ is that developed by Parzen (1959). As detailed in Spruill and Studden (1978) our experiment induces a stochastic process Z on the new parameter set $\Gamma = \{(x_1,1),(x_1,2),\dots,(x_1,n_1),\dots,(x_k,n_k)\} \times T$ whose covariance kernel is B. Denoting by H(B) the reproducing kernel

Hilbert space generated by B the linear estimators are $\{\langle Z,g\rangle_B : g\in H(B)\}$. The particular way in which the estimator is to be robust was introduced and studied by Speckman (1979) when the observations are scalars. There is a given bounded linear operator T which maps Θ into a separable Hilbert space $\mathcal{H}$ and a given $\varepsilon > 0$. The associated minimax linear estimator $\langle Z,g_0\rangle_B$ satisfies $\inf_g \sup_{||T\theta||\le\varepsilon} V(\tau,\theta,g) = \sup_{||T\theta||\le\varepsilon} V(\tau,\theta,g_0)$ where $V(\tau,\theta,g) = E_\theta(\langle Z,g\rangle_B-(\tau,\theta))^2$. The first theorem provides a concise and useful formula for $\inf_g \sup_{||T\theta||\le 1} V(\tau,\theta,g)$, where we have assumed without any loss of generality that ε has been absorbed into the definition of T. The proof of the theorem proceeds exactly along the lines of Speckman's proof. Define the mapping $m: \Theta \to H(B)$ by $m\theta(\gamma) = (m_x\theta)(t)$ where $\gamma = (x,v,t) \in \Gamma$. We shall denote the adjoint of any bounded linear operator A by A^*, the range of A by $\mathcal{R}(A)$, and the null space of A by $\mathcal{N}(A)$. Define

$$M = m^*m + T^*T.$$

Theorem 2.1. If $\mathcal{R}(M)$ is closed then whenever $\tau \in \mathcal{R}(M)$

$$\inf_{g\in H(B)} \sup_{||T\theta||\le 1} V(\tau,\theta,g) = \tau M^{\#}\tau \tag{2.1}$$

where $M^{\#}$ is the Moore-Penrose inverse relative to the ordinary orthogonal projections. If $\tau \notin \mathcal{R}(M)$ then the expression on the ℓ.h.s. of (2.1) is $+\infty$. If $\tau \in \mathcal{R}(M)$ and $\tilde{\theta}$ is any solution to the equation $M\tilde{\theta} = \tau$ then $m(\tilde{\theta}) = g_0$ yields the unique minimax linear estimator of $\tau(\theta)$.

Corollary 2.1. If $\mathcal{R}(M)$ is closed then for all $\tau \in \Theta$

$$\inf_g \sup_{||T\theta||\le 1} V(\tau,\theta,g) = \sup_{\theta\in N} \left(\frac{(\tau,\theta)^2}{||m(\theta)||^2+||T\theta||^2} \right)$$

where $N = \{\theta: ||m(\theta)||^2+||T\theta||^2 > 0\}$.

Proof. The proof proceeds as in lemma 2.3 of Spruill and Studden (1978) (except that our M here is a linear operator on Θ and τ replaces c) and will not be given here. See Nashed and Votruba (1976) Section 5.3 to verify that all the needed characteristics of the proof of lemma 2.3 are present in this case. □

We shall require at some points an alternative characterization of the minimax linear estimator proven by Speckman using different methods and under stronger conditions than those assumed in the next lemma.

Lemma 2.1. If $H(K) = \mathbb{R}^1$ and the range of the mapping (m,T) from Θ into the product Hilbert space $\mathbb{R}^N \times \mathcal{H} = W$ is closed and m and T are bounded then there is a Θ- valued random variable $\bar{\theta}(Y)$ minimizing for each Y in $\mathbb{R}^N$ the expression $||Y-m\theta||^2_{\mathbb{R}^N} + ||T\theta||^2_{\mathcal{H}}$ and the minimax linear estimator of (τ,θ) is $(\hat{\tau,\theta}) = (\tau,\bar{\theta}(Y))$.

Proof. One can show that $(m,T)^*(m,T) = M = m^*m + T^*T$ has closed range and is bounded. Introducing the Moore-Penrose inverse $(m,T)^{\#}$ of (m,T) we have

$$(m,T)^{\#} = M^{\#}(m,T)^*$$

as is verified in Beutler and Root (1976) by equation (1.44) and the material which it precedes. An element w of W is of the form $w = (w_1, w_2)$ where $w_1' = (w_{11},\ldots,w_{1N})$ and its norm squared is $||w||^2_W = \Sigma w^2_{1i} + ||w_2||^2_{\mathcal{H}}$. Since $(Y,0) \in W$ the usual properties of the Moore-Penrose inverse show that $\inf_{\Theta} ||(Y,0)-(m,T)\theta||^2_W$ is achieved for

$$\bar{\theta}(Y) = (m,T)^{\#}(Y,0) = M^{\#}m^*Y.$$

By theorem 2.1 the formula

$$(\hat{\tau,\theta}) = m\tilde{\theta}'Y$$

holds for Speckman's estimator whenever $M\tilde{\theta} = \tau$. Since $\tilde{\theta} = M^{\#}\tau$ satisfies the equation and since theorem 1.6 of Beutler and Root shows that $(M^{\#})^* = (M^*)^{\#} = M^{\#}$ we have $(\tau,\hat{\theta}) = (M^{\#}\tau, m^*Y) = (\tau, M^{\#}m^*Y) = (\tau, \tilde{\theta}(Y))$ proving the lemma. □

3. Some theorems on optimal designs. The experiment above has an associated probability measure ξ defined by $\xi(x_i) = \frac{n_i}{N}$, $i=1,\ldots,k$. We wish now to examine the dependence of $\inf_g \sup_{||T\theta||\leq 1} V(\tau,\theta,g)$ on the design measures ξ.

We assume that the set X has a topology and denote the minimal σ-field containing the open sets by $\mathcal{B}_X$, the Borel σ-field on X. We shall denote by Ξ an arbitrary, fixed, non-empty convex collection of Borel probability measures on X and for each $\xi \in \Xi$, $S(\xi)$ will denote the support of ξ. The set $S(\xi)$ will be regarded as having measurable subsets $S(\xi) \cap \mathcal{B}_X$. The set $V = H(K) \times \mathcal{H}$ is a Hilbert space with inner product $[(u_1,v_1),(u_2,v_2)] = (u_1,v_1)_K + (u_2,v_2)_{\mathcal{H}}$ and the collection of measurable functions f from $S(\xi)$ to V satisfying $\int_{S(\xi)} ||f(x)||_V^2 d\xi(x) < \infty$ is $L^2(\xi)$ and is a Hilbert space with inner product $(f_1,f_2)_\xi = \int_{S(\xi)} (f_1(x),f_2(x))_V d\xi(x)$ as is shown in Dunford and Schwarz (1958).

(A1) We make the assumption that for every $\theta \in \Theta$ and every $\xi \in \Xi$ the functions $m_x\theta$ of x are measurable on $S(\xi)$.

In light of this we define, for each $\xi \in \Xi$, the linear mappings $L_\xi : \Theta \to L^2(\xi)$ given by $L_\xi\theta(x) = (m_x\theta, T\theta)$ for $x \in S(\xi)$. It is easily checked that if (A2) is satisfied then L_ξ are bounded since T is assumed to be bounded.

(A2) For all $\xi \in \Xi$, $\int_{S(\xi)} ||m_x||^2 d\xi(x) < \infty$ where $||m_x|| = \sup_{||\theta||=1} ||m_x\theta||_K$.

Define $\mathfrak{F}$ to be the collection of measurable functions $\phi: X \to V$ satisfying $||\phi(x)||_V \leq 1$. Note that under (A1) the function $L_x^*\phi(x)$ is measurable from X into Θ whenever $\phi \in \mathfrak{F}$ and under (A2) $\int ||L_x\theta||_V^2 d\xi(x) < \infty$ and the Bochner integral $\int L_x^*\phi(x)d\xi(x) \in \Theta$. Define

$$\mathfrak{R} = \{\int L^*_x\phi(x)d\xi(x): \ \phi \in \mathfrak{F}, \ \xi \in \Xi\}$$

and

$$d_T(\tau,\xi) = \sup_{\theta \in N} \frac{(\tau,\theta)^2}{\int ||L_x\theta||_V^2 d\xi(x)} \tag{3.1}$$

where $N = \{\theta: \int ||L_x\theta||_V^2 d\xi(x) > 0\}$.

(A3) For each $\xi \in \Xi$ L_ξ is bounded and $\mathfrak{R}(L_\xi)$ is closed in $L^2(\xi)$.

We shall also use the operator $M(\xi) = L_\xi^*L_\xi$ and point out here that $\mathfrak{R}(M(\xi))$ is closed if and only if $\mathfrak{R}(L_\xi^*)$ is closed and in this case $\mathfrak{R}(L_\xi^*) = \mathfrak{R}(M(\xi))$.

(A4) There is a proper closed supporting hyperplane to $\mathfrak{R}$ at each of its boundary points.

(A5) For each $\theta \neq 0$, $\sup_X ||L_x\theta|| > 0$.

Let $v_0 = \inf_{\xi \in \Xi} d_T(\tau,\xi)$.

<u>Theorem 3.1</u>. Under conditions (A1) - (A5) if $\tau \in \mathfrak{R}(M(\xi))$ for some $\xi \in \Xi$ then $d_T(\tau,\xi_0) = v_0$ and $\xi_0 \in \Xi$ if and only if there is a function $\phi \in \mathfrak{F}$ such that $||\phi(x)||_V \equiv 1$ and $\int L^*_x\phi(x)d\xi_0(x)$ is

i) proportional to τ and

ii) in $\mathfrak{R} \cap \partial\mathfrak{R}$.

