MINIMUM DISTANCE ESTIMATION

AND COMPONENTS OF GOODNESS OF FIT STATISTICS

by

William C. Parr
William R. Schucany

Technical Report No. 139
Department of Statistics ONR Contract

September 1980

DIVISION OF MATHEMATICAL SCIENCES
Department of Statistics
Southern Methodist University
Dallas, Texas 75275

MINIMUM DISTANCE ESTIMATION
AND COMPONENTS OF GOODNESS OF FIT STATISTICS

William C. Parr
Institute of Statistics
Texas A&M University
College Station, Texas 77843 USA

William R. Schucany
Department of Statistics
Southern Methodist University
Dallas, Texas 75275 USA

SUMMARY

The relationship of minimum distance (MD) estimation to other methods
of estimation is considered. M-estimation is viewed as a special case.
with interesting interpretations in terms of the defining ψ - function
as related to components of goodness-of-fit statistics and modified
Fourier approximations to the efficient score. Applications to the compo-
site and simple goodness-of-fit problems are considered.

1. Introduction, Definitions, and Consistency

Robust estimation has received much attention in recent statistical literature, with a comprehensive survey given by Huber (1977). The problem considered is as follows; a random sample $X_1, \ldots, X_n$ is observed from some unknown distribution G, where it is presumed (although not necessarily true) that $G \in \Gamma = \{F_\theta, \theta \in \Omega\}$, where the model Γ is a parametrized family of distribution functions. The goal of robust estimation is to estimate θ with an estimator $T[G_n]$, such that T is nearly fully efficient when $G \in \Gamma$, i.e. when the model is correct, and which estimates a _meaningful_ quantity with reasonable efficiency when $G \notin \Gamma$, but G is close to Γ in an appropriate topology on the space of distribution functions.

Minimum distance estimation was first subjected to comprehensive study in a series of papers culminating in Wolfowitz (1957), and has since been considered as a method for deriving robust estimators by Knüsel (1969) and Parr and Schucany (1980). An extensive bibliography is given by Parr (1980). The basic philosophy of minimum distance (MD) estimation is to match the empirical distribution function G_n to an element, F_T, of the model Γ as closely as possible. Thus, for a suitably chosen "distance function" $\delta(.,.)$ measuring the discrepancy between two distribution functions, an MD-estimator of θ based on G_n and with respect to the model Γ and the discrepancy $\delta(.,.)$ is given by a value T such that

$$(1) \qquad \delta(G_n, F_T) = \inf_{\theta \in \Omega} \delta(G_n, F_\theta) .$$

Due to possible nonuniqueness of the value T achieving the infimum or to nonattainability of the infimum in Γ, we are forced for generality to the following definition.

<u>Definition</u>. A sequence of random variables $\{T_n\}_{n=1}^{\infty}$ is a sequence of asymptotic minimum distance estimators based on $\{G_n\}_{n=1}^{\infty}$ with respect to the model Γ and the discrepancy $\delta(.,.)$ if and only if

i) $T_n \in \Omega$ for all $n \geq 1$

and ii) there exists a nonnegative function $K(n)$ with $\lim\limits_{n \to \infty} K(n) = 0$ such that for all $n \geq 1$

$$(2) \quad \delta(G_n, F_{T_n}) \leq \inf_{\theta \in \Omega} \delta(G_n, F_\theta) + K(n) \ .$$

Some natural choices for the discrepancy $\delta(.,.)$ would include the Kolmogorov discrepancy,

$$D(K,L) = \sup_{-\infty < x < \infty} |K(x) - L(x)| \ ,$$

the Kuiper discrepancy,

$$V(K,L) = \sup_{-\infty < a < b < \infty} |\{K(b) - K(a)\} - \{L(b) - L(a)\}| \ ,$$

and the class of discrepancies given by

$$(3) \quad H_{a,b}^2(K,L) = a\int_{-\infty}^{\infty} \{K(x) - L(x)\}^2 dL(x) + b[\int_{-\infty}^{\infty} \{K(x) - L(x)\} dL(x)]^2 \ ,$$

considered by Sahler (1970), which includes the Cramér-von Mises discrepancy W^2 for $a = 1$, $b = 0$; the Watson U^2 discrepancy for $a = 1$, $b = -1$; and the Chapman discrepancy, C^2, for $a = 0$, $b = 1$. We assume

here and henceforth that all distribution functions in Γ are absolutely
continuous. Actual choice of which discrepancy to use for a specific
situation would depend upon i) which aspects of the sampled population
one desires to match, ii) efficiency considerations, and iii) robust-
ness considerations. The connections of MD-estimation with other
methods discussed should provide some insight into the trade-off among
those competing criteria.

It is of interest to determine conditions under which a sequence
of asymptotic MD estimators is consistent. The following theorem
(a generalization of Theorem 1 of Parr and Schucany (1980)) provides
suitable (if somewhat stringent) restrictions on Γ, $\delta(.,.)$, and the
sampling situation.

<u>Theorem 1:</u>   Let $\{G_n\}_{n=1}^{\infty}$ be a sequence of random distribution functions
on R^1, and $\{T_n\}_{n=1}^{\infty}$ be a sequence of asymptotic MD-estimators based
on $\{G_n\}_{n=1}^{\infty}$ with respect to $\Gamma = \{F_\theta, \theta \in \Omega\}$ and $\delta(.,.)$. If the
following hold:

   i) there exists a metric $||\cdot||$ on $\tilde{F}$ (where $\bar{F}$ is the
       space of one-dimensional distribution functions) such that
       $||G_n - G|| \longrightarrow 0$ with probability one,

   ii) the class of functions $\{\delta(\cdot, F_\theta) , \theta \in \Omega\}$ is equicontinuous
        at G (with respect to the metric $||\cdot||$),

   iii) there exists a point $\theta_0 \in \Omega$ such that $\delta(G, F_{\theta_0}) < \delta(G, F_\theta)$
         for $\theta \neq \theta_0$, $\theta \in \Omega$, and

   iv) for any sequence $\{\theta_k\}_{k=1}^{\infty}$ of elements of Ω, $\lim_{k \to \infty} \delta(G, F_{\theta_k}) =$
        $\delta(G, F_{\theta_0})$ implies $\lim_{k \to \infty} \theta_k = \theta_0$,

then $T_n \longrightarrow \theta_0$ with probability one.

<u>Proof</u>: The proof is trivial and hence omitted.

Notes:

1) Conditions iii) and iv) are designed to insure <u>uniqueness</u> of the minimum of $\delta(G,F_\theta)$ and a <u>reasonable</u> <u>parametrization</u> of Γ, respectively.

2) Condition i) is the only restriction on the sampling situation. While it is easily satisfied for "small" choices of $||\cdot||$, ii) competes by being easily satisfied by "big" $||\cdot||$. A typical choice might be the L^∞ metric (Kolmogorov discrepancy). For such a choice and random sampling, if Γ is a translation family with θ the translation parameter $(F_\theta(x) = F_o(x - \theta)$ for all (x,θ) and $G \in \Gamma)$, the conditions are satisfied for all discrepancies mentioned in this section. They are also satisfied for the above discrepancies when $G \notin \Gamma$ if iii) and iv) hold.

3) Condition ii) can be omitted if $\delta(.,.)$ is a metric on $\tilde{F}$.

