

A CONSISTENT ESTIMATE OF A
NONPARAMETRIC SCALE PARAMETER

by

Gerald L. Sievers

Department of Mathematics
Western Michigan University

Department of Statistics
University of North Carolina

SUMMARY

A consistent estimate is proposed for the scale parameter $\int f^2$ in the model $Y_i = \mu_i + e_i$, $1 \leq i \leq n$, where the μ_i are unknown location parameters and the e_i are independent, identically distributed random errors with density function f . This parameter arises in the variance formula for rank estimates of location. The proposed estimate is based on differences of residuals $Y_i - \hat{\mu}_i$, where $\hat{\mu}_i$ is an estimate of μ_i . When the μ_i follow the structure of the general linear model, the estimate is shown to be consistent under the usual assumptions on the design matrix. The estimate does not require the symmetry of the density f .

Key Words: Nonparametric, scale parameter, rank estimates, linear model

This research was supported by the Office of Naval Research Contract No. N00014-78-C-0637 (NR 042-407). Distribution of this document is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

1. INTRODUCTION

Let n observations Y_1, Y_2, \dots, Y_n follow the model $Y_i = \mu_i + e_i$, $1 \leq i \leq n$, where the μ_i are unknown location parameters and the e_i are independent, identically distributed random variables with density f . Of particular interest here is the case where the μ_i follow the structure of a linear model.

Consider the problem of estimating the scale parameter $\gamma = \int f^2(x) dx$. This parameter arises in various statistical procedures based on ranks. For instance, in one-sample, two-sample and linear model problems it appears in the variances of rank estimates when Wilcoxon scores are used. In linear model problems, Hettmansperger and McKean (1976), (1977) proposed test statistics that are based on the drop in a dispersion function between full and reduced models and this difference must be scaled by γ .

For the one-sample and two-sample problems, Lehmann (1963) proposed a consistent estimate of γ based on the length of a distribution-free confidence interval for a shift parameter. For the linear model problem, under the assumption that f is symmetric about zero, Hettmansperger and McKean (1976), (1977) proposed using Lehmann's estimate of γ computed from the residuals and proved it was consistent.

In this paper an estimate of γ is proposed that is applicable for linear models and more general situations without assuming symmetry for the density f . To describe this estimate consider residuals $\hat{Z}_i = Y_i - \hat{\mu}_i$, where $\hat{\mu}_i$ is an estimate of μ_i , $1 \leq i \leq n$. Let $\hat{H}_n(t)$ be a weighted empirical cdf of the absolute values of differences of residuals,

$$\hat{H}_n(t) = \sum_{i < j} a_{ij} \phi(|\hat{Z}_j - \hat{Z}_i|, t)$$

where $\phi(u, v) = 0, \frac{1}{2}, 1$ as $u > v, u = v, u < v$ and $a_{ij} \geq 0$ with $\sum_{i < j} a_{ij} = 1$. The proposed estimate of γ is

$$(1.1) \quad \hat{\gamma} = \hat{H}_n(\hat{t}/\sqrt{n}) / (2\hat{t}/\sqrt{n})$$

where \hat{t} is the α -th quantile of $\hat{H}_n(t)$, $0 < \alpha < 1$.

Another similar estimate is

$$(1.2) \quad \gamma^* = \hat{H}_n(W_k) / 2W_k = k / 2M W_k$$

where $W_1 \leq \dots \leq W_M$ are the $M = \binom{n}{2}$ ordered absolute differences $|\hat{Z}_j - \hat{Z}_i|$ for $i < j$ and k is chosen so that $\sqrt{n} W_k = O_p(1)$.

The unweighted (or equal weight) case provides the basic estimate while weights can be used for special purposes. For example, the use of 0-1 weights would restrict the number of differences used in \hat{H}_n and reduce computational effort. In analysis of variance models and stratified sampling models where the data arise naturally in groups, the use of weights $a_{ij} = 0$ if \hat{Z}_i, \hat{Z}_j are from different groups would leave the estimate dependent only on the within-group variation of the data. This intuitively appealing estimate has apparently not been considered in the literature.

For the one-sample problem with $\mu_i = \mu$ for all i and under the assumption of a symmetric distribution, Antille (1974) considered an estimate of γ based on the number of averages $(\hat{Z}_j + \hat{Z}_i) / 2$ in the interval $(0, 1/\sqrt{n})$ for $i < j$. This estimate is basically a window estimate of the density of $(Y_1 + Y_2) / 2$ at μ . The window is not symmetric about zero and the window width does not adjust to the spread of the data.

Schweder (1975) proposed a Lehmann-type estimate of γ (in formula (3.1)) which is basically a window estimate and recommended a uniform window symmetric about zero. Assuming $\mu_i = 0$ for all i so that no nuisance parameters need be estimated, his estimate is

$$\begin{aligned} & \sum_{i,j} \phi(|Y_j - Y_i|, h_n/2) / n^2 h_n \\ &= ((n-1)/n) H_n(h_n/2) / h_n + (1/n h_n) \\ &\doteq H_n(h_n/2) / h_n, \end{aligned}$$

where h_n is the width of a uniform window and H_n is the empirical cdf of $|Y_j - Y_i|$ for $i < j$. The estimate (1.1) is seen to be similar to this type of estimate with a window width of $O_p(n^{-1/3})$. In proving consistency and asymptotic normality, Schweder assumed the window width satisfies $\sqrt{n} h_n \rightarrow \infty$, ruling out the $O(n^{-1/3})$ cases, and recommended $h_n = O(n^{-1/3})$.

The estimate $\hat{\gamma}$ differs from Schweder's estimate by allowing for the estimation of nuisance parameters in using residuals rather than iid variables. Also it uses a window width estimated from the data. In an actual application of Schweder's estimate the choice of an appropriate s for $h_n = s n^{-1/3}$ is difficult to make and his Monte Carlo study indicates that the bias of the estimate is sensitive to the choice made.