Proof. When the proper replacements are made (L_x for m_x, etc.) the same arguments as in Theorem 3.1 of Spruill (1980) apply. As is not made clear in that proof the G-inverse $M^{\#}$ is relative to the orthogonal projectors. □

Theorem 3.2 of Spruill also goes through. However we have a much improved version which we prove below in its stead. Let $\Delta = \{\theta \in \Theta : (\tau,\theta) = 1\}$ and recall that $L_x\theta = (m_x, T\theta)$, an ordered pair in $V = H(K) \times \mathcal{H}$ for all $x \in X$ and $\theta \in \Theta$.

Theorem 3.2. Suppose there is a point $\delta_0 \in \Delta$ and a design $\xi_0 \in \Xi$ satisfying

i) $S(\xi_0) \subset \{x: ||L_x\delta_0||_V = \sup_X ||L_x\delta_0||_V\}$

ii) $\int L_x^* L_x \delta_0 d\xi_0(x) = \alpha\tau$ for some $\alpha > 0$, and

iii) $\int ||L_x\theta||_V^2 d\xi_0(x) = 0$ entails $(\tau,\theta) = 0$.

Then ξ_0 satisfies $d_T(\tau,\xi_0) = \inf_\Xi d_T(\tau,\xi)$ and

iv) $\inf_\Delta \sup_X ||L_x\delta||_V^2 = \sup_X ||L_x\delta_0||_V^2$.

The conditions required are (A1) and (A2). Conversely if (A1) - (A5) hold and there is a $\xi_0 \in \Xi$ satisfying $d_T(\tau,\xi_0) = v_0 < \infty$ then a point $\delta_0 \in \Delta$ may be found satisfying conditions i) through iv).

Proof. Clearly

$$d(\tau,\xi) \geq [\inf_{\theta \in N \cap \Delta} \int ||L_x\theta||_V^2 d\xi(x)]^{-1}. \tag{3.2}$$

Since $\inf_{N\cap\Delta} \int ||L_x\theta||^2 d\xi \leq \inf_{N\cap\Delta} \sup_X ||L_x\theta||^2 \leq \sup_X ||L_x\delta_0||^2 = s$, we have $d(\tau,\xi) \geq s^{-1}$ for all $\xi \in \Xi$. Using ii) we have for ξ_0

$$d(\tau,\xi_0) = \sup_N \frac{\alpha^{-2}[\int (L_x\theta, L_x\delta_0)_V d\xi_0(x)]^2}{\int ||L_x\theta||_V^2 d\xi_0(x)}$$

$$\leq \sup_N \frac{s\alpha^{-2}[\int ||L_x\theta|| d\xi_0(x)]^2}{\int ||L_x\theta||^2 d\xi_0(x)} \leq s\alpha^{-2} = s^{-1}.$$

Since by (3.2) $d(\tau,\xi_0) \geq s^{-1} \geq [\inf_{N\cap\Delta} \sup_X ||L_x\theta||_V^2]^{-1}$ we have $\inf_{N\cap\Delta} \sup_X ||L_x\theta||^2 \geq \sup_X ||L_x\delta_0||^2$. By iii) $\Delta \cap N = \Delta$ and we have shown that if (A1) and (A2) hold then i), ii), and iii) entail $d(\tau,\xi_0) = \inf_{\Xi} d(\tau,\xi)$ and iv). Now suppose that (A1) - (A5) hold and that $\xi_0 \in \Xi$ minimizes $d_T(\tau,\xi)$. By theorem 3.1 there is a function $\phi: X \to H(K) \times \mathcal{H}$ such that $||\phi(x)|| \equiv 1$,

(3.3) $$\int L_x^* \phi(x) d\xi_0(x) = \beta\tau ,$$

and $\beta\tau \in \partial\mathcal{R}$. By (A4) there is a $\lambda \neq 0$, $\lambda \in \Theta$ such that $(\lambda, r) \leq \beta(\lambda,\tau)$ for all $r \in \mathcal{R}$. Since by (A5) $\sup_X ||L_x\lambda|| > 0$ we may find a sequence of points $\{x_n\}$ in X satisfying $||L_{x_n}\lambda|| \uparrow \sup_X ||L_x\lambda||$ and $||L_{x_n}\lambda|| > 0$. Set $r_n = L_{x_n}^* \frac{L_{x_n}\lambda}{||L_{x_n}\lambda||}$. Then $r_n \in \mathcal{R}$ for all n and since (3.3) holds

$$\lim_{n\to\infty} ||L_{x_n}\lambda|| \leq \int (L_x^*\phi(x), \lambda) d\xi_0(x)$$

$$\leq \sup_X ||L_x\lambda||$$

with strict inequality unless $||L_x\lambda|| \equiv \sup_X ||L_x\lambda||$ a.e. ξ_0. Set $\delta_0 = \frac{\lambda}{(\tau,\lambda)}$ ($(\tau,\lambda) \neq 0$ since $\beta(\tau,\lambda) > 0$). Clearly i) is satisfied. From above we also conclude that $\phi(x) = k_x L_x \lambda$ a.e. ξ_0. This in turn implies that $\phi(x) = \frac{L_x\lambda}{||L_x\lambda||}$ a.e. ξ_0. Therefore

$$\int L_x^*\phi(x) d\xi_0(x) = [\int L_x^* L_x \delta_0 d\xi_0(x)][\sup_X ||L_x\lambda||]^{-1}$$

and we see that ii) is also satisfied. If iii) is not satisfied then there is a sequence θ_n such that $\int||L_x\theta_n||^2 d\xi_0(x) \to 0$ and $(\tau,\theta_n) \to t \neq 0$. This implies $d_T(\tau,\xi_0) = +\infty$ which contradicts our assumptions. We conclude that iii) is satisfied and consequently that iv) also is satisfied. □

When Θ is finite dimensional, X is compact, and the mappings $L_x\theta$ are continuous in x for each fixed θ then one may prove much stronger theorems than 3.1 and 3.2. We state, without proof, one such theorem. We shall need the following conditions.

(B1) The mappings L_x: $\Theta \to V$ are linear for each $x \in X$.

(B2) There is a topology on X for which X is compact, one point subsets are Borel measurable, and Ξ is the collection of all finitely supported probability measures.

(B3) For each fixed $\theta \in \Theta$ the mappings $L_x\theta$: $X \to V$ are continuous in x.

(B4) Θ is a finite dimensional Hilbert space.

Theorem 3.3. Under conditions (A5) and (B1) - (B4) there is an optimal design $\xi_0 \in \Xi$ for estimating (τ,θ) whose support contains no more than dim Θ points. In addition conditions (A1) - (A4) are satisfied so that a point δ_0 exists in $\Delta = \{\theta: (\tau,\theta)=1\}$ satisfying i) - iv) of Theorem 3.2.

Finally one can prove the following theorem, which is the analogue of Theorem 4.1 in Spruill (1980), using the same techniques as employed there.

Theorem 3.4. If assumptions (A1) and (A2) hold and if there is a constant $k > 0$ such that for all θ

$$\sup_X ||L_x\theta||_V \geq k||\theta||$$

then (A4) and (A5) also hold.

4. <u>Finding good exact designs.</u> In the previous section few assumptions were made concerning the set Ξ of design measures. In this section we shall also assume that $\Xi \supset \bigcup_{N \geq 1} \Xi_N$ where Ξ_N is the collection of exact (or rational) designs on N or fewer points

$$\Xi_N = \{\text{prob. meas. } \xi \text{ on } X\text{: } \#S(\xi) \leq N \text{ and } N\xi(x_i) \in \mathbb{N} \;\; \forall\, x_i \in S(\xi)\}\ ,$$

where $\mathbb{N} = \{0,1,2,\ldots\}$ is the set of natural numbers.

We shall demonstrate in the last two sections the effectiveness of the theorems above in producing designs which minimize, for a given T and τ, the function $d_T(\tau,\xi)$ over Ξ . However for a given N,T, and τ we really would like to minimze over Ξ_N the expression in (2.1) which may be written as

$$N^{-1} d_{N/\sqrt{T}}(\tau,\xi). \tag{4.1}$$

In the usual design problem using the best linear unbiased estimator the operator T is the zero operator. In that case one may employ the optimal approximate theory design ξ in Ξ to pass for each N sufficiently large to a good design in Ξ_N. Fedorov (1972) gives such a procedure and inequalities which provide a measure of the departure from optimality of the design so constructed. In our present case we observe the possibility that the designs which minimize $d_{T/\sqrt{N}}(\tau,\xi)$ will depend upon N. Thus a change in the routine of passing from the approximate theory to the exact theory has been introduced. Moreover, since the assumptions employed by Fedorov in providing the inequalities do not hold it is not clear what procedure should be employed in the construction of a good exact design from the optimal approximate theory design. We shall prove in this section that Fedorov's procedure continues in the present case to provide as good a method of finding a good exact design from the approximate as it does when unbiased linear estimators are employed.

There is a reasonable alternative to the modus operandi upon which we have embarked which bears mentioning. If in the definition of the linear estimator we measure the allowable deviation from the model relative to the size of the sample we may circumvent some ensuing difficulties. Specifically, if we take the supremum over θ satisfying $||T\theta|| \leq N^{-\frac{1}{2}}$ the minimax mean square error is $N^{-1}d_T(\tau,\xi)$. For this estimator the approximate theory optimal design (if one exists) will not depend upon N. Still the assumptions employed by Fedorov in bounding the error in passing from the optimal approximate theory design to a good design in Ξ_N do not hold for the functional $d_T(\tau,\xi)$.