4) The theorem is really a statement of continuity of the functional $T[G_n] = T_n$ at G with respect to the metric $||\cdot||$.

5) Condition ii) could (at a sacrifice of simplicity) clearly be relaxed to requiring equicontinuity of the $\delta(\cdot,F_\theta)$ at G only for θ in a neighborhood $U(\theta_o)$ of θ_o, and that

$$\inf_{\theta \in \Omega - U(\theta_o)} \delta(H,F_\theta) \geq M + 0(||H - G||)$$

for some $M > \delta(G,F_{\theta_o})$.

The statement of general results for asymptotic distribution theory proves to be much less succinct. MD estimators divide into two basic types: 1) those based upon "integral-type" discrepancies such as $H_{a,b}^2$ or weighted versions thereof, and 2) those based upon "sup-type" discrepancies such as D,V, or weighted versions thereof. The first type are asymptotically normal under suitable conditions (Sahler (1970), Parr and Schucany (1980), Parr and DeWet (1979) and Boos (1980)), while the second type are typically not asymptotically normal (Bolthausen (1977)) even in the simplest and smoothest cases. Littell and Rao (1975), and Pollard (1980) are also good references in this area.

As we shall see in the following sections, frequency-domain analyses of MD-procedures can yield a great amount of insight into the proper choice of the discrepancy $\delta(.,.)$, based upon the competing goals of deriving an estimator with high efficiency when $G \in \Gamma$ and of maintaining robustness when $G \notin \Gamma$. The first criterion (efficiency) will require high fidelity of a particular tapered Fourier approximation to the efficient score, while the second (robustness) will amount to use of a low-pass filter to dampen out high frequency components of the same approximate.

2. Components of Goodness-of-fit Statistics and MD Estimation

Durbin and Knott (1972) introduced the idea of interpreting the quantities z_{nj}^2 in the orthogonal representation of the Cramér-von Mises statistic

$$(4) \quad H_{1,0}^2 = \sum_{j=1}^{\infty} \frac{z_{nj}^2}{j^2\pi^2}$$

as components representing different aspects of the discrepancy between G_n and F_θ, where

$$(5) \quad Z_{nj} = \sqrt{(2n)} j\pi \int_{-\infty}^{\infty} \left\{ G_n(x) - F_\theta(x) \right\} \sin\left\{ j\pi F_\theta(x) \right\} dx$$

$$= \sqrt{(2/n)} \sum_{i=1}^{n} \cos\left\{ j\pi F_\theta(X_i) \right\} .$$

Here, $X_1, \ldots, X_n$ is a random sample of size n from some distribution G and the statistic $H_{1,0}^2$ is being used to test whether or not $G \equiv F_\theta$. For "smooth" alternatives, the main source of the discrepancy between G_n and F_θ should be in the first few components. For a null hypothesis of X_i standard normal, Durbin and Knott found that

    i) Z_{n1} contained most of the information about pure location shifts, having an asymptotic power of .93 against contiguous location changes when the t-test had power .95.

    ii) Z_{n2} was similarly efficient against scale changes.

    iii) Z_{n1} was orthogonal to contiguous scale changes and Z_{n2} orthogonal to contiguous location shifts.

This suggests that a suitable reweighting of the components in (4) or a similar test might result in a higher efficiency. This program of study is carried out in Schoenfeld (1977). To discuss such extensions we need the following notation:

Let $\{\zeta_0(1), \zeta_1(1), \ldots\}$ be a complete orthonormal basis for the space of square integrable functions on $[0,1]$. Require $\zeta_0(u) = 1$, $0 \leq u \leq 1$ so that $\int_0^1 \zeta_i(u) du = 0$, $i \neq 0$. Let $\Gamma_\theta = \{F_\theta, \theta \in \Omega\}$. be a parametrized family of distribution functions (called the "model") and G_n and G be as above. Further paralleling (5), define the random functions

$$(6) \quad d_{nj}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \zeta_j \Big(F_\theta(X_i)\Big) \ .$$

When $\zeta_j(u) = \sqrt{2}\cos(j\pi u)$, $\sqrt{n}\, d_{nj}(\theta)$ is Z_{nj}, the jth component of the Cramér-von Mises statistic as in (5). Thus

$$(7) \quad H^2_{1,0} = \sum_{j=1}^{\infty} \frac{nd^2_{nj}(\theta)}{j^2 \pi^2} \quad \text{in this special case.}$$

More generally, it is expressed as a weighted sum of the squared $d_{nj}(\theta)$.

Since the $\zeta_j(\cdot)$ will usually have a frequency-type interpretation, different weightings of the squared components will correspond to the creation of a goodness-of-fit test sensitive to departures from the null hypotheses having specific frequency interpretations. As we shall see from the following, similar interpretations will be possible for MD-estimators related to these tests.

In the context of (possibly robust) minimum distance estimation, this suggests a broadening of the class of estimators. Define the random functions which are candidates for useful new discrepancy measures between F_θ and G_n,

$$(8) \quad K(\theta,a;G_n) = \left\{ \sum_{j=1}^{\infty} a_j d_{nj}(\theta) \right\}^2$$

for some fixed sequence $\{a_j\}$ such that

$$\sum_{j=1}^{\infty} a_j^2 < \infty \ ;$$

and

$$(9) \quad L(\theta,b;G_n) = \sum_{j=1}^{\infty} b_j d^2_{nj}(\theta)$$

for $\sum_{j=1}^{\infty} |b_j| < \infty \ .$

We are now able to consider MD-estimation utilizing the discrepancies $K(\theta,a;G_n)$ and $L(\theta,b;G_n)$.

<u>Case 1</u>: Estimators minimizing $K(\theta,a;G_n)$

$K(\theta,a;G_n)$ may be written as a functional of G_n in the form

$$K(\theta,a;G_n) = \left\{ \int_{-\infty}^{\infty} \sum_{j=1}^{\infty} a_j \zeta_j \left(F_\theta(x) \right) dG_n(x) \right\}^2 .$$

Observe that

$$K(\theta,a;F_\theta) = \left\{ \int_{-\infty}^{\infty} \sum_{j=1}^{\infty} a_j \zeta_j \left(F_\theta(x) \right) dF_\theta(x) \right\}^2 = 0 , \text{ so that}$$

when $G \in \Gamma$, the estimand is a root of the equation

$$(10) \quad \int_{-\infty}^{\infty} \sum_{j=1}^{\infty} a_j \zeta_j \left(F_\theta(x) \right) dG(x) = 0 .$$

It follows that the value $T_K[G_n]$, which minimizes $K(\theta,a;G_n)$ as a function of θ, is (subject to the conditions of Theorem 2) a root of

$$(11) \quad \int_{-\infty}^{\infty} \sum_{j=1}^{\infty} a_j \zeta_j \left(F_\theta(x) \right) dG_n(x) = 0 ,$$

i.e. that $T_K[G_n]$ is an M-estimator with defining ψ-function given by

$$(12) \quad \psi(x;\theta) = \sum_{j=1}^{\infty} a_j \zeta_j \left(F_\theta(x) \right) .$$

Thus, the usual theory for M-estimation is applicable.

