The Annals of Statistics
1993, Vol. 21, No. 1, 434–446

MINIMAX REGRESSION DESIGNS UNDER UNIFORM DEPARTURE MODELS[1]

BY DEI-IN TANG

Nathan S. Kline Institute for Psychiatric Research

Model robustness in optimal regression design is studied by introducing a family of nonparametric models, which are defined as neighborhoods of classical parametric models in terms of the uniform norm. Optimal designs are sought under a minimax criterion for estimating linear functionals on such models that may be put as integrals using measures of finite support. A set of conditions equivalent to design optimality is derived using a Lagrangian principle applicable when the dimension is infinite and the function is not everywhere differentiable. From these conditions various optimal designs follow. Among them is the classical extrapolation design of Kiefer and Wolfowitz for Chebyshev regression, which is therefore model-robust against uniform departure. The conditions also shed light on other classical results of Kiefer and Wolfowitz and of others.

1. Introduction. We address a problem that arises in designing an experiment for estimating a regression parameter. In such a situation Kiefer and Wolfowitz (1959) show that the efficiency of estimation can be greatly improved by using an optimal design over a naive one.

The regression setting is assumed as follows. A regression function f is defined on a set T. Let T_0 be a subset of T. At each $t \in T_0$ uncorrelated random variables $Y_i(t)$ may be observed which satisfy $E(Y_i(t)) = f(t)$ and $\mathrm{Var}(Y_i(t)) = \sigma^2$. It is known that $f \in \Phi$, a specified class of functions on T.

In classical design theory, Φ is a linear space of finite dimension, consisting of continuous functions such as polynomials. In this case it is standard to estimate a regression parameter by the least-squares method, assuming that the design is reasonable so that the method will produce an unbiased estimator. A natural criterion for design optimality is therefore the variance of the least-squares estimator.

However, it is likely that $f \notin \Phi$. Further, as shown in Box and Draper (1959), even for small model departure a variance minimizing design can on the other hand introduce a large estimation bias, thus also a large mean square error (MSE). In this sense some classical optimal designs are not model-robust.

The problem of finding optimal designs that are more model-robust than the classical designs has since received considerable attention. Typically Φ is

Received May 1988; revised April 1992.

[1]Partially supported by NIMH Grant 42959 and NCI Grant 54708.

AMS 1991 *subject classifications*. Primary 62K05; secondary 62J02, 41A50.

Key words and phrases. Chebyshev polynomials, model-robustness, optimal designs, uniform-departure models.

somewhat larger than a classical model. Initially it remained finite-dimensional. Infinite-dimensional Φ's appeared in a later trend. The motivation for such Φ's is given, for example, in Sacks and Ylvisaker (1984). However, since they use Sobolev norms in defining Φ, their optimality criteria can be difficult to work with. For simplicity we propose to use the uniform norm. Although nonsmooth f's are admitted, the results are reasonable, many having counterparts in Sacks and Ylvisaker (1984). As argued by Fabian (1988), to use the uniform norm in regression is natural.

To define our Φ we assume for generality that T is a first countable compact Hausdorff space, such as the interval $[0, 1]$. Let $C(T)$ be the space of all continuous functions on T under the uniform norm. Let Φ_0 be a classical model, that is, a finite-dimensional linear subspace of $C(T)$, of which let $\{f_i\}_{i=0}^r$ be a basis. For any $\varepsilon > 0$, we define a uniform-departure model as the ε neighborhood of Φ_0 in $C(T)$, that is,

$$\Phi_\varepsilon = \{f \in C(T) \mid d(f, \Phi_0) \leq \varepsilon\},$$

where $d(f, \Phi_0)$ denotes a distance defined in the usual way. Note that when, for example, $T = [0, 1]$, Φ_ε with $\varepsilon > 0$ is of infinite dimension. In any application ε needs to be chosen. However, this problem will not be discussed. Our aim is to find optimal designs under any Φ_ε. Such knowledge should be useful for choosing ε.

Following Sacks and Ylvisaker (1984), we view a regression parameter under Φ_ε as a functional defined on Φ_ε. However, we develop a theory only for the estimation of the general linear functional

$$\Gamma(f) = \sum_{i=1}^{q} \gamma_i f(x_i),$$

where γ_i are known real coefficients and x_i are known points in T and $q < \infty$. Hereafter this general form of Γ is assumed unless specifically otherwise given. We may put $\Gamma(f)$ as an integral of f in terms of a (signed) measure and identify Γ with that measure.

Also following Sacks and Ylvisaker (1984), we use among linear estimators one that minimizes the maximum MSE. A design, consisting of design points and a rule for allocating observations, is optimal if it minimizes the maximum MSE of the best linear estimator.

For a more detailed discussion on related literature and the motivation of the above formulation see Sacks and Ylvisaker (1984). It should be noted that they also treat linear functionals that are induced by continuous measures.

In Section 2 we reduce the problem to the minimization of a function of signed measures subject to a constraint. In Section 3 we give the main result which is a set of conditions equivalent to design optimality. In Section 4 we discuss the notion of model-robustness. Then we give optimal designs in various situations. We also illustrate a continuity principle useful for computation. In Section 5 we recast some classical results.