In the following $A = A_0^*A_0$, $B = B_0^*B_0$, and $D = D_0^*D_0$ where A_0, B_0, and D_0 are all bounded linear operators from Θ into Hilbert spaces and all their ranges are closed. For $\tau \in \Theta$ fixed define

$$\tilde{L}_D(A) = \left[\sup_{\theta \in N(A,D)} \frac{(\tau,\theta)^2}{\theta(A+D)\theta}\right]^{-1}$$

where we take $[+\infty]^{-1} = 0$ and $N(A,D) = \{\theta: \theta(A+D)\theta > 0\}$. One can show (see Spruill (1982)) that for all scalars $k \geq 1$ $k\tilde{L}_D(A) \geq \tilde{L}_D(kA)$ and $\tilde{L}_D(A+B) \geq \tilde{L}_D(A)$.

Let N be a fixed positive integer and T, as above, be a bounded linear operator from Θ into $\mathcal{H}$. Suppose that $\xi^* \in \Xi$ minimizes (4.1) over Ξ and that $\#S(\xi^*) = r < N$. Also suppose that $\xi_N^* \in \Xi_N$ minimizes (4.1) over Ξ_N. Define the measure $\tilde{\xi}_N \in \Xi_N$ from ξ^* as follows. Denoting by [ ] the least integer function defined by [x] = smallest integer greater than or equal to x, assign $[(N-r)\xi^*(x_i)]$ observations at $x_i \in S(\xi^*)$. Since

$[(N-r)\xi(x_i)] < (N-r)\xi(x_i) + 1$ this uses $\leq \sum_{i=1}^{r} ((N-r)\xi(x_i) + 1 = N - r + r = N$ observations. Assign the other observations in any manner and denote the proportion of the total N at x_i by $\tilde{\xi}_N(x_i)$.

Theorem 4.1. The design $\tilde{\xi}_N$ constructed from ξ^* as above satisfies

$$0 \leq 1 - \frac{d_{T/\sqrt{N}}(\tau, \xi_N^*)}{d_{T/\sqrt{N}}(\tau, \tilde{\xi}_N)} \leq \frac{r}{N-r}$$

and

$$0 \leq 1 - \frac{d_{T/\sqrt{N}}(\tau, \xi^*)}{d_{T/\sqrt{N}}(\tau, \xi_N^*)} \leq \frac{r}{N}$$

whenever $M(\xi) = \int L_X^* L_X d\xi(x)$ has closed range for all $\xi \in \Xi$.

Proof: Both relationships above are a consequence of

(4.2) $\tilde{L}_D[M(\xi^*)] \geq \tilde{L}_D[M(\xi_N^*)] \geq \tilde{L}_D[M(\xi_N)] \geq (1-\frac{r}{N})\tilde{L}_D[M(\xi^*)]$

where $D = T^*T/_N$ so we show (4.2). The left-most inequality follows from $\Xi \supset \Xi_N$. The middle inequality follows from the fact that $\tilde{\xi}_N \in \Xi_N$ by definition of ξ_N^*. The proof of the right-most inequality uses the facts given above. Since $\frac{N}{N-r} > 1$

$$\frac{N}{N-r} \tilde{L}_D[M(\tilde{\xi}_N)] \geq \tilde{L}_D(\frac{NM(\tilde{\xi}_N)}{N-r})$$

$$= \tilde{L}_D[\frac{\sum_{i=1}^{n} ([(N-r)\xi^*(x_i)] + \alpha_i)L_{x_i}^* L_{x_i}}{N-r}],$$

where $N\tilde{\xi}_N(x_i) = [(N-r)\xi^*(x_i)] + \alpha_i$. Since $\frac{[(N-r)\xi^*(r_i)]}{N-r} \geq \frac{(N-r)\xi^*(x_i)}{N-r}$ we have

$$\frac{N}{N-r} \tilde{L}_D[M(\tilde{\xi}_N)] \geq \tilde{L}_D[M(\xi^*) + \sum_{i=1}^{r} \gamma_i L_{x_i}^* L_{x_i}]$$

where $\gamma_i = \frac{[(N-r)\xi^*(x_i)]+\alpha_i}{N-r} - \xi^*(x_i) \geq \frac{\alpha_i}{N-r} \geq 0$. Since $\tilde{L}_D[M(\xi^*)+\Sigma\gamma_i L^*_{x_i} L_{x_i}] \geq \tilde{L}_D[M(\xi^*)]$ the proof of theorem is concluded.□

Note that r itself may conceivably depend upon N. We do not investigate this general question here. In the examples below r remains bounded for all N so that by making N sufficiently large the relative loss which comes about by using $\tilde{\xi}_N$ instead of ξ^*_N can be made as small as desired.

5. Examples. The basic strategy in our examples is

a) to find if possible a point $\delta_0 \in \Delta = \{\theta \in \Theta : (\tau,\theta)=1\}$ satisfying $\inf_\Delta \sup_X ||L_x\theta||^2 = \sup_X ||L_x\delta_0||^2$, and

b) to find a probability measure ξ_0 in the collection Ξ of all finitely supported measures satisfying $\int(L_x\theta, L_x\delta_0)d\xi_0(x) \overset{\theta}{\equiv} \sup_X ||L_x\delta_0||^2(\tau,\theta)$ and $S(\xi_0) \subset \{x: ||L_x\delta_0||^2 = \sup_X ||L_x\delta_0||^2\}$. If such can be found then ξ_0 is optimal in the approximate theory and if it is exact then the minimax linear estimator is $[N \sup_X ||L_x\delta_0||^2]^{-1} < Z, m\delta_0 >$.

In our first example we require the notion of a band limited function. The reader is referred to Slepian and Pollack (1961) for details. Denote the space of Lebesgue measurable complex valued functions $f: \mathbb{R}^1 \to \mathbb{C}$ satisfying $\int|f|^p < \infty$ by L^p. There is an isometry $F: L^2 \to L^2$, called the Fourier-Plancherel transform, which satisfies for all $f \in L^1 \cap L^2$, $w \in \mathbb{R}^1$

$$F(f)(w) = \int_{-\infty}^{\infty} f(t)e^{-iwt}dm(t),$$

where m is Lebesgue measure normalized by $(2\pi)^{-\frac{1}{2}}$ (see Rudin 1966)). If $a > 0$ then the space $\mathcal{B}_a$ of band limited functions is that subset of L^2 whose Fourier-Plancherel transforms vanish off of the interval $[-a,a]$.

We denote this set by $\mathcal{B}_a$. We have, defining the operators B_a by $B_a f = fI_{[-a,a]}$,

$$\mathcal{B}_a = \{f \in L^2 : f = (F^{-1}B_a F)f\}.$$

The operators $F^{-1}B_aF$ may be verified to be orthogonal projections onto $\mathcal{B}_a$ which are closed subspaces of L^2. For $f \in L^2$ denote by $f_{\mathcal{B}_a}$ the projection of f onto $\mathcal{B}_a$.

Example 1. For each finite collection of points $\{x_1,\ldots,x_N\}, x_j \in (0,x_0]$ for all j, we may observe the uncorrelated stochastic processes $\{Y(x_i,t): t \in (-\infty,+\infty),\ i=1,\ldots,N\}$ where

$$Y(x,t) = m_x(\theta)(t) + \varepsilon_x(t), \quad t \in (-\infty,+\infty),$$

ε_x is a zero mean stationary process with spectral density $f_x(\lambda) = I_{[-x,x]}(\lambda)$, $\theta \in L^2$ is unknown, and $m_x = F^{-1}B_xF$. Setting $T\theta = \theta_{\mathcal{B}^\perp}$ find the optimal design using Speckman's estimator with $||T\theta|| \le \alpha$ for estimating $(u,\theta) = \int_{-\infty}^{+\infty} u(s)\theta(s)ds$, where $u \in L^1 \cap L^2$ is continuous.

We shall prove that the optimal design places all mass at x_0 regardless of the function u and give the resulting best minimax linear estimator. One should note that in this problem the error process depends upon the design variable x. Our methods still apply if we replace $||m_x(\theta)||_K^2$ by $||m_x(\theta)||_{K_x}^2$ in theorem 3.2.

One can show that $m_x(\theta) \in H(K_x)$ for all $x \in (0,x_0]$. Fixing N we seek $\bar{\theta}$ satisfying

$$\sup_{0<x\le x_0} ||m_x\bar{\theta}||_{K_x}^2 + \frac{1}{N\alpha^2}||T\bar{\theta}||_2^2 =$$

(5.1)

$$\inf_{\{\theta:(u,\theta)=1\}} \sup_{0<x\le x_0} ||m_x\theta||_{K_x}^2 + \frac{1}{N\alpha^2}||T\theta||^2$$

and claim that

(5.2) $$\bar{\theta} = \frac{u_{\beta}+N\alpha^2 u_{\beta}^{\perp}}{||u_{\beta}||^2+N\alpha^2||u_{\beta}^{\perp}||^2}.$$

In order to see this introduce

$$\Delta_1 = \{\theta: ||u_{\beta}||\ ||\theta_{\beta}|| + ||u_{\beta}^{\perp}||\ ||\theta_{\beta}^{\perp}|| \geq 1\}$$

Then $\Delta_0 = \{\theta: (u,\theta) = 1\} \subset \Delta_1$ and we show that $\bar{\theta} \in \Delta_1$ minimizes

$$[\sup_x ||m_x\theta||^2 + \frac{1}{N\alpha^2}||T\theta||^2] .$$

Since $\bar{\theta}$ is also in Δ_0 it satisfies (5.1). Since $||m_x\theta||_{K_x}^2 = ||B_x F\theta||_2^2$ $= \int_{-x}^{x} |F\theta(t)|^2 dm(t)$ we have $\sup_{0<x\leq x_0} ||m_x\theta||_{K_x}^2 = ||B_{x_0}F\theta||_2^2 = ||F^{-1}B_{x_0}F\theta||_2^2$. Thus the problem for $\theta \in \Delta_1$ becomes minimize $A^2 + \eta B^2$ subject to $aA + bB \geq 1$, where $a = ||u_{\beta}||, b = ||u_{\beta}^{\perp}||$, and $\eta = (N\alpha^2)^{-1}$. The solution, using calculus methods, is $A_0 = a/2\gamma$, $B_0 = b/2\eta\gamma$, and $\gamma = \frac{a^2}{2} + \frac{b^2}{2\eta}$. Choosing $\bar{\theta}$ as in (5.2) we see that indeed these norm conditions are met. Obviously $\bar{\theta} \in \Delta_0 \cap \Delta_1$ proving the claim in (5.1) that $\bar{\theta}$ is the minimizer.