The estimate γ^* of (1.2) can be viewed as a nearest neighbor type estimate based on the absolute values of the differences of residuals. Nearest neighbor density function estimates are discussed in Moore and Yackel (1976), (1977).

Cheng and Serfling (1981) considered estimates of efficiency-related functionals, including γ as a special case.

2. MAIN RESULTS

The main results are presented in this section. The proofs can be found in the Appendix.

For each integer $n = 1, 2, \dots$, assume

$$(2.1) \quad \underline{Y} = \underline{\mu}_n + \underline{e}_n,$$

where $\underline{Y}_n = (Y_{1n}, \dots, Y_{nn})'$ is an $n \times 1$ random vector, $\underline{\mu}_n = (\mu_{1n}, \dots, \mu_{nn})'$ is an $n \times 1$ parameter vector and \underline{e}_n is an $n \times 1$ vector of independent, identically distributed random variables, each with density function f . At times the n will be suppressed to simplify the notation. For an $n \times 1$ vector $\underline{\eta}$, define an $n \times 1$ vector

$$\begin{aligned} \underline{Z}(\underline{\eta}) &= \underline{Y} - \underline{\eta} \\ &= (Z_1(\underline{\eta}), \dots, Z_n(\underline{\eta}))' . \end{aligned}$$

Let an indicator function be defined by $\phi(u, v) = 0, \frac{1}{2}, 1$ as $u > v$, $u = v$, $u < v$. Suppose for each integer $n = 1, 2, \dots$ a set of weights is given $a_{ijn} = a_{ij}$ for $1 \leq i < j \leq n$ with $\sum_{i < j} a_{ij} = 1$. Consider the weighted empirical cdf of the absolute values of the differences in $\underline{Z}(\underline{\eta})$ given by

$$H_n(t, \underline{\eta}, \underline{Y}) = \sum_{i < j} a_{ij} \phi(|Z_j(\underline{\eta}) - Z_i(\underline{\eta})|, t) .$$

This is not quite a proper cdf since $\phi = \frac{1}{2}$ in case of ties.

REMARK 2.1

- (i) $H_n(t, \underline{\eta}, \underline{Y}) = H_n(t, \underline{\eta} - \underline{\mu}, \underline{Z}(\underline{\mu}))$
- (ii) $H_n(t, \underline{\eta}, \underline{Y}) \Big|_{\underline{\mu}} \stackrel{L}{=} H_n(t, \underline{\eta} - \underline{\mu}, \underline{Y}) \Big|_{\underline{0}}$
- (iii) $H_n(t, \hat{\underline{\mu}}(\underline{Y}), \underline{Y}) \Big|_{\underline{\mu}} \stackrel{L}{=} H_n(t, \hat{\underline{\mu}}(\underline{Y}), \underline{Y}) \Big|_{\underline{0}}$

if $\hat{\underline{\mu}}(\underline{Y})$ is an estimate of $\underline{\mu}$ based on \underline{Y} such that $\hat{\underline{\mu}}(\underline{Y} - \underline{\mu}) = \hat{\underline{\mu}}(\underline{Y}) - \underline{\mu}$.

The remark shows that with a translation-invariant location estimate, it is sufficient to deal with the case $\mu = 0$.

For each $n = 1, 2, \dots$ let M_n be a measurable subset of n -dimensional Euclidean space and consider the following assumptions for the sequence $\{M_n\}_{n=1}^{\infty}$.

ASSUMPTION (A₁)

$$\sup_{\mu_n \in M_n} \max_{1 \leq i < j \leq n} |\eta_{jn} - \eta_{in}| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

ASSUMPTION (A₂) For any $\delta > 0$, there exists integers N_1 and L and partitions $M_n(1), \dots, M_n(L)$ of M_n such that for each $\ell=1, \dots, L$, if

$$d_{ij}^U(\ell) = \sup\{\mu_{jn} - \mu_{in} : \mu_n \in M_n(\ell)\}$$

$$d_{ij}^L(\ell) = \inf\{\mu_{jn} - \mu_{in} : \mu_n \in M_n(\ell)\}$$

then $(1/n) \sum_{i < j} (d_{ij}^U(\ell) - d_{ij}^L(\ell))^2 < \delta$ for all $n \geq N_1$.

Two other assumptions that will be needed are as follows.

ASSUMPTION (A₃) There exists constants N_0 and B_0 such that

$$n^2 \sum_{i < j} a_{ij}^2 \leq B_0^2 \quad \text{for all } n \geq N_0.$$

ASSUMPTION (A₄) If $G(t)$ denotes the cdf of the difference of two independent random variables, each with density f , then G has a density $g = G'$ with $g(t)$ continuous at $t = 0$ and $0 < g(0) < \infty$.

$$\text{Note } g(0) = \int f^2 = \gamma.$$

To study the asymptotic properties of the estimate in (1.1) consider the function

$$R_n(t, \eta_n, Y_n) = \sqrt{n} H_n(t/\sqrt{n}, \eta_n, Y_n) - 2t\gamma .$$

LEMMA 2.1 Let $\{M_n\}$ be a sequence of subsets satisfying assumption (A_1) and let assumptions (A_3) and (A_4) hold. Then for any $\eta_n \in M_n$, $n = 1, 2, \dots$ and any real number $t > 0$

$$R_n(t, \eta_n, Y_n) \Big|_0 \xrightarrow{P} 0 \text{ as } n \rightarrow \infty .$$

This lemma shows that $\sqrt{n} H_n(t/\sqrt{n}, \eta_n, Y_n)/2t$ converges in probability to γ but the result is not strong enough since an appropriate t must be chosen to fit the data and the nuisance location parameter μ_n must be estimated when μ_n is not zero. The following theorem shows that the convergence in the lemma is uniform and this will allow the construction of a suitable estimate of γ .

