Outlier Resistant Design-Foundations

by

R. D. Cook and L. A. Thibodeau*

Technical Report #290

Department of Applied Statistics

*L. A. Thibodeau is currently an Assistant Professor, Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115.

June, 1977

SUMMARY

A general framework based on the influence curve is developed for outlier resistant design. Several criteria for outlier resistant design are discussed. It is shown that one criterion results in a class of resistant designs which is a proper subset of the set of classical admissible designs.

Some key words: Design Admissibility, Influence curve, Invariant design, Measures of resistance, Optimal design, Outliers.

0. INTRODUCTION

Experimental design for regression problems has been a concern of scientists since the development of least squares regression. Regression design problems particularly for polynomial regression were discussed in papers by Laplace and Gergonne in the early nineteenth century. Since this time a vast body of literature has accumulated which can aid in the choice of an adequate design. However, most of the past work assumes that the experiment will produce "ideal" data. Recently some attention has shifted towards the investigation of characteristics of optimal designs in the presence of contaminated data. This attention is probably due, in part, to a contemporary interest in robust estimation and to the realization that rarely are data ideal.

The choice and justification of design selection criteria is dependent on the estimation procedure and the overall goal of the experiment. For example, when least squares estimation is used D-optimal designs produce confidence ellipsoids with minimal volume. However, the study of robust or resistant design requires a more formal (i.e. functional) connection between the design and its effects on the estimation procedure than has previously been the case.

The influence curve has proven an effective tool in the study of robust estimators. Generally, in this paper we employ the influence curve to study how the design influences the effects of an outlier on the least squares estimator. Section 1 is devoted to a brief review of optimal regression design. Invariance is considered briefly since it is often regarded as a property of fundamental importance and has special relevance to outlier resistance.

In section 2 we define outliers and discuss the sensitivity of least squares to outlying observations. The influence function (first Von-Mises derivative) for least squares is considered as a function of the design.

Our consideration of the influence function leads to several possible measures of outlier resistance which are examined in Section 3. Design admissibility with respect to optimality and outlier resistance is examined in Sections 4 and 5. It is shown that generally the goals of optimal and resistant design are in accord while there are situations in which the usual optimal designs are not resistant. The proofs of all theorems and lemmas have been relegated to an appendix.

1. REGRESSION DESIGN

Let $f' = (f_1, f_2, \ldots, f_p)$ be a vector in R^p defined by the set of linearly independent functions f_i which are assumed to be continuous on some compact space χ. An experiment consists of selecting any $x \in \chi$ and observing a random variable $Y(x)$ with regression function $E(Y|x) = f'\theta$ and constant variance σ^2. The functions f_i are assumed known while $\theta' = (\theta_1, \theta_2, ,,, \theta_p)$, an element of R^p and σ^2 are unknown. Further, assume that for two experiments x_i and x_j $(i \neq j)$ that $Y(x_i)$ and $Y(x_j)$ are uncorrelated.

An experimental design is defined by any probability measure on χ. A design problem is specified by a pair (f, χ). Exact designs for experiments of size N concentrate mass $\xi_N(x_i)$ at points x_i, $i = 1, 2 \ldots r$, subject to the restriction that $N\xi_N(x_i) = n_i$ be integral for all i. An exact design specifies that the experimenter is to take N uncorrelated observations, n_i at each x_i. The selection of an exact design is usually a difficult combinatoric problem often solvable only by an exhaustive search of a large subset of all possible exact designs. Approximate designs are not constrained by the requirement that $N\xi(x_i)$ be integral for all i. The set of all designs on χ, Ξ_χ, is the space of all probability measures on the design space χ. The support of a design, ξ, will be denoted by $S(\xi)$. In this paper we restrict our results to approximate designs. Results for exact designs will be considered in a future paper.

The least squares estimator of θ, $\hat{\theta}$, is the principal estimator considered. It is well-known that $\hat{\theta}$ is the minimum variance unbiased linear estimator of θ.

The covariance matrix of $\hat{\theta}$ is of the form

$$\mathrm{COV}(\hat{\theta}) = (\sigma^2/N)\ M^{-1}(\xi)$$

where $M(\xi)$ is of the information matrix of the design ξ;

$$M(\xi) = \int_{\chi} f(x)f'(x)\ d\xi(x).$$

M defines a map from Ξ_χ to a space information matrices, $\underline{M}(f,\chi)$, where

$$\underline{M}(f,\chi) = \{M | M = M(\xi) \text{ for some } \xi \in \Xi_\chi\}.$$

Since each $M(\xi)$ in $\underline{M}(f,\chi)$ depends only on a finite set of moments of ξ this map is not in general one to one.

Many criteria have been proposed for optimizing the selection of a design for the design problem (f,χ). Generally the criteria specify the selection of a design which minimizes some functional of the information matrix, $M(\xi)$. Justification of such criteria is often based on the properties of the resulting least squares estimator $\hat{\theta}$. Kiefer (1975) introduced a large class of such measures defined by the functional Φ_q;

$$\Phi_q(\xi) = [\mathrm{tr}\ \{M^{-q}(\xi)\}/p]^{1/q} \qquad 0 < q < \infty$$

$$\Phi_o(\xi) = \lim_{q \downarrow o} \Phi_q(\xi) = \{\det M^{-1}(\xi)\}^{1/p}$$

$$\Phi_\infty(\xi) = \lim_{q \to +\infty} \Phi_q(\xi) = \max_i \lambda_i^{-1}(\xi)$$

where $\lambda_i(\xi)$ is the i-th eigenvalue of $M(\xi)$. It is well-known [Keifer (1959), Fedorov (1972)] that these criteria are relevant to estimation and hypothesis testing. Two classical criteria are D-optimality (q=0) and A-optimality (trace-optimality, q=1).

Design criteria based on prediction variances have also been widely used. The predicted value at a point x_o is

$$\hat{Y}(x_o) = f'(x_o)\hat{\theta} \text{ and}$$

$$\text{Var }(\hat{Y}(x_o)) = (\sigma^2/N)\ f'(x_o)M^{-1}(\xi)f(x_o).$$

For notational convenience let

$$d(x_o,\xi) = f'(x_o)M^{-1}(\xi)f(x_o)$$

and refer to $d(x_o,\xi)$ as the variance function for ξ at x_o. A minimax design is a design which minimizes the maximum of the variance function over χ.

Invariance will be an important concern in our development of criteria for outlier resistant design. In the remainder of this section we consider some invariance characteristics of the Φ_q criteria. Let G be a group of transformations on χ such that, for each $g \in G$, there is a corresponding linear transformation $\bar{g}$ on R^p which can be represented as a pxp nonsingular matrix such that for all x and θ,

$$f'(x)\ \theta = f'(gx)\ \bar{g}\theta = f'(x)\ \bar{g}^{-1}\bar{g}\theta.$$

Thus, the elements of G define the mappings

$$(f,\chi) \xrightarrow{g} (f,g\chi)$$

and

$$(f(x),\ M^{-1}(\xi)) \xrightarrow{g} ((\bar{g}^{-1})'f(x),\ \bar{g}\ M^{-1}(\xi)\bar{g}').$$

This group includes, among others, scale and location transformations.

Note that the D-optimal design is invariant under G and thus D-optimal for all $g \in G$. However, in general for the Φ_q-optimal design for (f,χ), say ξ_q, there may exist $g \in G$ such that $g\xi_q$ is not Φ_q-optimal for $(f,g\chi)$. In particular let G^* be the group acting on χ such that $\bar{G}^*$ is the group of all pxp non-singular matrices then for $q \neq 0$ there exists g such that $g\xi_q$ is not Φ_q optimal. Lemma 1.1 gives necessary and sufficient conditions for a design measure to be invariant under such a group.