2. Mathematical preliminaries. Consider a general design with design points $t_i \in T_0$, $i = 1, \dots, k$, and n_i observations allocated at t_i subject to that $N \equiv \sum n_i$ is fixed. Let $Y_j(t_i)$ be the random variables to be observed. Let the estimator be $\sum_{i=1}^{k} c_i \bar{Y}(t_i)$, where $\bar{Y}(t_i) = \sum_{j=1}^{n_i} Y_j(t_i)/n_i$. Then,

$$\mathrm{MSE}(f; \{c_i\}, \{t_i\}, \{n_i\}) \equiv E\Big(\sum c_i \bar{Y}(t_i) - \Gamma(f)\Big)^2$$

$$= \sigma^2 \sum c_i^2/n_i + \Big(\sum c_i f(t_i) - \sum \gamma_j f(x_j)\Big)^2.$$

The two linear functionals in the bias term above may be put as $\int f dD$ and $\int f d\Gamma$ with $D = \sum c_i \delta_{t_i}$ and $\Gamma = \sum \gamma_i \delta_{x_i}$, where δ_x denotes the unit measure concentrated at x. The best linear estimator can be found by minimizing with respect to c_i the maximum of the above MSE over $f \in \Phi_\varepsilon$. To find an optimal design we need to further minimize the resulting minimum with respect to t_i and n_i. However, the mathematical problem is simply to maximize the MSE with respect to f and then minimize the maximum with respect to t_i, n_i and c_i. As in classical theory [see Kiefer and Wolfowitz (1959)], the difficulty of minimizing over integers is here compromised by allowing n_i to take on arbitrary nonnegative numbers (subject to the constraint on their total). Thus it can be shown that if an optimal design exists then the optimal n_i and the best c_i satisfy

$$n_i/n_j = |c_i|/|c_j|. \tag{2.1}$$

Substituting (2.1) into the above MSE, the problem is reduced to

$$\min_{\{c_i, t_i\}} \left\{ \sigma^2 \Big(\sum |c_i|\Big)^2 / N + \max_{\{f \in \Phi_\varepsilon\}} \left(\int f d(D - \Gamma) \right)^2 \right\}. \tag{2.2}$$

Hereafter we denote $\sum |c_i|$ by $\|D\|$, for it is both the total variation of D as a measure and the norm of D as a linear functional on $C(T)$.

Note that a measure D that solves (2.2) contains the t_i of an optimal design and the corresponding c_i for estimation. The corresponding n_i can then be obtained from (2.1). Thus any D considered in (2.2) may be referred to for convenience as a design. Other than in such cases a design mentioned hereafter remains consisting of design points and an allocation rule, which in classical theory are summarized together as a probability measure. Note that in classical theory the estimation is done a priori by using the least-squares estimator.

We now simplify the max in (2.2). First, we restrict attention to unbiased designs, that is, those D's such that

$$\int f dD = \int f d\Gamma \quad \text{for } f \in \Phi_0. \tag{2.3}$$

This is because otherwise the max is $+\infty$. Condition (2.3) is satisfied in classical approaches because the least-squares estimator is unbiased. For an unbiased D we can show that

$$\max_{\{f \in \Phi_\varepsilon\}} \left(\int f d(D - \Gamma) \right)^2 = \varepsilon^2 \|D - \Gamma\|^2$$

with the aid of Tietze's extension theorem [Royden (1963)]. Thus, (2.2) is reduced to the problem of minimizing

$$Q_\rho(D) \equiv \rho \|D\|^2 + \|D - \Gamma\|^2$$

over unbiased designs, where $\rho \equiv N^{-1}\sigma^2\varepsilon^{-2}$. When $\varepsilon = 0$, implying that $\rho = \infty$, the expression for Q_ρ does not make sense. In this case $Q_\infty(D)$ is defined as $\|D\|^2$. Note that there is, as should be, a sense of continuity for Q_ρ even at $\rho = \infty$. This is based on the fact that for $0 < \rho < \infty$ a design minimizes Q_ρ if and only if it minimizes $\rho^{-1}Q_\rho(D) = \|D\|^2 + \rho^{-1}\|D - \Gamma\|^2$. For convenience we also consider the case $\rho = 0$, for which it is obvious that if each support point of Γ belongs to the design space T_0, then $D = \Gamma$ is the unique optimal design. In the following discussion the general range for ρ is $0 \le \rho \le \infty$. As with ε, in actual applications ρ is usually unknown. However, we consider below only the minimization of Q_ρ with ρ arbitrary, for which D_ρ^* will denote an optimal solution.

Since function Q_ρ is derived from (2.2), extending its domain to include measures of arbitrary support does not make practical sense. Mathematically, to minimize over more general D's may be more difficult, as the set of finitely supported measures is not dense in the space of all measures under the norm of total variation.

From any results for Φ_ε with $\varepsilon > 0$ corresponding results for Φ_0 may be obtained by letting $\varepsilon \to 0$ (equivalently $\rho \to \infty$) or by following similar arguments. Certain optimal designs we obtain under Φ_0 are actually classical designs. However, as we reduce the problem in a different way, our approach serves to shed new light on, rather than prove the classical results.

3. General results. For any measure D, let supp(D) denote its support. For any sets A and B, let $A - B$ denote the set of elements in A but not in B, and let $|A|$ denote the cardinality of A. Let $D(A)$ and $|D|(A)$ denote, respectively, the mass and total variation of A under D. Further, for any real number x let $\operatorname{sgn}(x) = 1, -1$ or 0 according to $x >, <$ or $= 0$.

LEMMA 3.1. *For any unbiased design D, there exists an unbiased design D' such that $Q_\rho(D') \le Q_\rho(D)$ for any ρ and $|\operatorname{supp}(D') - \operatorname{supp}(\Gamma)| \le \dim(\Phi_0)$.*

PROOF. By the discussion in the end of Section 2, we treat only the case $0 < \rho < \infty$. Recall that $\dim(\Phi_0) = r + 1$. Let $D = \sum_{i=1}^k c_i \delta_{t_i}$ with $\{t_i\}_{i=1}^m = \operatorname{supp}(D) - \operatorname{supp}(\Gamma)$, where all t_i are distinct, all $c_i \ne 0$ and $m \le k$. Suppose $m = r + 2$. The general case $m > r + 1$ follows by induction.