For the following result we further require that the $\zeta_j(\cdot)$ be individually continuous and uniformly bounded. Such orthonormal bases for $L^2[0,1]$ do exist, such as $\delta_j(u) = 2^{1/2} \cos(j\pi u)$, $j = 1, 2, \ldots$. These conditions are stronger than necessary, but serve to reduce the mathematical complexity of the results.

Theorem 2: If

i) $\sum\limits_{j=1}^{\infty} |a_j| < \infty$, and

ii) $\dfrac{\partial}{\partial\theta} \int\limits_{-\infty}^{\infty} \sum\limits_{j=1}^{\infty} a_j \zeta_j\Big(F_\theta(x)\Big) dG(x)\Big|_{\theta=\theta_o}$ is finite and nonzero,

where θ_o is a root of

$$\int\limits_{-\infty}^{\infty} \sum\limits_{j=1}^{\infty} a_j \zeta_j\Big(F_\theta(x)\Big) dG(x) = 0 \ ,$$

then there exists a sequence $\{T_K[G_n]\}_{n=1}^{\infty}$ of roots of equation (11) such that

$$T_K[G_n] \longrightarrow \theta_o \quad \text{with probability one.}$$

Proof: The result is a simple corollary of elementary consistency theorems for M-estimators, since i) θ_o is an isolated root, and ii) $\sum\limits_{j=1}^{\infty} a_j \zeta_j\Big(F_\theta(x)\Big)$ is continuous and bounded.

Theorem 3: If in addition to the assumptions of Theorem 2,

$$\frac{\partial}{\partial\theta} \sum\limits_{j=1}^{\infty} a_j \zeta_j\Big(F_\theta(x)\Big)\Big|_{\theta_o}$$

is uniformly continuous in x, then any sequence $\{T_K[G_n]\}$ satisfying $T_K[G_n] \longrightarrow \theta_o$ also satisfies

$$\sqrt{n}(T_K[G_n] - \theta_o) \xrightarrow{d} N(0,\sigma_K^2) \ ,$$

with

(13) $$\sigma_K^2 = \frac{E_G[\psi^2(X;\theta_o)]}{\left(E_G\left[\dfrac{\partial\psi(X;\theta)}{\partial\theta}\Big|_{\theta_o}\right]\right)^2}$$

Proof: This is also a simple consequence of standard normality theorems for the associated M-estimator.

Note: If $G \in \Gamma$ and $\dfrac{\partial^2 F_\theta(x)}{\partial\theta\partial x} = \dfrac{\partial^2 F_\theta(x)}{\partial x\partial\theta}$ almost everywhere, with

$$\lim_{x \to \pm\infty} \frac{\partial F_\theta(x)}{\partial \theta} = 0 \quad \text{for all} \quad \theta, \text{ then}$$

$$(14) \qquad \sigma_K^2 = \frac{\sum\limits_{j=1}^{\infty} a_j^2}{\left\{\sum\limits_{j=1}^{\infty} a_j \int\limits_{-\infty}^{\infty} \zeta_j\big(F_\theta(x)\big)\frac{\partial f_\theta(x)}{\partial \theta}dx\right\}^2}$$

$$= \frac{\sum\limits_{j=1}^{\infty} a_j^2}{\left(\sum\limits_{j=1}^{\infty} a_j C_j\right)^2} \quad ,$$

where $\quad C_j = \int\limits_{-\infty}^{\infty} \zeta_j\big(F_\theta(x)\big)\frac{\partial \ln f_\theta(x)}{\partial \theta} f_\theta(x)dx$

is the jth Fourier coefficient in the expansion of the score
function

$$(15) \quad J(u) = \frac{\partial}{\partial \theta}\ln f_\theta(x)\Big|_{x=F_\theta^{-1}(u)} = \sum_{j=1}^{\infty} C_j \zeta_j(u) \quad .$$

(The expansion is valid if the Fisher information is finite.)
For notational convenience, the derivative of a function with respect
to its argument will be denoted by a prime $(\acute{})$, its derivative with
respect to θ denoted by a dot $(\cdot)$. For example, $\dot{F}_\theta(x) = \dfrac{\partial F_\theta(x)}{\partial \theta}$ and
$F_\theta^{\acute{}}(x) = f_\theta(x)$. Second order derivatives with respect to θ will be
denoted by two dots $(\cdot\cdot)$. Also, we write $\ell_\theta(x) = \ln f_\theta(x)$. Hence,

$$(16) \qquad \sigma_K^2 = \frac{\sum\limits_{j=1}^{\infty} a_j^2 \sum\limits_{j=1}^{\infty} C_j^2}{\left(\sum\limits_{j=1}^{\infty} a_j C_j\right)^2} \cdot \frac{1}{\int\limits_{-\infty}^{\infty}\left\{\dot{\ell}_\theta(x)\right\}^2 f_\theta(x)dx}$$

and efficiency of T_K is related to the "correlation" of the a and c
sequences. A by product of the proof is the fact that

$$(17) \quad \sqrt{n}\left\{ T_K[G_n] - \theta_o - \frac{1}{n}\sum_{i=1}^{n} IC_{T_K,F_{\theta_o}}(X_i)\right\} \xrightarrow{P} 0 \quad ,$$

where

$$(18) \quad IC_{T_K,F_\theta}(w) = \frac{\sum_{j=1}^{\infty} a_j \zeta_j\left(F_\theta(w)\right)}{-\sum_{j=1}^{\infty} a_j C_j} \quad .$$

Note further that, for the case of an <u>unbounded</u> score function, full efficiency of the estimator is inconsistent with absolute summability of the a_j . However, when the a_j are absolutely summable, we have that $IC_{T_K,F_{\theta_o}}(w)$ is bounded and continuous, and hence $T_K[G_n]$ is robust in that sense. (In fact, $\sum_{j=1}^{\infty} a_j \zeta_j(u)$ is <u>uniformly</u> continuous.)

<u>Case 2</u>: Estimators minimizing $L(\theta,b;G_n)$

Computing the influence curve for T_K yields

$$(19) \quad IC_{T_K,G}(x) = \frac{\sum_{j=1}^{\infty} a_j \zeta_j\left(F_{\theta_o}(x)\right)}{-\int_{-\infty}^{\infty} \sum_{j=1}^{\infty} a_j \zeta_j'\left(F_{\theta_o}(y)\right) \dot{F}_{\theta_o}(y)\, dG(y)}$$

$$= \frac{\sum_{j=1}^{\infty} a_j \zeta_j\left(F_{\theta_o}(x)\right)}{-\sum_{j=1}^{\infty} a_j C_j} \quad .$$

The second equality is true with C_j as defined in the note to Theorem 4, under the conditions to be stated in that note.

Now, computing the influence curve for T_L, we obtain

$$(20) \quad IC_{T_L,G}(x) = \frac{\sum_{j=1}^{\infty} b_j e_j(\theta_o) \zeta_j\left(F_{\theta_o}(x)\right)}{-\sum_{j=1}^{\infty} b_j e_j^2(\theta_o)} \quad ,$$

where $\quad e_j(\theta_o) = E_G\left[\zeta_j'\left(F_{\theta_o}(X)\right) \dot{F}_{\theta_o}(X)\right] \quad .$

Thus, the choice of $b_j(\theta_o) = \dfrac{a_j}{e_j(\theta_o)}$ yields locally the same influence curve for the two methods of MD-estimation. In fact, the following stronger result holds.