Let ξ^* place all mass at $x = x_0$. One can verify

(5.3) $$m_{x_0}^* m_{x_0}\bar{\theta} + \frac{1}{N\alpha^2} T^*T\bar{\theta} = (||u_{\beta}||^2 + N\alpha^2||u_{\beta}^{\perp}||^2)^{-1}u$$

by taking the inner product with an arbitrary θ on both sides. Therefore ξ^* is optimal. By (5.3) $\tilde{\theta} = N^{-1}(||u_{\beta}||^2+N\alpha^2||u_{\beta}^{\perp}||^2)\bar{\theta} = N^{-1}u_{\beta} + \alpha^2 u_{\beta}^{\perp}$ and $m_{x_0}(\tilde{\theta}) =$ $N^{-1}u_{\beta}$. The corresponding estimator is $N^{-1} < Y(x_0), u_{\beta} >$. Since $u_{\beta}(\cdot)$ $= \int_{-\infty}^{+\infty} u(t)K_x(\cdot,t)dm(t)$, the minimax linear estimator of (u,θ) is

$\frac{1}{N} \sum_{j=1}^{N} \int_{-\infty}^{\infty} Y_j(x_0,t)u(t)dm(t)$. This estimator has a maximum mean square error of $\frac{||u_{\mathcal{B}}||^2}{N} + \alpha^2||u_{\mathcal{B}}^{\perp}||^2$ over θ with $||\theta_{\mathcal{B}}^{\perp}|| \leq \alpha$. If $u \in \mathcal{B}_{x_1}$ where $0 < x_1 < x_0$ then one can verify that any design whose support is contained in $[x_1,x_0]$ will be optimal.

The following result is used below. Suppose that $f_1,\ldots,f_m$ is a T-system on $[a,c]$ and $b \in (a,c)$ as above. In general the subspace $\{\sum_{i=1}^{m} \alpha_j f_j : \alpha' d=0\}$, where d is a fixed arbitrary vector, does not have the Chebyshev property.

Lemma 5.1. The collection of functions $\mathfrak{F} = \{\sum_{i=1}^{m} \beta_j f_j : \beta' f(c)=0\}$ has the Chebyshev property on $[a,b]$. In particular at least one of $f_j(c)$, $j \in \{1,\ldots,m\}$, say $f_1(c)$, must be non-zero and the system $\{g_i\}_{i=2}^{m}$ is a T-system for $\mathfrak{F}$, where

$$(5.4) \qquad g_i(x) = f_i(x) - \frac{f_i(c)}{f_1(c)} f_1(x).$$

Proof. It is easily checked that the span of $\{g_i\}_{i=2}^{m}$ is $\mathfrak{F}$ if $f_1(c) \neq 0$. Let $a \leq \tau_1 < \tau_2 < \ldots < \tau_{m-1} \leq b$ and form the determinants

$$D = \begin{vmatrix} g_2(\tau_1)\ldots g_2(\tau_{m-1}) \\ \vdots \\ g_m(\tau_1)\ldots g_m(\tau_{m-1}) \end{vmatrix} \quad \text{and} \quad A = \begin{vmatrix} f_1(\tau_1)\ldots f_1(c) \\ \vdots \\ f_m(\tau_1)\ldots f_m(c) \end{vmatrix} .$$

Since $A = (-1)^m f_1(c)D$ and the determinants A do not change sign the determinants D must not change sign. □

Example 2. We assume in this, as in all the examples except the first, that the observable random variables are scalar valued. The mean function is assumed to be a polynomial of degree m, $m_x\theta = \sum_{j=0}^{m} \theta_j x^j$ for $x \in [a,c]$.

Writing $f_j(x) = x^j$ and $f(x)$ for the column vector of functions f_j we write $m_x\theta = \theta'f(x)$. Our objective is the estimation of the value $\theta'f(c)$ based upon observations in the interval $[a,b]$ where $b \in (a,c)$ is fixed and under the assumption that $\theta_{k+1},\ldots,\theta_m$ are small. Specifically we use Speckman's minimax estimator with $\Theta = \mathbb{R}^{m+1}$, $\mathcal{H} = \mathbb{R}^{m-k}$, and $T\theta = \varepsilon^{-1}(\theta_{k+1},\ldots,\theta_m)'$. We fix a number N larger than $k+1$ and minimize $d_{TN^{-Y2}}(\tau,\xi)$ over ξ in the collection Ξ of all finitely supported probability measures on $[a,b]$ where τ is the $m+1$ vector $f(c)$.

We prove that there are points $a \leq x_0^* < x_1^* < \ldots < x_k^* \leq b$ which constitute the support of the optimal design ξ^* and that $\xi^*(x_i^*) = |\phi_{x_i^*}(c)|/ \sum_{i=0}^{k} |\phi_{x_i^*}(c)|$ for $i=0,\ldots,k$, where $\phi_{x_i^*}(x)$ is the Lagrange interpolation polynomial of degree k at x_i^*. We do not prove here the fact that the optimal estimator of $\theta'f(c)$ is the value at c of the kth degree polynimial which passes through the $k+1$ points $\{x_i^*,\bar{y}(x_i^*)\}_{i=0}^{k}$ but mention that its proof is a consequence of lemma 2.1 and parallels that of theorem 6.3 below.

One can easily verify that the conditions of theorem 3.3 are satisfied. Thus a design ξ_0 and a point $\delta_0 \in \mathbb{R}^{m+1}$ exist satisfying i) - iv) of theorem 3.2 where $L_x\theta = (\theta'f(x), \frac{1}{\varepsilon\sqrt{N}} T\theta) \in \mathbb{R}^1 \times \mathbb{R}^{m-k}$. From lemma 5.1 it follows that the collection of functions $\{\theta_0+\theta_1x+\ldots+\theta_kx^k: \sum_{j=0}^{k} \theta_jc^j=0\}$ has the Chebyshev property on $[a,b]$. Using an argument like that in theorem 6.1 one can show that $\delta_0'f(x)$ equioscillates on $[a,b]$ at $k+1$ points $a \leq x_0^* < \ldots < x_k^* \leq b$ and is the unique function satisfying iv) of theorem 3.2 above. Furthermore these points must be isolated or $\delta_0'f(x)$ is a constant violating the equioscillation. Thus the support of ξ_0 is discrete. By ii) of theorem 3.2 we know that for all $\theta \in \mathbb{R}^m$

$$\sum_{j=0}^{k} \xi_0(x_j^*)\theta' f(x_j^*)\delta_0' f(x_j^*) + \frac{1}{N\alpha^2} \sum_{j=k+1}^{m} \theta_j \delta_{0j} = \alpha\theta' f(c)$$

for some $\alpha > 0$. Setting θ to satisfy $\theta_j = 0$ for $j = k+1,\dots,m$ and $\sum_{j=0}^{k} \theta_j f_j(x) = \phi_{x_i^*}(x)$ we have

$$\xi_0(x_i^*)|\delta_0' f(x_i^*)|\,\mathrm{sign}(\delta_0' f(x_i^*)) = \alpha\phi_{x_i^*}(c) > 0. \tag{5.5}$$

Therefore sign $\delta_0' f(x_i^*) = (-1)^{i-k}$ and since (5.5) holds for $i=0,\dots,k$, $\xi_0(x_i^*) \equiv \xi^*(x_i^*)$.

An argument like that in theorem 6.2 also proves that the design is unique if $m = k + 1$. An obvious question remains. Are these designs precisely the same as the usual unbiased extrapolation designs when the mean function is $\sum_{j=0}^{k} \theta_j x^j$? We do not know the general answer but we can show in the case $k=1$ and $m=2$ that the answer is dependent upon the value of $N\varepsilon^2$. A detailed examination in the latter case of the general solution $\delta_0' f(x)$ which may be shown to be, setting $\eta = (N\varepsilon^2)^{-1}$,

$$\delta_0' f(x) = (s^2+\eta z^2)^{-1}\Big[z\eta \sum_{j=0}^{k} \phi_{x_j}(x)(-1)^{j-m} + \sum_{j=k+1}^{m} \alpha_j\Big(x^j - \sum_{i=0}^{k} \phi_{x_i}(x)x_i^j\Big)\Big],$$

where $\alpha_j = c^j - \sum_{i=0}^{k} \phi_{x_i}(c)x_i^j$, $s^2 = \sum_{j=k+1}^{m} \alpha_j^2$, and $z = \sum_{i=0}^{k} |\phi_{x_i}(c)|$, reveals that if $N\varepsilon^2 \le \eta_0^{-1} = \left\{\left[\frac{4(c-b)}{(b-a)^3} + \frac{2}{(b-a)^2}\right]^{-1}(c-b)(c-a)\right\}^{-1}$ then the designs coincide with the usual and if $N\varepsilon^2 > \eta_0^{-1}$ is sufficiently close to η_0^{-1} the design is supported on $\{x_1, b\}$ where x_1 is the unique solution in (a,b) to

$$\frac{4\eta(c-b)}{(b-x_1)^3} + \frac{2\eta}{(b-x_1)^2} = (c-b)(c-x_1).$$

6. Robust nonparametric extrapolation. This section is devoted in its entirety to the proof of the assertions in the introduction concerning the solutions to the extremal problems P_η and statistical problems which we shall here abstract and call S_η. More specifically let Ξ be the collection of all probability measures on $[a,b]$, $-\infty < a < b < +\infty$, with the Borel subsets as the σ-field. Denote by D^m the mth derivative operator defined on the Sobolev space $W_m^2[a,c]$ of all functions whose $m-1^{\underline{st}}$ derivatives are absolutely continuous with respect to Lebesgue measure and whose $m^{\underline{th}}$ derivatives are in $L^2[a,c]$. Let $\eta > 0$ be arbitrary and define the statistical problems

S_η: Find $\xi_0 \in \Xi$ if such exists minimizing $d_{\sqrt{\eta}\,D^m}(e_c,\xi)$.