There exists a nontrivial solution $\{d_i^*\}_{i=1}^{r+2}$ to the equations,

$$\sum_{i=1}^{r+2} d_i f_j(t_i) = 0 \quad \text{for } j = 0, \ldots, r.$$

For convenience let $d_i^* = 0$ for $i = r+3, \ldots, k$. Let $F = \Sigma_{i=1}^k d_i^* \delta_{t_i}$. Then $D + sF$ is unbiased for any scalar s. When $|s| \neq 0$ is small enough, we have $c_i(c_i + sd_i^*) > 0$ for all i, from which $\|D + sF\| = \|D\| + s\Sigma d_i^* \operatorname{sgn}(c_i)$. Assume $\Sigma d_i^* \operatorname{sgn}(c_i) > 0$; other possibilities can be similarly treated. By continuity there exists $s_0 < 0$ such that for all i, $c_i(c_i + s_0 d_i^*) \geq 0$ and equality holds for at least one $i \leq r+2$. Let $D' = D + s_0 F$. Then $\|D'\| = \Sigma(c_i + s_0 d_i^*)\operatorname{sgn}(c_i) = \|D\| + s_0 \Sigma d_i^* \operatorname{sgn}(c_i) < \|D\|$. As D' differs from D only on design points not in $\operatorname{supp}(\Gamma)$, $\|D' - \Gamma\| < \|D - \Gamma\|$ also holds. Thus $Q_\rho(D') < Q_\rho(D)$ and $|\operatorname{supp}(D') - \operatorname{supp}(\Gamma)| \leq r + 1 = \dim(\Phi_0)$. □

A sequence D_n of measures of finite support on T is said to converge *weakly in norm* to measure D_0 of finite support on T if, assuming $\operatorname{supp}(D_0) = \{x_1, \ldots, x_e\}$ with distinct x's, for any $\delta > 0$ and any open sets O_i, $i = 1, \ldots, e$ in T with $x_i \in O_i$, there exists k such that for any $n > k$,

$$|D_n|(T - \cup O_i) + \sum_{i=1}^{e} |(D_n - D_0)(O_i)| < \delta.$$

Note that D_0 is necessarily unique and $\liminf \|D_n\| \geq \|D_0\|$. In addition, if all D_n are unbiased designs, so is D_0.

LEMMA 3.2. *For any design sequence D_n such that both $|\operatorname{supp}(D_n)|$ and $\|D_n\|$ are bounded in n, there exists a subsequence which converges weakly in norm to a design satisfying the same bounds.*

PROOF. Write $D_n = \Sigma_{i=1}^M c_{i,n} \delta_{t_{i,n}}$ for a finite fixed M, where some $c_{i,n}$'s may be 0. There exists a subsequence, say $\{n_j\}$, of $\{n\}$ such that $t_{i,n_j} \to t_{i,0}$ and $c_{i,n_j} \to c_{i,0}$ for all $i = 1, \ldots, M$ as $n_j \to \infty$. It follow that $D_{n_j} \to \Sigma_{i=1}^M c_{i,0} \delta_{t_{i,0}}$ weakly in norm. □

The next theorem follows from Lemmas 3.1 and 3.2.

THEOREM 3.1. *For any ρ, there exists an optimal design D_ρ^* satisfying*

$$|\operatorname{supp}(D_\rho^*) - \operatorname{supp}(\Gamma)| \leq \dim(\Phi_0).$$

An optimal design may not be unique, as the norm $\|D\|$ is not a strictly convex function. In view of Theorem 3.1 we hereafter consider only designs D satisfying $|\operatorname{supp}(D) - \operatorname{supp}(\Gamma)| \leq \dim(\Phi_0)$.

The next lemma will be needed for Theorem 4.1.

LEMMA 3.3. *If $\rho_n \to \rho$ (including $\rho = \infty$) and D_ρ^* is the unique optimal design under ρ, then for any optimal design $D_{\rho_n}^*$ under ρ_n, $D_{\rho_n}^* \to D_\rho^*$ weakly in norm.*

PROOF. Any subsequence of $D^*_{\rho_n}$ contains a sub-subsequence which converges weakly in norm to a design D_ρ. This D_ρ is optimal under ρ. By uniqueness, $D_\rho = D^*_\rho$. □

We now use the theory of Lagrange multiplier to derive conditions equivalent to design optimality. Since the function to be minimized is not everywhere differentiable and its domain is of infinite dimension, we follow an argument in Whittle [(1971), page 52]. Let ρ, $0 < \rho < \infty$, be fixed. Define B to be the set of those points $(x_1, \dots, x_{r+2})$ each of which satisfies that for a design D (not necessarily unbiased), $x_{i+1} = \int f_i \, d(D - \Gamma)$, $i = 0, \dots, r$ and $x_{r+2} \geq Q_\rho(D)$. Clearly B is a convex set. For an optimal design D^*_ρ we consider the point $b \equiv (0, \dots, 0, Q_\rho(D^*_\rho))$. This point must lie in the boundary of B. Otherwise there would exist $x^{**} < Q_\rho(D^*_\rho)$ such that $(0, \dots, 0, x^{**}) \in B$ and consequently also an unbiased design D^{**} such that $Q_\rho(D^{**}) \leq x^{**} < Q_\rho(D^*_\rho)$, contradicting the optimality of D^*_ρ. Finally, the existence of a supporting hyperplane at b implies that for some $\lambda_0, \dots, \lambda_{r+1}$,

$$\sum_{i=0}^{r} \lambda_i \int f_i \, d(D - \Gamma) + \lambda_{r+1} Q_\rho(D) \geq \lambda_{r+1} Q_\rho(D^*_\rho)$$

for any design D. As λ_{r+1} must be > 0, it may be chosen to be 1. Thus, D^*_ρ minimizes $Q^\lambda_\rho(D) \equiv Q_\rho(D) + \sum_{i=0}^r \lambda_i \int f_i \, d(D - \Gamma)$ among all designs D. A necessary condition of optimality is therefore that

$$(d/d\varepsilon) Q^\lambda_\rho(D^*_\rho + \varepsilon D)\big|_{\varepsilon = 0+} \geq 0 \quad \text{for all } D, \tag{3.1}$$

which can be seen to be also sufficient among unbiased designs. To find the derivative in (3.1) we use the formula,

$$\begin{aligned}(d/d\varepsilon)\|D^*_\rho + \varepsilon D\|\big|_{\varepsilon = 0+} &= \sum_{t \in \operatorname{supp}(D)} D(t) \operatorname{sgn}\big(D^*_\rho(t)\big) \\ &\quad + \sum_{t \in \operatorname{supp}(D)} |D(t)| \big(1 - \big|\operatorname{sgn}\big(D^*_\rho(t)\big)\big|\big).\end{aligned} \tag{3.2}$$