Theorem 4: Let $\{\zeta_j, j = 0,1,\ldots\}$ be a complete orthonormal basis for $L^2[0,1]$, where the $\zeta_j(\cdot)$ are individually continuous and uniformly bounded. Let $\Gamma, \{L(\theta,b;G_n)\}_{n=1}^{\infty}$, $\{T_L(G_n)\}_{n=1}^{\infty}$, be such that $\{T_L(G_n)\}_{n=1}^{\infty}$ minimizes $\{L(\theta,b,G_n)\}_{n=1}^{\infty}$ componentwise and θ_o minimizes $\dot{L}(\theta,b;G)$. Then if

i) $T_L(G_n) \longrightarrow \theta_o$ in probability

ii) for some $\varepsilon > 0$, $\theta \in (\theta_o - \varepsilon, \theta_o + \varepsilon)$ implies

$$E_G\left[\left|\frac{\partial \zeta_j\left(F_\theta(x)\right)}{\partial \theta}\right|\right] < h_j, \text{ where}$$

$$\sum_{j=1}^{\infty} |b_j|\,||\zeta_j(\cdot)||_V h_j < \infty \quad, \text{ with } ||\cdot||_V \text{ the total variation}$$

norm, and

iii) $\dfrac{\partial}{\partial \theta}\, \zeta_j\left(F_\theta(x)\right)$ is a continuous function of θ at θ_o for a set of x having G-probability one for each j, then

$$\sqrt{n}\{T_L(G_n) - \theta_o\} \xrightarrow{d} N\left(0, \sigma_L^2(\theta_o)\right) \quad,$$

where

$$\sigma_L^2(\theta_o) = \frac{4E_G\left[\left\{\sum_{j=1}^{\infty} b_j e_j(\theta_o)\zeta_j\left(F_{\theta_o}(x)\right)\right\}^2\right]}{\{\ddot{L}(\theta_o,b;G)\}^2} \quad \text{is assumed finite.}$$

Hence, subject to the conditions of these two theorems, the estimators $T_K(G_n)$ and $T_L(G_n)$ have the same asymptotic distribution if $b_j(\theta_o)e_j(\theta_o) = a_j$.

Proof: The proof is given in the appendix.

Notes:

1) If $G = F_{\theta_o}$, or more generally $\int \zeta_j\left(F_{\theta_o}(x)\right) dG(x) = 0$ for $j=1,2, \ldots$, then $\left\{\overset{..}{L}(\theta_o,b;G)\right\}^2 = \left\{2 \overset{\infty}{\underset{j=1}{\sum}} b_j e_j^2(\theta_o)\right\}^2$.

2) Condition ii) is the crucial one for the proof of this result. It is not necessary as can be seen by considering the weights $b_j = 1/j^2\pi^2$ and $\zeta_j(u) = 2^{1/2}\cos(j\pi u)$, for which $L(\theta,b,G_n)$ is the Cramér-von Mises statistic, the desired asymptotic normality holding in the location case if $G \in \Gamma$ and the population density is cube integrable. However, ii) fails in this case.

Thus, estimators derived from minimizing $L(\theta,b;G_n)$ can duplicate the behavior of those derived from $K(\theta,a;G_n)$. Examining the case of $K(\theta,a;G_n)$ in (for simplicity) the location case, we have

$$IC_{T_K,F_{\theta_o}}(w) = \frac{\overset{\infty}{\underset{j=1}{\sum}} a_j \zeta_j\left(F_{\theta_o}(w)\right)}{-\overset{\infty}{\underset{j=1}{\sum}} a_j C_j} \quad .$$

Thus, if we have the ζ_j ordered according to a "frequency" idea, i.e. perhaps $\zeta_j(u) = 2^{1/2}\cos(j\pi u)$, we see that

i) Since the ζ_j are uniformly bounded, each is continuous on $[0,1]$, and $\overset{\infty}{\underset{j=1}{\sum}} |a_j| < \infty$, $IC_{T_K,F_{\theta_o}}(x)$ is uniformly continuous (a robustness property).

ii) The extent to which the weights a_j "taper off" as $j \to \infty$ will correspond to the degree of differentiability of $IC_{T_K,F_{\theta_o}}(x)$, and hence to the degree to which T_K possesses additional robustness properties.

iii) The inner product $\dfrac{\left(\sum\limits_{j=1}^{\infty} a_j c_j\right)^2}{\sum\limits_{j=1}^{\infty} a_j^2 \; \sum\limits_{j=1}^{\infty} c_j^2}$ determines the efficiency of

the estimation procedure. Thus, for instance, if the a_j's taper off fast to achieve i) and ii) and in doing so fail to maintain a high correlation with the c_j's, the efficiency of the estimator will be low.

The desire to perfectly duplicate the high frequency aspects of $\dot{\ell}_\theta(x)$ must produce non-robust estimators by violating i) and ii) to achieve iii). Many results in robustness may in fact be viewed as means of tapering the sequence $\{a_j\}$ to eliminate or minimize high frequency components of $\dot{\ell}_\theta(x)$ while maintaining high fidelity. Therefore, we have seen that both M- and MD-estimation may be linked to a (tapered) Fourier expansion of $\dot{\ell}_\theta(x)$

Figures 1 and 2 illustrate this phenomenon. They give the <u>truncated</u> Fourier approximates to $\dot{\ell}_\theta(F_\theta^{-1}(u))$ of the form $\sum\limits_{j=1}^{M} C_j \varsigma_j(u)$ for $M = 1,3,5,7,$ and ∞. (We take $\theta = 0$ without loss of generality.) (Only $u \geq .5$ is shown, since both the functions and their approximates are odd in $F_\theta^{-1}(u)$, and hence $C_{2k} = 0$, $k = 1,2, \ldots$).

For the normal density, inclusion of more terms (increasing M) allows greater fidelity to $\phi^{-1}(u) = \dot{\ell}_0(F_0^{-1}(u))$, at the price of increasing the supremum of the approximate. For the Laplace density, the added terms improve the approximation near the discontinuity in $\dot{\ell}(F_0^{-1}(u)) = 1 - 2I(u < .5)$, at the expense of making the approximate's derivatives larger. Relative efficiencies attained by K-type estimators using the various truncations of the efficient scores are given in Table 1.

3. Connections of Minimum Distance
to Other Estimation Methods

Several interesting connections exist between MD and other methods
of estimation. Estimation of θ_o based upon minimizing $K(\theta, a; G_n)$ is
easily seen to be equivalent to defining the estimator $\hat{\theta}_n$ to be a root of

$$\sum_{j=1}^{\infty} a_j d_{nj}(\hat{\theta}_n) = 0 \; ,$$

i.e.

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{\infty} a_j \zeta_j\Big(F_{\hat{\theta}_n}(x_i)\Big) \right\} = 0 \; .$$

Hence, this MD estimator is an M-estimator with defining ψ-function

$$\psi(x;\theta) = \sum_{j=1}^{\infty} a_j \zeta_j\Big(F_\theta(x)\Big) \; ,$$

as noted in Section 2.