Here $(e_c,\theta) = \theta(c)$ is the evaluation functional on $W_m^2[a,c]$ and is a bounded linear functional, $L_x\theta = (\theta(x),\sqrt{\eta}\,\theta^{(m)}) \in \mathbb{R}^1 x L^2[a,c]$, $\forall\, x \in [a,b]$, where $\theta^{(m)} = D^m\theta$, and

$$d_{\sqrt{\eta}\,D^m}(\tau,\xi) = \sup_{\theta\in N} \frac{[\theta(c)]^2}{\int ||L_x\theta||_V^2 d\xi(x)} ,$$

where $N = \{\theta: \int ||L_x\theta||_V^2 d\xi(x) > 0\}$. The integral $\int ||L_x\theta||_V^2 d\xi(x)$ is well defined as is $\int L^*_x\phi(x)d\xi(x)$ (a Bochner integral) as a consequence of assumptions (A1) and (A2) which hold in the present case. Let N and $\varepsilon > 0$ be given and set $N\varepsilon^2 = \eta^{-1}$. When the observable random variables are $Y(x) = \theta(x) + \varepsilon$, $E(\varepsilon) = 0$, $E(\varepsilon^2) = 1$, for $x \in [a,c]$ where $\theta \in W_m^2[a,c]$ is unknown and when $\theta(c)$ is to be estimated based only upon x's in $[a,b]$ using Speckman's minimax estimator with $||\theta^{(m)}||_2^2 = \int_a^c (\theta^{(m)}_{(t)})^2 dt \le \varepsilon^2$ then the solution to S_η, if it concentrates on $r < N$ points may be used as described in section 4 and in more detail prefatory to theorem 6.3 to construct an experiment using N observations which is within $\frac{r}{N-r}$ of the optimal experiment using that number of observations. In this sense S_η is a statistical problem.

The extremal problems are

P_η: Minimize over $\theta \in W_m^2[a,c]$ satisfying $\theta(c) = 1$ the functional

$$\rho(\theta) = ||\theta||_\infty^2 + ||\theta^{(m)}||_2^2 .$$

(Recall the abuse of notation above, $||\cdot||_\infty$ is restricted to the subinterval $[a,b]$ while $||\cdot||_2$ is on $[a,c]$.) The problems P_η are not standard since ρ is not differentiable (see Spruill (1982)). Nevertheless the next two theorems reveal a great deal about the problems P_η and their solutions. The first theorem applies more generally. Let $D^m + a_{m-1}D^{m-1} + \dots + a_0$, where a_j is j times continuously differentiable, $G(x,s)$ satisfy $LG(\cdot,s) = 0$ on $[s,c]$ subject to $\frac{d^i}{dx^i} G(x,s)\Big|_{x=s} = \delta_{i,m-1}$ (the Kronecker delta) $G(x,s) = 0$ for $x < s$ and $\{\psi_{x_i}\}_{i=1}^{m}$ be the Lagrange interpolation polynomials for the null space of L on $[a,c]$ to the points $\{x_1,\dots,x_m\}$. Define

$$h_x(s) = G(x,s) - \sum_{i=1}^{m} \psi_{x_i}(x)G(x_i,s).$$

For $\eta > 0$ the problem $\tilde{P}_\eta$ is

$\tilde{P}_\eta$: Minimize

$$\tilde{\rho}(\theta) = ||\theta||_\infty^2 + \eta||L\theta||_2^2$$

over θ in $W_m^2[a,c]$ satisfying $\theta(c) = 1$.

<u>Theorem 6.1</u>. The problem $\tilde{P}_\eta$ has a solution $\theta_0 \in W_m^2[a,c]$. If the null space of L has the Chebyshev property then there exist points $a \le x_1 < x_2$

$< \dots < x_m \le b$ and $q=0$ or 1 such that

$$\theta_0(x_i) = (-1)^{i+q} \sup_{a\le x\le b} |\theta_0(x)|$$

and the solution is unique if $\eta > 0$.

Proof. If $\eta = 0$ clearly $\tilde{\rho}_\eta(\theta_0) = 0$ and there are many solutions. If $\eta > 0$ let $\theta_n \in W_m^2[a,c]$ be such that $\tilde{\rho}_\eta(\theta_n) \downarrow \inf \tilde{\rho}_\eta(\theta)$. The sequence $L\theta_n$ satisfies $\eta||L\theta_n||_2^2 \le \tilde{\rho}_\eta(\theta_1)$. Since $||h||_2 \le k$ is weakly sequentially compact in $L_2[a,c]$ there is an element $u \in L_2[a,c]$ and a subsequence n' such that $L\theta_{n'} \xrightarrow{w} u$. We may write, for $a \le \tau_1 < \tau_2 < \dots < \tau_m \le b$,

$$\theta_{n'}(x) = \sum_{i=1}^{m} \theta_{n'}(\tau_i)\psi_{\tau_i}(x) + \int_a^c h_x(s)L\theta_{n'}(s)ds.$$

The integrals $\int_a^c h_x(s)L\theta_{n'}(s)ds$ converge to $\int_a^c h_x(s)u(s)ds$ for all $x \in [a,c]$. Since $\sup_{a\le x\le b} |\theta_{n'}(x)| \le \tilde{\rho}_\eta(\theta_1)$ all of the m sequences $\{\theta_{n'}(\tau_i)\}_{i=1}^{m}$ are bounded. Appealing again to sequential compactness of $\mathbb{R}^1$ we may assume that the sequence n' has been chosen to satisfy $\theta_{n'}(\tau_i) \to \alpha_i$ also. Define the function $\theta_0 \in W_m^2[a,c]$ by

$$\theta_0(x) = \sum_{i=1}^{m} \alpha_i\psi_{\tau_i}(x) + \int_a^c h_x(s)u(s)ds.$$

Since $||u||_2 \le \underline{\lim}||L\theta_{n'}||_2$ it is clear that $\tilde{\rho}_\eta(\theta_0) \le \inf \tilde{\rho}_\eta(\tau)$. Since $\theta_n(x) \to \theta_0(x)$ for all x, $\theta_0(c) = 1$. The first assertion of the theorem has been proven.

For the following arguments we shall employ the notation of Karlin and Studden (1966a) in counting the zeros of a continuous function on $[a,b]$.

If $x_0 \in (a,b)$ is an isolated zero of f and f does not change sign at x_0 then x_0 is termed a non-nodal zero. All other zeros, including zeros at the endpoints are nodal zeros. For any such function $\tilde{Z}(f)$ is the number of zeros in [a,b] counting one for each nodal zero and two for each non-nodal zero. Suppose that g is a continuous function on [a,b] and there are points $a \leq x_1 < x_2 < \dots < x_m \leq b$ and $q \in \{0,1\}$ such that $g(x_i) = (-1)^{i+q}||g||_\infty$, $i=1,\dots,m$. If there exists a point $x_0 \in \{x_1,\dots,x_m\}$ and a continuous function h such that $h(x_0) = g(x_0)$ then $||h||_\infty \leq ||g||_\infty$ entails $\tilde{Z}(g-h) \geq m - 1$. Consider the collection of functions on $[a,c]$ $\{\sum_{j=1}^{m} \alpha_j f_j \colon \alpha' f(c)=0\}$ where $f_1,\dots,f_m$ is a T system for the null space of L. Their restriction to [a,b] is spanned by the Chebyshev system $g_2,\dots,g_m$ which may be defined, if $f_1(c) \neq 0$, except for the sign of one of them, from (5.4). By Bernstein's theorem (see Karlin and Studden (1966) there are constants $\beta_1,\dots,\beta_m$ such that $\beta' f(c) = 0$, and the function $\sum_{j=1}^{m} \beta_j f_j = g_0$ is the minimax approximant to θ_0 on [a,b]. Therefore there exist m points $a \leq x_1 < \dots < x_m \leq b$ and a $q \in \{0,1\}$ such that

(6.1) $$(g_0(x_i)-\theta_0(x_i))(-1)^{i+q} = ||g_0-\theta_0||_\infty \quad ,$$

where $||g_0-\theta_0||_\infty = \sup_{a \leq x \leq b} |g_0(x)-\theta_0(x)|$. Also

$$||g_0-\theta_0||_\infty = \inf\{||\Sigma\beta_j f_j-\theta_0||_\infty \colon \beta' g(c)=0\} .$$

Since $Lg_0 = 0$ on [a,c] and $\beta' f(c) = 0$ entails $(\theta_0-g_0)(c) = 1$ we must have $||\theta_0||_\infty = ||\theta_0-g_0||_\infty$. Furthermore there must be a point $x \in [a,b]$ at which $\theta_0(x) = \theta_0(x) - g_0(x) = \pm||\theta_0||_\infty$. We conclude $\tilde{Z}(\theta_0-(\theta_0-g_0)) \geq m-1$ on [a,b]. However if $\theta_0 - (\theta_0-g_0) = g_0$ is a non-trivial polynomial in the system $g_2,\ldots,g_m$ we must have by Theorem 4.2 of Karlin and Studden (1966a) $\tilde{Z}(g_0) \leq m-2$. We conclude that g_0, the best approximant from $\mathfrak{F}$ on [a,b] is zero. Consequently θ_0 itself equioscillates in the sense of (6.1).