THEOREM 3.2. *For $0 < \rho < \infty$, let $\alpha = \rho\|D^*_\rho\| / (\rho\|D^*_\rho\| + \|D^*_\rho - \Gamma\|)$. Then D^*_ρ is optimal if and only if it is unbiased and there exists $p \in \Phi_0$ such that*:

(a) $$|p(t)| \leq 1 \quad \text{for } t \in T_0$$

(b) $$p(t) = -\operatorname{sgn}\big(D^*_\rho(t)\big) \quad \text{for } t \in \operatorname{supp}(D^*_\rho) - \operatorname{supp}(\Gamma).$$

(c) $$-\alpha + (1 - \alpha)\operatorname{sgn}(\Gamma(t)) \leq p(t) \leq \alpha + (1 - \alpha)\operatorname{sgn}(\Gamma(t)) \quad \text{for } t \in \big(\operatorname{supp}(\Gamma) - \operatorname{supp}(D^*_\rho)\big) \cap T_0.$$

(d) $$\begin{aligned} &-\alpha \operatorname{sgn}\big(D^*_\rho(t)\big) + (1 - \alpha)\big\{-1 + \big|\operatorname{sgn}\big((D^*_\rho - \Gamma)(t)\big)\big| - \operatorname{sgn}\big((D^*_\rho - \Gamma)(t)\big)\big\} \\ &\leq p(t) \leq -\alpha \operatorname{sgn}\big(D^*_\rho(t)\big) + (1 - \alpha)\big\{1 - \big|\operatorname{sgn}\big((D^*_\rho - \Gamma)(t)\big)\big| - \operatorname{sgn}\big((D^*_\rho - \Gamma)(t)\big)\big\} \\ &\qquad \text{for } t \in \operatorname{supp}(D^*_\rho) \cap \operatorname{supp}(\Gamma).\end{aligned}$$

PROOF. (Necessity.) By (3.1) and (3.2),

$$
\begin{aligned}
&\rho\|D_\rho^*\|\Big\{\sum D(t)\operatorname{sgn}\big(D_\rho^*(t)\big)+\sum|D(t)|\big(1-\big|\operatorname{sgn}\big(D_\rho^*(t)\big)\big|\big)\Big\}\\
&\quad+\|D_\rho^*-\Gamma\|\Big\{\sum D(t)\operatorname{sgn}\big((D_\rho^*-\Gamma)(t)\big)\\
&\qquad\qquad+\sum|D(t)|\big(1-\big|\operatorname{sgn}\big((D_\rho^*-\Gamma)(t)\big)\big|\big)\Big\}\\
&\quad+2^{-1}\sum_{i=0}^{r}\lambda_i\int f_i\,dD\geq 0
\end{aligned}\tag{3.3}
$$

for some λ_i and any D, where the range for t in all the summations is supp(D). Plugging $D=\pm\delta_t$, for each $t\in T_0$, into (3.3), (a) to (d) follow with $p(t)=2^{-1}(\rho\|D_\rho^*\|+\|D_\rho^*-\Gamma\|)^{-1}\Sigma_{i=0}^{r}\lambda_i f_i(t)$.

(Sufficiency.) From the existence of the described p follows (3.3), hence also (3.1). □

For the case $\rho=\infty$ a corresponding result can be obtained by taking $\alpha=1$.

COROLLARY . *D_∞^* is optimal if and only if it is unbiased and there exists $p\in\Phi_0$ such that*

(i) $$|p(t)|\leq 1\quad \text{for } t\in T_0.$$

(ii) $$p(t)=-\operatorname{sgn}(D_\infty^*(t))\quad\text{for } t\in\operatorname{supp}(D_\infty^*).$$

This corollary suggests a simple way for obtaining optimal designs under Φ_0. Pick a function $p\in\Phi_0$ and normalize it so that $|p(t)|\leq 1$ for $t\in T_0$. Find D that satisfies (ii) in the corollary. This D is optimal for any Γ for which it is unbiased. However, when Γ is given first it may not be easy to find an optimal D in this way.

In the special case where Γ is a positive measure, $1\in\Phi_0$ and $T_0=T$, the design $D=\Gamma$ is optimal for all ρ and is unique except for $\rho=\infty$. This is because $\|\Gamma\|=\int d\Gamma=\int dD\leq\|D\|$ and $Q_\rho(D)\geq\rho\|D\|^2\geq\rho\|\Gamma\|^2=Q_\rho(\Gamma)$ for all D.

4. Special results. Theorem 3.2 reminds one of the theory of Chebyshev polynomials. We review here some relevant results [see Karlin and Studden (1966)].

A set of continuous functions $\{g_i\}_{i=0}^m$ on the interval $[a,b]$ is said to be a Chebyshev system (C-system) if any $\Sigma_{i=0}^m c_i g_i$ with c_i not all zero has at most m distinct zeros in $[a,b]$. For this C-system the following results hold.

1. There exists a unique function $p_m=\Sigma c_i^* g_i$ such that $|p_m(t)|\leq 1$, $t\in[a,b]$ and $p_m(t_{m-i})=(-1)^i$ for some t_i, $i=0,\dots,m$, with $a\leq t_0<t_1<\cdots<t_m\leq b$. We refer to p_m as the C-function and the t_i's as a set of C-points.
2. If there exist c_i's such that $\Sigma c_i g_i(t)=1$ for all t, then the set $\{t\mid p_m^2(t)=1\}$ consists of $m+1$ distinct points, including a and b.
3. For all sets $\{s_i\}_{i=0}^m$ with $s_i<s_{i+1}$ and $s_i\in[a,b]$, the determinants $|g_j(s_i)|$ are either all positive or all negative.