Hodges and Lehmann (1963) obtained robust estimators of location
via the "inversion" of rank test. For a random sample $X_1, \ldots, X_n$
from G and any value $-\infty < \theta < \infty$, they define the mirror image of the
sample about θ by $Y_i(\theta) = 2\theta - X_i$, $i = 1, \ldots, n$. If the distribu-
tion G is symmetric about θ_o, then the $\{Y_i(\theta_o)\}$ have the same distri-
bution as the $\{X_i\}$. Then, taking h to be any two-sample rank test
statistic for shift having the property that when the two populations
are identical, the distribution of h is symmetric about some value μ,
they define a rank (R) estimate of location by

$$\hat{\theta} = [\sup\{\theta : h(X_1, \ldots, X_n; Y_1(\theta), \ldots, Y_n(\theta))$$

$$+ \inf\{\theta : h(X_1, \ldots, X_n; Y_1(\theta), \ldots, Y_n(\theta)) < \mu\}]/2$$

Thus, an R-estimator of location "inverts" a rank test in the sense that it selects as an estimator that value, $\hat{\theta}$ such that the rank test based upon the statistic $h(\underline{X},\underline{Y}(\hat{\theta}))$ finds it hardest to reject the hypothesis of symmetry. (A similar development of R-estimators is possible from one-sample rank tests.)

Similarly, in an obvious fashion, MD-estimators invert goodness-of-fit tests. This similarity provides a heuristic method for choosing highly efficient MD-estimators. In R-estimation, Hodges and Lehmann (1963) found that rank tests possessing high power against location shifts yielded, upon inversion, extremely efficient R-estimators. Similarly, minimization of a goodness-of-fit discrepancy, which is highly powerful against location shifts, yields a MD-estimator with a good efficiency. This motivation could in fact lead to the fully efficient MD-estimators discussed in Section 2, which coincide with the optimal M-estimators. There is, however, the added quirk that most goodness-of-fit tests are not asymptotically normal, and thus the formal theory developed by Hodges and Lehmann (1963) does not directly apply to the inversion of typical goodness-of-fit tests. (But see Parr and DeWet (1979) and Boos (1980) for development of optimum weighting schemes for MD-estimation.)

As a method, adaptive estimation is somewhat difficult to characterize. The spirit of the method as developed by Hogg (1974) and others is, however, straightforward to describe. Consider for simplicity the case of location estimation for symmetric populations. The statistician examines a characteristic of the sample data which measures, perhaps, tailweight (naively kurtosis, but more likely one of the subsequent tailweight measures discussed in Hogg (1974) which involve ratios of scale estimators which are linear functions of order statistics). Based

upon the value of this statistic, an estimator is chosen from a (possibly infinite) set which is expected to perform well for distributions with tailweight of the estimated order. Thus, adaptive estimation is a two-step process:

1) Based upon some characteristic of the data, select an estimator which is believed to work well for the apparent class of parent populations (or, otherwise stated, select a model which appears to be an adequate approximation to the data)

and 2) Use that estimator (or, use an estimation procedure expected to be competitive at or near the expected model).

A procedure of this sort arises naturally in MD-estimation of a location parameter. Instead of considering Γ = {a specific location/ scale family of distribution functions} and minimizing $\delta(G_n, F_\theta)$ over $F_\theta \in \Gamma$, we could as well let $\Gamma = \Gamma_1 \cup \Gamma_2 \cup, \ldots, \cup \Gamma_K$, where each Γ_i is a distinct location/scale family, generated perhaps by prototype t-distributions with $v_1, v_2, \ldots, v_K$ degrees of freedom. Finding an MD-estimator with respect to δ and Γ is precisely equivalent to choosing the sub-model Γ_i such that $\inf\limits_{F_\theta \in \Gamma_i} \delta(G_n, F)$ is the smallest, and then using the MD-estimator with respect to δ and Γ_i. Thus, the MD-estimator adaptively selects the closest location/scale family (in the sense of δ) and then estimates based upon a projection into that family.

4. Goodness-of-fit Tests

In this section we consider the goodness-of-fit problem both in the simple null case and for composite nulls when the estimator of the unknown parameters possesses an asymptotically linear structure. This will enable us to examine simultaneous model-dependent MD-estimation and goodness-of-fit tests.

In the case of a simple null hypothesis, i.e. testing $H_o : G = F_{\theta_o}$ versus $H_A : G \neq F_{\theta_o}$, we consider as test statistics

$$(21) \quad K(\theta_o, r; G_n) = \left\{ \sum_{j=1}^{\infty} r_j d_{nj}(\theta_o) \right\}^2$$

and

$$(22) \quad L(\theta_o, s; G_n) = \sum_{j=1}^{\infty} s_j d^2_{nj}(\theta_o) \ , \quad s_j > 0 \quad \text{for all } j \ .$$

Schoenfeld (1977) examines $K(\theta_o, r; G_n)$ in detail for $\sum_{j=1}^{\infty} r_j^2 < \infty$ both under H_o and under contiguous alternative densities of the form

$$(23) \quad P_n(x) = f_{\theta_o}(x) + \frac{1}{\sqrt{n}} h\left(F_{\theta_o}(x)\right) + \frac{1}{n} k_n\left(F_{\theta_o}(x)\right) \ ,$$

with $\int_0^1 h^2(u) du < \infty$ and $|k_n(u)| < m(u)$ for all n with $\int_0^1 m^2(u) du < \infty$.

(Actually he studies the signed square root of $K(\theta_o, r; G_n)$.)

He derives an asymptotically optimal choice of the r_j to be of the form

$$(24) \quad r_j^{opt} = \int h(u) \zeta_j(u) du, \ j = 1, 2, \ldots, \ . \quad \text{Under } H_o, n K(\theta_o, r; G_n) \text{ is}$$

of course asymptotically distributed as a chi-square with one degree of freedom, when divided by $\sum_{j=1}^{\infty} r_j^2$.

Under the uniform boundedness condition on the $\zeta_j(\cdot)$, we obtain

the null distribution of $nL(\theta_o, s; G_n)$ as

$$nL(\theta_o, s; G_n) \xrightarrow{\ d\ } \sum_{j=1}^{\infty} s_j Z_j^2 \ ,$$

where the Z_j are iid unit normal random variables. Under alternatives of the form (23),

$$nL(\theta_o, s; G_n) \xrightarrow{\ d\ } \sum_{j=1}^{\infty} s_j \{Z_j + \int_0^1 h(u)\zeta_j(u)du\}^2 \ .$$

The null distribution can be approximated using the results of Solomon and Stephens (1977) or, most simply, Gregory (1980, p. 121).