Lemma 1.1. If $\bar{G}^*$ is the group of all pxp non-singular matrices then a maximal invariant with respect to G^* acting on $(f(x), M^{-1}(\xi))$ is

$$d(x,\xi) = f'(x)M^{-1}(\xi)\, f(x).$$

The principal consequence of Lemma 1.1 is that if ξ is Ψ-optimal then ξ is invariant under G^* if and only if there exists a function H such that

$$\Psi(\xi) = H(d(x,\xi)).$$

That is, Ψ is a function only of the variance function $d(x,\xi)$. Consequently a minimax design is also an invariant design under G^*. The Equivalence Theorem proved by Kiefer and Wolfowitz (1960) established the implied duality between D-optimal and minimax designs; a design is D-optimal if and only if it is a minimax design. Other groups G may yield a larger class of invariant designs and thus additional equivalences.

2. OUTLIERS AND INFLUENCE.

An essential first step in any development of robust theory is to describe the deviations from assumptions we wish to guard against. Much of the original work in robustness has been concerned with symmetric heavy tailed alternatives to the normal distribution. For our purposes, a somewhat simpler approach to outliers will suffice. Suppose that $y(x)$, an observation on some model $f'(x)\theta$, has an "aberration" c added to it so we observed $(y(x) + c)$, an outlier. It is assumed that the occurrence and magnitude of an outlier at a design point x are independent of x. Clearly c cannot be directly measured since $y(x)$ is a realization of a random variable $Y(x)$. Thus we must be concerned with "wild" observations; that is, those observations which markedly deviate from their expected values.

In what follows, we employ the influence curve to measure the effects of an abberation on the least squares estimator and to aid in understanding how the design of an experiment may be used to lessen its effects. As background, we first briefly review the influence curve.

Hampel (1974) discusses the use of the influence curve in the study of robust estimators. (See also Andrews, et. at (1972)). Let R be the real line, T be a real-valued functional defined on some subset of the set of all probability measures on R, and let F denote a probability measure on R for which T is defined. Then the <u>influence curve</u>, $IC_{T,F}(\cdot)$, of the functional ("estimator") T at F (the underlying probability distribution on R) is defined pointwise as

$$IC_{T,F}(y) = \lim_{\varepsilon \downarrow 0} \{T((1-\varepsilon)F + \varepsilon\delta(y)) - T(F)\}/\varepsilon$$

if this limit is defined for every y in R. $\delta(y)$ is the probability measure with mass 1 at y.

The influence curve $IC_{T,F}(y)$ measures the influence an observation of magnitude y has on the estimator T. For example, a simple least squares estimator is the arithmetic mean which may be defined by the functional $T(F) = \int_R y \, dF(y)$ for all probability measures F with first moments. If the mean of F is μ then the influence curve of T is

$$\begin{aligned} IC_{T,F}(y) &= \lim_{\varepsilon \downarrow 0} \{(1-\varepsilon)\mu + \varepsilon y - \mu\}/\varepsilon \\ &= y - \mu \end{aligned}$$

Thus the mean T has a linear influence curve; each observation y influences T by $y - \mu$, an influence linear in the "error" in y. Clearly the influence function for the mean is unbounded and thus the potential influence of an observation is similarily unbounded. Robust estimators of location such as Windsorized or trimmed means alter the influence curve by truncation and, thus, bound the gross error sensitivity (see Hampel, 1974).

Least squares estimation for a linear model $f'(x)\theta$ results in a vector valued estimator necessitating the extension of the influence curve to vector-valued functionals: Let T be a vector-valued mapping (estimator) from a subset of probability measures on, say, R^n into R^p. Further let F be a probability measure in the domain of T and y be a point in R^n. Then the vector-valued influence function for T at F is defined pointwise as above. For least squares Hinkley (1977) states the following lemma without proof:

Lemma 2.1. Let F be the conditional distribution of Y(x) such that

$$E_F(Y|x) = f'(x)\theta.$$

Further let H be the joint distribution function of the design point x and the response variable Y, $H = \xi x F$, such that

$$E_H\{ \begin{matrix} f(x) \\ y \end{matrix} (f'(x), y)\} = \begin{bmatrix} M(H) & \gamma(H) \\ \gamma'(H) & \tau(H) \end{bmatrix} = \begin{bmatrix} M(\xi) & \gamma(H) \\ \gamma'(H) & \tau(H) \end{bmatrix}$$

(Note that x may have design or probability measure. For the purposes of this paper x has design measure, ξ.)

Define the least squares functional θ by

$$\theta(H) = M^{-1}(\xi)\gamma(H).$$

Then the influence function of θ at (x,y) for $x \in S(\xi)$ is

$$I_\xi(e,x;\theta) = eM^{-1}(\xi)f(x)$$

where $e = y - f'(x)\theta$. (By assumption e is independent of x and θ)

Clearly the influence function for least squares is linear in e, the error in y(x) at x, for each estimator $\hat{\theta}_j$, $j = 1, ,\ldots p$. However, the influence function is a function of both x, the design point at which the observation is taken, and the design ξ. Therefore, the influence that an error of magnitude e has on $\hat{\theta}_j$ depends on both the design point x where e occurs and the design points specified by ξ. However, the influence of an error of magnitude e at x does not depend on the errors at the other design points. It follows that the influence function is a measure of sensitivity not constrained to the single outlier case. Note the implicit assumption that the design point x is determined without error. The following example illustrates the role of the influence curve in experimental design.

<u>Example.</u> For linear regression with

$$f'(x) = (1,x), \ \chi = [-1,1],$$

and

$$E(Y|x) = \theta_o + \theta_1 x$$

consider the designs $\xi_{a,\alpha}$ of the form

$$\xi_{a,\alpha}(-1) = \xi_{a,\alpha}(1) = \frac{1-\alpha}{2}, \ \xi_{a,\alpha}(-a) = \xi_{a,\alpha}(a) = \alpha/2$$

where $0 \leq \alpha \leq 1$ and $0 \leq a \leq 1$.

For such designs,

$$M^{-1}(\xi_{a,\alpha}) = \begin{bmatrix} 1 & 0 \\ 0 & [(1-\alpha) + a^2\alpha]^{-1} \end{bmatrix}$$

and

$$I_{\xi a,\alpha}(e,x;\theta) = e\begin{bmatrix} 1 \\ x[(1-\alpha) + a^2\alpha]^{-1} \end{bmatrix}.$$

Clearly for all (a,α) and $x \in S(\xi_{a,\alpha})$ the influence function for θ_o is unchanged. In fact $I_\xi(e,x;\theta_o)$ has the same influence function as the sample mean for all ξ symmetric on $[-1,1]$. Consider now the influence function for θ_1 on the support of the design $\xi_{a,\alpha}$. For x = a, 1

$$I_{\xi a,\alpha}(e,x;\theta_1) = ex((1-\alpha) + a^2\alpha)^{-1} = -I_{\xi a,\alpha}(e,-x;\theta_1).$$

Thus if $a = k^{-1}$ then an error of magnitude e at the point x = +1, -1 will be k times as influential as the same error at x = +a,-a. Moreover as $a \to 0$ the influence on $\hat{\theta}_1$ at x = a goes to 0; that is, observations at x = 0 have no influence on the least squares estimator $\hat{\theta}_1$.

Clearly some designs are preferrable to others in terms of influence. Selecting a design that is outlier resistant requires that the design selected be relatively insensitive to outlying observations. Some measures that may be optimized (minimized) over (a,α) are

(i) $\sup_{x \in S(\xi_{a,\alpha})} I_{\xi_{a,\alpha}}(e,x;\theta_1)$,

(ii) $I_{\xi_{a,\alpha}}(e,1;\theta_1) - I_{\xi_{a,\alpha}}(e,a;\theta_1)$,

and

(iii) $\sum_{x \in S(\xi_{a,\alpha})} \{I_{\xi_{a,\alpha}}(e,x;\theta_1)\}^2 \xi_{a,\alpha}(x)$.

Clearly the D,A-optimal design $(a=1)$ minimizes each of these measures. This concludes the example.