We now study the situation in which there exists a design that is optimal under Φ_0 and is model-robust against uniform departure in a strong sense. According to the literature a design optimal under Φ_ε for one $\varepsilon > 0$ may be said to be model-robust. This definition seems weak in the sense that to employ such a design the departure parameter ε should be carefully chosen to achieve both efficiency and robustness. The strongest definition is that the design is optimal under Φ_ε for any $\varepsilon \geq 0$. However, such designs simply may not exist in most cases. We explore a definition that lies in between. Namely, an optimal design D_∞^* is model-robust if there exists an optimal design D_ρ^* for each $\rho > 0$, such that $\operatorname{supp}(D_\infty^*) \supseteq \operatorname{supp}(D_\rho^*)$ for all $\rho > 0$ and $D_\rho^* \to D_\infty^*$ weakly in norm as $\rho \to \infty$. The continuity condition seems reasonable. The motivation for the condition of inclusion arises from the opinion that the design points are the most vital part of a design and that a model-robust design should have sufficiently many design points so that any degree of model departure can be detected. However, the following result says that this definition is no different from the strongest one. As the proof shows, the condition that the inclusion relationship holds for all $\rho > 0$ is very restrictive. In applications one may be able to specify a realistic finite upper bound on ε so that for ε beyond this bound Φ_ε is too wide to be a suitable model. Thus, one may want to relax the inclusion condition from "for all $\rho > 0$" to "for all ρ greater than a specified positive number." However, no characterization of optimal designs model-robust in this relaxed sense is known.

THEOREM 4.1. *Assume $T_0 = T = [a, b]$, $\{f_i\}_{i=0}^r$ is a C-system on T and $1 \in \Phi_0$. Then D_∞^* is model robust if and only if $D_\infty^* = \Gamma$, in which case it is optimal under Φ_ε for any $\varepsilon > 0$.*

PROOF. (Sufficiency.) It suffices to prove the optimality of $D_\infty^* = \Gamma$ under Φ_ε for any $\varepsilon > 0$, which follows immediately from Theorem 3.2 and its corollary.

(Necessity.) Assume $\int 1\, d\Gamma \geq 0$. Let $D_\rho^* \to D_\infty^*$ weakly in norm with $\operatorname{supp}(D_\infty^*) \supseteq \operatorname{supp}(D_\rho^*)$. By Lemma 3.3 we have $D_\rho^* \to \Gamma$ weakly in norm as $\rho \to 0$, since $D \equiv \Gamma$ is the only optimal design under $\rho = 0$. From $\operatorname{supp}(D_\infty^*) \supseteq \operatorname{supp}(D_\rho^*)$ follows $\operatorname{supp}(D_\infty^*) \supseteq \operatorname{supp}(\Gamma)$.

We claim that $D_\infty^* = \Gamma$. Suppose not. Let $G = D_\infty^* - \Gamma$. Then $G(f) = 0$ for all $f \in \Phi_0$. By result (3) of C-system this is possible for a measure $G \neq 0$ only if $|\operatorname{supp}(G)| > \dim(\Phi_0)$. Since $\operatorname{supp}(D_\infty^*) \supseteq \operatorname{supp}(G)$, we have $|\operatorname{supp}(D_\infty^*)| > \dim(\Phi_0)$. Applying the corollary of Theorem 3.2 to D_∞^*, there exists $p \in \Phi_0$ such that $|p(t)| \leq 1$ for all t with equality for $k > \dim(\Phi_0)$ distinct t's. A standard argument using the definition of C-system implies that p must be a constant, which must then be either 1 or -1. Accordingly, D_∞^* is either a positive or a negative measure. The latter is impossible due to $\int 1\, dD_\infty^* = \int 1\, d\Gamma$, which is greater than or equal to 0 by assumption. Hence D_∞^* is a positive measure. By continuity and $\operatorname{supp}(D_\infty^*) \supseteq \operatorname{supp}(D_\rho^*)$, D_ρ^* is a positive measure with $\operatorname{supp}(D_\rho^*) = \operatorname{supp}(D_\infty^*)$ for sufficiently large ρ. For each such ρ consider

$D'_\rho \equiv (1-\beta)D^*_\rho + \beta\Gamma$ for a $\beta > 0$ so small that D'_ρ remains a positive measure. This is possible by $\mathrm{supp}(D^*_\rho) \supseteq \mathrm{supp}(\Gamma)$. From $\|D'_\rho\| = \int 1\,dD'_\rho = \|D^*_\rho\|$, $D'_\rho - \Gamma = (1-\beta)(D^*_\rho - \Gamma)$ and $Q_\rho(D^*_\rho) \le Q_\rho(D'_\rho)$ follows $\|D^*_\rho - \Gamma\| = 0$. Thus, $\Gamma = D^*_\rho$, which leads, as $\rho \to \infty$, to the contradiction that $D^*_\infty = \Gamma$. □

By Theorem 4.1 the class of model-robust optimal designs for all Γ's is the same as the class of optimal designs under Φ_0 for all Γ's. This is because a design D optimal under Φ_0 for, say Γ_1, is model-robust for $\Gamma_2 \equiv D$.

We present some results for special cases of Φ_0.

Proposition 4.1. *Let $T_0 = T$ and Φ_0 be the constant regression model. For any ρ there exists D^*_ρ with* $\mathrm{supp}(\Gamma) \supseteq \mathrm{supp}(D^*_\rho)$ *and* $0 \le D^*_\rho(t)/\Gamma(t) \le 1$ *for* $t \in \mathrm{supp}(\Gamma)$.