Results on the asymptotic power of tests based upon statistics of the form of $nL(\theta_o, s; G_n)$ can also be easily obtained using the results of Gregory (1980). Let two discrepancies based on different sequences of positive weights be denoted by

$$(25) \quad nL(\theta_o, s; G_n) = \sum_{j=1}^{\infty} s_j \{\sqrt{n}\ d_{nj}(\theta_o)\}^2 \quad \text{and}$$

$$(26) \quad nL(\theta_o, s^*; G_n) = \sum_{j=1}^{\infty} s_j^* \{\sqrt{n}\ d_{nj}(\theta_o)\}^2 \ .$$

Also let $j(i)$ denote the index of the ith largest s_j, i.e. $s_{j(1)} \geq s_{j(2)} \geq \cdots > 0$ and similarly define $j^*(i)$. Let $n_i(n_i^*)$ be the multiplicity of the ith largest distinct value in $\{s_j\}(\{s_j^*\})$. If we denote the limiting power of a size α test against the sequence in (23) using $nL(\theta_o, s; G_n)$ $(nL(\theta_o, s^*; G_n))$ by $p(\alpha)$ $(p^*(\alpha))$, then by Theorem 2.5 of Gregory (1980), with $a_j = \int_0^1 h(u)\delta_j(u)du$, $A^2 = \sum_{i=1}^{n_1} a_{j(i)}^2$, $A^{*2} = \sum_{i=1}^{n_1^*} a_{j^*(i)}^2$,

$$\lim_{\alpha \to 0} p(\alpha)/p^*(\alpha) = 0 , \qquad \text{if } A^* > A \quad \text{or}$$

$$A^* = A \quad \text{and} \quad n_1^* < n_1$$

$$= \infty , \qquad \text{if } A > A^* \quad \text{or}$$

$$A^* = A \quad \text{and} \quad n_1 < n_1^*$$

$$= \frac{\exp\left\{\frac{1}{2} \sum_{k=n_1+1}^{\infty} a_j^2(k) s_j(k) / (s_j(1) - s_j(k))\right\}}{\exp\left\{\frac{1}{2} \sum_{k=n_1^*+1}^{\infty} a_{j*}^2(k) s_{j*}^*(k) / (s_{j*}^*(1) - s_{j*}^*(k))\right\}} ,$$

$$\text{if } A = A^* \quad \text{and} \quad n_1 = n_1^* .$$

Several observations can be made from this result concerning the relative efficiency of these tests.

i) If $n_1 = n_1^* = 1$, then the condition $A^* > A$ is just

$$a_{j*(1)}^2 > a_{j(1)}^2, \quad \text{i.e.,} \quad \{\int_0^1 h(u)\zeta_{j*(1)}(u)du\}^2 > \{\int_0^1 h(u)\zeta_{j(1)}(u)du\}^2 .$$

Thus, if distinct weights s_i are to be used, the largest weight should be given to the ζ_j with the largest coefficient in the expansion

$$h(u) \sim \sum_{j=1}^{\infty} t_j \zeta_j(u) , \qquad 0 \le u \le 1$$

for the direction of the alternatives in (23).

ii) We see furthermore from this that the ideal basis $\{\zeta_j(u), j = 1, \ldots\}$ would have $h(u)$ as a member. Hence, no test of the form $n \sum_{j=1}^{\infty} s_j d_{nj}^2(\theta_0)$ will be as efficient as a test of the form $\left\{\frac{1}{\sqrt{n}} \sum_{j=1}^{n} h(F_{\theta_0}(x_i))\right\}^2$, unless $s_j = 0$ for $j \ge 2$ and $\zeta_1(u) = h(u)$ almost everywhere with respect to Lebesque measure.

iii) Thus, in general tests of the type $nL(\theta_o, s; G_n)$ are a poor choice if there is much knowledge concerning the likely alternatives to be encountered, although they do possess (if $s_j > 0$ for all j) the desirable omnibus property of being consistent against alternatives of the form of (23), and are hence appropriate for poorly specified alternatives.

Seldom, however, are null goodness-of-fit hypotheses simple. Most involve the estimation of one or more parameters. Schoenfeld (1979) discusses the behavior of tests of the form of $nK(\hat{\theta}_n, r; G_n)$ when the estimation is asymptotically first-order-efficient. He shows that if the constants $\{r_j, j = 1, 2, \ldots\}$ are chosen to be orthogonal to a specified subspace of dimension equal to that of θ, the null distribution of $nK(\hat{\theta}_n, r; G_n)$ is, under some regularity conditions, the same as that of $nK(\theta_o, r; G_n)$.

Similarly, let θ be, for the sake of simplicity, one-dimensional, and let $\hat{\theta}_n$ be sufficiently regular that, when sampling from F_{θ_o},

$$\sqrt{n}\left\{\hat{\theta}_n - \theta_o - \frac{1}{n}\sum_{i=1}^{n}\psi(X_i; \theta_o)\right\} \xrightarrow{P} 0 \ .$$

Here, $\psi(\cdot; \theta_o)$ is usually termed the influence curve (see Hampel (1974)) of $\hat{\theta}_n$. If $\psi(\cdot; \theta_o)$ is such that $\int_o^1 \psi^2\left(F_o^{-1}(u); \theta_o\right)du < \infty$, it then possesses a Fourier expansion of the form

$$(27) \quad \psi\!\left(F_{\theta_o}^{-1}(u);\theta_o\right) = \sum_{j=1}^{\infty} a_j \zeta_j(u) \ .$$

If the r_j are such that $a_j r_j \equiv 0$, then $\sqrt{n}(\hat{\theta}_n - \theta_o)$ and $nL(\hat{\theta}_n,r;G_n)$ are asymptotically independent and $n\{L(\hat{\theta}_n,r;G_n) - L(\theta_o,r;G_n)\} \xrightarrow{P} 0$, permitting use of the critical points of $nL(\theta_o,r;G_n)$ in conducting the composite hypothesis tests. This also makes it possible to estimate the unknown parameter and conduct an asymptotically independent and parameter-free test of the fit of the proposed model. An estimator with an odd influence curve would thus, under the appropriate smoothness conditions, be asymptotically independent of a test which put nonzero weight only on the even components (looking for tailweight or kurtosis departures from a conjectured symmetric model).

5. Summary

Minimum distance methods provide a large class of estimation procedures possessing interesting analogies to other estimation methods. A frequency domain analysis can suggest which aspects of the efficient score should and should not be copied to balance the competing criteria of efficiency and robustness. MD estimation leads to a natural class of goodness-of-fit tests, and provides (once again via frequency domain insight) a general method for asymptotically parameter-free goodness-of-fit test construction in the composite goodness-of-fit problem.

BIBLIOGRAPHY

Bolthausen, E. (1977). Convergence in distribution of minimum-distance estimators. _Metrika_, _24_, 215-227.

Boos, D. D. (1980). Minimum distance estimators for location and goodness of fit. To appear in _Journal of the American Statistical Association_.

Durbin, J. and Knott, M. (1972). Components of Cramer-von Mises statistics. _Journal of the Royal Statistical Society_, _B_, 34, 290-307.

Gregory, G. G. (1980). On efficiency and optimality of quadratic tests. _Annals of Statistics_, _8_, 116-131.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. _Journal of the American Statistical Association_, _69_, 383-393.

Hodges, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests. _Annals of Mathematical Statistics_, _34_, 598-611.

Hogg, R. V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. _Journal of the American Statistical Association_, _69_, 909-927.

Huber, P. J. (1977). Robust Statistical Procedures. Bristol, England. Society for Industrial and Applied Mathematics.

Knusel, L. F. (1969). Uber minimum-distance-schatzungen. Ph.D. dissertation, Eidgenossischen Technischen Hochscule, Zurich.

Littell, R. C. and Rao, P. V. (1975). Estimation based on the Kolmogorov-Smirnov goodness-of-fit statistic. Technical Report 98, University of Florida, Department of Statistics.

Parr, W. C. and DeWet, T. (1979). Minimum CVM-norm parameter estimation. Unpublished manuscript.

Parr, W. C. and Schucany, W. R. (1980). Minimum distance and robust estimation. To appear in _Journal of the American Statistical Association_.