THEOREM 2.1 Let $\{M_n\}$ be a sequence of subsets satisfying assumptions (A_1) and (A_2) and let assumptions (A_3) and (A_4) hold. Let t_0 be a positive number. Then

$$\sup_{\substack{0 \leq t \leq t_0 \\ \eta_n \in M_n}} |R_n(t, \eta_n, Y_n)| \Big|_0 \xrightarrow{P} 0 \text{ as } n \rightarrow \infty .$$

Suppose that $\hat{\mu}_n = \hat{\mu}_n(Y_n)$ is an estimate of μ_n satisfying the following assumptions.

ASSUMPTION (A_5) For all $n = 1, 2, \dots$

$$\hat{\mu}_n(Y_n - \eta_n) = \hat{\mu}_n(Y_n) - \eta_n .$$

ASSUMPTION (A₆) For all $\epsilon > 0$ there exists a sequence of subsets $\{M_n\}$ satisfying (A₁) and (A₂) and an integer N such that

$$P_0(\hat{\mu}_n \in M_n) > 1 - \epsilon \quad \text{for all } n \geq N.$$

THEOREM 2.2 Let assumptions (A₃), (A₄), (A₅) and (A₆) hold. Let t_0 be a positive number. Then

$$\sup_{0 \leq t \leq t_0} |R_n(t, \hat{\mu}_n, Y_n)| \Big|_{\mu_n} \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty .$$

Let the $n \times 1$ residual vector be $\hat{z}_n = Y_n - \hat{\mu}_n$ and the empirical cdf of the absolute differences in \hat{z}_n be $\hat{H}_n(t) = \sum_{i < j} a_{ij} \phi(|\hat{z}_j - \hat{z}_i|, t)$. The quantile of order α for \hat{H}_n will be denoted by \hat{t}_α . Similarly the quantile of order α for $H(t) = G(t) - G(-t)$, $t > 0$, will be denoted by t_α .

LEMMA 2.2 Let assumptions (A₃), (A₅) and (A₆) hold. Then for $0 < \alpha < 1$,

$$\hat{t}_\alpha \Big|_{\mu_n} \xrightarrow{P} t_\alpha \quad \text{as } n \rightarrow \infty .$$

For a given $\alpha \in (0,1)$ there would exist t_0 so that $\hat{t}_\alpha \leq t_0$ with arbitrarily high probability for large n . Then Remark 2.1 (iii), Theorem 2.2 and Lemma 2.2 yield the main result that follows.

COROLLARY 2.1 Let assumptions (A₃), (A₄), (A₅) and (A₆) hold. Let $0 < \alpha < 1$ and

$$\hat{\gamma} = \hat{H}_n(\hat{t}_\alpha / \sqrt{n}) / (2\hat{t}_\alpha / \sqrt{n}) .$$

Then $\hat{\gamma} \Big|_{\mu_n} \xrightarrow{P} \gamma$ as $n \rightarrow \infty$.

The estimate $\hat{\gamma}$ arises from $R_n(t, \hat{\mu}_n, Y_n)$ by using $t = \hat{t}_\alpha = W_{[M\alpha]+1}$,

where $W_1 \leq \dots \leq W_M$ denote the $M = \binom{n}{2}$ ordered absolute differences $|\hat{z}_j - \hat{z}_i|$. The estimate γ^* of (1.2) arises by using $t = \sqrt{n} W_k$, where $k=k_n$ converges to zero at an appropriate rate. Specifically, if $0 < a_1 < a_2$ and k_n satisfies $a_1 < \sqrt{n} k_n/M < a_2$ for all n , then $\sqrt{n} W_{k_n}$ is bounded away from zero and from above in probability and Theorem 2.2 gives the following result.

COROLLARY 2.2 Let assumptions (A_3) , (A_4) , (A_5) and (A_6) hold. Assume there are positive constants a_1, a_2 such that $0 < a_1 < \sqrt{n} k_n/M < a_2$ for all n . Let

$$\gamma^* = k_n / 2M W_{k_n} .$$

Then $\gamma^* \Big|_{k_n} \xrightarrow{P} \gamma$ as $n \rightarrow \infty$.

3. APPLICATION TO LINEAR MODELS

In this section it is shown that the assumptions of the previous section are satisfied for a linear model under mild conditions and as a result, $\hat{\gamma}$ or γ^* can be used as a consistent estimate of γ . Thus when using rank procedures, a consistent estimate of γ is available without assuming symmetry for the error distribution.

Consider the model (2.1) with

$$(3.1) \quad u_n = \beta_0 \underline{1} + \underline{X} \underline{\Delta} / \sqrt{n} ,$$

where $\underline{1}$ is an $n \times 1$ vector of ones, $\underline{X} = (x_{ij})$ is an $n \times p$ design matrix and $\underline{\Delta} = (\Delta_1, \dots, \Delta_p)'$ is a $p \times 1$ parameter vector. Let Δ_0 be a positive number and $\mathcal{D} = \{\underline{\Delta} : |\Delta_k| \leq \Delta_0 \text{ for } k = 1, \dots, p\}$. For each integer $n = 1, 2, \dots$, let

$$(3.2) \quad M_n = \{\beta_0 \underline{1} + \underline{X} \underline{\Delta} / \sqrt{n} : \underline{\Delta} \in \mathcal{D}\} .$$

ASSUMPTION (B₁) For each $k = 1, \dots, p$,

$$\max_{1 \leq i \leq n} |x_{ik} - \bar{x}_k| / \sqrt{n} \rightarrow 0 \text{ as } n \rightarrow \infty ,$$

where \bar{x}_k is the average of the k^{th} column of \underline{X} .

ASSUMPTION (B₂)

$$(1/n) \underline{X}'_c \underline{X}_c \rightarrow \underline{\Sigma} \text{ as } n \rightarrow \infty ,$$

where $\underline{\Sigma}$ is a $p \times p$ positive definite matrix and $\underline{X}_c = (\underline{I} - (1/n)\underline{J})\underline{X}$ is the centered design matrix.

REMARK 3.1 Let assumptions (B_1) and (B_2) hold. Then the sequence of subsets M_n in (3.2) satisfies assumptions (A_1) and (A_2) .

Proof: For $\eta_n \in M_n$ write $\eta_{jn} - \eta_{in} = \sum_{k=1}^p (x_{jk} - x_{ik}) \Delta_k / \sqrt{n}$. Then $\max_{1 \leq i < j \leq n} |\eta_{jn} - \eta_{in}| \leq 2\Delta_0 \sum_{k=1}^p \max_{1 \leq i \leq n} \{|x_{ik} - \bar{x}_k| / \sqrt{n}\}$, if $\Delta \in \mathcal{D}$. Then (B_1) implies (A_1) .

Next let $\delta' > 0$. Then there is a partition $\mathcal{D}_1, \dots, \mathcal{D}_L$ of \mathcal{D} such that for all $\ell = 1, \dots, L$, $\Delta, \Delta' \in \mathcal{D}_\ell$ implies $|\Delta_k - \Delta'_k| < \delta'$ for each $k = 1, \dots, p$. Then $M_n(\ell) = \{\beta_0 \underline{1} + \sum_{k=1}^p \Delta_k / \sqrt{n} : \Delta \in \mathcal{D}_\ell\}$ for $\ell = 1, \dots, L$, is a partition of M_n .

Now

$$\begin{aligned} & (1/n) \sum_{i < j} [d_{ij}^U(\ell) - d_{ij}^L(\ell)]^2 \\ &= (1/n) \sum_{i < j} \left[\sup_{\Delta, \Delta' \in \mathcal{D}_\ell} \sum_{k=1}^p (x_{jk} - x_{ik}) (\Delta_k - \Delta'_k) / \sqrt{n} \right]^2 \\ &\leq (1/n^2) \sum_{i < j} \sup_{\Delta, \Delta' \in \mathcal{D}_\ell} \left[\sum_{k=1}^p (x_{jk} - x_{ik})^2 \sum_{k=1}^p (\Delta_k - \Delta'_k)^2 \right] \\ &\leq (p\delta'^2/n^2) \sum_{k=1}^p \sum_{i < j} (x_{jk} - x_{ik})^2 \\ &= (p\delta'^2/n^2) \sum_{k=1}^p n \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2 \\ &= p\delta'^2 \text{ trace } (X_c' X_c / n) . \end{aligned}$$

Since δ' is arbitrary and (B_2) holds, it follows that (A_2) holds. \square

Suppose that $\hat{\Delta} = \hat{\Delta}(Y)$ is an estimate of Δ . The intercept parameter β_0 has no effect on matters here since it can be absorbed in the error variables of model (2.1).

ASSUMPTION (B₃) Let $\hat{\Delta}$ satisfy $\hat{\Delta}(\underline{Y} - \underline{X} \underline{\Delta} / \sqrt{n}) = \hat{\Delta}(\underline{Y}) - \underline{\Delta}$ and let $\hat{\Delta}$ be bounded in probability when $\underline{\Delta} = \underline{0}$ in model (3.1).

This assumption is satisfied for the usual rank estimates of $\underline{\Delta}$ proposed by Jaeckel (1972), Hettmansperger and McKean (1977) and others.

REMARK 3.2 Let assumptions (B₁) and (B₂) hold and let $\hat{\Delta}$ satisfy assumption (B₃). Then assumption (A₅) holds for $\underline{\mu}_n$ in the column space of \underline{X} and assumption (A₆) holds with M_n given by (3.2).

Proof: Assumption (A₅) follows directly from (B₃). For $\epsilon > 0$, the boundedness in probability of $\hat{\Delta}$ implies that a sufficiently large Δ_0 can be chosen in defining \mathcal{D} for (3.2) so that $P_0(\hat{\Delta} \in \mathcal{D}) > 1 - \epsilon$ for large n . Then (A₆) follows since $\hat{\Delta} \in \mathcal{D}$ implies $\hat{\underline{\mu}}_n = \beta_0 \underline{1} + \underline{X} \hat{\Delta} / \sqrt{n} \in M_n$. \square

These remarks show that (B₁), (B₂) and (B₃) can replace (A₅) and (A₆) in Corollaries 2.1 or 2.2 in the case of linear model. Thus under typical assumptions, $\hat{\gamma}$ or γ^* provides a consistent estimate of γ .

4. LEHMANN'S TWO-SAMPLE ESTIMATE

Lehmann (1963) proposed a consistent estimate of γ for the two-sample problem based on the length of a confidence interval for a shift parameter. It is interesting to note that the estimates $\hat{\gamma}$ and γ^* are computationally similar to Lehmann's estimate if the set of residuals is duplicated to form a second sample. Lehmann's work showing consistency is based on the assumption of two independent samples and would not apply in the present context.

To describe Lehmann's estimate let two samples be denoted by Y_1, \dots, Y_m and X_1, \dots, X_n . Define $U(\Delta) = \sum_i \sum_j \phi(X_i, Y_j - \Delta)$. Define Δ_U and Δ_L by $U(\Delta_U) \doteq d_\alpha - (1/2)$ and $U(\Delta_L) \doteq mn - d_\alpha + (1/2)$, where $d_\alpha = (mn/2) - z_{\alpha/2} (mn(m+n)/12)^{1/2}$ and z_α is the upper α^{th} quantile of the standard normal distribution. Then Lehmann's estimate of γ is $2z_{\alpha/2} / \sqrt{3(m+n)} (\Delta_U - \Delta_L)$.

Consider the case where the two samples are identical, that is $m = n$ and $Y_i = X_i$. Then $U(-\Delta) = n^2 - U(\Delta)$, $\Delta_L = -\Delta_U$ and Lehmann's estimate becomes

$$(4.1) \quad (2z_{\alpha/2} / \sqrt{6n}) / 2\Delta_U .$$

Now suppose that the residuals \hat{Z}_i of section 2 are duplicated to form two "samples". Use equal weights $a_{ij} = 1/\binom{n}{2}$. Then for $t > 0$,

$$\begin{aligned} \hat{H}_n(t) &= 1 - \binom{n}{2}^{-1} \sum_i \sum_j \phi(t, \hat{Z}_j - \hat{Z}_i) \\ &= 1 - \binom{n}{2}^{-1} U(t) . \end{aligned}$$

Using this in the equation defining Δ_U yields the equation

$$\hat{H}_n(\Delta_U) \doteq (-1/n) + (2n/(n-1)) z_{\alpha/2} / \sqrt{6n} .$$

Thus for large n , Δ_U satisfies

$$(4.2) \quad \hat{H}_n(\Delta_U) \doteq 2z_{\alpha/2} / \sqrt{6n}$$

and with (4.1) it follows that Lehmann's estimate is approximately $H_n(\Delta_U)/2\Delta_U$. This is the same form as the estimate $\hat{\gamma}$ in Corollary 2.1. Also note that in the two-sample model, Lehmann showed $\sqrt{n} \Delta_U$ converges in probability to a constant and this matches the behavior of \hat{t}_α in $\hat{\gamma}$. The probability levels α used in each estimate are not related in any useful way.

In Corollary 2.2, suppose k_n is defined by $k_n/M = 2z_{\alpha/2} / \sqrt{6n}$. Then (4.2) shows that Δ_U corresponds to W_{k_n} and the estimate γ^* is then directly in the form of Lehmann's estimate (4.1).

5. APPENDIX

Proof of LEMMA 2.1

Write

$$R_n(t, \eta_n, Y_n) = \sqrt{n} \sum_{i < j} a_{ij} (W_{ij} - 2t\gamma/\sqrt{n})$$

where $W_{ij} = \phi(|Y_j - Y_i - d_{ij}|, t/\sqrt{n})$ and $d_{ij} = \eta_{jn} - \eta_{in}$. Using

$$\begin{aligned} E_0(W_{ij}) &= G(d_{ij} + t/\sqrt{n}) - G(d_{ij} - t/\sqrt{n}) \\ &= g(\xi_{ij})(2t/\sqrt{n}) \end{aligned}$$

where $|\xi_{ij} - d_{ij}| < t/\sqrt{n}$ and $\gamma = g(0)$, it follows that

$$E_0(R_n(t, \eta_n, Y_n)) = 2t \sum_{i < j} a_{ij} (g(\xi_{ij}) - g(0)).$$

Since $\eta_n \in M_n$, assumption (A₁) implies that $\max_{1 \leq i < j \leq n} |d_{ij}| \rightarrow 0$ and then

$\max_{1 \leq i < j \leq n} |\xi_{ij}| \rightarrow 0$ as $n \rightarrow \infty$. The continuity of g at zero gives

$|g(\xi_{ij}) - g(0)| < \varepsilon$ for all $1 \leq i < j \leq n$ for any $\varepsilon > 0$ if n is sufficiently large. Hence for any $\varepsilon > 0$, $|E_0(R_n(t, \eta_n, Y_n))| < 2t\varepsilon$ for large n and it follows that $E_0(R_n(t, \eta_n, Y_n)) \rightarrow 0$ as $n \rightarrow \infty$.

Using assumption (A₃) and $|W_{ij}| \leq 1$, a standard argument shows that $\text{var}_0(R_n(t, \eta_n, Y_n)) \rightarrow 0$ as $n \rightarrow \infty$. With the mean and variance tending to zero the lemma follows. \square

Proof of THEOREM 2.1

This proof is somewhat long and it will be divided into pieces with some lemmas. Begin by choosing any $\varepsilon > 0$. Choose $\delta > 0$ sufficiently small so that $4g(0)B_0\sqrt{\delta} < \varepsilon/6$, where B_0 is specified in (A₃). For this δ we have

N_1 (take $N_1 > N_0$ of (A_3)), the partitions $M_n(\ell)$, $\ell=1, \dots, L$ and the extreme differences $d_{ij}^U(\ell)$ and $d_{ij}^L(\ell)$ of (A_2) . Choose $\delta' > 0$ so that $6g(0)\delta' < \epsilon/6$. Partition the interval $T = [0, t_0]$ into subintervals T_1, \dots, T_M of width less than δ' . For each $m = 1, \dots, M$ let

$$t_m^U = \sup\{t: t \in T_m\} \text{ and } t_m^L = \inf\{t: t \in T_m\} .$$

Then for each $m = 1, \dots, M$

$$(5.1) \quad |t_m^U - t_m^L| < \delta' .$$

For each $\ell=1, \dots, L$ and $m = 1, \dots, M$, define

$$S_{ij}(\ell, m) = \begin{cases} 1, & \text{if } Y_j - Y_i \in (d_{ij}^L(\ell) - t_m^U/\sqrt{n}, d_{ij}^U(\ell) - t_m^L/\sqrt{n}) \\ & \text{or } Y_j - Y_i \in (d_{ij}^L(\ell) + t_m^L/\sqrt{n}, d_{ij}^U(\ell) + t_m^U/\sqrt{n}) \\ 0, & \text{otherwise} \end{cases}$$

for $1 \leq i < j \leq n$ and $Q_{\ell, m} = \sqrt{n} \sum_{i < j} a_{ij} S_{ij}(\ell, m)$.

LEMMA 5.1 For each $\ell = 1, \dots, L$ and $m = 1, \dots, M$

$$\sup_{\substack{t, t' \in T_m \\ \eta, \eta' \in M_n(\ell)}} |R_n(t, \eta, Y) - R_n(t', \eta', Y)| \leq Q_{\ell, m} + 2\gamma\delta'$$

Proof: Fix ℓ and m and write

$$\begin{aligned} & R_n(t, \eta, Y) - R_n(t', \eta', Y) \\ &= \sqrt{n} \sum_{i < j} a_{ij} (W_{ij} - W_{ij}') - 2\gamma(t - t') \end{aligned}$$

where $W_{ij} = \phi(|Y_j - Y_i - d_{ij}|, t/\sqrt{n})$, $W_{ij}' = \phi(|Y_j - Y_i - d_{ij}'|, t'/\sqrt{n})$, $d_{ij} = \eta_{jn} - \eta_{in}$ and $d_{ij}' = \eta_{jn}' - \eta_{in}'$. Inspection of possible cases shows

that $|W_{ij} - W'_{ij}| \leq S_{ij}(\ell, m)$ for all $1 \leq i < j \leq n$, all $\eta, \eta' \in M_n(\ell)$ and all $t, t' \in T_m$. Then the lemma follows. \square

Using Assumption (A_4) , there is $\delta'' > 0$ such that $|d| < \delta''$ implies $g(d) < 2g(0)$. Then from Assumption (A_1) there exists $N_2 > N_1$ such that $|\eta_{jn} - \eta_{in}| < \delta''/2$ for all $\eta_n \in M_n$, all $1 \leq i < j \leq n$ and all $n \geq N_2$. This implies

$$(5.2) \quad (-\delta''/2) \leq d_{ij}^L(\ell) \leq d_{ij}^U(\ell) \leq (\delta''/2)$$

for all $\ell = 1, \dots, L$, all $1 \leq i < j \leq n$ and all $n \geq N_2$.

LEMMA 5.2 For all $\ell = 1, \dots, L$ and $m = 1, \dots, M$,

(a) there exists an integer N_3 such that

$$E_0(Q_{\ell, m}) \leq (\epsilon/6) + 4g(0)\delta' \quad \text{for all } n \geq N_3,$$

(b) $\text{var}_0(Q_{\ell, m}) \rightarrow 0$ as $n \rightarrow \infty$.

Proof:

$$\begin{aligned} E_0(Q_{\ell, m}) &= \sqrt{n} \sum_{i < j} a_{ij} E_0(S_{ij}(\ell, m)) \\ &= \sqrt{n} \sum_{i < j} a_{ij} \{ [G(d_{ij}^U(\ell) - t_m^L/\sqrt{n}) - G(d_{ij}^L(\ell) - t_m^U/\sqrt{n})] \\ &\quad + [G(d_{ij}^U(\ell) + t_m^U/\sqrt{n}) - G(d_{ij}^L(\ell) + t_m^L/\sqrt{n})] \} \\ &= \sqrt{n} \sum_{i < j} a_{ij} \{ g(\xi_{ij}(\ell, m)) [d_{ij}^U(\ell) - d_{ij}^L(\ell) + (t_m^U - t_m^L)/\sqrt{n}] \\ &\quad + g(v_{ij}(\ell, m)) [d_{ij}^U(\ell) - d_{ij}^L(\ell) + (t_m^U - t_m^L)/\sqrt{n}] \} \end{aligned}$$

where $d_{ij}^L(\ell) - t_m^U/\sqrt{n} \leq \xi_{ij}(\ell, m) \leq d_{ij}^U(\ell) - t_m^L/\sqrt{n}$ and $d_{ij}^L(\ell) + t_m^L/\sqrt{n} \leq v_{ij}(\ell, m) \leq d_{ij}^U(\ell) + t_m^U/\sqrt{n}$ for all $1 \leq i < j \leq n$.

Now there is an integer $N_3 > N_2$ such that $t_0/\sqrt{n} < \delta''/2$ for all $n \geq N_3$. This along with (5.2) implies

$$-\delta'' < \xi_{ij}(\ell, m) < \delta''$$

$$-\delta'' < v_{ij}(\ell, m) < \delta''$$

for all $1 \leq i < j \leq n$ and all $n \geq N_3$. Then by the choice of δ'' it follows that

$$\begin{aligned} E_0(Q_{\ell, m}) &\leq \sqrt{n} \ 4 \ g(0) \sum_{i < j} a_{ij} [d_{ij}^U(\ell) - d_{ij}^L(\ell) + (t_m^U - t_m^L)/\sqrt{n}] \\ &\leq 4 \ g(0) \ \sqrt{n} \sum_{i < j} a_{ij} [d_{ij}^U(\ell) - d_{ij}^L(\ell)] + 4 \ g(0) \delta' \end{aligned}$$

from (5.1) for all $n \geq N_3$. Applying the Cauchy-Schwarz inequality and assumptions (A_2) , (A_3) yields

$$\begin{aligned} E_0(Q_{\ell, m}) &\leq 4 \ g(0) \ [n^2 \sum_{i < j} a_{ij}^2 \ (1/n) \sum_{i < j} (d_{ij}^U(\ell) - d_{ij}^L(\ell))^2]^{1/2} + 4 \ g(0) \delta' \\ &\leq 4 \ g(0) \ B_0 \sqrt{\delta} + 4 \ g(0) \ \delta' \\ &\leq (\varepsilon/6) + 4 \ g(0) \ \delta' \end{aligned}$$

from the choice of δ .

The second part of this lemma follows from assumption (A_3) and the fact that $|S_{ij}(\ell, m)| \leq 1$ with a standard variance argument. \square

Choose an arbitrary $\varepsilon' > 0$. For each $\ell = 1, \dots, L$ and $m = 1, \dots, M$ choose a sequence $\tilde{n}_n^\ell \in M_n(\ell)$, $n = 1, 2, \dots$ and a number $t_m \in T_m$.

LEMMA 5.3 There exists an integer N_4 such that

$$P_0(|R_n(t_m, \tilde{n}_n^\ell, \tilde{y}_n)| > \varepsilon/3) \leq \varepsilon'/2LM$$

for all $\ell = 1, \dots, L$, all $m = 1, \dots, M$ and all $n \geq N_4$.

Proof: Apply Lemma 2.1 for each of the finite number of pairs (ℓ, m) . \square

With these preliminaries completed, the proof of the theorem will now be completed. For each $\ell = 1, \dots, L$ and $m = 1, \dots, M$ write

$$\begin{aligned}
 & P_0 \left(\sup_{\substack{t \in T_m \\ \eta_n \in M_n(\ell)}} |R_n(t, \eta_n, Y_n)| \geq \varepsilon \right) \\
 \leq & P_0 \left(\sup_{\substack{t \in T_m \\ \eta_n \in M_n(\ell)}} |R_n(t, \eta_n, Y_n) - R_n(t_m, \eta_n^\ell, Y_n)| + |R_n(t_m, \eta_n^\ell, Y_n)| \geq \varepsilon \right) \\
 \leq & P_0(Q_{\ell, m} + 2\gamma\delta' + |R_n(t_m, \eta_n^\ell, Y_n)| \geq \varepsilon) \\
 = & P_0(Q_{\ell, m} - E_0(Q_{\ell, m}) + E_0(Q_{\ell, m}) + |R_n(t_m, \eta_n^\ell, Y_n)| \geq \varepsilon) \\
 \leq & P_0(Q_{\ell, m} - E_0(Q_{\ell, m}) + |R_n(t_m, \eta_n^\ell, Y_n)| \geq 2\varepsilon/3) \\
 \leq & P_0(Q_{\ell, m} - E_0(Q_{\ell, m}) \geq \varepsilon/3) + P_0(|R_n(t_m, \eta_n^\ell, Y_n)| \geq \varepsilon/3) \\
 \leq & (9/\varepsilon^2) \text{Var}_0(Q_{\ell, m}) + (\varepsilon'/2LM) \quad , \\
 \leq & \varepsilon'/LM
 \end{aligned}$$

where Lemma 5.1 was used in the third line, Lemma 5.2(a) and the choice of δ' were used in the fifth line, Chebyshev's inequality and Lemma 5.3 were used in the seventh line, and the last line follows from Lemma 5.2(b), since the first term in the seventh line above can be made less than $\varepsilon'/2LM$ for all ℓ, m for n sufficiently large, say for $n \geq N_5 > N_4$. Finally,

$$\begin{aligned}
 & P_0 \left(\sup_{0 \leq t \leq t_0} |R_n(t, \eta_n, \tilde{y}_n)| \geq \varepsilon \right) \\
 & \quad \eta_n \in M_n \\
 & \leq \sum_{m=1}^M \sum_{\ell=1}^L P_0 \left(\sup_{t \in T_m} |R_n(t, \eta_n, \tilde{y}_n)| \geq \varepsilon \right) \\
 & \quad \eta_n \in M_n(\ell) \\
 & \leq \sum_{m=1}^M \sum_{\ell=1}^L \varepsilon' / LM = \varepsilon'
 \end{aligned}$$

for all $n \geq N_5$ and this completes the proof. \square

Proof of THEOREM 2.2

With assumption (A_5) and Remark 2.1 it is sufficient to prove the result when $\mu_n = 0$. Let $\varepsilon > 0$. Then

$$\begin{aligned}
 & P_0 \left(\sup_{0 \leq t \leq t_0} |R_n(t, \hat{\mu}_n, \tilde{y}_n)| > \varepsilon \right) \\
 & \leq P_0 \left(\sup_{0 \leq t \leq t_0} |R_n(t, \hat{\mu}_n, \tilde{y}_n)| > \varepsilon, \hat{\mu}_n \in M_n \right) + P_0(\hat{\mu}_n \notin M_n) \\
 & \leq P_0 \left(\sup_{0 \leq t \leq t_0} |R_n(t, \eta_n, \tilde{y}_n)| > \varepsilon \right) + P_0(\hat{\mu}_n \notin M_n). \\
 & \quad \eta_n \in M_n
 \end{aligned}$$

Then Theorem 2.1 and assumption (A_6) imply that this can be made arbitrarily small for n sufficiently large and the result follows. \square

Proof of LEMMA 2.2

The proof is basically straightforward and brief details will be given here. Assumption (A_5) implies that it is sufficient to deal with the case

$\mu_n = 0$. Noting that $\hat{Z}_j - \hat{Z}_i = (Y_j - Y_i) - (\hat{\mu}_{jn} - \hat{\mu}_{in})$, assumption (A₆) (only the (A₁) part is needed) implies that $\max_{1 \leq i < j \leq n} |\hat{\mu}_{jn} - \hat{\mu}_{in}|$ will be arbitrarily

small with arbitrarily high probability for large n and the empirical cdf of $\hat{Z}_j - \hat{Z}_i$ can then be bounded by empirical cdf's of $Y_j - Y_i$. Assumption (A₃) is used to show the weighted empirical cdf of $Y_j - Y_i$ converges in probability.

REFERENCES

ANTILLE, A. (1974). A linearized version of the Hodges-Lehmann Estimator. *Ann. Statist.* 2, 1308-1313.

CHENG, K.F. and Serfling, R.J. (1981). On estimation of a class of efficiency-related parameters. *Scand. Act. J.*, 83-92.

HETTMANSPERGER, T.P. and McKean, J.W. (1977). A robust alternative based on ranks to least squares in analyzing linear models. *Technometrics* 19, 275-284.

JAECKEL, L.A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* 43, 1449-1458.

LEHMANN, E.L. (1963). Nonparametric confidence intervals for a shift parameter. *Ann. Math. Statist.* 34, 1507-1512.

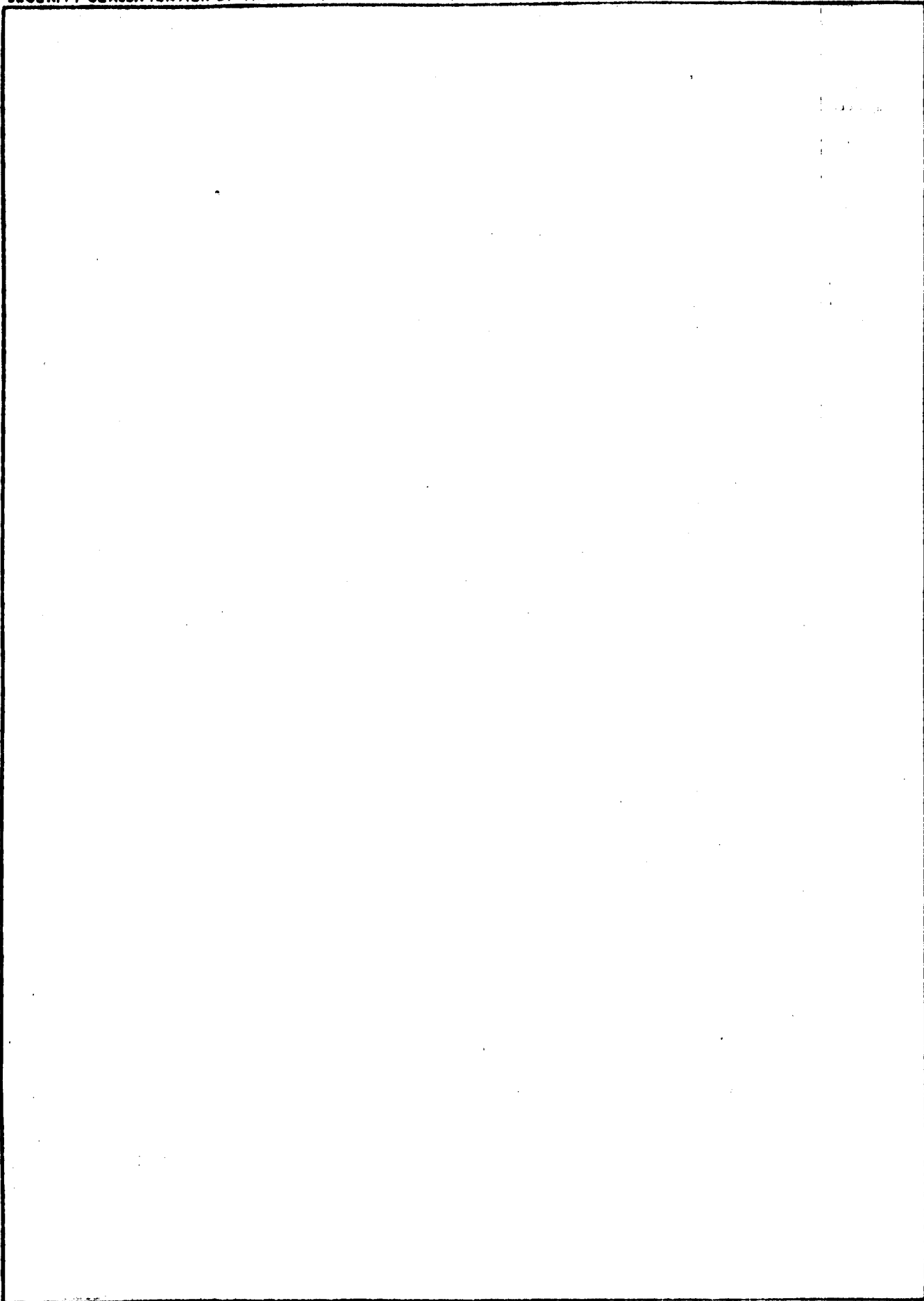
MCKEAN, J.W. and Hettmansperger, T.P. (1976). Tests of hypothesis based on ranks in the general linear model. *Comm. Statist.* A5, 693-709.

MOORE, D.S. and Yackel, J.W. (1976). Large sample properties of nearest neighbor density function estimators. *Statistical Decision Theory and Related Topics*. Academic Press, New York.

MOORE, D.S. and Yackel, J.W. (1977). Consistency properties of nearest neighbor density function estimates. *Ann. Statist.* 5, 143-154.

SCHWEDER, T. (1975). Window estimation of the asymptotic variance of rank estimators of location. *Scand. J. Statist.* 2, 113-126.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Consistent Estimate of a Nonparametric Scale Parameter		5. TYPE OF REPORT & PERIOD COVERED Technical Report 1981
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Gerald L. Sievers		8. CONTRACT OR GRANT NUMBER(s) N00014-78-C-0637
9. PERFORMING ORGANIZATION NAME AND ADDRESS Western Michigan University Kalamazoo, Michigan 49008		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-407
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program		12. REPORT DATE May 1982
		13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release: Distribution Unlimited.		
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) Nonparametric, scale parameter, rank estimates, linear models		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Please See Next Page		



20. ABSTRACT

A consistent estimate is proposed for the scale parameter $\int f^2$ in the model $Y_i = \mu_i + e_i$, $1 \leq i \leq n$, where the μ_i are unknown location parameters and the e_i are independent, identically distributed random errors with density function f . This parameter arises in the variance formula for rank estimates of location. The proposed estimate is based on differences of residuals $Y_i - \hat{\mu}_i$, where $\hat{\mu}_i$ is an estimate of μ_i . When the μ_i follow the structure of the general linear model, the estimate is shown to be consistent under the usual assumptions on the design matrix. The estimate does not require the symmetry of the density f .