Proof. Among all optimal designs let D^*_ρ be one such that $|\mathrm{supp}(D^*_\rho) - \mathrm{supp}(\Gamma)|$ is a minimum. Suppose there exists $t^* \in \mathrm{supp}(D^*_\rho) - \mathrm{supp}(\Gamma)$. Pick any x from $\mathrm{supp}(\Gamma)$ and define $D' \equiv D^*_\rho + D^*_\rho(t^*)(\delta_x - \delta_{t^*})$. Clearly D' remains optimal, but $|\mathrm{supp}(D') - \mathrm{supp}(\Gamma)| < |\mathrm{supp}(D^*_\rho) - \mathrm{supp}(\Gamma)|$, a contradiction. Next, suppose that $0 \le D^*_\rho(t)/\Gamma(t) \le 1$ does not hold for $t_1 \in \mathrm{supp}(\Gamma)$. Then $D^*_\rho(t_1)(D^*_\rho(t_1) - \Gamma(t_1)) > 0$. By unbiasedness there exists $t_2 \in \mathrm{supp}(\Gamma)$ such that $D^*_\rho(t_1)(D^*_\rho(t_2) - \Gamma(t_2)) < 0$. Then, assuming $D^*_\rho(t_1) > 0$, $D'' \equiv D^*_\rho - c(\delta_{t_1} - \delta_{t_2})$ for a sufficiently small $c > 0$ is a better design than D^*_ρ, a contradiction. □

Proposition 4.2. *Assume $T_0 = T = [a,b]$ and Φ_0 is the linear regression model. For any ρ there exists D^*_ρ with* $\{a,b\} \supseteq \mathrm{supp}(D^*_\rho) - \mathrm{supp}(\Gamma)$ *and* $0 \le D^*_\rho(t)/\Gamma(t) \le 1$ *for* $t \in \mathrm{supp}(\Gamma) \cap (a,b)$.

Proof. Assume $0 < \rho < \infty$. Let D^*_ρ be an optimal design. Suppose there exists $t^* \in \mathrm{supp}(D^*_\rho) - \mathrm{supp}(\Gamma)$ with $a < t^* < b$. By (a) and (b) in Theorem 3.2, the function p must be one of ± 1, say -1. By (b) $D^*_\rho(t) > 0$ for $t \in \mathrm{supp}(D^*_\rho) - \mathrm{supp}(\Gamma)$. By (c) $\Gamma(t) < 0$ for $t \in \mathrm{supp}(\Gamma) - \mathrm{supp}(D^*_\rho)$. By (d) $D^*_\rho(t) \ge \Gamma(t)$ for $t \in \mathrm{supp}(D^*_\rho) \cap \mathrm{supp}(\Gamma)$. Consequently $\int 1\,dD^*_\rho > \int 1\,d\Gamma$, a contradiction. Next, suppose that $0 \le D^*_\rho(t)/\Gamma(t) \le 1$ does not hold for $t_1 \in \mathrm{supp}(\Gamma) \cap (a,b)$. Then $D^*_\rho(t_1)(D^*_\rho(t_1) - \Gamma(t_1)) > 0$. By (d), $|p(t_1)| = 1$, implying that p is one of ± 1. A contradiction follows as above. □

In the next proposition let $\Gamma^+ - \Gamma^-$ be the Hahn decomposition of Γ.

Proposition 4.3. *Assume $T_0 = T$ and Φ_0 is the constant regression model. An optimal design under ρ is given as follows*:

(i) *When* $\int 1\,d\Gamma = 0$, $D^*_\rho = (1+\rho)^{-1}\Gamma$.

(ii) *When* $\int 1\,d\Gamma \ne 0$, *say* > 0 (*assume then* $\Gamma^- \ne 0$), $D^*_\rho \equiv c_1\Gamma^+ - c_2\Gamma^-$ *for* $\rho \le \rho_0$, *where* $\rho_0 = 2\int 1\,d\Gamma^-/\int 1\,d\Gamma$, $c_1 = ((1+\rho)^{-1}\|\Gamma\| + \int 1\,d\Gamma)/(\|\Gamma\| + \int 1\,d\Gamma)$ *and* $c_2 = ((1+\rho)^{-1}\|\Gamma\| - \int 1\,d\Gamma)/(\|\Gamma\| - \int 1\,d\Gamma)$ *and* $D^*_\rho = D^*_{\rho_0}$ *for* $\rho > \rho_0$.

PROOF. Invoke Theorem 3.2 with $p(t) = 1 - 2\alpha$. $\square$

The next example illustrates that an optimal design may not be unique.

EXAMPLE 4.1. Let $T_0 = T = [0, 1]$, Φ_0 be the constant regression model and $\Gamma = \delta_1 - \delta_{1/2} + \delta_0$. In addition to the design in Proposition 4.3 the following design is also optimal, $D_\rho = \delta_1 - s\delta_{1/2} + s\delta_0$ with $s = (2-\rho)/(2+2\rho)$ when $\rho \le 2$, and $D_\rho = \delta_1$ when $\rho > 2$.

The following continuity principle may be useful for obtaining optimal designs. Starting with $\rho = \infty$, find a D_∞^*. This requires checking (i) and (ii) in the corollary of Theorem 3.2. Next, check if the same design and p function continue to work when ρ is finite but large. This requires checking (a) to (d) in Theorem 3.2. Among them (a) and (b) will automatically hold. Usually $|p(t)| < 1$ for $t \in \operatorname{supp}(\Gamma) - \operatorname{supp}(D_\infty^*)$, so (c) will hold when α is close to 1, that is, when ρ is large. However, (d) will not hold if $0 < D_\infty^*(t)/\Gamma(t) < 1$ for some t. When this occurs, either try a different design to start over with or other methods, which may be based on results such as Propositions 4.1 and 4.2, are in need. Fortunately, this situation seems rare. So we assume here that a D_∞^* is found so that for finite but large ρ, (d) either is satisfied or does not apply. According to the above explanation, this design continues to be optimal for finite but large ρ, until a value of ρ, say ρ_1, is reached for which (c) holds only critically, that is, it no longer holds for $\rho < \rho_1$. We may then enlarge $\operatorname{supp}(D_\infty^*)$ by including those points of $\operatorname{supp}(\Gamma)$ at which (c) holds critically. New coefficients as well as a new $p(t)$ may be needed. However, the results obtained for large ρ may provide clues on how to find these things for ρ in the next range. The same procedure may be repeated until $\rho = 0$ is reached. We illustrate the continuity principle via the following example.