Parr, W. C. (1980). Minimum distance estimation: Survey and bibliography. Unpublished manuscript.

Pollard, D. (1980). The minimum distance method of testing. _Metrika_, _27_, 43-70.

Sahler, W. (1970). Estimation by minimum discrepancy methods. *Metrika*, 16, 85–106.

Schoenfeld, D. (1977). Asymptotic properties of tests based on linear combinations of the orthogonal components of the Cramér-von Mises statistic. *Annals of Statistics*, 5, 1017–1026.

Schoenfeld, D. (1979). Analyzing goodness of fit using the orthogonal components of the Cramér-von Mises statistic with estimated parameters. To appear in *Annals of Statistics*.

Solomon, H. and Stephens, M. A. (1977). Distribution of a sum of weighted chi-square variables. *Journal of the American Statistical Association*, 72, 881–885.

Wolfowitz, J. (1957). The minimum distance method. *Annals of Mathematical Statistics*, 28, 75–88.

Appendix

We present a brief sketch of the proof of Theorem 4. Define

$$h(t) = \frac{\dot{L}(t,b;G) - \dot{L}(\theta_o,b;G)}{t - \theta_o}$$

$$= \frac{\dot{L}(t,b;G)}{t - \theta_o} \qquad \text{for} \qquad t \neq \theta_o \; ,$$

$$= \ddot{L}(\theta_o,b;G) \qquad \text{for} \qquad t = \theta_o \; .$$

Hence, we have

$$T_L[G_n] - \theta_o = \frac{\dot{L}(T_L[G_n],b;G)}{h(T_L(G_n))} \; .$$

Further define the function

$$H(G_n) = \frac{\ddot{L}(\theta_o,b;G)}{h(T_L(G_n))}$$

and the differential

$$D(G_n - G) = \frac{\displaystyle\int_{-\infty}^{\infty} \sum_{j=1}^{\infty} b_j e_j(\theta_o)\zeta_j(F_{\theta_o}(x))d(G_n - G)(x)}{-\displaystyle\sum_{j=1}^{\infty} b_j e_j^2(\theta_o)} \; ,$$

which is the "average" of $IC_{T_L,G}(x)$ over the data points, our proposed linear approximation for $T_L[G_n] - \theta_o$.

Using the above, we write

$$T_L(G_n) - \theta_o - H(G_n)D(G_n - G)$$

$$= \frac{1}{h(T_L[G_n])}\left\{\dot{L}(T[G_n],b;G) - 2\int_{-\infty}^{\infty}\sum_{j=1}^{\infty}b_j e_j(\theta_o)\zeta_j(F_{\theta_o}(x))d(G_n - G)(x)\right\}$$

$$= \frac{2}{h(T_L(G_n))}\sum_{j=1}^{\infty}b_j\left\{\int_{-\infty}^{\infty}\zeta_j(F_{T_L[G_n]}(x))dG(x)\int_{-\infty}^{\infty}\zeta_j'(F_{T_L[G_n]}(x))\dot{F}_{T_L[G_n]}(x)dG(x)\right.$$

$$\left. - e_j(\theta_o)\int_{-\infty}^{\infty}\zeta_j(F_{\theta_o}(x))d(G_n - G)(x)\right\}.$$

This can be shown by lengthy algebra (utilizing repeatedly the fact that $\dot{L}(T_L(G_n),b;G_n) = \dot{L}(\theta_o,b;G) = 0$) to be equal to

$$\frac{2}{h(T_L[G_n])}\sum_{j=1}^{\infty}b_j\left[\int_{-\infty}^{\infty}\{\zeta_j(F_{\theta_o}(x)) - \zeta_j(F_{T_L[G_n]}(x))\}d(G_n - G)(x)\int_{-\infty}^{\infty}\zeta_j'(F_{T_L[G_n]}(x))\dot{F}_{T_L[G_n]}(x)dG(x)\right.$$

$$\left. + \int_{-\infty}^{\infty}\zeta_j(F_{\theta_o}(x))d(G_n - G)(x)\int_{-\infty}^{\infty}\{\zeta_j'(F_{\theta_o}(x))\dot{F}_{\theta_o}(x) - \zeta_j'(F_{T_L[G_n]}(x))\dot{F}_{T_L[G_n]}(x)\}dG(x)\right].$$

Consider the two terms inside the outer brackets separately. The contribution of the second can be bounded in absolute value by

$$\frac{2}{|h(T_L(G_n))|}\sum_{j=1}^{\infty}|b_j|\sup_x|G_n(x) - G(x)| \; ||\zeta_j(F_{\theta_o}(\cdot))||_V \; .$$

$$\int_{-\infty}^{\infty}|\zeta_j'\{F_{\theta_o}(x)\}\dot{F}_{\theta_o}(x) - \zeta_j'\{F_{T_L[G_n]}(x)\}\dot{F}_{T_L[G_n]}(x)|dG(x)$$

$$\leq \frac{2 \sup_x |G_n(x) - G(x)|}{h(T_L(G_n))} \sum_{j=1}^{\infty} |b_j| \, ||\zeta_j(F_{\theta_o}(\cdot))||_V \cdot o_p(1)$$

by condition ii) and known properties of the Kolmogorov-Smirnov statistic. The contribution of the first term is disposed of similarly (using condition iii)), giving

$$T_L(G_n) - \theta_o - H(G_n)D(G_n - G) = o_p(n^{-\frac{1}{2}})$$

since $H(G_n) \xrightarrow{p} 1$. Hence, by the Lindeberg-Levy central limit theorem for iid summands,

$$\sqrt{n}\{T_L(G_n) - \theta_o\} \xrightarrow{d} N(0, \sigma_L^2(\theta_o)).$$

Table 1

Relative Efficiences of Location Estimators
T_k Based on Truncated Fourier Approximates

Order of Approximation	Normal	Laplace
1	.90	.79
3	.96	.89
5	.98	.92
7	.99	.94
∞	1.00	1.00