Recall that the influence function $I_\xi(e,x;\theta)$ is unbounded and linear in e for all designs ξ. Consequently the design determines the slope of each of the influence curves $I_\xi(e,x;\theta_j), j = 1,2,\ldots,p$, as a function of x. Measures of influence may, thus, be based on these slopes which will be represented by

$$I_\xi(x;\theta) = M^{-1}(\xi)f(x) \quad (p\times 1).$$

This is the only part of the influence function that can be controlled by the design. To distinguish between $I_\xi(x;\theta)$ and $I_\xi(e,x;\theta)$ we shall refer to the former as the design influence function (DIF). The following lemmas give some relevant properties of the DIF.

Lemma 2.2. The DIF for the least squares predictor $\hat{Y}(x) = f'(x)\hat{\theta}$ is

$$I_\xi(x, f'(x)\theta) = d(x,\xi)$$

It follows from lemmas 1.1 and 2.2 that only measures of the form

$$H(ed(x,\xi)) = H(I_\xi(e,x;f'(x)\theta))$$

are invariant measures under G*. A consequence of invariance is the restriction of outlier resistance measures to those which depend only on the DIF for the specific linear combinations $f'(x)\theta$, $(x \in S(\xi))$.

Recall that the occurrance and magnitude of an outlier are assumed to be independent of the design; Given the presence of an outlier the probability that it occurs at $x \in S(\xi)$ is proportional to $\xi(x)$. Letting $I_\xi(\theta)$ denote the average design influence of a single randomly occurring outlier we have

$$I_\xi(\theta) = \int_\chi I_\xi(x;\theta)d\xi(x).$$

The following lemma gives the form of $I_\xi(\theta)$ when the model contains a constant term:

Lemma 2.3. If for some $j, f_j(x) = 1$ for all $x \in \chi$ then for all $\xi \in \Xi_\chi$ for which $M(\xi)$ is positive definite

$$I_\xi(\theta_i) = \delta_{ij}.$$

Where

$$I_\xi(\theta)' = (I_\xi(\theta_1), I_\xi(\theta_2), \ldots, I_\xi(\theta_p))$$

and δ_{ij} is the Kronecker delta-

$$\delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j. \end{cases}$$

This result implies that if the same design is used repeatedly then on the average the effects of the outlying observations will cancel. Of course, this is of little consequence for a single replicate of the design. This result is not directly extendable to general models without overall means. The next lemma gives the "covariance" matrix for $I_\xi(x;\theta)$.

Lemma 2.4. $\int_{\mathcal{X}} (I_\xi(x;\theta))(I_\xi(x;\theta))' d\xi(x) = M^{-1}(\xi)$.

In view of Lemma 2.3, the diagonal elements of $M^{-1}(\xi)$ measure both the variance and magnitude of the influence of a randomly occurring outlier on the elements of $\hat{\theta}$.

3. MEASURES OF RESISTANCE.

In this section we consider measures of outlier resistance as derived from the influence curve.

3.1 BOX-DRAPER AND HUBER CRITERIA.

Past formulations of outlier resistant measures have resulted directly from the use of the residuals , $y(x) - \hat{y}(x)$, to detect outlying observations. These residuals are invariant under G* and thus measures of their resistance to outlying observations are invariant.

Box and Draper (1975) consider one such invariant measure which they describe as a measure of the discrepancy in $\hat{y}(x)$ caused by an outlying observation. The measure they consider is equivalent to

$$r = \int_{\mathcal{X}} \{d(x,\xi)\}^2 d\xi(x) \quad ,$$

the squared variance function averaged with respect to the design measure.

Their justification of this measure is based on exact design considerations and will not be reproduced here. By Lemma 2.2, r is proportional to the averaged squared influence of $f'(x)\hat{\theta}$. Of course, Box and Draper's goal is to minimize r by an appropriate design selection.

Huber (1975) mentions an invariant measure of outlier resistance which is also based on $\hat{y}(x)$. Huber suggests the measure

$$\bar{r} = \sup_{x \in S(\xi)} d(x,\xi),$$

the maximum of the variance function over the support of ξ.

Again, by Lemma 2.2, $\bar{r}$ is the maximum design influence for $f'(x)\hat{\theta}$.

<u>Lemma 3.1</u>. If ξ^* is a D-optimal design for (f,χ) then

$$1) \quad \inf_{\xi \in \Xi_\chi} r = \int_\chi \{d(x,\xi^*)\}^2 \, d\xi^*(x) \doteq p^2$$

and

$$2) \quad \inf_{\xi \in \Xi_\chi} \bar{r} = \sup_{x \in S(\xi^*)} d(x,\xi^*) = p.$$

Consequently both measures r and $\bar{r}$ are minimized by a D-optimal design. In fact, if a measure of resistance is invariant and defined by its properties on the support of ξ then such measures would be expected to be optimized by the D-optimal design.

The measures r and $\bar{r}$ of ξ are, as noted above, measures of the influence of $f'(x)\hat{\theta}$ only for those x values in the support of ξ. The principal criticism of r and $\bar{r}$ is that they concentrate on the influence of $\hat{y}(x)$ while potentially ignoring elements of $\hat{\theta}$. A design ξ with $f_i(x)$ small for all x in $S(\xi)$ may result in $I_\xi(e,x;\theta_i)$ large in absolute value with negligible effect or r or $\bar{r}$.

It is important to note that a fundamental difference between optimal design and outlier resistant design is that in the former a measure is evaluated over χ whereas in the latter a measure is evaluated only over $S(\xi)$. This difference is a consequence of the assumption that the design points x are determined without error and thus observations are only taken at the points in $S(\xi)$.

3.2 ALTERNATIVE CRITERIA

A large set of resistance measures similar to the Box-Draper and Huber criteria can be generated by the inner products

$$R(\xi,C,x) = (I_\xi(x;\theta))'C(I_\xi(x;\theta))$$
$$= f'(x)M^{-1}(\xi)CM^{-1}(\xi)f(x)$$

for C non-negative definite. Specifically, consider the measures

$$R(\xi,C) = \int_\chi R(\xi,C,x)d\xi(x)$$
$$= \operatorname{tr} M^{-1}(\xi)C$$

and

$$\bar{R}(\xi,C) = \sup_{x \in S(\xi)} R(\xi,C,x).$$

The selection of C depends on the importance of the resistance properties of individual estimators $\hat{\theta}_j$, $j = 1,\ldots,p$. The following lemma shows the existence of optimal designs for these measures.

Lemma 3.2. Let g be an element of G* such that $\bar{g}' = B$ where B is any pxp non-singular matrix such that $B'B = C$. If $g^{-1}\xi_c$ is A- optimal for the problem $(f, g\chi)$ then

1) $\inf_{\xi \in \Xi_\chi} R(\xi, C) = R(\xi_c, C)$

and

2) $\inf_{\xi \in \Xi_\chi} \bar{R}(\xi, C) = \bar{R}(\xi_c, C)$.

Consequently, if either measure of influence, $R(\xi,C)$ or $\bar{R}(\xi,C)$, is consistent with the resistance objectives for the problem (f,χ), Lemma 3.2, shows that a solution exists. Moreover Lemma 3.2 gives a specific solution which is not difficult to find.(c.f. Fedorov (1972) page 137.)

Alternatively, one might consider design criteria which are scalar-valued functions of

$$\int_\chi I_\xi(x;\theta)\, I'_\xi(x;\theta)\, d\xi(x)$$

However, by Lemma 2.4, we are lead back to considering functions of the information matrix. Consequently, designs which are optimal with respect a measure based on the information matrix may also be regarded as outlier resistant with respect to that measure. Of course, it is of little help to be able to say that any Φ_q-optimal design is outlier resistant in some sense.