We now verify the uniqueness of the solution when the null space of L is spanned by a T-system $\{f_1,\ldots,f_m\}$. Because the norm on the Hilbert space $L_2[a,c]$ is strictly convex, if h_0 and h_1 are two solutions to the $\tilde{P}_\eta$ problem then $\tilde{\rho}_\eta(\alpha h_0+(1-\alpha)h_1) < \tilde{\rho}_\eta(h_0)$ unless $kLh_1 = Lh_0$, a.e. on [a,c] for some constant k. Since $||\alpha h_0+(1-\alpha)h_1||_\infty^2 \leq \alpha||h_0||_\infty^2 + (1-\alpha)||h_1||_\infty^2$, where $||h||_\infty = \sup_{a\leq x\leq b} |h(x)|$, we have $\tilde{\rho}_\eta(\alpha h_0+(1-\alpha)h_1) \leq B(\alpha)$ where the function B is defined by

$$B(\alpha) = \alpha||h_0||_\infty^2 + (1-\alpha)||h_1||_\infty^2 + \eta(k+\alpha(1-k))^2||Lh_0||_2^2 .$$

Note that $B(0) = B(1) = \tilde{\rho}_\eta(h_0)$ and $B''(\alpha) = 2\eta(1-k)^2||Lh_0||_2^2$. Thus, whether $||Lh_0|| = 0$ or $||Lh_0|| > 0$ we must have $||h_0||_\infty = ||h_1||_\infty$. Again it must be the case that h_0 and h_1 share a common extreme value at one of their points of equioscillation. Therefore $\tilde{Z}(h_0-h_1) \geq m-1$. If $Lh_0 = 0$ then $h_0 - h_1 \in \mathfrak{F}$ and of necessity $\tilde{Z} \leq m-2$ unless $h_0 = h_1$. If $||Lh_0|| > 0$ then $k=1$ so that again, $h_0 - h_1$ is in $\mathfrak{F} = \{\Sigma\beta_j f_j : \beta' f(c)=0\}$ and $\tilde{Z} \leq m-2$. We conclude that in any case $h_0 = h_1$ proving that the solution is unique. □

When $L = D^m$ the Lagrange interpolation polynomials which we now write as $\{\phi_{x_i}\}_{i=1}^m$ are explicitly $\phi_{x_i}(x) = \prod_{j\neq i}(x-x_j)/\prod_{j\neq i}(x_i-x_j)$ and

$$G(x,s) = \frac{(x-s)_+^{m-1}}{(m-1)!} .$$

One can prove that if ξ is a probability measure on [a,b] then the operator $L_\xi: \Theta \to L^2(\xi)$ defined by $L_\xi\theta = (\theta, \sqrt{n}\,\theta^{(m)}) \in L^2(\xi)$ is bounded and has closed range and hence satisfies assumption (A3).

Lemma 6.1. Conditions (A4) and (A5) are satisfied.

Proof. It suffices to prove the existence of a constant $k > 0$ satisfying $\sup_X ||L_X\theta|| \geq k||\theta||$ for all θ in $W_m^2[a,c]$ because of theorem 3.4. Suppose that no such constant can be found. Then there is a sequence $\{\theta_n\}$ satisfying for all $n \geq 1$ $||\theta_n||^2 \geq n[||\theta_n||_\infty^2 + n||\theta_n^{(m)}||^2]$. Since $\frac{\theta_n}{||\theta_n||}$ satisfies the same inequality we may and do assume that $||\theta_n|| \equiv 1$.

Thus $||\theta_n||_\infty$ and $||\theta_n^{(m)}||_2$ both tend to zero as n becomes large. Write for $t_0 \in [a,c]$ arbitrary

$$(6.2) \qquad \theta_n(t) = \sum_{j=1}^{m} \theta_n^{(j)}(t_0) \frac{(t-t_0)^j}{j!} + \int_a^c \frac{(t-s)_+^{m-1}}{(m-1)!} \theta_n^{(m)}(s)ds.$$

Equation (6.2) shows that $||\theta_n^{(j)}||_\infty \to \infty$ for $j=0,1,\ldots,m-1$. Since

$$|\theta_n^{(j)}(t)| \leq ||\theta_n^{(j)}||_\infty + \sqrt{b-a}\, ||\theta_n^{(j+1)}||_2$$

we have

$$(6.3) \qquad ||\theta_n^{(j)}||_2^2 \leq K'[||\theta_n^{(j)}||_\infty^2 + ||\theta_n^{(j+1)}||_2^2]$$

so that starting with $j = m - 1$ in (6.3) and proceeding to successively smaller values we find that $\lim_{n\to\infty} ||\theta_n|| = 0$. This contradiction establishes the lemma. □

Lemma 6.2. If the probability measures ξ_n in Ξ converge weakly to $\xi \in \Xi$ then $||M(\xi_n)-M(\xi)|| \to 0$ as $n \to \infty$.

Proof. Consider for $||\theta|| = 1$

$$|(\theta,[M(\xi_n)-M(\xi)]\theta)| = |\int_a^b ||L_x\theta||^2 d\xi_n(x) - \int_a^b ||L_x\theta||^2 d\xi(x)|.$$

Since $M(\xi_n) - M(\xi)$ is self adjoint $||M(\xi_n)-M(\xi)|| = \sup_{||\theta||=1} |(\theta,[M(\xi_n)-M(\xi)]\theta)|$. Suppose that $||M(\xi_n)-M(\xi)||$ does not converge to zero as $n \to \infty$. Then there is a subsequence $\{\xi_{n_k}\}_{k\geq 1}$ of measures and a sequence $\{\theta_k\}_{k\geq 1}$ and an $\varepsilon > 0$ such that $|(\theta_k,[M(\xi_{n_k})-M(\xi)]\theta_k)| \geq \varepsilon$ for all k. Since the functions $||L_x\theta_k||$ of x are uniformly bounded and equicontinuous on [a,b] there is a subsequence k' and a continuous function h such that $||L_x\theta_{k'}||^2$ converges to h(x) uniformly on [a,b]. By theorem 5.5 of Billingsley (1968) $\lim_{k'\to\infty} \int_a^b ||L_x\theta_{k'}||^2 d\xi_{n_{k'}} = \int_a^b h(x)d\xi(x)$. We also have

$\lim_{k'\to\infty} \int_a^b ||L_x\theta_{k'}||^2 d\xi(x) = \int h(x)d\xi(x)$. Thus $\lim_{k'\to\infty} |(\theta_{k'},[M(\xi_{n_{k'}})-M(\xi)]\theta_{k'})|$ $= 0$. The contradiction establishes the lemma. □

If $\xi_n \xrightarrow{w} \bar{\xi}$ and the supports of all contain more than m - 1 points then $\lim_{n\to\infty} d(\theta,\xi_n) = d(\theta,\bar{\xi})$ for if $S(\xi)$ has more than m-1 points of support then $M^{-1}(\xi)$ exists and is bounded and

$$||M^{-1}(\xi_n)-M^{-1}(\bar{\xi})|| \leq \frac{||M^{-1}(\bar{\xi})||^2||M(\bar{\xi})-M(\xi_n)||}{1-||M^{-1}(\bar{\xi})||\ ||M(\bar{\xi})-M(\xi_n)||}$$

whenever $||M(\bar{\xi})-M(\xi_n)|| < ||M^{-1}(\bar{\xi})||^{-1}$. Lemma 6.2 and

$$|d(\tau,\xi_n)-d(\tau,\bar{\xi})| \leq ||\tau||^2||M^{-1}(\xi_n)-M^{-1}(\bar{\xi})||$$

show that $\lim_{n\to\infty} d(\theta,\xi_n) = d(\theta,\bar{\xi})$.

Theorem 6.2. Let $\eta > 0$ be arbitrary.

i) There is a unique solution ξ_0 to S_η. Its support $S(\xi_0)$ consists of m distinct points $\{x_1,\ldots,x_m\}$ in [a,b] and

(6.4) $$\xi_0(x_j) = |\phi_{x_j}(c)| / \sum_{i=1}^{m} |\phi_{x_i}(c)|$$

for j=1,...,m.

ii) There is a unique solution δ_0 to P_η. It is given by the formula

(6.5) $$\delta_0(x) = (s^2+\eta z^2)^{-1}[\eta z \sum_{i=1}^{m} (-1)^{i-m}\phi_{x_i}(x) + \int_a^c h_x(s)h_c(s)ds],$$

where

$$h_x(s) = \frac{(x-s)_+^{m-1}}{(m-1)!} - \sum_{i=1}^{m} \phi_{x_i}(x) \frac{(x_i-s)_+^{m-1}}{(m-1)!},$$

$$s^2 = ||h_c||_2^2, \text{ and } z = \sum_{i=1}^{m} |\phi_{x_i}(c)| .$$

iii) Parts i) - iv) of theorem 3.2 are satisfied by ξ_0 and δ_0.