EXAMPLE 4.2. Let $T_0 = T = [0, 1]$ and Φ_0 be the linear regression model. Consider $\Gamma = \delta_{x_1} - \delta_{x_2}$ with $0 < x_1 < x_2 < 1$, $x_1 + x_2 = 1$. According to Proposition 4.2, we can find an optimal design based on $D = c_0\delta_0 + c_1\delta_{x_1} + c_2\delta_{x_2} + c_3\delta_1$. However, for $\rho = \infty$ it may suffice to only consider $D = c_0\delta_0 + c_3\delta_1$. In this case there is a unique unbiased design, of which $c_0 = -c_3 = x_1 - x_2$. It turns out this design, denoted by D_∞, is optimal by the corollary of Theorem 3.2 with $p(t) = p_\infty(t) \equiv 1 - 2t$. For $\rho < \infty$, D_∞ and p_∞ clearly satisfy (a) and (b) in Theorem 3.2. Moreover, (d) does not apply. Finally, (c) requires $\alpha \ge x_2$, which holds if $\rho \ge \rho_1 \equiv 2x_2^2/x_1(x_2 - x_1)$. Therefore D_∞ remains optimal for $\rho \ge \rho_1$. For $\rho < \rho_1$ we consider $D_\rho = c_0\delta_0 + c_1\delta_{x_1} + c_2\delta_{x_2} + c_3\delta_1$. In this case c_1 determines the other c's through $-c_0 = c_3 = (x_2 - x_1)(1 + c_1)$ and $c_2 = -c_1$. As (c) no longer applies, it remains to verify (d). By Proposition 4.2 we require $0 > c_1 \ge -1$. Then express α as $\rho(c_3 - c_1)/(\rho(c_3 - c_1) + 1 + c_1 + c_3)$. Condition (d) is satisfied if $p = p_\infty$ and $\alpha = x_2$. Thus, coefficient c_1 can be determined in terms of ρ. Finally, it is necessary to verify if c_1 obtained this way satisfies $0 > c_1 \ge -1$. Due to $\rho < \rho_1$ the answer is yes. Hence an optimal design for $\rho < \rho_1$ is found. Notice that this design is supported on four points, although as $\rho \to 0$ it converges weakly in norm to Γ, supported only on two points.

We turn to extrapolation problems, which means that $T_0 \cap \operatorname{supp}(\Gamma) = \varnothing$.

THEOREM 4.2. *For any extrapolation problem, an optimal design under* Φ_0 *is optimal under* Φ_ε *for any* $\varepsilon > 0$.

PROOF. Because $\|D - \Gamma\| = \|D\| + \|\Gamma\|$, the minimization criterion $Q_\rho(D)$ under Φ_ε is equivalent to that under Φ_0, that is, $Q_\infty(D) \equiv \|D\|^2$. □

Hoel and Levine (1964) consider an extrapolation problem in polynomial regression. Their results are generalized to C-system by Kiefer and Wolfowitz (1965). The question of model robustness for the design of Hoel and Levine has been studied by Spruill (1985) under Sobolev-departure models.

5. Classical optimal designs. In this section the model is restricted to be Φ_0. In this case a parameter Γ may be identified with the vector $\Gamma(\mathbf{f}) \equiv (\Gamma(f_0), \dots, \Gamma(f_r))$. We demonstrate that our method can reproduce many classical designs.

THEOREM 5.1. *When* $T_0 = [a, b]$, $\Gamma = \delta_x$ *for some* $x > b$, $\{f_i\}_{i=0}^r$ *is a* C-*system on* $[a, x]$, *and* $\{t_i\}_{i=0}^r$ *is a set of* C-*points on* T_0, *then the unique unbiased design supported on* $\{t_i\}_{i=0}^r$ *is optimal.*

PROOF. The unique unbiased design supported on $\{t_i\}_{i=0}^r$ is given by $D_\infty^* = \sum_{i=0}^r c_i^* \delta_{t_i}$ with $c_i^* = A^{-1}| f_j(s_{k,i})|_{j,k=0}^r$, where $A = | f_j(t_m)|_{j,m=0}^r$ and $s_{k,i} = t_k$ for all k except that $s_{i,i} = x$. By result (3) of C-system, $\operatorname{sgn}(c_i^*)$ alternates between ± 1. The optimality follows from the corollary of Theorem 3.2 with p chosen appropriately between the C-function on T_0 and its negative. □

Essentially the same results appear in Hoel and Levine (1964) for polynomials, and in Hoel (1966) and Kiefer and Wolfowitz (1965) for C-systems. When $1 \in \Phi_0$, the optimal design in Theorem 5.1 is unique. This is a special case of the next theorem. Note that when $1 \in \Phi_0$, the set of C-points on T_0 is unique by result (2) of C-system. This set will be denoted again by $\{t_i\}_{i=0}^r$. Let $\mathbf{v_i} = (f_0(t_i), \dots, f_r(t_i))$, $R^* = \{\sum_{i=0}^r (-1)^i c_i \mathbf{v_i} \mid c_i$ all positive or all negative$\}$, and $S^* = \{\sum_{i=0}^r c_i \mathbf{v_i} \mid c_i$ all positive or all negative$\}$. Thus, design $D = \sum_{i=0}^r c_i \delta_{t_i}$ is unbiased if and only if

$$\sum_{i=0}^{r} c_i \mathbf{v_i} = \Gamma(\mathbf{f}). \tag{5.1}$$

The following condition will be assumed in Theorem 5.2:

(5.2) Each $f \in \Phi_0$ either has at most $r - 1$ local extrema excluding boundary points or else is constant in T.

THEOREM 5.2. *Suppose* $T_0 = T = [a, b]$, $\{f_i\}_{i=0}^r$ *is a C-system on* T, $1 \in \Phi_0$ *and* (5.2) *holds. Then there exists an optimal design supported on* $\{t_i\}_{i=0}^r$ *if and only if* $\Gamma(\mathbf{f}) \in R^* \cup S^*$. *In addition, the optimal design is unique if* $\Gamma(\mathbf{f}) \in R^*$, *and not unique if* $\Gamma(\mathbf{f}) \in S^*$.