The previous discussion illustrates the ways in which resistant design criteria might be constructed from the DIF. Apart from r and $\bar{r}$, these criteria are rather arbitary and, without further refinement, are of limited practical use. The Box-Draper and Huber criteria may prove useful for experiments in which prediction is of primary goal. However, they seem less desirable for the purpose of estimation. (Recall that the principal criticism of these measures is that the potentially ignore esitmates, $\hat{\theta}_j$, that may be very sensitive to outliers.) In the next section we discuss a useful broad criterion that may be used to check the robustness characteristics of a design, particularly when estimation is the goal of the experiment.

4. ADMISSIBILITY.

The choice of a single criterion (measure) for design selection, whether optimal or outlier resistant, results in an unknown degree of specificity. Kiefer (1975) and others have noted that a Ψ-optimal design may be very inefficient with respect to another measure, say Ψ', even when Ψ and Ψ' are apparently consistent in formulation. This concern with specificity is directly applicable to outlier resistant design. An alternative to the selection of a single design is to partition the design space Ξ_χ using broad principles of design goodness. The resulting admissible design space may be regarded as a space of designs which are efficient in a wide sense.

The classical definition of design admissibility (Elfving, 1958) follows from the fact that most optimality criteria are nonincreasing on $\underline{M}(f,\chi)$.

Definition 4.1. A design ξ is admissible for (f,χ) if and only if there does not exist $\xi^* \in \Xi_\chi$ such that

$$M(\xi^*) \geq M(\xi).$$

($A \geq B$ if $A - B$ is non-negative definite and $A - B \neq 0$.) Note that if $M(\xi) > 0$ then $M(\xi^*) \geq M(\xi)$ if and only if $M^{-1}(\xi) \geq M^{-1}(\xi^*)$. Thus, if a design ξ is inadmissible and $M(\xi) > 0$ there exists a contrast vector c and an admissible design ξ^* such that;

i) $c'M^{-1}(\xi)c > c'M^{-1}(\xi^*)c$

and

ii) for all possible contrasts $\tilde{c}$

$\tilde{c}'M^{-1}(\xi)\tilde{c} \geq \tilde{c}'M^{-1}(\xi^*)\tilde{c}$.

Note also that the set of Φ_q, $0 \leq q \leq \infty$, optimal designs are all admissible.

The concept of admissibility partitions both the design space Ξ_χ and $\underline{M}(f,\chi)$. Let $\underline{M}^*(f,\chi)$ denote the subspace of $\underline{M}(f,\chi)$ corresponding to the space of admissible designs. Specifically, $M^*(f,\chi) = \{M(\xi)\,|\,M(\xi) \in \underline{M}(f,\chi)$ and ξ is admissible$\}$. $M^*(f,\chi)$ is a space of non-negative definite matrices (not necessarily positive definite) with the properties:

Theorem 4.1.

1. $\underline{M}^*(f,\chi)$ is a boundary set of $\underline{M}(f,\chi)$.
2. If M is an element of $\underline{M}^*(f,\chi)$ then there exists an admissible design ξ such that $M = M(\xi)$ and the support of ξ has no more than $p(p+1)/2$ points.
3. For all $g \in G^*, M(\xi) \in \underline{M}^*(f,\chi)$ if and only if $M(g\xi) \in \underline{M}^*(f,g\chi)$.

The principle result of this theorem is that $\underline{M}^*(f,\chi)$ is an invariant boundary set of $\underline{M}(f,\chi)$.

Recall that $M^{-1}(\xi)$ is a measure of the "covariance" of the vector $I_\xi(x;\theta)$ and thus admissible designs are preferred designs with respect to this measure. It may be desireable for outlier resistance to have $M^{-1}(\xi)$ diagonal, however beyond this it is unclear what structures for $M^{-1}(\xi)$ are to be preferred. An omnibus measure of outlier resistance is the averaged squared influence vector

$$ASI_{\xi}(\theta) = (m^{11}(\xi),\dots,m^{pp}(\xi))$$
$$= \text{diagonal } \{M^{-1}(\xi)\}.$$

$m^{ii}(\xi)$ is a measure of both the magnitude of $I_{\xi}(x;\theta_i)$ and its variability over $S(\xi)$; if a design ξ results in a estimator $\hat{\theta}_i$ that is extremely sensitive to outlying observations then $m^{ii}(\xi)$ will be large.

Clearly the design space Ξ_{χ} or equivalently the information matrix space $M(f,\chi)$ cannot be completely ordered by $ASI_{\xi}(\theta)$. However $ASI_{\xi}(\theta)$ does induce a partial ordering of these spaces.

Definition 4.2. If ξ_1 and ξ_2 are elements of Ξ_{χ} then

1) ξ_1 is preferred to ξ_2, $(\xi_1 > \xi_2)$ if

$$m^{ii}(\xi_1) \le m^{ii}(\xi_2)$$

for all i, i=1,...p, with strict inequality for at least one i.

2) ξ_1 is not preferred to ξ_2, $(\xi_1 \approx \xi_2)$ if neither

$\xi_1 > \xi_2$ nor $\xi_2 > \xi_1$.

This preference ordering is invariant only under changes of scale and permutation of the parameter vector . If an ordering is required to be invariant under larger transformation groups then such a preference ordering cannot depend on the particular parameterization. Generally, invariant orderings cannot be sensitive to the outlier resistance of individual least squares estimators. Consequently, if C' is a pxp matrix of contrasts then the preferrence ordering of Ξ_{χ} for (f,χ) may not be the same as the preference ordering of Ξ_{χ} for $(C^{-1}f,\chi)$. Examples in which $\hat{\theta}$ is apparently resistant to outliers while linear combinations of the elements of $\hat{\theta}$ are sensitive to outliers are easily constructed. This means that a model must be parameterized in terms of the parameters of interest prior to any resistance considerations based on $ASI_{\xi}(\theta)$.

The preference ordering $>$ provides a method of selecting an outlier resistant subset of Ξ_χ:

Definition 4.2. A design $\xi \in \Xi_\chi$ is robust admissible or R-admissible for (f,χ) if there does not exist $\xi' \in \Xi_\chi$ such that $\xi' > \xi$.

Admissibility partitions Ξ_χ and induces a partition of $\underline{M}(f,\chi)$. R-admissibility results in a similar subspace of $\underline{M}(f,\chi)$. Let $\underline{M}^R(f,\chi)$ denote the space of information matrices corresponding to the R-admissible designs. Designs corresponding to elements of $\underline{M}^R(f,\chi)$ are outlier resistant designs. Throughout this paper outlier resistance has been shown to be associated with optimality criteria. Box and Draper (1975) expressed concern that robustness considerations might be inconsistent with classical optimality considerations. For outlier resistance (robustness) these concerns are consistent and in harmony with the classical optimality criteria as shown by the following theorem.

Theorem 4.2. For any (f,χ) all R-admissible designs are admissible. However, all admissible designs are not necessarily R-admissible:

$$\underline{M}^R(f,\chi) \subset \underline{M}^*(f,\chi).$$

Therefore R-admissible designs are both outlier resistant and admissible. However optimal designs are not always outlier resistant. Consequently while an optimal design must be verified outlier resistant all outlier resistant designs are optimal. In emphasis, we do not propose that any one measure of outlier resistance is preferrable in all problems. Rather, we would suggest, as with optimality criteria, that a broad range of criteria are applicable and available for the evaluation of a particular design with respect to outlier resistance. R-admissibility is defined to encompass many of the measures of relevance. The next section is devoted to conditions for and some further thoughts on R-admissibility.

5. OUTLIER RESISTANT DESIGNS.

The following lemma shows that there always exists at least one R-admissible design.

Lemma 5.1. If ξ_A is A-optimal for (f,χ) then ξ_A is R-admissible.

Consequently, since Φ_1 is a linear-optimal criterion, the trace-optimal design is, for most problems (f,χ), easily generated by an iterative algorithm. However, the principal thrust of our development of outlier resistance is that some measure of resistance should always be used to evaluate a design choice. If for (f,χ) a classical optimality criteria is applicable then it is necessary to determine if the resulting optimal design is R-admissible. If not, then other optimal designs might be examined to find one that satisfies both optimality and robustness considerations. The remainder of this section will be devoted to methods of determining if a specific design is R-admissible.

Lemma 5.2. If ξ is both D-optimal and orthogonal for (f,χ) then ξ is R-admissible.

The consequences of Lemma 5.2 are clear when considering the classical balanced design situation. It is well known that most balanced othogonal designs are D-optimal. See for example Kiefer [(1958), (1975)]. Hence these classical balanced designs are R-admissible.

The following example illustrates the specificity necessary to determine R-admissibility even when the designs satisfy the same optimality criteria.

Example. Consider polynomial regression on $\chi = [-1,1]$ with

$f'(x) = (1,x,\ldots,x^p)$.

1) For $p = 1$ ξ_1 is D-optimal if

$$M(\xi_1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = M^{-1}(\xi_1).$$

By Lemma 2.8, ξ_1 is R-admissible since it is both D-optimal and orthogonal.

2) For p = 2 ξ_2 is D-optimal if

$$M^{-1}(\xi_2) = \begin{bmatrix} 3 & 0 & -3 \\ & 1.5 & 0 \\ & & 4.5 \end{bmatrix}.$$

It can be shown that if there exists a design ξ such that $\xi > \xi_2$ then ξ has support (-1,b,1) for some $b \in (-1,1)$. We have computationally verified that there does exist such a ξ. Thus, ξ_2 is R-admissible.

3) For p = 3 ξ_3 is D-optimal if

$$\xi_3(-1) = \xi_3(-1/\sqrt{5}) = \xi_3(1/\sqrt{5}) = \xi_3(1) = 1/4.$$

for which

$$M^{-1}(\xi_3) = \begin{bmatrix} 3.25 & 0 & -3.75 & 0 \\ & 15.75 & 0 & -16.25 \\ & & 6.25 & 0 \\ & & & 18.75 \end{bmatrix}.$$

Consider the design ξ a linear combination of ξ_3 and the point mass $\delta(x)$ at $x = -.34$; $\xi = .94\, \xi_3 + .06\, \delta(-.34)$. Then

$$M^{-1}(\xi) = \begin{bmatrix} 3.02 & .71 & -3.52 & -.72 \\ & 15.54 & -.75 & -16.05 \\ & & 6.15 & .76 \\ & & & 18.69 \end{bmatrix}.$$

Clearly $\xi > \xi_3$. Thus the D-optimal design is not R-admissible. This concludes the example.

In the previous example a single point augmentation of ξ_3 was used to show that ξ_3 is not R-admissible. Typically procedures for iterative construction of better designs with respect some criteria depend on single point augmentation of an initial design. Clearly if there exists x an element of χ and α, $0 < \alpha < 1$, such that $\xi_{x,\alpha}$ is preferred to ξ, $\xi_{x,\alpha} > \xi$, where

$$\xi_{x,\alpha} = (1-\alpha)\xi + \alpha\delta(x),$$

then ξ is not R-admissible. Moreover it seems reasonable to suspect that if there does not exist a pair (x,α) such that $\xi_{x,\alpha} > \xi$ then ξ is R-admissible. Consider the following two lemmas:

<u>Lemma 5.3</u>. If ξ is an element of Ξ_χ such that $M(\xi)$ is positive definite for (f,χ) then ξ is not R-admissible if

$$\max_{x\varepsilon\chi} \min_i K_\alpha(x,i,\xi) > 1-\alpha \text{ for some } \alpha,\ 0 < \alpha < 1.$$

Where

$$K_\alpha(x,i,\xi) = \frac{\{m^i(\xi)' f(\chi)\}^2}{m^{ii}(\xi)} - \alpha\, d(x,\xi)$$

and

$$m^i(\xi)' = (m^{i1}(\xi), m^{i2}(\xi), \ldots, m^{ip}(\xi)).$$

<u>Lemma 5.4.</u> If ξ and ξ^* are elements of Ξ_χ and $\xi > \xi^*$ then for all α, $0 \leq \alpha < 1$

$$(1-\alpha)\xi + \alpha\xi^* > \xi^*.$$

Lemma 5.3 provides sufficient conditions for the existence of a single point augmented design $\xi_{x,\alpha}$ preferred to ξ. Moreover by Lemma 5.4 if there exists a preferred design ξ^* then all linear combinations of ξ^* and ξ are preferred to ξ. The following lemma partially ties the R-admissibility of a design ξ to the existance of a preferred single point design.

<u>Lemma 5.5</u>. If χ is closed and convex and $K_\alpha(x,i,\xi)$ is concave in x for all α and i ($0 \leq \alpha < 1$, $i = 1,2...p$), then ξ is R-admissible if and only if there does not exists a pair (x^*,α^*), $x^* \in \chi$ and $0 \leq \alpha^* < 1$, such that ξ_{x^*,α^*} is preferred to ξ.

Lemma 5.5 is a fundamental result as it delineates the rather severe restrictions necessary for the use of single point augmentation. The assumption that $K_\alpha(x,i,\xi)$ is concave in x for all α and each i is very difficult to satisfy. In general, given the structures of this problem it may seem reasonable to conjecture that if a preferred design exists then there exists a preferred design $\xi_{x,\alpha}$. As a counterexample consider simple linear regression on $\chi = [-1,1]$. The D-optimal design ξ_1 with $\xi_1(-1) = \xi_1(1) = \frac{1}{2}$ is preferred to ξ where $\xi(-1) = \xi(-1/\sqrt{5}) = \xi(1/\sqrt{5}) = \xi(1) = 1/4$. However, there does not exist a design $\xi_{x,\alpha}$ preferred to ξ.

In the previous sections we have developed foundational principles for outlier resistant design and discussed a consequence of these principles, R-admissibility. In this section, while we have shown that it is always possible to find an R-admissible design, we have also shown that it may be very difficult to establish the R-admissibility of an arbitrary design. In fact probably the most generally useful result of Lemma 5.5 is that single point augmentation, the foundation of all optimal design generating algorithms, will usually not work. Obviously the need here is for efficient algorithms to generate the elements of $\underline{M}^R(f,\chi)$, or preferrably, theoretical insights into the general structure of $\underline{M}^R(f,\chi)$. While these objectives are yet to be met, for arbitrary (f,χ), it must be noted that usually in a particular problem insights into structure of $\underline{M}^R(f,\chi)$ can be derived from known results in design theory.

Utilizing these insights it is often a simple matter to characterize $\underline{M}^R(f,\chi)$ or minimally $\underline{M}^*(f,\chi)$.

APPENDIX

Proofs

Lemma 1.1

For $g \in G^*$,

$$d(gx, g\xi) = f'(x)\bar{g}^{-1}\bar{g} \; M^{-1}(\xi)\bar{g}'(\bar{g}^{-1})'f(x)$$
$$= d(x,\xi).$$

Thus $d(x,\xi)$ is invariant. To show that $d(x,\xi)$ is maximal it is necessary to show that if

$$d(x,\xi) = d(\tilde{x},\tilde{\xi})$$

or equivalently

$$f'(x)M^{-1}(\xi)f(x) = f'(\tilde{x})M^{-1}(\tilde{\xi})f(\tilde{x})$$

then there exists $\bar{g}$ an element of $\bar{G}^*$ such that

$$((\bar{g}^{-1})'f(x), (\bar{g}^{-1})'M(\xi)\bar{g}^{-1}) = (f(\tilde{x}), M(\tilde{\xi})).$$

Let T and $\tilde{T}$ denote two lower triangular matrices such that $T\,T' = M(\xi)$ and $\tilde{T}\,\tilde{T}' = M(\tilde{\xi})$. Consider

$$(\bar{g}^{-1})' = \tilde{T}\,\Gamma\,T^{-1} \text{ for some } \Gamma \text{ a pxp orthogonal matrix.}$$

Then

$$(\bar{g}^{-1})M(\xi)\bar{g}^{-1} = (\tilde{T}\,\Gamma\,T^{-1})(TT')(\tilde{T}\,\Gamma\,T^{-1})'$$
$$= \tilde{T}\,\tilde{T}' = M(\tilde{\xi}).$$

However

$$(\bar{g}^{-1})'f(x) = \tilde{T}\,\Gamma\,T^{-1}f(x)$$

and we require $(\bar{g}^{-1})'f(x) = f(\tilde{x})$. Thus we must have

$$\Gamma\,T^{-1}f(x) = \tilde{T}^{-1}f(\tilde{x}).$$

Such a Γ clearly exists since $d(x,\xi) = d(\tilde{x},\tilde{\xi})$ or equivalently $\| \Gamma T^{-1} f(x) \| = \| \tilde{T}^{-1} f(\tilde{x}) \|$. Therefore there exists g an element of G^* corresponding to $\tilde{g}' = T \Gamma \tilde{T}^{-1}$ such that $f(gx) = f(\tilde{x})$ and $M(g\xi) = M(\tilde{\xi})$. Therefore $d(x,\xi)$ is maximal invariant.