Proof. As we have indicated assumptions (A1) through (A5) hold. We know that if a sequence of probability measures ξ_n converges weakly to a probability measure ξ and all are supported on more than m-1 points then $d(\tau,\xi_n)$ converges to $d(\tau,\xi)$. This entails the existence of a probability measure ξ^* which minimizes $d(e_c,\xi)$ over $\xi \in \Xi$ for let ξ_n satisfy $d(e_c,\xi_n) \to \inf_\Xi d(e_c,\xi)$. Then there is a probability measure ξ^* and subsequence $\xi_{n'}$ satisfying $\xi_{n'} \xrightarrow{W} \xi^*$. We know that $\inf_\Xi d(e_c,\xi) < \infty$ because the design ξ' which places equal mass at any m distinct points in [a,b] satisfies $d(e_c,\xi) = (e_c, M^{-1}(\xi')e_c)$. If a design $\xi \in \Xi$ is supported on fewer than m points then $d(e_c,\xi) = +\infty$ as can be seen from equation (2.1) by choosing an arbitrary point

$y \in [a,b] - S(\xi)$ and selecting a sequence θ_n converging to ϕ_y. Hence we conclude that eventually the supports $S(\xi_{n'})$ all contain at least m points. On the other hand, if $S(\xi^*)$ contains fewer than m points then choosing $z \in [a,b] - S(\xi^*)$ yields

$$\lim_{n\to\infty} \frac{|\phi_z(c)|^2}{(\phi_z, M(\xi_{n'})\phi_z)} = +\infty .$$

Since $d(e_c,\xi_{n'}) \geq \frac{|\phi_z(c)|^2}{(\phi_z, M(\xi_{n'})\phi_z)}$ this contradicts the convergence of $d(e_c,\xi_{n'})$ to $\inf_{\Xi} d(e_c,\xi)$. Clearly $d(e_c,\xi^*) = \inf_{\Xi} d(e_c,\xi)$.

Now we may apply theorem 3.2 to conclude that an element δ_0 of $W_m^2[a,c]$ may be found satisfying i) through iv) of theorem 3.2. The function δ_0 is the solution of P_n and, as has been shown, δ_0 is the unique solution of P_n. It was also shown that there are points $a \leq x_1^* < x_2^* < \ldots < x_m^* \leq b$ and a $q = 0$ or 1 such that $\delta_0(x_j^*) = (-1)^{j+q}||\delta_0||_\infty$ for $j=1, \ldots, m$. The points $x^*_2,\ldots,x_m^*$ all have the property that $|\delta_0(x)| < ||\delta_0||$ for all x in some deleted neighborhood of each of the given points for suppose not. Then $\delta_0(x) \equiv (-1)^{j_0+q}||\delta_0||_\infty$ for some interval containing $x_{j_0}^*$. Thus $\delta_0^{(i)}(x_{j_0}^*) = 0$ for $i=1,2,\ldots,m$.

Consider the function $\delta_1(x) = \begin{cases} \delta_0(x) & x \geq x_{j_0}^* \\ (-1)^{j_0+q}||\delta_0||_\infty & a \leq x \leq x_{j_0}^* \end{cases}$. Clearly $\rho(\delta_1) \leq \rho(\delta_0)$. The unicity of the minimizer of ρ then shows that $\delta_1 = \delta_0$ and contradicts the fact that δ_0 must oscillate m times. Let

$$x_1 = \sup\{t: \delta_0(x)=\delta_0(a) \text{ for } x\in [a,t]\} > 0.$$

By using ii) of theorem 3.2 one can prove that ξ_0 puts no mass on $[a,x_1)$ if $a < x_1$. Furthermore, by using again the relationship ii) of theorem 3.2, choosing θ in succession to be ϕ_{x_j}, $j=1,\ldots,m$, one concludes that the sign of $\delta_0(x_j)$ is the same as that of $\phi_{x_j}(c)$, namely $(-1)^{j-m}$, and that

(6.6) $$\xi_0(x_j) = |\phi_{x_j}(c)| / \sum_{j=1}^{m} |\phi_{x_j}(c)| > 0$$

for $j=1,\ldots,m$. The unicity of ξ_0 will be verified following our demonstration that $\delta_0^{(m)}$ must be proportional to h_c. Let $\xi_n \in \Xi_n$ satisfy $\xi_n(x_j) = \frac{p_n(j)}{n}$ where $\{p_n(j)\}_{j=1}^{m}$ are positive integers for each n, $\sum_{j=1}^{m} p_n(j) \equiv n$, and $\lim_{n\to\infty} \frac{p_n(j)}{n} = \xi_0(x_j)$. Consider the estimation problem, using Speckman's minimax linear estimator based upon n uncorrelated observations $\{Y_1(x_1),\ldots,Y_{p_n(1)}(x_1),\ldots,Y_{p_n(m)}(x_m)\}$ where $E(Y(x)) = \theta(x)$ and $T = \sqrt{n\eta}\, D^m$. We know that the minimax mean square error is

$$\inf_g \sup_{||T\theta||\leq 1} V_{\xi_n}(\tau,\theta,g) = n^{-1} d_{\sqrt{\eta}\, D^m}(e_c,\xi_n)$$

$$= n^{-1} d(e_c,\xi_n)$$

and that $g_n = m(\tilde{\theta}_n)$ yields the best estimator where $\tilde{\theta}_n$ is any solution to $\left(\sum_{j=1}^{m} m_{x_j}^* m_{x_j} + n\eta(D^m)^* D^m\right)\tilde{\theta}_n = e_c$. Setting $M(\xi_n) = \int_a^b L_x^* L_x \, d\xi_n(x)$, and noting that whenever the support of a measure ξ contains m or more points $M^{-1}(\xi)$ exists we have $\tilde{\theta}_n = M^{-1}(\xi_n) \frac{e_c}{n}$. Now write

$$\inf_g V_{\xi_n}(\tau,\theta,g) = V_{\xi_n}(\tau,\theta,g_n) = E_\theta[<Z,g_n>_B - \theta(c)]^2$$

$$= E\Big(\sum_{j=1}^{m}\sum_{i=1}^{p_n(j)} Y_i(x_j)\tilde{\theta}_n(x_j)-\theta(c)\Big)^2$$

(6.7) $$= \sum_{j=1}^{m} p_n(j)\tilde{\theta}_n^2(x_j) + \Big(\sum_{j=1}^{m} \tilde{\theta}_n(x_j)p_n(j)\theta(x_j)-\theta(c)\Big)^2.$$

Using the representation

$$\theta(x) = \sum_{j=1}^{m} \phi_{x_j}(x)\theta(x_j) + \int_a^c h_x(s)\theta^{(m)}(s)ds$$

in (6.7) we see that $p_n(j)\tilde{\theta}_n(x_j) \equiv \phi_{x_j}(c)$ and that $\sup\limits_{||T\theta||\leq 1} \inf\limits_g V_{\xi_n}(\tau,\theta,g)$ occurs for θ_n satisfying $\theta_n^{(m)} = \gamma_n h_c$ where $\gamma_n = (\sqrt{n\eta}\ ||h_c||)^{-1}$. According to the proof of theorem 2.1 if $\sqrt{n\eta}\ ||\theta_n^{(m)}||_2 > 0$ then $\theta_n = \frac{1}{\sqrt{n\eta}} \frac{\tilde{\theta}_n}{||\tilde{\theta}_n^{(m)}||_2}$.

We claim that for n sufficiently large we must have $||\tilde{\theta}_n^{(m)}||_2 > 0$. The reason is that $\tilde{\theta}_n = M^{-1}(\xi_n)\frac{e_c}{n}$ so that $n\tilde{\theta}_n$ converges in $W_m^2[a,c]$ to $M^{-1}(\xi_0)e_c = \frac{1}{\alpha}\delta_0$. Since $||\delta_0^{(m)}||_2 > 0$ the conclusion follows. We therefore have, for n sufficiently large, $\tilde{\theta}_n^{(m)} = k_n h_c$ for the constants $k_n = \sqrt{n\eta}\ ||\tilde{\theta}_n^{(m)}||_2\gamma_n$ hence $n\tilde{\theta}_n^{(m)} = nk_n h_c$ converges in $L^2[a,c]$ to $\frac{1}{\alpha}\delta_0^{(m)}$. Clearly we must have $\lim\limits_{n\to\infty} nk_n = k$ where $|k| < \infty$ and we conclude that $\delta_0^{(m)}$ is indeed proportional to h_c.

Now one can again employ ii) to prove that the constants given in the asserted form of δ_0 are as claimed.

The design ξ_0 must be unique for suppose otherwise. Let ξ_1 solve S_n. Let $S(\xi_0) = \{x_1,\ldots,x_m\}$ and $S(\xi_1) = \{\bar{x}_1,\ldots,\bar{x}_m\}$ and let $j_0 = \max\{j: x_j \neq \bar{x}_j\}$. Clearly, because the same supports imply the same weights, j_0 is well defined. Both $S(\xi_0)$ and $S(\xi_1)$ determine via (6.6) the same function δ_0 because of the unicity of the solution to P_η. Thus on $s \in [a,c]$

$$0 \equiv \sum_{j=1}^{m} [\phi_{x_j}(c)(x_j-s)_+^{m-1} - \phi_{\bar{x}_j}(c)(\bar{x}_j-s)^{m-1}].$$

Assume without loss of generality that $x_{j_0} < \bar{x}_{j_0}$. Then for $s \in (x_{j_0}, \bar{x}_{j_0})$ $0 \equiv -\phi_{\bar{x}_{j_0}}(c)(\bar{x}_j-s)^m$. This is impossible because of (6.8). This concludes the proof of the theorem. □

In the following theorem we assume as above that for every finite set of distinct points $\{z_1,\ldots,z_k\}$ in the closed bounded interval $[a,b]$ and for every collection of positive integers $\{n_1,\ldots,n_k\}$ we are able to observe the uncorrelated random variables $\{Y_1(z_1),\ldots,Y_{n_1}(z_1),\ldots,Y_{n_k}(z_k)\}$. We assume that for every x in the closed bounded interval $[a,c]$, $c > b$, the expectations $E[Y^2(x)]$ are finite $E[Y(x)-E(Y(x)]]^2 \equiv 1$, and that the function of x, $\theta(x) = E[Y(x)]$ though unknown has $m-1$ absolutely continuous derivatives and an $m^{\underline{th}}$ derivative $\theta^{(m)}$ which is in $L^2[a,c]$. Let $N > m > 1$ be an arbitrary integer and $\varepsilon > 0$ be an arbitrary number. Set $\eta = (N\varepsilon^2)^{-1}$ and let ξ_0 solve S_η. It has been proven that ξ_0 is unique and has support $S(\xi_0)$ consisting of exactly m points which we denote by $\{x_1,\ldots,x_m\}$. Let $\tilde{n}_1,\ldots,\tilde{n}_m$ be any one of the non-empty collection of integers satisfying

i) $\tilde{n}_i \geq \left\lceil \dfrac{(N-m)|\phi_{x_i}(c)|}{\sum_{j=1}^{m} |\phi_{x_j}(c)|} \right\rceil$ for $i=1,\ldots,m$, where $[x]$ is the smallest integer greater than or equal to x,

and ii) $\sum_{i=1}^{m} \tilde{n}_i = N.$

Let $\hat{\theta}(c)$ denote the value at c of the polynomial of degree m - 1 which passes through the m points $\{(x_i, \sum_j y_j(x_i)/\tilde{n}_i)\}_{i=1}^{m}$.

Every linear estimator of $\theta(c)$ based upon N observations is of the form $\ell'y = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \ell_{ij} y_j(z_i)$ where $\{\ell_{ij}\}$ is a collection of real numbers and $\{z_1,\ldots,z_k\}$ and $\{n_1,\ldots,n_k\}$ are as above with the additional restriction that $k \leq N$ and $\sum_1^k n_i = N$. Define

$$I(N) = \inf \sup_{||\theta^{(m)}||_2 \leq \varepsilon} E_\theta(\ell'Y-\theta(c))^2$$

where the infimum extends over all linear estimators based upon N observations.

Theorem 6.3.

$$0 \leq 1 - I(N)\Big(\sup_{||\theta^{(m)}||_2 \leq \varepsilon} E_\theta(\hat{\theta}(c)-\theta(c))^2\Big)^{-1} \leq \frac{m}{N-m}.$$

Proof. Because of theorem 4.1 it suffices to prove that $\hat{\theta}(c)$ is Speckman's estimator based upon observations at $\{x_1,\ldots,x_m\}$ in the quantities $\{\tilde{n}_1,\ldots,\tilde{n}_m\}$. It can be shown that the assumptions of lemma 2.1 are satisfied in our present case. Thus Speckman's estimator is the value at c of the function $\bar{\theta}(Y)$ in $W_m^2[a,c]$ which minimizes

$$\sum_{i=1}^{m} \sum_{j=1}^{\tilde{n}_i} (y_j(x_i)-\bar{\theta}(x_i))^2 + \frac{1}{\varepsilon^2} \int_a^c (\bar{\theta}^{(m)}(t))^2 dt.$$

Since $\sum_{i=1}^{m} \sum_{j=1}^{\tilde{n}_i} (y_j(x_i)-\bar{\theta}(x_i))^2 \geq \sum_{i=1}^{m} (y_j(x_i)-\bar{y}(x_i))^2$,

$$\hat{\theta}(x_i) = \bar{y}(x_i),\ \hat{\theta} \text{ is in } W_m^2[a,c],$$

and $\hat{\theta}^{(m)}(t) \equiv 0$ we have proven the theorem. □

When $m = 2$ more can be said. By straightforward but tedious computation, setting $x_2 = b$ in (6.5), one can show that

$$\delta_0'(x) = \frac{4\eta[c-(\frac{x_1-b}{2})]}{(b-x_1)^2} - \frac{(b-x_1)(c-b)}{b} + \int_a^x h_c(s)ds.$$

Setting

$$\eta_0 = \frac{(b-a)^2}{24}\left[\frac{1}{b-a} + \frac{1}{2(c-b)}\right]^{-1}$$

and defining the function $x_1(\eta)$ to be a when $\eta \geq \eta_0$ and to be the unique real solution to

$$(6,8) \qquad \frac{(b-x)^2}{24\eta} - \frac{1}{2(c-b)} = \frac{1}{b-x}$$

when $\eta \in (0,\eta_0)$ one can check (see also Spruill (1981a and b)) that the conditions i), ii), and iii) of theorem 3.2 are satisfied by δ_0 and ξ_0 in (6.4) and (6.5) when x_1 is as above and $x_2 = b$. Thus the unique optimal design places mass $\frac{c-b}{c-b+c-x_1}$ at $x_1((N\xi^2)^{-1})$ and $\frac{c-x_1}{c-b+c-x_1}$ at b.

Huber (1975) employed somewhat different assumptions and arrived at similar conclusions. Huber assumed that observations were available on $x \in [0,\infty)$, $\theta(-1)$ was to be estimated by a minimax linear estimator, and the contamination was $\sup_{-1\leq x<+\infty} |\theta''(x)| \leq \varepsilon$. The same results obviously hold if the interval is $(-\infty,0]$ and $\theta(1)$ is to be estimated by replacing his $1/\gamma$ with $-1/\gamma$ as the location of the left-most point of support in the optimal design. In this context, the right-most point is 0. As

Huber points out, his optimal designs also assign weights corresponding to those of the usual optimal designs on $[-\frac{1}{\gamma}, 0]$ to extrapolate to 1. Using only i), ii), and iii) of theorem 3.2 reveals that our results continue to be valid of $a = -\infty$ by setting $\frac{1}{n_0} = 0$. We shall compare the two optimal designs at $N\varepsilon^2$ values. In Huber's notation $n = N$, and we take $\sigma^2 = 1$. Denote the location of our left-most point by $x_1(N\varepsilon^2) = \frac{-1}{\gamma_1}$ and his by $x(N\varepsilon^2) = -\frac{1}{\gamma}$. Using his equation (6.22) and our equation (6.8) we see that $1 = \frac{12\gamma_1^2(1+2\gamma_1)}{8\gamma^2(1+2\gamma)(\gamma/(1+\gamma))}$. As $N\varepsilon^2 \to \infty$ both γ_1 and γ become large and we have $\frac{\gamma_1}{\gamma} \to (\frac{2}{3})^{1/3}$. Thus our right-most point is to the right of his by a factor of roughly $(3/2)^{1/3} - 1$ for large N and equal ε^2.

How does one find, for general m, the points $\{x_1,\ldots,x_m\}$? One method which suggests itself involves the problems P_n. Since the form of the solution to P_n is known one should be able to combine numerical searches for maxima with sequential selection of the points $\{x_1,\ldots,x_m\}$ to get a sequence converging to the solution of P_n and a sequence of sets converging to the correct support. We leave these matters and others for future consideration.

7. Acknowledgement. Among those with whom I have discussed these problems I would like to thank especially L. A. Karlovitz, Ker-Chau Li and W.J. Studden.

References

Billingsley, Patrick (1968). Convergence of Probability Measures. Wiley, New York.

Beutler, F. J. and Root, W. L. (1976). The operator pseudoinverse in control and systems identification, in Generalized Inverses and Applications, M. Z. Nashed Ed., Academic Press, New York.

Dunford, N. and Schwartz, J. T. (1964). Linear Operators Part I: General Theory, Wiley Interscience, New York.

Fedorov, V. V. (1972). Theory of Optimal Experiments. Academic Press, New York.

Hoel, P. G. and Levine, A. (1964). Optimal spacing and weighting in polynomial prediction. Ann. Math. Statist. 33, 1553-1560.

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

Karlin, S. and Studden, W. J. (1966a). Tschebycheff Systems: with Applications in Analysis and Statistics. Interscience, New York.

Karlin, S. and Studden, W. J. (1966b). Optimal experimental designs. Ann. Math. Statist. 37, 783-815.

Kiefer, J. (1980). Designs for extrapolation when bias is present. Multivariate Analysis V. North-Holland, Amsterdam, 79-93.

Li, Ker-Chau (1981). Robust regression designs when the design space consists of finitely many points. Purdue Mimeo. Series #81-45.

Li, K. C. and Notz, W. (1980). Robust designs for nearly linear regression. Purdue Mimeo. Series #80-3.

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

Nashed, M. Z. and Votruba, G. F. (1976). A unified operator theory of generalized inverses in Generalized Inverses and Applications, M.Z. Nashed Ed., Academic Press, New York.

Notz, W. (1980). Optimal robust designs for some regression problems. Purdue Mimeo. Series #80-29.

Parzen, Emanuel (1959). Statistical inference on time series by Hilbert space methods, in Technical Report No. 23, Stanford University.

Pesotchinsky, L. (1979). Robust designs and optimality of least squares for regression problems. Manuscript.

Slepian, D. and Pollak, H. O. (1961). Prolate spheroidal wave functions Fourier analysis and uncertainty - I. Bell System Technical Journal.

Speckman, P. (1979). Minimax estimates of linear functionals in a Hilbert space. To appear in Ann. Statist.

Spruill, Carl (1980). Optimal designs for second order processes with general linear means. Ann. Statist. 8, 652-663.

Spruill, M. C. (1981a). Traversing the smallest possible corridor under velocity and energy constraints. Purdue Mimeo. Series #81-52.

Spruill, M. Carl (1981b). Optimal designs using Speckman's minimax linear estimator. Purdue Mimeo. Series #81-49.

Spruill, M. Carl (1982). A method for the determination of optimal model robust designs in the estimation of a linear functional. Technical Report.

Spruill, Carl and Studden, W. J. (1978). Optimum designs when the observations are second order processes. J. Multivariate Analysis.

Rudin, W. (1966). Real and Complex Analysis. McGraw-Hill, New York.