PROOF. (Sufficiency.) The existence of an optimal design for $\Gamma(\mathbf{f}) \in R^*$ or for $\Gamma(\mathbf{f}) \in S^*$ follows from the corollary of Theorem 3.2 with p chosen, respectively, between the C-function and its negative or between ± 1.

(Necessity.) If $D^* = \sum_{i=0}^r c_i^* \delta_{t_i}$ is optimal, the corollary of Theorem 3.2 gives $p^* \in \Phi_0$ which attains extremum, 1 or -1, at t_i for $i = 0, \dots, r$. When $p^*(t_i)$ alternates, so does $\operatorname{sgn}(c_i^*)$. It follows from (5.1) that $\Gamma(\mathbf{f}) \in R^*$. If $p^*(t_i)$ does not alternate, then p^* has at least r local extrema. Consequently by (5.2) p^* must be one of ± 1, implying $\Gamma(\mathbf{f}) \in S^*$.

That an optimal design is not unique when $\Gamma(\mathbf{f}) \in S^*$ can be seen by perturbing $\{t_i\}_{i=0}^r$ so little that the signs of the coefficients are preserved. Finally, to show the uniqueness when $\Gamma(\mathbf{f}) \in R^*$ let D_∞^* and D_0 both be optimal with $\operatorname{supp}(D_\infty^*) = \{t_i\}_{i=0}^r$. Suppose $D_\infty^* \neq D_0$. It follows that $\operatorname{supp}(D_0)$ is not contained in $\{t_i\}_{i=0}^r$. By convexity every $D_s = sD_\infty^* + (1-s)D_0$ with $0 \le s \le 1$ is also optimal. When s is close but not equal to 1, we have $\operatorname{supp}(D_s) = \operatorname{supp}(D_\infty^*) \cup \operatorname{supp}(D_0)$. Hence, $|\operatorname{supp}(D_s)| > r + 1$. Let p_s be given for D_s by the corollary of Theorem 3.2. For some number b near 1 or -1, $p_s(t) - b$ ($\in \Phi_0$) has more than r distinct zeros but is not identically zero, contradicting the definition of C-system. □

Essentially the same results appear in Kiefer and Wolfowitz (1965) and in Studden (1968). An application of Theorem 5.2 is the estimation of θ_r when $f = \sum_{i=0}^r \theta_i f_i$. Denote the corresponding linear functional by Γ_r, that is, $\Gamma_r(f) = \theta_r$. It is easy to see that if $\{f_i\}_{i=0}^{r-1}$ is also a C-system on $[a, b]$ then $\Gamma_r(\mathbf{f}) \in R^*$. This result first appeared in Kiefer and Wolfowitz (1959) for polynomials. See Studden (1968) for more such results.

Application of Theorem 3.2 to T of dimension greater than 1 is limited by the lack of theories similar to C-system. Studden (1971) extends Theorem 5.1 to the case where T_0 is a compact convex subset with nonempty interior in a Euclidean space. Studden's optimal design, which is essentially one-dimensional, also follows from the corollary of Theorem 3.2 when a polynomial in Studden (1971) is used as the p function. We give an example to illustrate this result.

EXAMPLE 5.1. Let $T = \{\mathbf{t} \mid \mathbf{t}$ is any m vector$\}$ for a fixed m and let Φ_0 be the set of all quadratic polynomials $\{f \mid f(\mathbf{t}) = \mathbf{t}'A\mathbf{t} + \mathbf{b}'\mathbf{t} + c$ for a matrix A, a vector $\mathbf{b}$ and a number $c\}$. Let $T_0 = \{\mathbf{t} \mid \mathbf{t}'H\mathbf{t} \le 1\}$ for some matrix $H > 0$ and $\Gamma = \delta_{\mathbf{x}}$ for some $\mathbf{x} \notin T_0$. By the corollary of Theorem 3.2 with $p(\mathbf{t}) = 1 - 2\mathbf{t}'H\mathbf{t}$, an optimal design can be found supported on $\{\mathbf{0}, \pm \mathbf{x}/(\mathbf{x}'H\mathbf{x})^{1/2}\}$.

Acknowledgments. This work is based on the author's Ph.D. dissertation at University of California, Los Angeles. The author is grateful to

Professor Donald Ylvisaker for his encouragement, guidance and support. The presentation benefits greatly from the comments and suggestions of the referees, in addition to the editorial help kindly provided by Dr. H. B. Kushner.

REFERENCES

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.

FABIAN, V. (1988). Polynomial estimation of regression functions with the supremum norm error. *Ann. Statist.* **16** 1345–1368.

HOEL, P. G. (1966). A simple solution for optimal Chebyshev regression extrapolation. *Ann. Math. Statist.* **37** 720–725.

HOEL, P. G. and LEVINE, A. (1964). Optimal spacing and weighting in polynomial prediction. *Ann. Math. Statist.* **35** 1553–1560.

KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. Wiley, New York.

KIEFER, J. and WOLFOWITZ, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist.* **30** 271–294.

KIEFER, J. and WOLFOWITZ, J. (1965). On a theorem of Hoel and Levine on extrapolation. *Ann. Math. Statist.* **36** 1627–1655.

ROYDEN, H. L. (1963). *Real Analysis*. Macmillan, New York.

SACKS, J. and YLVISAKER, D. (1984). Some model robust designs in regression. *Ann. Statist.* **12** 1324–1348.

SPRUILL, M. C. (1985). Model robustness of Hoel–Levine optimal designs. *J. Statist. Plann. Inference* **11** 217–225.

STUDDEN, W. J. (1968). Optimal designs on Tchebyshev points. *Ann. Math. Statist.* **39** 1435–1447.

STUDDEN, W. J. (1971). Optimal designs for multivariate polynomial extrapolation. *Ann. Math. Statist.* **42** 828–832.

WHITTLE, P. (1971). *Optimization Under Constraints*. Wiley, New York.

NATHAN S. KLINE INSTITUTE
FOR PSYCHIATRIC RESEARCH
ORANGEBURG, NEW YORK 10962