Lemma 2.1

Let $z' = (f'(x),y)$ and $\delta(z)$ be the probability measure with mass 1 at z. Then by definition

$$IC_{\theta,H}(z) = \lim_{\varepsilon \downarrow 0} \{\theta((1-\varepsilon)H + \varepsilon\delta(z)) - \theta(H)\}/\varepsilon.$$

Now,

$$\theta((1-\varepsilon)H + \varepsilon\delta(z)) = M^{-1}((1-\varepsilon)\xi + \varepsilon\delta(x))\gamma((1-\varepsilon)H + \varepsilon\delta(z)),$$

$$M((1-\varepsilon)\xi + \varepsilon\delta(x)) = (1-\varepsilon)M(\xi) + \varepsilon f(x)f'(x)$$

and

$$\gamma((1-\varepsilon)H + \varepsilon\delta(z)) = (1-\varepsilon)\gamma(H) + \varepsilon\, yf(x).$$

Using the following identity given by Fedorov (1972), page 106,

$$M((1-\varepsilon)\xi + \varepsilon\delta(x))^{-1} = (1-\varepsilon)^{-1}\{M^{-1}(\xi) - \frac{\varepsilon M^{-1}(\xi)f(x)f'(x)M^{-1}(\xi)}{1-\varepsilon + \varepsilon f'(x)M^{-1}(\xi)f(x)}\}$$

it follows that

$$\lim_{\varepsilon \downarrow 0} \{\theta((1-\varepsilon)H + \varepsilon\delta(z)) - \theta(H)\}/\varepsilon = M^{-1}(\xi)f(x)y - M^{-1}(\xi)f(x)f'(x)\theta(H)$$

$$= M^{-1}(\xi)f(x)(y - f'(x)\theta)$$

$$= eM^{-1}(\xi)f(x).$$

Thus completing the proof.

Lemma 2.2

The proof of this lemma is to show that the influence function is a linear function of θ, a property that is clear from the definition.

Lemma 2.3

Take $\xi \in \Xi_\chi$ with $M(\xi)$ positive definite and let

$$M(\xi) = (m_1(\xi), \ldots, m_p(\xi))$$

where

$$m_i(\xi) = \int_\chi f(x) f_i(x) d\xi(x) \qquad (p\times 1)$$

and

$$M^{-1}(\xi) = (m^1(\xi), \ldots, m^p(\xi)) .$$

Since

$$I_\xi(x;\theta_i) = m^i(\xi)' f(x) ,$$

it follows that

$$I_\xi(\theta_i) = \int_\chi I_\xi(x;\theta_i) d\xi(x)$$

$$= m^i(\xi)' \int_\chi f(x) d\xi(x) .$$

Suppose $f_j(x) = 1$, for all $x \varepsilon \chi$, then

$$m_j(\xi) = \int_\chi f(x) d\xi(x) .$$

Consequently

$$I_\xi(\theta_i) = m^i(\xi)' m_j(\xi) .$$

However $M^{-1}(\xi)M(\xi) = \{\delta_{ik}\} = \{m^i(\xi)' m_k(\xi)\}$

and thus $I_\xi(\theta_i) = \delta_{ij}$ as claimed.

Lemma 2.4

The proof of this lemma is evident from the form of the integral.

Lemma 3.1

First we note that for all ξ with $M(\xi)$ positive definite

$$\int_{\chi} d(x,\xi)d\xi(x) = \text{tr } M^{-1}(\xi)M(\xi) = p .$$

Thus by the Cauchy-Schwarz inequality for all such ξ

$$r = \int_{\chi} \{d(x,\xi)\}^2 d\xi(x) \geq \{\int_{\chi} d(x,\xi)d\xi(x)\}^2 = p^2 .$$

Furthermore since

$$\int_{\chi} d(x,\xi)d\xi(x) = p$$

it follows that

$$\max_{x \in S(\xi)} d(x,\xi) \geq p .$$

By the Equivalence Theorem if ξ^* is D-optimal then

$$\max_{x \in \chi} d(x,\xi^*) = p$$

from which it follows that

$$\max_{x \in S(\xi^*)} d(x,\xi^*) = p .$$

Thus the D-optimal design ξ^* minimizes $\bar{r}$.

Furthermore since $r \geq p^2$ and $d(x,\xi^*) = p$ for all x in $S(\xi^*)$ it follows that

$$\int_{\chi} \{d(x,\xi^*)\}^2 d\xi^*(x) = p^2 ,$$

proving that ξ^* minimizes r.

Lemma 3.2

Fedorov (1972), p. 125, presents and proves a theorem that is equivalent to:

The following assertions:

(1) ξ^* minimizes $\operatorname{tr} M^{-1}(\xi^*)$,

(2) ξ^* minimizes $\sup_{x \in \chi} f'(x)\, M^{-2}(\xi^*) f(x)$,

(3) $\sup_{x \in \chi} f'(x)\, M^{-2}(\xi^*) f(x) = \operatorname{tr} M^{-1}(\xi^*)$

are equivalent for the problem (f,χ). Note ξ^* assigns measure one to a set of x where the supremum in (3) is achieved.

Now for $C = I_p$ we have

$R(\xi, I_p) = \operatorname{tr} M^{-1}(\xi)$

and

$\overline{R}(\xi, I_p) = \sup_{x \in S(\xi)} f'(x)\, M^{-2}(\xi) f(x)$.

Thus by the result above the lemma is proved for $C = I_p$.

In general for $g \in G^*$ corresponding to C we note that

$R(\xi, C) = R(g\xi, I_p)$

and

$\overline{R}(\xi, C) = \overline{R}(g\xi, I_p)$.

Then applying the argument above to the problem $(f, g\chi)$ completes the proof of this lemma.

Theorem 4.1

1. For $M \in M^*(f,\chi)$, assume that M is an interior point of the convex set $\underline{M}(f,\chi)$. Then there exists a positive number α such that the matrix $M_\alpha = (1+\alpha)M$ is also an element of $\underline{M}(f,\chi)$. However $M_\alpha > M$, which is a contradiction. Thus M is a boundary point for all $M \in \underline{M}^*(f,\chi)$.
2. Since M is a boundary point of $\underline{M}(f,\chi)$ we can apply Caratheodory's Theorem and the result follows.
3. Take $g \in G^*$ and $M \in \underline{M}^*(f,\chi)$. Then g acts on M by

$$M \overset{g}{\rightarrow} (\bar{g}^{-1})' M \bar{g}^{-1} = M_g \ .$$

Assume that M_g is not an element of $\underline{M}^*(f,g\chi)$.

Then there exists $\tilde{M}_g$ an element of $\underline{M}^*(f,g\chi)$ such that

$$\tilde{M}_g - M_g > 0$$

or

$$(\bar{g}^{-1})'[\tilde{M} - M]\bar{g}^{-1} > 0 \ .$$

Consequently $\tilde{M} - M > 0$ which is a contradiction. Thus M_g is an element of $M^*(f,g\chi)$ for all M and g.

Theorem 4.2

Take M an element of $\underline{M}^R(f,\chi)$ and assume that M is not an element of $\underline{M}^*(f,\chi)$. Then there exists $\tilde{M}$ an element of $\underline{M}^*(f,\chi)$ such that $\tilde{M} \geq M$.

By the definition of $\underline{M}^R(f,\chi)$, M is positive definite. Thus, since $\tilde{M} - M \geq 0$ and $\tilde{M}$ is positive definite, it follows that $M^{-1} \geq \tilde{M}^{-1}$. However $M^{-1} \geq \tilde{M}^{-1}$ implies $m^{ii} = \tilde{m}^{ii}$ for all i, i=1,...,p. Consequently either $m^{ii} = \tilde{m}^{ii}$ for all i or $\tilde{M}$ is not an element of $\underline{M}^*(f,\chi)$.

Claim: If $m^{ii} = \tilde{m}^{ii}$ for all i, i=1,...,p and $M^{-1} - \tilde{M}^{-1} \geq 0$, then $M^{-1} = \tilde{M}^{-1}$.

Proof: Let $A = M^{-1} - \tilde{M}^{-1} \geq 0$, $A = \{a_{ij}\}$, $a_{ii} = 0$ for all i, i=1,...,p. To show that $a_{ij} = 0$ for all i,j=1,...,p, take x an element of R^p; $x' = (x_1, x_2, \ldots, x_p)$. Then

$$x'Ax = \sum_{i=1}^{p} x_i^2 a_{ii} + \sum_{i=1}^{p} \sum_{j \neq i}^{p} x_i x_j a_{ij}$$

$$= \sum_{i=1}^{p} \sum_{j \neq i}^{p} x_i x_j a_{ij} \geq 0 .$$

Let $x_i = 1$, $x_j = -1$ and $x_k = 0$ $\quad k = 1, 2, \ldots, p$, $k \neq i$, $k \neq j$

then $x'Ax = -2a_{ij} \geq 0$

or

$a_{ij} \leq 0$ for all i,j .

Let $x_i = x_j = 1$, $x_k = 0$ $\quad k \neq i$, $k \neq j$

then $x'Ax = 2a_{ij} \geq 0$

or

$a_{ij} \geq 0$ for all i,j .

Therefore $a_{ij} = 0$ for all i,j.

Consequently either $M = \tilde{M}$ or $\tilde{M}$ is not an element of $M^*(f,\chi)$ which are both contradictions.

Therefore

$\underline{M}^R(f,\chi) \subseteq \underline{M}^*(f,\chi)$. To show that $M^R(f,\chi) \subset \underline{M}^*(f,\chi)$ it is only necessary to show there exists an admissible design that is not R-admissible. See part 3 of the example in Section 4.

Lemma 5.1

The proof of this lemma is evident from the conditions for A-optimality.

Lemma 5.2

Let ξ_D be D-optimal with $M(\xi_D)$ diagonal and assume that ξ_D is not R-admissible. Then there exists ξ^* an element of Ξ_χ such that

$$m^{ii}(\xi^*) \leq m^{ii}(\xi_D) \text{ for all } i = 1, 2, \ldots, p$$

with strict inequality for at least one i.

Thus

$$\prod_{i=1}^{p} m^{ii}(\xi^*) < \prod_{i=1}^{p} m^{ii}(\xi_D) = \det M^{-1}(\xi_D)$$

since $M^{-1}(\xi_D)$ is diagonal.

By Hadamard's inequality, $\det M^{-1}(\xi^*) \leq \prod_{i=1}^{p} m^{ii}(\xi^*)$,

and equality holds if and only if $M^{-1}(\xi^*)$ is diagonal.

Then

$$\det M^{-1}(\xi^*) \leq \prod_{i=1}^{p} m^{ii}(\xi^*) < \prod_{i=1}^{p} m^{ii}(\xi_D) = \det M^{-1}(\xi_D) ,$$

which is a contradiction since ξ_D is D-optimal.

Lemma 5.3

Fedorov (1972), page 106, presents and proves the following identity.

Let $\xi_{x,\alpha} = (1-\alpha)\xi + \alpha\delta(x)$ then for all x and α, $0 \leq \alpha < 1$,

$$m^{ii}(\xi) = (1-\alpha)^{-1}\{m^{ii}(\xi) - \frac{\alpha[m^{i}(\xi)'f(x)]^2}{1 - \alpha + \alpha d(x,\xi)}\} \quad .$$

Now by applying this identity a sufficient condition for the existence of a preferred design $\xi_{x,\alpha}$ can be established. From which a sufficient condition to show that a design is not R-admissible follows.

If $\xi_{x,\alpha}$ is preferred to ξ, then for all i,

$$m^{ii}(\xi_{x,\alpha}) \leq m^{ii}(\xi)$$

with strict inequality for at least one i, or using the identity above,

$$m^{ii}(\xi) - \frac{\alpha\{m^{i}(\xi)'f(x)\}^2}{1-\alpha + \alpha d(x,\xi)} \leq (1-\alpha)m^{ii}(\xi)$$

or equivalently

$$\frac{\{m^{i}(\xi)'f(x)\}^2}{m^{ii}(\xi)} - \alpha d(x,\xi) \geq 1-\alpha .$$

Let

$$K_{\alpha}(x,i,\xi) = \frac{\{m^{i}(\xi)'f(x)\}^2}{m^{ii}(\xi)} - \alpha d(x,\xi)$$

then if $\xi_{x,\alpha}$ is preferred to ξ it follows that

$$K_{\alpha}(x,i,\xi) \geq 1-\alpha$$

for all i with strict inequality for at least one i.

Thus if there exists a pair (x^*,α^*) such that

$$K_{\alpha^*}(x^*,i,\xi) > 1-\alpha \quad \text{for all } i ,$$

then ξ_{x^*,α^*} is preferred to ξ. The condition

$$\max_{x\epsilon\chi} \min_{i} K_{\alpha}(x,i,\xi) > 1-\alpha \quad \text{for some } \alpha$$

is a sufficient condition for the existence of at least one such pair (x^*,α^*). Therefore if the condition holds there exists at least one design ξ_{x^*,α^*} preferred to ξ proving that ξ is not R-admissible.

Lemma 5.4

By definition of the preference ordering $M(\xi)$ and $M(\xi^*)$ are positive definite. Let

$$\xi_\alpha = (1-\alpha)\xi + \alpha\xi^* .$$

Then

$$M(\xi_\alpha) = (1-\alpha)M(\xi) + \alpha M(\xi^*)$$

and it is well known that

$$M^{-1}(\xi_\alpha) \leq (1-\alpha)M^{-1}(\xi) + \alpha M^{-1}(\xi^*)$$

with equality if and only if $M(\xi) = M(\xi^*)$. Thus the inequality in this case is strict and it follows that

$$(1-\alpha)m^{ii}(\xi) + \alpha m^{ii}(\xi^*) > m^{ii}(\xi_\alpha) .$$

However $\xi > \xi^*$, consequently

$$m^{ii}(\xi^*) \geq m^{ii}(\xi) \quad \text{for all } i$$

or for all α and i

$$m^{ii}(\xi^*) \geq (1-\alpha)m^{ii}(\xi) + \alpha m^{ii}(\xi^*) .$$

Therefore $m^{ii}(\xi^*) > m^{ii}(\xi_\alpha)$ for all α and i. Thus $\xi_\alpha > \xi^*$.

Lemma 5.5

Clearly if there exists (x^*,α^*) such that ξ_{x^*,α^*} is preferred to ξ then ξ is not R-admissible. Consequently if ξ is R-admissible there cannot exist such a pair. This proves sufficiency.

Suppose ξ is not R-admissible; that is, there exists ξ^* such that ξ^* is preferred to ξ. Then it must be shown that there exists a pair (x^*,α^*) such that ξ_{x^*,α^*} is preferred to ξ for which a sufficient condition is

$$\max_{x\in\chi} \min_{i} K_\alpha(x,i,\xi) > 1-\alpha \quad \text{for some } \alpha .$$

Let Q be the p-dimensional probability simplex, then for q in Q, $q'=(q_1,\ldots,q_p)$ and $\sum_{i=1}^{p} q_i=1$ Q is the convex hull of the px1 indicator vectors, ε_i, $i = 1,2,\ldots p$.

Then

$$\min_{i} K_\alpha(x,i,\xi) = \min_{q\in Q} \sum_{i=1}^{p} K_\alpha(x,i,\xi)q_i$$

and consequently

$$\max_{x\in\chi} \min_{i} K_\alpha(x,i,\xi) = \max_{x\in\chi} \min_{q\in Q} \sum_{i=1}^{p} K_\alpha(x,i,\xi)q_i .$$

Q and χ are closed convex sets. Moreover, $\sum_{i=1}^{p} K_\alpha(x,i,\xi)q_i$ is convex in q (linear) and by assumption concave in x. Then by the Minimax Theorem,

$$\max_{x\in\chi} \min_{q\in Q} \sum_{i=1}^{p} K_\alpha(x,i,\xi)q_i = \min_{q\in Q} \max_{x\in\chi} \sum_{i=1}^{p} K_\alpha(x,i,\xi)q_i .$$

Recall that Ξ_χ is the convex hull of the point measures $\delta(x)$, $x\in\chi$. It follows that

$$\max_{x\in\chi} \sum_{i=1}^{p} K_\alpha(x,i,\xi)q_i = \max_{\delta(x)} \sum_{i=1}^{p} K_\alpha(x,i,\xi)q_i \; \delta(x)$$

$$= \max_{\tilde{\xi}\in\Xi_\chi} \int_\chi \sum_{i=1}^{p} K_\alpha(x,i,\xi)q_i \; d\tilde{\xi}(x)$$

and

$$\max_{x\in\chi} \sum_{i=1}^{p} K_\alpha(x,i,\xi)q_i \geq \int_\chi \sum_{i=1}^{p} K_\alpha(x,i,\xi)q_i \; d\xi^*(x)$$

$$= \sum_{i=1}^{p} q_i\{\int_\chi K_\alpha(x,i,\xi) \; d\xi^*(x)\} .$$

Thus it is enough to show that there exists an α, $0 < \alpha < 1$, such that for all i,

$$\int_{\mathcal{X}} K_\alpha(x,i,\xi)\, d\xi^*(x) > 1-\alpha .$$

Now by definition

$$K_\alpha(x,i,\xi) = \frac{\varepsilon_i' M^{-1}(\xi) f(x) f'(x) M^{-1}(\xi)\varepsilon_i}{m^{ii}(\xi)} - \alpha f'(x) M^{-1}(\xi) f(x) .$$

Thus

$$\int_{\mathcal{X}} K_\alpha(x,i,\xi)\, d\xi^*(x) = \{m^{ii}(\xi)\}^{-1}\varepsilon_i' M^{-1}(\xi) M(\xi^*) M^{-1}(\xi)\varepsilon_i - \alpha \operatorname{tr} M^{-1}(\xi) M(\xi^*) .$$

Note that $0 < \operatorname{tr} M^{-1}(\xi)M(\xi^*) < \infty$ since $M^{-1}(\xi)$ is positive definite.

Let $\operatorname{tr} M^{-1}(\xi)M(\xi^*) = c$ then

$$\int_{\mathcal{X}} K_\alpha(x,i,\xi)\, d\xi^*(x) > 1 - \alpha$$

implies that

$$\{m^{ii}(\xi)\}^{-1}\varepsilon_i' M^{-1}(\xi) M(\xi^*) M^{-1}(\xi)\varepsilon_i > 1 - \alpha + c\alpha .$$

Consequently it is sufficient to show that if ξ^* is preferred to ξ then for all i

$$\{m^{ii}(\xi)\}^{-1}\varepsilon_i' M^{-1}(\xi) M(\xi^*) M^{-1}(\xi)\varepsilon_i > 1$$

so

$$\varepsilon_i' M^{-1}(\xi) M(\xi^*) M^{-1}(\xi)\varepsilon_i > m^{ii}(\xi) = \varepsilon_i' M^{-1}(\xi)\varepsilon_i$$

or

$$\varepsilon_i' M^{-1}(\xi) M(\xi^*) M^{-1}(\xi)\varepsilon_i - \varepsilon_i' M^{-1}(\xi)\varepsilon_i > 0 .$$

Claim: If ξ^* is preferred to ξ then for all i

$$\varepsilon_i' M^{-1}(\xi) M(\xi^*) M^{-1}(\xi)\varepsilon_i - \varepsilon_i' M^{-1}(\xi)\varepsilon_i > \varepsilon_i'\{M^{-1}(\xi) - M^{-1}(\xi^*)\}\varepsilon_i .$$

Proof: Let $D = M^{-1}(\xi) - M^{-1}(\xi^*)$; then the claim is that for all i,

$$\varepsilon_i' M^{-1}(\xi)M(\xi^*)D\varepsilon_i > \varepsilon_i' D\varepsilon_i \ ;$$

so

$$\varepsilon_i'\{M^{-1}(\xi)M(\xi^*)D - D\}\varepsilon_i > 0 \ .$$

Now,

$$\begin{aligned}\varepsilon_i'\{M^{-1}(\xi)M(\xi^*)D - D\}\varepsilon_i &= \varepsilon_i'\{M^{-1}(\xi)M(\xi^*) - I\}D\varepsilon_i \\ &= \varepsilon_i'\{M^{-1}(\xi) - M^{-1}(\xi^*)\}M(\xi^*)D\varepsilon_i \\ &= \varepsilon_i' DM(\xi^*)D\varepsilon_i \ .\end{aligned}$$

Since $M(\xi^*)$ is positive definite,

$$\varepsilon_i' DM(\xi^*)D\varepsilon_i > 0 \text{ for all } i$$

and the claim follows. Moreover since ξ^* is preferred to ξ,

$$m^{ii}(\xi^*) \leq m^{ii}(\xi) \text{ for all } i \ .$$

Therefore

$$\varepsilon_i'\{M^{-1}(\xi) - M^{-1}(\xi^*)\}\varepsilon_i = m^{ii}(\xi) - m^{ii}(\xi^*) \geq 0$$

for all i. Thus there exists a pair (x^*, α^*) such that ξ_{x^*,α^*} is preferred to ξ.

REFERENCES

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. and Tukey, J.W. (1972). Robust Estimates of Location: Survey and Advances. Princeton, New Jersey: Princeton University Press.

Box, G.E.P. and Draper, N.R. (1975). Robust designs. Biometrika 62, 347-352.

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

Hampel, F.R. (1974). The influence curve and its role in robust estimation. J. Amer. Statist. Assoc. 69, 383-393.

Hinkley, D.V. (1977). On jackknifing in unbalanced situations. To appear in Technometrics.

Huber, P.J. (1975). Robustness and designs. In A Survey of Statistical Design and Linear Models. J.N. Srivastava editor. Amsterdam: North Holand.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. Canad. J. Math. 12, 363-366.

Kiefer, J. (1958). On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. Ann. Math. Statist. 29, 675-699.

Kiefer, J. (1959). Optimum experimental designs. J. Roy. Statist. Soc. Ser. B. 21, 272-319.

Kiefer, J. (1975). Optimal designs: Variation in structure and performance under change of criterion. Biometrika 62, 277-288.

Thibodeau, L.A. (1977). Robust Design for Regression Problems. Ph.D. Thesis, University of Minnesota.