Figure 1: Plot of Fourier Approximations to Fisher Score for Normal Density
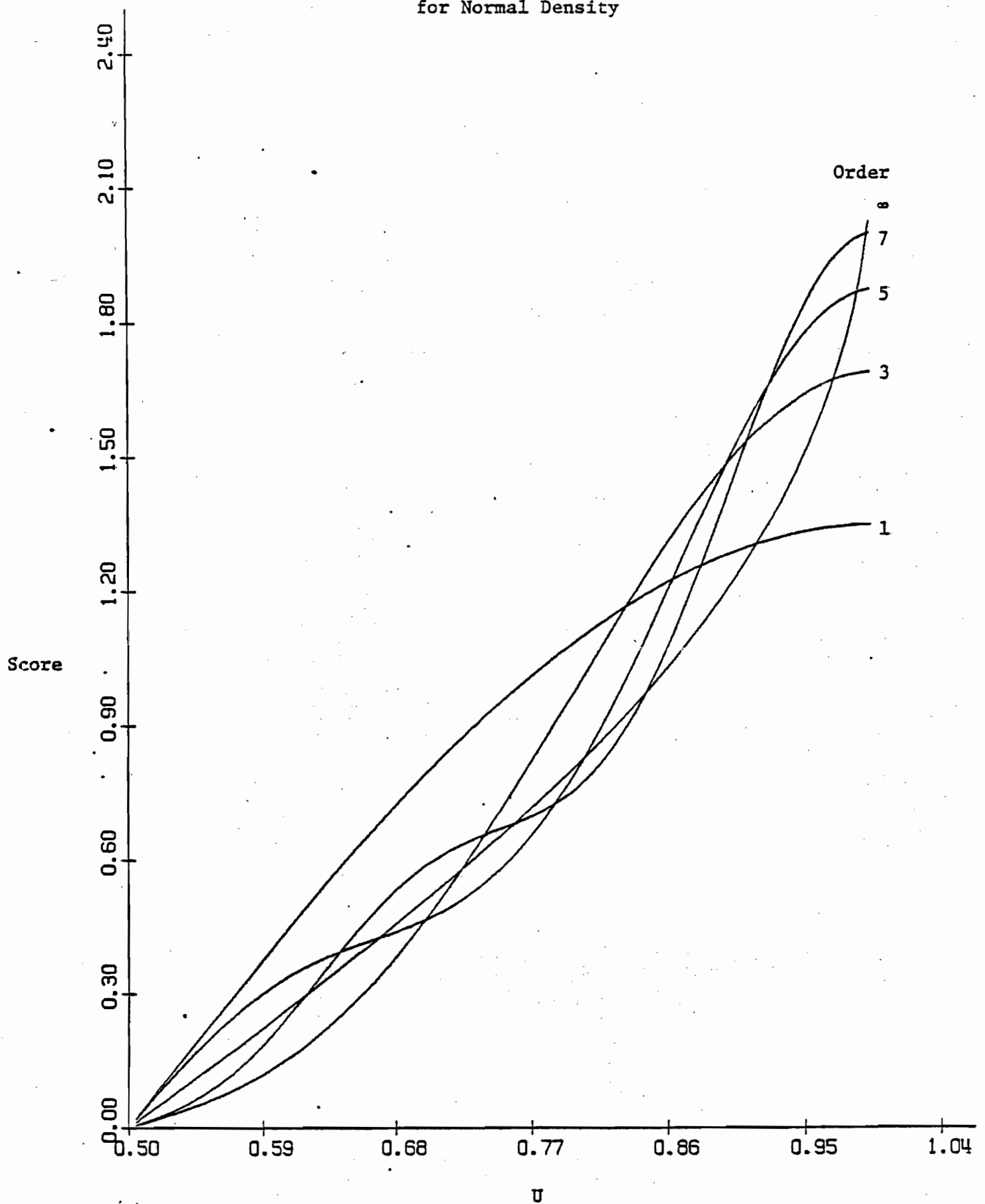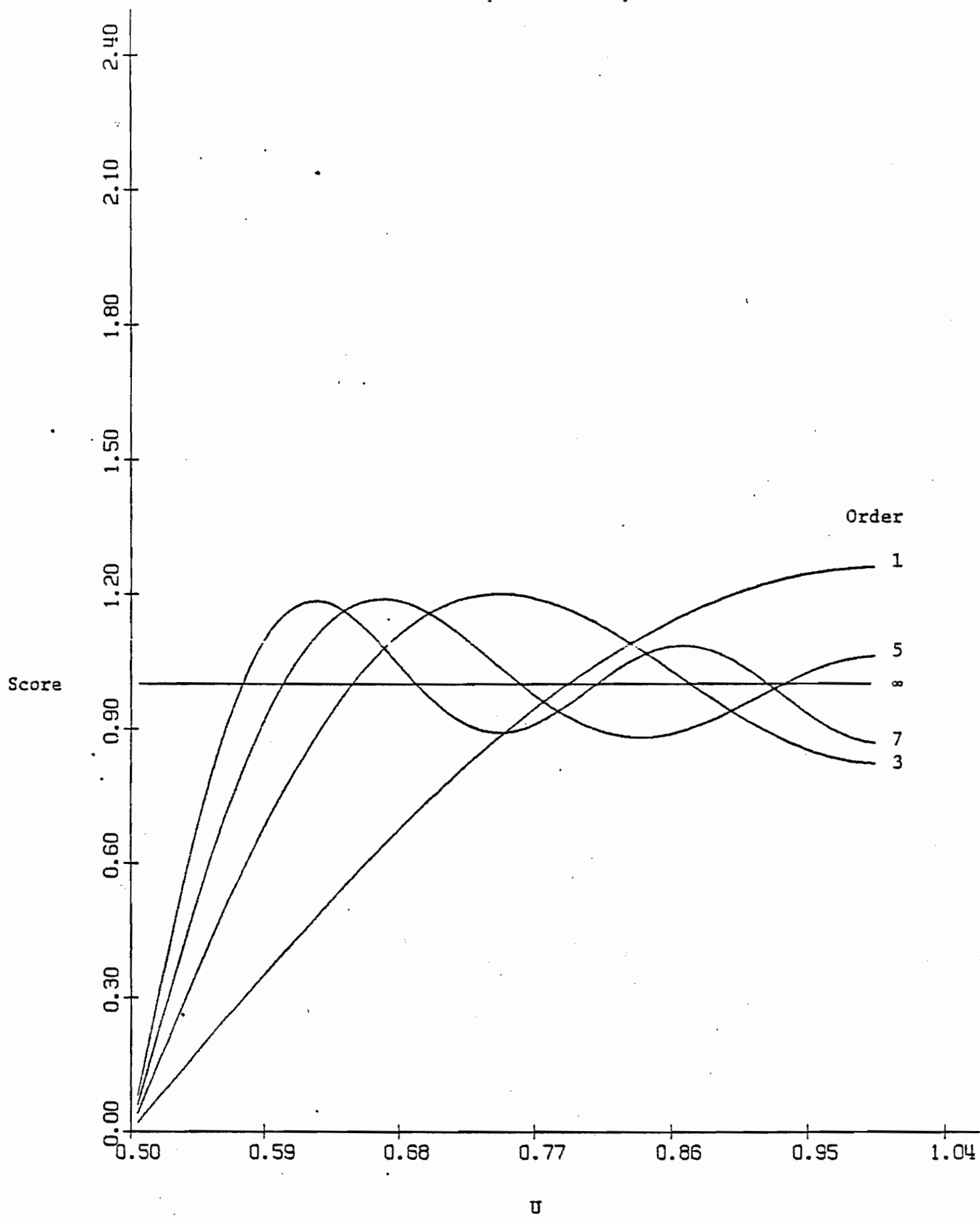
Figure 2: Plot of Fourier Approximations to Fisher Score
for Laplace Density

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 139	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Minimum Distance Estimation and Components of Goodness of Fit Statistics		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER 139
7. AUTHOR(s) William C. Parr William R. Schucany		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0439
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southern Methodist University Dallas, Texas 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042 280
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE Sept , 1980
		13. NUMBER OF PAGES 30
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Components of goodness-of-fit statistics; Harmonic analysis; M-estimation; Minimum distance estimation; Robustness.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The relationship of minimum distance (MD) estimation to other methods of estimation is considered. M-estimation is viewed as a special case, with interesting interpretations in terms of the defining ψ - function as related to components of goodness-of-fit statistics and modified Fourier approximations to the efficient score. Applications to the composite and simple goodness-of-fit problems are considered.

DD FORM 1473 1 JAN 73   EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |