The Annals of Statistics
1973 Vol. 1, No. 4, 597-616

ON SOME ANALOGUES TO LINEAR COMBINATIONS OF ORDER STATISTICS IN THE LINEAR MODEL[1]

BY P. J. BICKEL

University of California, Berkeley and Princeton University

We consider the general linear model with independent symmetric errors. In this context we propose and examine the large sample behavior of some estimates of the regression parameters. For the location model these statistics are linear combinations of order statistics. In general they depend on a preliminary estimate and the ordered residuals based on it. The asymptotic efficiency of these procedures is independent of the design matrix. Specifically analogues of the median and trimmed and Winsorized means are proposed.

1. Introduction. Linear combinations of order statistics have long been of interest as estimates of location (and scale) parameters. The reasons for this appear to be three-fold. First, they are easy to compute ([24]), second, by a suitable choice of weights they can be made asymptotically efficient if the shape of the underlying population is known ([2], [17], [8]), and third, suitable members of the family such as the median and the trimmed mean are robust estimates of location ([30], [3], [7], [9], [16]). Our aim is to generalize this class of procedures to the linear model.

Specifically, we consider the problem of estimating the regression parameters of a linear model as the number of observations becomes large and the number of regression parameters remains fixed. That is, we want to estimate $\boldsymbol{\beta}$ when we observe $\mathbf{X} = (X_1, \cdots, X_n)$ where

$$\mathbf{X} = \boldsymbol{\beta} C + \mathbf{E} \tag{1.1}$$

$\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$ is a vector of unknown regression parameters, $\mathbf{E} = (E_1, \cdots, E_n)$ is a vector of errors and $C = ||c_{ij}||_{p\times n}$ is a matrix of known regression constants of rank p. (Of course the coordinates of $\mathbf{X}$ as well as C depend on n. For convenience we suppress n dependence.) Throughout we shall suppose the E_i are independent and identically distributed with common cdf F and density f with respect to Lebesgue measure. If F is Gaussian with mean 0 the appropriate procedure to use is, of course, the least squares estimate

$$\hat{\boldsymbol{\beta}}_G = \mathbf{X}C'[CC']^{-1} . \tag{1.2}$$

It is well known that in the location model ($p = 1$, $c_{ij} \equiv 1$) the sample mean can be a very poor estimate when F is not Gaussian. For this location submodel,

Received February 1971; revised September 1972.

[1] This work was supported, in part, by a grant to the author from the John Simon Guggenheim Memorial Foundation and in part by Grant N00014-67-A0114-0004 from the Office of Naval Research.

three different classes of procedures from which good members can be selected have been considered. The first is the class of (M) estimates (Huber's [14] nomenclature). These procedures are the solutions $\hat{\beta}$ of equations of the form

$$\sum_{j=1}^{n} \psi(X_j - \hat{\beta}) = 0 . \tag{1.3}$$

(As is usual in asymptotic theory we shall speak of an "estimate" when we mean a rule which produces a number for each possible sample $x_1, \cdots, x_n$ and $n = 1, 2, \cdots$ and we are interested in the behavior of this rule for n large.) If f is smooth and $\psi = -f'/f$ then the maximum likelihood estimate is an (M) estimate. Obviously, not all maximum likelihood estimates are members of (M) and there are (M) estimates which are not maximum likelihood procedures for any f. However, this class seems adequate for all practical purposes.

The asymptotic theory of these procedures has been considered from two essentially different points of view, f known, and f unknown. If f is known, it follows from the work of LeCam and others that under suitable regularity conditions the appropriate (M) estimate is asymptotically normal, achieves the information inequality lower bound, and is efficient in various senses. A good review of the various notions of efficiency which apply may be found in Hàjek [10]. If f is unknown the optimal choice $\psi = -f'/f$ is not possible and the general class (M) is of interest. Under successively milder regularity conditions on ψ and F, Huber showed in [13] and [14] that such $\hat{\beta}$ were consistent and asymptotically normal with mean β and variance $K(\psi, F)/n$ where

$$K(\psi, F) = \frac{[\int_{-\infty}^{\infty} \psi^2(t)\, dF(t)]}{[\int_{-\infty}^{\infty} \psi'(t)\, dF(t)]^2} . \tag{1.3'}$$

The second class of estimates considered in the location problem is that of linear combinations of order statistics. If $X_{(1)} < \cdots < X_{(n)}$ are the ordered observations, these are estimates of the form

$$\hat{\beta} = \sum_{j=1}^{n} \lambda_j X_{(j)} . \tag{1.4}$$

If the measures Λ_n assigning mass λ_j to $j/(n+1)$, $1 \leqq j \leqq n$, tend suitably to some probability measure Λ such that $\int_0^1 F^{-1}(t)\, d\Lambda(t) = 0$ and various other regularity conditions are satisfied ([4], [8], [28], [29], [23]) then $n^{\frac{1}{2}}(\hat{\beta} - \beta)$ is asymptotically normal with mean 0 and variance

$$V(\Lambda, F) = \int_0^1 \int_0^1 \frac{(\min(s, t) - st)}{f(F^{-1}(s))f(F^{-1}(t))}\, d\Lambda(s)\, d\Lambda(t) . \tag{1.5}$$

If F is known then

$$\Lambda'(t) = \frac{[f'/f]'(F^{-1}(t))}{\int_{-\infty}^{\infty} [f'/f]'(x)\, dF(x)}$$

yields an efficient estimate.

Of particular interest from the point of view of robustness are the α trimmed means corresponding to $\Lambda'(t) = 1/(1-2\alpha)$, $\alpha \leqq t \leqq 1-\alpha$, $= 0$ otherwise and the median corresponding to Λ placing point mass at $\frac{1}{2}$.

A third class of estimates (R) based on rank tests was proposed by Hodges and Lehmann [12]. (The nomenclature (L) and (R) is due to Jaeckel [16].) It has recently been remarked by Jaeckel [16] and van Eeden [39] that this class too can be drawn on for efficient estimates when F is known.

The class (M) has been generalized by Relles [26], and Huber [15] in an obvious way to the model we are dealing with. An (M) estimate $\hat{\boldsymbol{\beta}}$ is now any solution of the system of equations

$$\sum_{j=1}^{n} c_{ij}\psi(X_j - \sum_{k=1}^{p} c_{kj}\hat{\beta}_k) = 0\,, \qquad i = 1, \cdots, p\,. \tag{1.6}$$

Again if $\psi = -f'/f$ these are maximum likelihood estimates and if f is Gaussian with mean 0 we obtain the least squares estimate $\hat{\boldsymbol{\beta}} = \mathbf{X}C'[CC']^{-1}$. Under various regularity conditions the above authors have shown that $\hat{\boldsymbol{\beta}}$ is asymptotically normal with mean $\boldsymbol{\beta}$ and variance $K(\psi, F)[CC']^{-1}$. Thus relative efficiencies of estimates are independent of the design matrix and all robustness results carry over from the location problem.

Various generalizations of class (R) have been studied by Adichie [1], Jurečková [18] and Koul [19] and the same property appeared. A simplification of Koul's approach closely related to our work has been proposed by Kraft and van Eeden in [20] [21] and [22].

The procedures we introduce in Sections 2 and 3 generalize the class (L) as follows. In the location model they coincide with the procedures defined by (1.4). In general they are asymptotically normal with covariance matrices $V(\Lambda, F)[CC']^{-1}$. Our procedures share with the Kraft-van Eeden estimates and a simplification of the (M) estimate discussed in [5] the advantage of simplicity. They also share the disadvantage that they require a preliminary "reasonable" estimate of $\hat{\boldsymbol{\beta}}$; this will be made precise in Sections 2 and 3. Monte Carlo calculations (see [5]) suggest that with a fairly robust starting point the price paid is small. If the least squares estimate is used as a starter several iterations should probably be performed.

The paper is organized as follows. In Section 2 we introduce the estimates and state our theorems and regularity conditions for the case $p = 1$ (regression through the origin) in which all technical difficulties already appear. The general case is discussed more briefly in Section 3. Comments and questions are given at the end of these sections. Finally Section 4 contains the proofs of the theorems stated in Sections 2 and 3.

2. The estimates and their properties for $p = 1$. Write $c_j = c_{1j}$. Let $\{d_j\}$ $1 \leqq j \leqq n$ be nonnegative constants, $\sum_{j=1}^{n} d_j > 0$. Define,

$$Y_j(t) = X_j - c_j t \tag{2.1}$$

and let

$$Q_n(s, t) = \frac{1}{\sum_{j=1}^{n} d_j} \sum_{j=1}^{n} d_j I_{[Y_j(t) \leq s]}\,. \tag{2.2}$$

If β^* is an estimate of β then $Q_n(\cdot, \beta^*)$ is the cdf of the measure which assigns mass $d_k / \sum_{j=1}^n d_j$ to the residual $Y_k(\beta^*)$. Measures of this type arise naturally. For example, if $p = 1$, all the c_j are $\geqq 0$ and $d_j = c_j$ for all j then equation (1.6) can be read $\int_{-\infty}^{\infty} \psi(s)\, dQ_n(s, \hat{\beta}) = 0$. Define for $0 < w < 1$,

$$Q_n^{-1}(w, t) = \inf\{s : Q_n(s, t) \geqq w\}\,. \tag{2.3}$$

In the location problem ($c_j = d_j = 1$), $Q_n^{-1}(jn^{-1}, \beta^*)$ are the ordered $X_1 - \beta^*, \cdots, X_n - \beta^*$ and $\beta^* + Q_n^{-1}(jn^{-1}, \beta^*)$ is just $X_{(j)}$. Our theory is based on the behavior of Q_n^{-1} for suitable choices of d_j. Suppose first that all the c_j are of the same sign, $\sum_{j=1}^n |c_j| > 0$. Let $d_j = |c_j|$ in (2.2). The "right" analogue of $X_{(i)}$ turns out to be (cf. Propositions 4.1—4.3),

$$\beta^* + \frac{\sum_{j=1}^n |c_j|}{\sum_{j=1}^n c_j^2}\, Q_n^{-1}\left(\frac{i}{n}, \beta^*\right). \tag{2.4}$$

This naturally suggests that we define (L) estimates as follows. Let Λ be the distribution function of a finite signed measure on (0, 1) such that $\Lambda(1) = 1$. We say $\hat{\beta}$ is an (L) estimate of type 1 if,

$$\hat{\beta} = \beta^* + \frac{\sum_{j=1}^n |c_j|}{\sum_{j=1}^n c_j^2} \int_0^1 Q_n^{-1}(w, \beta^*)\, \Lambda(dw)\,. \tag{2.5}$$

What do these estimates look like? Let $Y_{(1)} < \cdots < Y_{(n)}$ be the ordered $Y_i(\beta^*)$. Define $(D_1, \cdots, D_n)$ by,

$$Y_{(j)} = Y_{D_j}(\beta^*)\,. \tag{2.6}$$

Let

$$W_k = Q_n(Y_{(k)}, \beta^*) = \frac{\sum_{r=1}^k |c_{D_r}|}{\sum_{r=1}^n |c_r|}\,. \tag{2.7}$$

Then,

$$\hat{\beta} = \beta^* + \frac{\sum_{j=1}^n |c_j|}{\sum_{j=1}^n c_j^2} \{\textstyle\sum_{r=1}^n (X_{D_r} - \beta^* c_{D_r})[\Lambda(W_r) - \Lambda(W_{r-})]\}\,. \tag{2.8}$$

If Λ has a derivative λ it is natural to replace Λ appropriately by the random measure which has density proportional to $\lambda(W_r)$ on the interval $[W_{r-1}, W_r)$ and define an (L) estimate of type 2 by,

$$\hat{\beta} = \frac{\sum_{j=1}^n \lambda(W_j) c_{D_j} X_{D_j}}{\sum_{j=1}^n \lambda(W_j) c_{D_j}^2}\,. \tag{2.9}$$

If all the c_j are not of the same sign we can define (L) estimates as follows. Let $c_j^+ = \max(c_j, 0)$, $c_j^- = c_j^+ - c_j$. Let $Q_n^\pm$ be the Q_n function formed with $d_j = c_j^\pm$, $1 \leqq j \leqq n$. (Whenever we use the $\pm$ notation we are giving two definitions in one—all the pluses go together as do all the minuses.) For Λ as above we shall say $\hat{\beta}$ is an (L) estimate of type 1 if

$$\hat{\beta} = \beta^* + \frac{1}{\sum_{j=1}^n c_j^2} [(\textstyle\sum_{j=1}^n c_j^+) \int_0^1 [Q_n^+]^{-1}(w)\, \Lambda(dw) - (\sum_{j=1}^n c_j^-) \int_0^1 [Q_n^-]^{-1}(w)\, \Lambda(dw)]\,. \tag{2.10}$$

In general, define $W_k^{\pm}$ in terms of $c_j^{\pm}$ as in (2.7). Then,

$$\hat{\beta} = \beta^* + \frac{1}{\sum_{j=1}^n c_j^2} \{\sum_{r=1}^n (X_{D_r} - \beta^* c_{D_r})\{[\Lambda(W_r^+) - \Lambda(W_{r-1}^+)](\sum_{j=1}^n c_j^+) - [\Lambda(W_r^-) - \Lambda(W_{r-1}^-)](\sum_{j=1}^n c_j^-)\}\} . \tag{2.11}$$

If Λ has a continuous derivative λ define an (L) estimate of type 2 by

$$\hat{\beta} = \frac{\sum_{j=1}^n [c_{D_j}^+ \lambda(W_j^+) - c_{D_j}^- \lambda(W_j^-)] X_{D_j}}{\sum_{j=1}^n \{[c_{D_j}^+]^2 \lambda(W_j^+) + [c_{D_j}^-]^2 \lambda(W_j^-)\}} . \tag{2.12}$$

An interesting special case is obtained by taking $\lambda_\alpha(t) = 1/(1 - 2\alpha)$ if $\alpha \leqq t \leqq 1 - \alpha$ and 0 otherwise where $0 \leqq \alpha < 1/2$. This gives "trimmed means," describable as follows. "Order the residuals and associate with $Y_{(r)}$ a "position index" $W_r^{\pm}$ where the choice of $+$ or $-$ depends on the sign of c_{D_r}. (For location $W_r = r/n$.) Trim off all observations corresponding to residuals with $W_r^{\pm} < \alpha$ or $> 1 - \alpha$. Form the usual least squares estimate with the remaining observations." If we make the replacement

$$\int_{W_{r-1}^{\pm}}^{W_r^{\pm}} \lambda(s)\, ds$$

for $c_{D_r} \lambda(W_r^{\pm})$ we obtain an interesting variation in which observations corresponding to residuals with $W_{r-1}^{\pm} < \alpha \leqq W_r^{\pm}$ are weighted by $\sum_{j=1}^r c_{D_j}^{\pm} - \alpha \sum_{j=1}^n c_j^{\pm}$. This modification is significant for small sample sizes even in the location problem. The type 1 estimates can be thought of as corrections to the initial estimate β^* by a trimmed linear combination of residuals with weights appropriate to the least squares estimate. An interesting special case of type 1 is the "median". Determine $m^{\pm}$ by $W_{m^{\pm}-1}^{\pm} < \frac{1}{2} < W_{m^{\pm}}^{\pm}$. The median M is given by,

$$M = \beta^* + \frac{1}{\sum_{j=1}^n c_j^2} [[\sum_{j=1}^n c_j^+](X_{D_{m^+}} - c_{D_{m^+}} \beta^*) - [\sum_{j=1}^n c_j^-](X_{D_{m^-}} - c_{D_{m^-}} \beta^*)] . \tag{2.13}$$

Another solution (for general c_j) and F symmetric about 0 was proposed by J. W. Tukey (personal communication). Let,

$$Y_j^*(t) = Y_j(t) \operatorname{sgn} c_j , \tag{2.14}$$

$$Q_n^*(s, t) = \frac{1}{\sum_{j=1}^n |c_j|} \sum_{j=1}^n |c_j| I_{[Y_j^*(t) \leqq s]} . \tag{2.15}$$

Define $(A_1, \cdots, A_n)$ by

$$Y_{A_j}^* = Y_{(j)}^* \tag{2.16}$$

where $Y_{(1)}^* < \cdots < Y_{(n)}^*$ are the ordered Y_j^*.

$$W_j^* = \frac{\sum_{r=1}^j |c_{A_r}|}{\sum_{r=1}^n |c_r|} . \tag{2.17}$$

Define $\hat{\beta}$ of type 1′ and type 2′ by replacing Q_n^{-1} by $[Q_n^*]^{-1}$ in the definition of type 1 (with $c_j \geqq 0$) and replacing $(D_1, \cdots, D_n)$ by $(A_1, \cdots, A_n)$ and $(W_1, \cdots, W_n)$ by $(W_1^*, \cdots, W_n^*)$ in the definition of type 2.

We turn now to the statement of the asymptotic properties of these procedures. We shall need the following conditions.

Define,

$$b(n) = (\textstyle\sum_{j=1}^n c_j^2)^{\frac{1}{2}} . \tag{2.18}$$

G Suppose

$$\max\{c_j^\pm : 1 \leqq j \leqq n\} = o(\textstyle\sum_{j=1}^n [c_j^\pm]^2) . \tag{2.19}$$

$$b(n) \to \infty . \qquad \text{(By convention } 0/0 = 0.) \tag{2.20}$$

F_1 The density f is uniformly continuous, positive, and bounded.

F_2 $\int_{-\infty}^{\infty} |t| \, dF(t) < \infty$.

We are given a preliminary estimate β^* (usually based on the same data) satisfying,

B (i) β^* is invariant, i.e.,

$$\beta^*(x_1 + c_1 t, \cdots, x_n + c_n t) = \beta^*(x_1, \cdots, x_n) + t . \tag{2.21}$$

(ii) β^* is $b(n)$ consistent, i.e.,

$$\beta^* = O_{P_0}(b^{-1}(n)) . \tag{2.22}$$

(The subscript β here and in the future indicates calculation is carried out when β is true.) Note that the usual least squares estimate satisfies B if Var $(X_1) < \infty$ and G holds.

THEOREM 2.1. *Suppose that B, G, and F_1 hold and $0 < \alpha < \frac{1}{2}$. If either,*

I. *$\hat{\beta}$ is a type 1 estimate corresponding to Λ concentrating on $[\alpha, 1 - \alpha]$ such that $\int_0^1 F^{-1}(w) \, \Lambda(dw) = 0$, or*

II. *$\hat{\beta}$ is a type 2 estimate corresponding to λ satisfying a first order Lipschitz condition on $[\alpha, 1 - \alpha]$ such that the measure Λ with density λ satisfies the conditions of* I, *then,*

$$\mathscr{L}_\beta(b(n)(\hat{\beta} - \beta)) \to \eta(0, V(\Lambda, F)) . \tag{2.23}$$

($\eta(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2.)

THEOREM 2.2. *If the conditions of Theorem 2.1 hold appropriately and F is in addition symmetric about 0 then (2.23) holds for type 1′ and 2′ estimates as well.*

REMARKS. (1) If F is symmetric about 0 and Λ is symmetric about $\frac{1}{2}$ then $\int_0^1 F^{-1}(w) \, \Lambda(dw) = 0$.

(2) The analogues of the median, trimmed and Winsorized means, systematic statistics, etc. all come under the provisions of these theorems. For example M, given by (2.13), is asymptotically normal with mean β and variance $1/4f^2(0)b^2(n)$ if F is symmetric about 0.

(3) Theorems 2.1—2.2 continue to hold if we permit Λ or λ defining the

estimates to depend on n or even on the data. For instance, suppose that $\hat{\beta}$ is a type 1 or 1′ estimate defined by Λ_n where the Λ_n concentrate on $[\alpha, 1 - \alpha]$ and there is a meaure Λ on $[\alpha, 1 - \alpha]$ such that

(i) $\int_\alpha^t \Lambda_n(s)\, ds \to \int_\alpha^t \Lambda(s)\, ds$ in probability for each t.

(ii) The total variations of the Λ_n are bounded in probability.

(iii) $$\int_0^1 F^{-1}(w)\Lambda_n(dw) = 0\,.$$

Then, (2.23) holds provided the appropriate conditions of Theorems 2.1—2.2 hold. This remark becomes important when we want to use the trimmed mean in an adaptive way. If one follows the prescription of Tukey-McLaughlin [30] and Jaeckel [16], α is to be chosen optimally on the basis of the sample.

The theorems we have just stated deal adequately with trimmed and Winsorized means, percentiles, etc. However, there are estimates which may be reasonable placing some slight mass on all the order statistics. For example, the efficient estimate corresponding to $F(s) = 1/(1 + e^{-s})$, the logistic distribution, has,

(2.24) $$\lambda(w) = 6w(1 - w)\,, \qquad 0 < w < 1\,.$$

The conditions for our theorems dealing with such situations are awkward and much too stringent. The methods and conditions are variations and combinations of those appearing in [18], [19] and [25].

For simplicity we restrict to estimates of all types generated by a fixed Λ, with derivative λ. We impose the condition,

P The derivative λ has a bounded derivative λ' on $[0, 1]$.

THEOREM 2.3. *Suppose that conditions* F_1, F_2, B *and* G *hold. Suppose that* $\hat{\beta}$ *corresponds to* Λ *satisfying* (P). *If, in addition,* $\int_0^1 F^{-1}(t)\lambda(t)\, dt = 0$, $V(\Lambda, F) < \infty$, *and either,*

(i) $\hat{\beta}$ *is of type* 1 *or* 2, *or*

(ii) F *is symmetric about* 0, *and* $\hat{\beta}$ *is of type* 1′ *or* 2′, *then* (2.23) *holds.*

Comments on the section.

(1) We have stated the theorems to reflect our primary interest in the case F unknown but symmetric about 0. In general, the asymptotic bias of a type 1 or 2 $\hat{\beta}$ is,

$$\frac{(\sum_{j=1}^n c_j)}{b^2(n)} \int_0^1 F^{-1}(w)\Lambda(dw)$$

and $\hat{\beta}$ suitably centered has the limiting normal distribution of (2.23). Note that the bias vanishes if $\sum_{j=1}^n c_j = 0$. This corresponds to the fact that we can efficiently estimate the slope of a regression line even if the errors are asymmetric if the design is symmetric. The behavior of type 1′ and 2′ estimates is more complicated. Rather than giving general formulas we note that if $\sum_{j=1}^n c_j = 0$, $\sum_{j=1}^n [c_j^+]^2 = \sum_{j=1}^n [c_j^-]^2$ and Λ is symmetric about $\frac{1}{2}$ then type 1′ and 2′ estimates are also asymptotically normal with mean β and variance $\tilde{V}/b^2(n)$ where

$$\tilde{V}(\Lambda, F) = \int_0^1 \int_0^1 K(s, t, F)\, d\Lambda(s)\, d\Lambda(t) \tag{2.25}$$

and

$$K(s, t, F) = [\min(s, t) - \tfrac{1}{2}(F(\tilde{F}^{-1}(s))F(\tilde{F}^{-1}(t)) + \bar{F}(\tilde{F}^{-1}(s))\bar{F}(\tilde{F}^{-1}(t))]/\tilde{q}(s)\tilde{q}(t)$$

and

$$\bar{F}(x) = 1 - F(-x)\,, \qquad \tilde{F} = \tfrac{1}{2}(F + \bar{F})\,, \qquad \tilde{q} = \tilde{f}(\tilde{F}^{-1})$$

and $\tilde{f} = \tilde{F}'$. There seems no clear cut choice between type 1 and type 1′ estimates. For example, the ratio of the asymptotic variance of the type 1 to the type 1′ median is $f^2(0)/f^2(F^{-1}(\frac{1}{2}))$ which ranges from 0 to ∞ as F ranges over asymmetric distributions.

(2) It seems natural to ask what happens if we iterate the procedure of forming an (L) estimate. Is there a fixed point under α trimming, for example, and does it behave properly for n large? For $p = 1$ the answer can be shown to be yes at least for type 1 and 1′ estimates if Λ is a probability measure. This follows from the fact that $[Q_n^{\pm}]^{-1}(w, t)$ are continuous and monotone in t. Two special cases are of interest. If $\lambda(t) = 1$ the least squares estimate is the fixed point and is reached after one iteration. The fixed point type 1′ median is the left-hand endpoint of the interval of medians of the distribution which assigns mass proportional to $|c_j|$ to X_j/c_j, $1 \leqq j \leqq n$. (This procedure, the (M) estimate for the double exponential, was suggested to me by J. W. Tukey.)

(3) These techniques can be applied to nonlinear functions of Q_n^{-1} processes. For example, the natural estimate of the unknown factor in the variance of the "α trimmed mean" is a "Winsorized variance." In our case this is

$$s^2(\alpha) = \frac{1}{(1 - 2\alpha)^2}\{\int_\alpha^{1-\alpha} [Q_n^{-1}(w, \beta^*)]^2\, dw + \alpha\{[Q_n^{-1}(\alpha, \beta^*)]^2 + [Q_n^{-1}(1 - \alpha, \beta^*)]^2\}\}$$

where Q_n is the empirical cdf of the residuals. It follows readily by our methods that $s^2(\alpha) \to V(\Lambda_\alpha, F)$ uniformly for $0 < \alpha_0 \leqq \alpha \leqq \alpha_1 < 1$ where Λ_α is the uniform distribution on $[\alpha, 1 - \alpha]$. Thus, the results of [16] carry over to the linear model.

3. The general case.

In analogy to the case $p = 1$ define

$$Y_j(\mathbf{t}) = X_j - \sum_{i=1}^p c_{ij} t_i \tag{3.1}$$

for $\mathbf{t} = (t_1, \cdots, t_p)$ and, for d_j as in Section 2,

$$Q_n(s, \mathbf{t}) = \frac{1}{\sum_{j=1}^n d_j} \sum_{j=1}^n d_j I_{[Y_j(\mathbf{t}) \leqq s]}\,. \tag{3.2}$$

If $d_j = c_{ij} \geqq 0$, $1 \leqq j \leqq n$, we shall denote Q_n by Q_{ni}. If we define $c_{ij}^{\pm}$ naturally and $d_j = c_{ij}^{\pm}$, $1 \leqq j \leqq n$, we shall denote Q_n by $Q_{ni}^{\pm}$. Finally we define,

$$Y_{ij}^*(\mathbf{t}) = Y_j(\mathbf{t}) \operatorname{sgn} c_{ij}\,, \tag{3.3}$$

and

$$Q_{ni}^*(s, \mathbf{t}) = \frac{1}{\sum_{j=1}^n |c_{ij}|} \sum_{j=1}^n |c_{ij}| I_{[Y_{ij}^*(\mathbf{t}) \leq s]} . \tag{3.4}$$

Inverses are defined and denoted as before. For general p we state B and G as follows.

G (3.5)
$$\frac{CC'}{b(n)} \to A$$

where A is positive definite and $b(n) \to \infty$.

$$\max \{c_{ij}^{\pm} : 1 \leqq j \leqq n\} = o((\sum_{j=1}^n [c_{ij}^{\pm}]^2)^{\frac{1}{2}}) . \tag{3.6}$$

A fortiori, $\max \{|c_{ij}| : 1 \leqq i \leqq p,\ 1 \leqq j \leqq n\} = o(b(n))$.

We are given a preliminary estimate $\boldsymbol{\beta}^*$ such that,

B (3.7)
$$\boldsymbol{\beta}^*(\mathbf{x} + \mathbf{t}C) = \boldsymbol{\beta}^*(\mathbf{x}) + \mathbf{t}$$

for any $\mathbf{x}$, $\mathbf{t}$.

$$\boldsymbol{\beta}^* = O_{P_0}(b^{-1}(n)) .$$

Again the least squares estimate satisfies B if G holds and $E(X_1^2) < \infty$. We now define the estimates. Given Λ as in Section 2 $\hat{\boldsymbol{\beta}}$ is of type 1 if $\hat{\mathbf{r}} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$ satisfies

$$\begin{aligned}[\textstyle\sum_{j=1}^n c_{ij}^+] \textstyle\int_0^1 [Q_{ni}^+]^{-1}(w, \boldsymbol{\beta}^*)\Lambda(dw) + [\textstyle\sum_{j=1}^n c_{ij}^-] \\ \times \textstyle\int_0^1 [Q_{ni}^-]^{-1}(w, \boldsymbol{\beta}^*)\Lambda(dw) = \textstyle\sum_{k=1}^p \hat{\mathbf{r}}_k (\textstyle\sum_{j=1}^n c_{ij} c_{kj}) , \qquad 1 \leqq i \leqq p .\end{aligned} \tag{3.8}$$

Equivalently, if

$$\begin{aligned}\mathbf{L}^{\pm} = ([\textstyle\sum_{j=1}^n c_{ij}^{\pm}] \textstyle\int_0^1 [Q_{n1}^{\pm}]^{-1}(w, \beta^*)\Lambda(dw), \cdots, [\textstyle\sum_{j=1}^n c_{pn}^{\pm}] \\ \times \textstyle\int_0^1 [Q_{np}^{\pm}]^{-1}(w, \beta^*)\Lambda(dw))\end{aligned} \tag{3.9}$$

then,

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^* + (\mathbf{L}^+ - \mathbf{L}^-)[CC']^{-1} . \tag{3.10}$$

Type 1′ is defined more simply by replacing $\mathbf{L}^+ - \mathbf{L}^-$ in (3.10) by the vector whose ith component is $[\sum_{j=1}^n |c_{ij}|] \int_0^1 [Q_{ni}^*]^{-1}(w)\Lambda(dw)$.

To define type 2 we introduce,

$$W_{ij}^{\pm} = \frac{\sum_{r=1}^j c_{iD_r}^{\pm}}{\sum_{r=1}^n c_{ir}^{\pm}} \tag{3.11}$$

where $(D_1, \cdots, D_n)$ are defined as in Section 2 with the $Y_j(\boldsymbol{\beta}^*)$ playing the role of $Y_j(\beta^*)$. Then $\hat{\boldsymbol{\beta}}$ is the type 2 estimate for λ as in Section 2 if $\hat{\boldsymbol{\beta}}$ satisfies

$$\begin{aligned}\textstyle\sum_{j=1}^n [c_{iD_j}^+ \lambda(W_{ij}^+) - c_{iD_j}^- \lambda(W_{ij}^-)] X_{D_j} \\ = \textstyle\sum_{k=1}^p \hat{\beta}_k \textstyle\sum_{j=1}^n c_{kD_j} [c_{iD_j}^+ \lambda(W_{ij}^+) - c_{iD_j}^- \lambda(W_{ij}^-)] .\end{aligned} \tag{3.12}$$

To define type 2′ we need to introduce $(D_{i1}^*, \cdots, D_{ip}^*)$ given by,

$$Y_{iD_{ij}^*}^*(\boldsymbol{\beta}^*) = Y_{i(j)}^* \tag{3.13}$$

where $Y^*_{i(1)} < \cdots < Y^*_{i(n)}$ are the order statistics of $Y^*_{ij}(\boldsymbol{\beta}^*)$, $1 \leqq j \leqq n$. Then, if

$$W^*_{ij} = \frac{1}{\sum_{j=1}^n |c_{ij}|} \sum_{r=1}^j |c_{iD^*_{ir}}| , \tag{3.14}$$

the estimate $\hat{\boldsymbol{\beta}}$ of type 2′ corresponding to λ is defined to be the solution of,

$$\sum_{j=1}^n X_{D^*_{ij}} c_{iD^*_{ij}} \lambda(W^*_{ij}) = \sum_{k=1}^p \hat{\beta}_k \left(\sum_{j=1}^n c_{iD^*_{ij}} c_{kD^*_{ij}} \lambda(W^*_{ij})\right) . \tag{3.15}$$

(The solution is well defined for n sufficiently large.)

We shall prove,

THEOREM 3.1. *The assertion,*

$$\mathscr{L}_{\boldsymbol{\beta}}(b(n)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) \to \eta(\mathbf{0}, V(\Lambda, F)A^{-1}) \tag{3.16}$$

holds for appropriately defined $\hat{\boldsymbol{\beta}}$ *of types* 1 *and* 2 *under the conditions of Theorems* 2.1, 2.2 *and* 2.3. *If we add the requirement that* Λ *is symmetric about* $\frac{1}{2}$ *it holds for types* 1′ *and* 2′ *estimates under the conditions of Theorems* 2.2. *and* 2.3. *In all cases G refers to the condition of Section* 3.

REMARKS. (1) The type 1′ and 2′ estimates require p orderings of the $Y^*_{ij}(\boldsymbol{\beta}^*)$ rather than the single ordering of types 1 and 2.

(2) It is possible to drop the requirement of a rate $b(n)$ but statements become unduly complicated.

(3) It is easy to see that the bias of type 1 and 2 estimates is,

$$\left(\int_0^1 F^{-1}(w)\Lambda(dw)\right)\boldsymbol{\Delta}_n[CC']^{-1} \qquad \text{where}$$
$$\boldsymbol{\Delta}_n = \left(\sum_{j=1}^n c_{1j}, \cdots, \sum_{j=1}^n c_{pj}\right) .$$

Again if $\sum_{j=1}^n c_{ij} = 0$ for all i then (3.16) holds even if $\int_0^1 F^{-1}(w)\Lambda(dw) \neq 0$. More significantly if, as usual, $c_{ij} = 1$, $1 \leqq j \leqq n$, $\sum_{j=1}^n c_{ij} = 0$ for $i > 1$, then the estimates $(\hat{\beta}_2, \cdots, \hat{\beta}_p)$ are asymptotically unbiased and have the right covariance structure. The problem of bias for types 1′ and 2′ is complicated and not of great interest.

(4) Whether fixed points in the sense of comment (2) of the previous section exist in general is not known to me. However, the (M) estimate for the double exponential is the "median" fixed point.

(5) A disadvantage of these procedures is that they are not invariant under a reparametrization of the vector space V spanned by the rows of C. If the rows of a matrix M form another basis of V we may define new parameters $(\gamma_1, \cdots, \gamma_p) = \boldsymbol{\gamma}$ related to $\boldsymbol{\beta}$ by $\boldsymbol{\gamma}M = \boldsymbol{\beta}C$. If we now apply our theory to the model $\mathbf{X} = \boldsymbol{\gamma}M + \mathbf{E}$ it would appear that the estimates $\hat{\boldsymbol{\gamma}}$ do not in general satisfy,

$$\hat{\boldsymbol{\gamma}}M = \hat{\boldsymbol{\beta}}C . \tag{3.17}$$

I do not know whether (L) estimates having this property (in general) can be defined. (M) estimates and Koul's (R) estimates *do* have this property.

4. Proofs. We begin with the theorems of Section 2.

Let the constants d_j defining the Q_n of Sections 2 and 3 depend on n in such a way that they satisfy the following assumption.

D Let $r(n) = (\sum_{j=1}^n d_j^2)^{\frac{1}{2}}$. We require that

$$\max\{d_j : 1 \leqq j \leqq n\} = o(r(n)) \tag{4.1}$$

$$r(n) = O(b(n)) . \tag{4.2}$$

For convenience, define $\sigma_n = \sum_{j=1}^n d_j$, $\tau_n = \sum_{j=1}^n c_j d_j$, $\gamma_n = \sum_{j=1}^n |c_j d_j|$.

Theorems 2.1 and 2.2 are based on the following propositions which are of independent interest. For simplicity we state them for $p = 1$. Because of invariance considerations we can and shall make all calculations under the assumption $\boldsymbol{\beta} = \mathbf{0}$.

PROPOSITION 4.1. *Suppose that G, D, and F_1 hold. Then for every $M < \infty$, $0 < \alpha < \frac{1}{2}$,*

$$\sup\left\{\sigma_n\left[(Q_n^{-1}(w, t) - F^{-1}(w)) + \frac{(Q_n(F^{-1}(w), 0) - w)}{q(w)}\right] + t\tau_n| : |t| \leqq M/b(n), \alpha \leqq w \leqq 1 - \alpha\right\} = o_p(r(n)) , \tag{4.3}$$

where $q(w) = f(F^{-1}(w))$.

PROPOSITION 4.2. *If β^* satisfies B and G, D, and F_1 hold then,*

$$\sup\left\{\left|\beta^*\tau_n + \sigma_n\left[(Q_n^{-1}(w, \beta^*) - F^{-1}(w)) - \frac{(w - Q_n(F^{-1}(w), 0))}{q(w)}\right]\right| : \alpha \leqq w \leqq 1 - \alpha\right\} = o_p(r(n)) . \tag{4.4}$$

The next proposition is essentially contained in [19].

PROPOSITION 4.3. *If D and F_1 hold, $0 < \alpha < \frac{1}{2}$, and,*

$$Z_n(w) = \frac{\sigma_n}{r(n)} \frac{(w - Q_n(F^{-1}(w), 0))}{q(w)} , \tag{4.5}$$

then the processes $\{Z_n(\cdot)\}$ converge weakly in the Skorokhod topology on $D[\alpha, 1 - \alpha]$ to a Gaussian process Z with mean 0 and covariance structure,

$$\operatorname{Cov}(Z(v), Z(w)) = \frac{v(1 - w)}{q(v)q(w)} \qquad \text{if } v \leqq w . \tag{4.6}$$

This is the Brownian bridge divided by q (cf. [4] for instance).

PROOF OF PROPOSITION 4.3. This is a straightforward weak convergence result and holds under the sole assumption that F is continuous. It may be argued as in [19] page 1968 or [6] page 106.

PROOF OF PROPOSITION 4.1. We begin by stating a lemma proved in [9]. (The conditions stated in [19] are not quite correct but it is evident that G, D *and* F_1 suffice.) A similar argument is given in some detail for Theorem 2.3.

LEMMA 4.1. *If G, D and F_1 hold then,*

$$\sigma_n \sup\{|Q_n(s,t) - F_n(s,t) - Q_n(s,0) + F(s)| : |s| < \infty, \; |t| \leqq M/r(n)\} = o_p(r(n)), \tag{4.7}$$

where,

$$F_n(s,t) = E_0(Q_n(s,t)) = \sigma_n^{-1} \textstyle\sum_{j=1}^n d_j F(s + c_j t). \tag{4.8}$$

To complete the proof of the proposition begin by noting that by D,

$$\frac{\sigma_n}{r(n)} \geqq r(n)/\max\{d_j : 1 \leqq j \leqq n\} \to \infty. \tag{4.9}$$

Therefore by Lemma 4.1 and Proposition 4.3,

$$\sup\{|Q_n(s,t) - F_n(s,t)| : |s| < \infty, |t| \leqq M/b(n)\} = o_p(1). \tag{4.10}$$

By F_1, D, and (4.9),

$$\sup\{|F_n(s,t) - F(s)| : |s| < \infty, |t| \leqq M/b(n)\} = O\left(\frac{\gamma_n}{\sigma_n b(n)}\right) = O\left(\frac{r(n)}{\sigma_n}\right) = o(1). \tag{4.11}$$

Since

$$|Q_n(Q_n^{-1}(w,t),t) - w| \leqq \frac{\max d_j}{\sigma_n} = o\left(\frac{r(n)}{\sigma_n}\right) \tag{4.12}$$

$$\sup\{|Q_n^{-1}(w,t) - F^{-1}(w)| : \alpha \leqq w \leqq 1-\alpha, |t| \leqq M/b(n)\} \to_P 0 \tag{4.13}$$

for $0 < \alpha < \frac{1}{2}$.

Now, arguing as in [25],

$$\begin{aligned}(Q_n^{-1}(w,t) - F^{-1}(w)) &= (Q_n^{-1}(w,t) - F_n^{-1}(w,t)) + (F_n^{-1}(w,t) - F^{-1}(w)) \\ &= \frac{-1}{R_n(w,t)}\{(Q_n(Q_n^{-1}(w,t),t) - F_n(Q_n^{-1}(w,t),t)) \\ &\quad + (w - Q_n(Q_n^{-1}(w,t)))\} + (F_n^{-1}(w,t) - F^{-1}(w)),\end{aligned} \tag{4.14}$$

where $F_n^{-1}(\cdot, t)$ is the inverse of $F_n(\cdot, t)$ and

$$R_n(w,t) = \frac{F_n(Q_n^{-1}(w,t),t) - F_n(F_n^{-1}(w,t),t)}{Q_n^{-1}(w,t) - F_n^{-1}(w,t)}. \tag{4.15}$$

Note that,

$$\frac{\partial F_n(x,t)}{\partial x} \to f(x) \tag{4.16}$$

uniformly in $|t| \leqq M/b(n)$ and x bounded. In view of (4.13),

$$\sup\{|R_n(w,t) - f(F^{-1}(w))| : |t| \leqq M/b(n), \; \alpha \leqq w \leqq 1-\alpha\} \to_P 0. \tag{4.17}$$

Consider $F_n^{-1}(w, t)$. Since,

$$\frac{\partial F_n^{-1}(w, t)}{\partial t} = \frac{-\partial F_n(x, t)/\partial t}{\partial F_n(x, t)/\partial x}\bigg|_{x=F_n^{-1}(w,t)}, \tag{4.18}$$

it is easy to show that,

$$\sigma_n \frac{\partial F_n^{-1}(w, t)}{\partial t} = -\tau_n + o(\gamma_n) \tag{4.19}$$

uniformly in $\alpha \leqq w \leqq 1 - \alpha$, $|t| \leqq M/b(n)$. Therefore,

$$\sigma_n(F_n^{-1}(w, t) - F^{-1}(w) + t\tau_n) = o(\gamma_n/b(n)) = o(r(n)) \tag{4.20}$$

uniformly in $\alpha \leqq w \leqq 1 - \alpha$, $|t| \leqq M/b(n)$.

By (4.13), Lemma 4.1 and Proposition 4.3

$$\begin{aligned} &\sup\{\sigma_n[Q_n(Q_n^{-1}(w, t), t) - F_n(Q_n^{-1}(w, t), t) \\ &\quad - Q_n(F^{-1}(w), 0) + w] : \alpha \leqq w \leqq 1 - \alpha, |t| \leqq M/b(n)\} \\ &\quad = o_p(r(n)) . \end{aligned} \tag{4.21}$$

By applying (4.12), (4.16), (4.20) and (4.21) to (4.14) the proposition follows.

PROOF OF PROPOSITION 4.2. This follows immediately from Proposition 4.1 in view of property B.

PROOF OF THEOREM 2.1. If, using G, we apply Proposition 4.2 to Q_n^+ and $Q_n^-(d_j = c_j^\pm)$ we get,

$$\begin{aligned} &\beta^*(\textstyle\sum_{j=1}^n [c_j^\pm]) \pm (\sum_{j=1}^n c_j^\pm) \int_0^1 [Q_n^\pm]^{-1}(w, \beta^*)\Lambda(dw) \\ &\quad = (\textstyle\sum_{j=1}^n c_j^\pm)\left[\int_0^1 F^{-1}(w)\Lambda(dw) + \int_\alpha^{1-\alpha} \frac{(w - Q_n^\pm(w), 0))}{q(w)} \Lambda(dw)\right] \\ &\quad + o_p(b(n)) . \end{aligned} \tag{4.22}$$

Adding the two equations of (4.22) and dividing by $b^2(n)$ we get,

$$\hat\beta = \left[\frac{\sum_{j=1}^n c_j}{\sum_{j=1}^n c_j^2}\right] \int_\alpha^{1-\alpha} \frac{(w - Q_n(F^{-1}(w), 0))}{q(w)} \Lambda(dw) + o_p(b^{-1}(n)) . \tag{4.23}$$

The first part of the theorem now follows from Proposition 4.2. To prove the second part begin by writing,

$$\begin{aligned} \hat\beta = \beta^* &+ \{(\textstyle\sum_{j=1}^n c_j^+) \int_0^1 [Q_n^+]^{-1}(w, \beta^*)\, d\Lambda_n^+(w) \\ &- (\textstyle\sum_{j=1}^n c_j^-) \int_0^1 [Q_n^-]^{-1}(w)\, d\Lambda_n^{-1}(w)\} \\ &\times \{\textstyle\sum_{j=1}^n [[c_{Dj}^+]^2 \lambda(W_j^+) - [c_{Dj}^-]^2\lambda(W_j^-)]\}^{-1} \end{aligned} \tag{4.24}$$

where $\Lambda_n^\pm$ is the random measure with density $\lambda_n^\pm$ given by,

$$\lambda_n^\pm(w) = \lambda(W_j^\pm) \quad \text{on} \quad (W_{j-1}^\pm, W_j^\pm], \qquad 1 \leqq j \leqq n . \tag{4.25}$$

Now,

$$\sup_{\alpha \leqq t \leqq (1-\alpha)} |\lambda(t) - \lambda_n^\pm(t)| \leqq M'' \max \frac{c_j^\pm}{\sum_{j=1}^n c_j^\pm} \tag{4.26}$$

where $(\overline{1-\alpha})$ is the largest $W_j^{\pm} \leqq 1-\alpha$, and M'' is the Lipschitz constant for λ. Therefore in light of Propositions 4.2 and 4.3 it is easy to see that,

$$(4.27)\qquad \frac{\sum_{j=1}^n c_j^{\pm}}{\sum_{j=1}^n c_j^2} \left|\int_0^1 ([Q_n^{\pm}]^{-1}(w, \beta^*) - F^{-1}(w))\, d(\Lambda_n - \Lambda)\right| = o_p(1/b(n))$$

and

$$(4.28)\qquad \frac{\sum_{j=1}^n c_j^{\pm}}{\sum_{j=1}^n c_j^2} \left|\int_0^1 F^{-1}(w)\, d\Lambda_n(w)\right| = o_p(1/b(n)) .$$

(The Lipschitz condition (rather than just continuity) is needed for (4.28) only.)

Finally, note that,

$$(4.29)\qquad W_j^{\pm} = Q_n^{\pm}(Y_{(j)}, \beta^*) .$$

Therefore,

$$(4.30)\qquad \sup_j |\lambda(W_j^{\pm}) - \lambda(F(Y_{(j)}))| = o_p(1) .$$

Hence,

$$(4.31)\qquad \begin{aligned} \sum_{j=1}^n \{[c_{D_j}^+]^2\lambda(W_j^+) + [c_{D_j}^-]^2\lambda(W_j^-)\} &= \sum_{j=1}^n c_{D_j}^2 \lambda(F(X_{D_j})) + o_p(b^2(n)) \\ &= \int_0^1 \lambda(t)\, dt + o_p(b^2(n)) . \end{aligned}$$

Theorem 2.1 follows since (4.27), (4.28), and (4.31) show that the type 1 and type 2 estimates are equivalent.

PROOF OF THEOREM 2.2. This result follows if we apply our preceding results to the new model,

$$(4.32)\qquad X_j^* = (\operatorname{sgn} c_j)X_j = |c_j|\beta + E_j \operatorname{sgn} c_j .$$

Of course if F is symmetric about 0 the $E_j \operatorname{sgn} c_j$, $1 \leqq j \leqq n$, have the same distribution as the E_j, $1 \leqq j \leqq n$ and the type 1′, 2′ estimates are seen to be type 1 and type 2 estimates based on the X_j^*.

Remark (3) following Theorem 2.2 can easily be established with the use of a stochastic version of the theorem on page 120 in [27].

PROOF OF THEOREM 2.3. We begin with an elementary lemma which we state without proof.

LEMMA 4.2. *Let G and H be the distribution functions of probability measures, with G concentrating on* [0, 1]. *If Z has distribution function $G(\cdot)$ then $H^{-1}(Z)$ has distribution function $G(H(\cdot))$.*

Write,

$$(4.33)\qquad \begin{aligned} T_n(t) &= \int_0^1 [Q_n^{-1}(w, t) - F_n^{-1}(w, t)]\lambda(w)\, dw \\ &= \int_0^1 [Q_n^{-1}(w, t) - F_n^{-1}(w, t)]\Lambda(dw) \\ &= \int_{-\infty}^{\infty} z\{\Lambda(Q_n(dz, t)) - \Lambda(F_n(dz, t))\} \end{aligned}$$

by Lemma 4.2. Integrating by parts we obtain,

$$(4.34)\qquad T_n(t) = \int_{-\infty}^{\infty} [\Lambda(F_n(z, t)) - \Lambda(Q_n(z, t)]\, dz .$$

LEMMA 4.3. *If Λ satisfies (P) and conditions G, D, F_1 and F_2 hold then,*

$$(T_n(t/b(n)) - T_n(0)) = o_p(r(n)) \tag{4.35}$$

for each fixed t.

PROOF. By (4.34) and (P),

$$\begin{aligned} T_n\left(\frac{t}{b(n)}\right) = -\int_{-\infty}^{\infty} \lambda\left(F_n\left(z, \frac{t}{b(n)}\right)\right)\left(Q_n\left(z, \frac{t}{b(n)}\right)\right. \\ \left. - F_n\left(z, \frac{t}{b(n)}\right)\right) dz + R_n \end{aligned} \tag{4.36}$$

where

$$|R_n| \leqq M'' \int_{-\infty}^{\infty} \left(Q_n\left(z, \frac{t}{b(n)}\right) - F_n\left(z, \frac{t}{b(n)}\right)\right)^2 dz \tag{4.37}$$

and M'' is the upper bound on $|\lambda'|$. Now,

$$\begin{aligned} &E_0\left(\int_{-\infty}^{\infty}\left(Q_n\left(z, \frac{t}{b(n)}\right) - F_n\left(z, \frac{t}{b(n)}\right)\right)^2 dz\right) \\ &= \int_{-\infty}^{\infty} E_0\left(Q_n\left(z, \frac{t}{b(n)}\right) - F_n\left(z, \frac{t}{b(n)}\right)\right)^2 dz \\ &= \frac{\sum_{j=1}^n d_j^2}{(\sum_{j=1}^n d_j)^2} \int_{-\infty}^{\infty} F\left(z + \frac{t}{b(n)} c_j\right)\left(1 - F\left(z + \frac{t}{b(n)} c_j\right)\right) dz \\ &\leqq \left[E_0(|X_1|) + \frac{t \max |c_j|}{b(n)}\right] \frac{r^2(n)}{\sigma_n^2}. \end{aligned} \tag{4.38}$$

Therefore, in view of G, D, F_2 and (4.9),

$$\frac{\sigma_n}{b(n)} E_0(|R_n|) \to 0. \tag{4.39}$$

Now, let $\Lambda^{(n)}(z) = \int_z^\infty \lambda(F_n(s, t/b(n)) ds$, $\Lambda^{(0)}(z) = \int_z^\infty \lambda(F(s)) ds$. (The finiteness of $\Lambda^{(n)}$, $n \geqq 0$, follows from (P) and F_2.)

$$\begin{aligned} &\int_{-\infty}^{\infty} \lambda\left(F_n\left(s, \frac{t}{b(n)}\right)\right)\left(Q_n\left(s, \frac{t}{b(n)}\right) - F_n\left(s, \frac{t}{b(n)}\right)\right) ds \\ &= \frac{1}{\sigma_n} \sum_{j=1}^n d_j\left[\Lambda_n\left(X_j - \frac{tc_j}{b(n)}\right) - E_0\left(\Lambda_n\left(X_j - \frac{tc_j}{b(n)}\right)\right)\right]. \end{aligned} \tag{4.40}$$

In view of (4.39) a simple L_2 calculation shows that the lemma follows provided that,

$$\max_j E_0\left(\Lambda^{(n)}\left(X_j - \frac{tc_j}{b(n)}\right) - \Lambda^{(0)}(X_j)\right)^2 \to 0. \tag{4.41}$$

Now

$$\begin{aligned} \Lambda^{(n)}(z) &= \int_{F_n(z, t/b(n))}^1 \lambda(w)\, dF_n^{-1}\left(w, \frac{t}{b(n)}\right) \\ &= z\lambda\left(F_n\left(z, \frac{t}{b(n)}\right)\right) - \int_{F_n(z, t/b(n))}^1 F_n^{-1}\left(w, \frac{t}{b(n)}\right) \lambda'(w)\, dw. \end{aligned} \tag{4.42}$$

It is easy to see that,

$$(4.43)\qquad |\Lambda^{(n)}(z) - \Lambda^{(0)}(z)| \leqq M'' \int_{-\infty}^{\infty} \left|F_n\left(s, \frac{t}{b(n)}\right) - F(s)\right| ds$$
$$\leqq M'' \max_j \int_{-\infty}^{\infty} \left|F\left(s + \frac{tc_j}{b(n)}\right) - F(s)\right| ds .$$

Assertion (4.41) and the lemma now follow by the dominated convergence theorem and an application of the theorem on page 64 of [11] to the integral on the right-hand side of (4.43).

LEMMA 4.4. *If Λ satisfies (P), and G, D, F_1 and F_2 hold then,*

$$(4.44)\qquad \sigma_n \sup \left\{\left|T_n\left(\frac{t}{b(n)}\right) - T_n(0)\right| : |t| \leqq M\right\} = o_p(r(n)) .$$

PROOF. Without loss of generality suppose Λ is $\uparrow$. Let $-M = t_0 \leqq t_1 < \cdots < t_{N(\delta)} = M$ be a δ partition of $[-M, M]$. If $t_k \leqq t < t_{k+1}$ then

$$\left|T_n\left(\frac{t}{b(n)}\right) - T_n\left(\frac{t_k}{b(n)}\right)\right|$$
$$\leqq \int_{-\infty}^{\infty} \left\{\Lambda\left(Q_n\left(s, \frac{t}{b(n)}\right)\right) - \Lambda\left(Q_n\left(s, \frac{t_k}{b(n)}\right)\right)\right.$$
$$(4.45)\qquad \left. + \Lambda\left(F_n\left(s, \frac{t}{b(n)}\right)\right) - \Lambda\left(F_n\left(s, \frac{t_k}{b(n)}\right)\right)\right\} ds$$
$$\leqq M''' \left\{\int_{-\infty}^{\infty} \left\{\left(Q_n\left(s, \frac{t_{k+1}}{b(n)}\right) - Q_n\left(s, \frac{t_k}{b(n)}\right)\right)\right.\right.$$
$$\left.\left. + F_n\left(s, \frac{t_{k+1}}{b(n)}\right) - F_n\left(s, \frac{t_k}{b(n)}\right)\right\} ds\right\}$$

where M''' is an upper bound on $|\lambda|$. But the right-hand side of (4.45) equals $2M''' \tau_n/\sigma_n b(n) = o(r(n)/\sigma_n)$.

Therefore,

$$(4.46)\qquad \lim_{\delta\to 0} \limsup_n P_0\left[\max\left\{\frac{\sigma_n}{r(n)} |T_n(t/b(n)) - T_n(t_j/b(n)| :\right.\right.$$
$$\left.\left. t_j \leqq t \leqq t_{j+1}, 0 \leqq j \leqq N(\delta)\right\} \geqq \varepsilon\right] = 0 ,$$

for every $\varepsilon > 0$. The lemma follows from (4.46) and Lemma 4.3.

Finally we have,

LEMMA 4.5. *If (P), (G), D, F_1 and F_2 hold then,*

$$\sup \left\{\left|\left[\int_0^1 \left[Q_n^{-1}\left(w, \frac{t}{b(n)}\right) - F^{-1}(w)\right] \Lambda(dw)\right.\right.\right.$$
$$(4.47)\qquad \left.\left.\left. + \int_{-\infty}^{\infty} \lambda(F(s))(Q_n(s, 0) - F(s))\, ds\right] \sigma_n + \tau_n t\right| : |t| \leqq M\right\}$$
$$= o_p(r(n)) .$$

In view of Lemma 4.3 and (4.36) we need only check that

$$\sigma_n \left\{ \int_0^1 \left[F_n^{-1}\left(w, \frac{t}{b(n)}\right) - F^{-1}(w) \right] \Lambda(dw) \right\} = -t\tau_n + o_p(r(n)) . \tag{4.48}$$

But the second factor on the left of (4.48) equals,

$$\begin{aligned} &-\int_{-\infty}^{\infty} \left[\Lambda\left(F_n\left(s, \frac{t}{b(n)}\right)\right) - \Lambda(F(s)) \right] ds \\ &= \frac{-\sum_{j=1}^n d_j \int_{-\infty}^{\infty} \lambda(F(s))(F(s + c_j t/b(n)) - F(s))\, ds}{\sigma_n} + R_n' \end{aligned} \tag{4.49}$$

where R_n' is bounded in absolute value by,

$$\frac{M''}{\sigma_n^2} \sum_{j=1}^n d_j^2 \int_{-\infty}^{\infty} \left(F\left(s + \frac{c_j t}{b(n)}\right) - F(s)\right)^2 ds . \tag{4.50}$$

Now,

$$\lim_{h\to 0} \int_{-\infty}^{\infty} \left| \frac{F(s+h) - F(s)}{h} - f(s) \right| ds = 0 , \tag{4.51}$$

and since f is bounded

$$\lim_{h\to 0} \int_{-\infty}^{\infty} \left(\frac{F(s+h) - F(s)}{h} \right)^2 ds = \int_{-\infty}^{\infty} f^2(s)\, ds < \infty . \tag{4.52}$$

Applying (4.51) and (4.52) to (4.49) and (4.50) it is easy to see that,

$$\begin{aligned} &\sigma_n \int_{-\infty}^{\infty} \lambda(F(s))\left(F\left(s + \frac{c_j t}{b(n)}\right) - F(s)\right) ds - t\tau_n \\ &= t \sum_{j=1}^n c_j d_j \left\{ \int_{-\infty}^{\infty} \lambda(F(s)) \left[\frac{(F(s + c_j t/b(n)) - F(s)}{c_j t} - f(s) \right] ds \right\} \\ &= o\left(\frac{\gamma_n}{b(n)}\right) \\ &= o\,(r(n)) , \end{aligned} \tag{4.53}$$

and that,

$$R_n' = o\left(\frac{r^2(n)}{\sigma_n^2}\right) .$$

Lemma 4.5 follows.

The validity of Theorem 2.3 for $\hat\beta$ of type 1 or 1′ follows readily from Lemma 4.5. To prove the result for estimates of type 2 and 2′ we need only show that they are equivalent to the type 1 estimates. This will follow from

$$\frac{\sigma_n}{b(n)} \int_0^1 [|Q_n^{-1}(w, \beta^*)| + |F_n^{-1}(w, \beta^*)|] |\lambda_n(w) - \lambda(w)| dw \to_P 0 \tag{4.54}$$

where as in (4.25),

$$\lambda_n(w) = \lambda(Q_n(Y_{(j)}, \beta^*)) \quad \text{on} \quad [Q_n(Y_{(j-1)}, \beta^*), Q_n(Y_{(j)}, \beta^*)] , \quad 1 \leq j \leq n, (Y_{(0)} = -\infty) . \tag{4.55}$$

By construction,

$$(4.56)\qquad |\lambda_n(w) - \lambda(w)| \leqq \frac{\max_j d_j}{\sigma_n}.$$

By D (4.54) will follow if, $\int_0^1 (|Q_n^{-1}(w, \beta^*)| + |F_n^{-1}(w, \beta^*)|)\, dw$ are bounded in probability. But,

$$(4.57)\qquad \int_0^1 |Q_n^{-1}(w, \beta^*)|\, dw = \frac{1}{\sigma_n} \sum_{k=1}^n d_k |X_k - c_k \beta^*|$$
$$\leqq \int_0^1 |Q_n^{-1}(w, 0)|\, dw + \frac{\gamma_n}{\sigma_n} |\beta^*|.$$

Therefore $\int_0^1 |Q_n^{-1}(w, \beta^*)|\, dw$ is bounded in probability. A similar argument disposes of $\int_0^1 |F_n^{-1}(w, \beta^*)|\, dw$ and (4.54) and Theorem 2.3 follow.

REMARK. Even for location this elementary technique yields Moore's [24] result under slightly weaker conditions than the ones he assumed.

PROOF OF THEOREM 3.1. The basic step in proving this result is a generalization of Theorem 2.1. Write,

$$(4.58)\qquad Q_n(s, \mathbf{t}) = \frac{1}{\sum_{j=1}^n d_j} \sum_{j=1}^n d_j I_{[Y_j(\mathbf{t}) \leqq s]}$$

where the d_j satisfy D.

LEMMA 4.6. *If G and F_1 hold then for every $M < \infty$, $0 < \alpha < \frac{1}{2}$,*

$$(4.59)\qquad \sup \left\{ \left| \left[(Q_n^{-1}(w, \mathbf{t}) - F^{-1}(w)) + \frac{(Q_n(F^{-1}(w, \mathbf{0}) - w)}{q(w)} \right] \sigma_n \right.\right.$$
$$\left.\left. + \sum_{k=1}^p t_k \sum_{j=1}^n c_{kj} d_j \right| : |\mathbf{t}| \leqq \frac{M}{b(n)}, \alpha \leqq w \leqq 1 - \alpha \right\}$$
$$= o_p(r(n)).$$

The proof of this lemma is essentially the same as that of Proposition 4.1 and will not be given. Upon substituting $d_j = c_{ij}^{\pm}$, we readily obtain parts I and II of Theorem 3.1 for type 1 estimates.

Showing the equivalence of type 1 and type 2 estimates for these parts is also straightforward. For example, if all the c_{ij} are $\geqq 0$ then,

$$(4.60)\qquad \left(\sum_{j=1}^n c_{ij}\right)\left\{\int_0^1 [Q_{ni}^{-1}(w, \beta^*) - F^{-1}(w)](\Lambda - \Lambda_{ni})(dw)\right\} = o_p(b(n))$$

where Λ_{ni} has density λ_{ni} with

$$(4.61)\qquad \lambda_{ni}(w) = \lambda(W_{ij}) \quad \text{for} \quad W_{i(j-1)} < w \leqq W_{ij}, \qquad 1 \leqq j \leqq n.$$

Now, the type 2 estimate in this case satisfies

$$(4.62)\qquad \frac{\sum_{j=1}^n c_{ij}}{b(n)} \int_0^1 Q_{ni}^{-1}(w, \beta^*)\Lambda_{ni}(dw)$$
$$= \sum_{k=1}^p b(n)(\hat{\beta}_k - \beta_k^*)\left(\frac{1}{b^2(n)} \sum_{j=1}^n c_{iD_j} c_{kD_j} \lambda(W_{ij})\right).$$

It is easy to show that,

$$\frac{1}{b^2(n)} \sum_{j=1}^n c_{iD_j} c_{kD_k} \lambda(W_{ij}) \to_P a_{ij} \tag{4.63}$$

the ijth element of A, and

$$\frac{(\sum_{j=1}^n c_{ij})}{b(n)} \int_0^1 F^{-1}(w)\Lambda_{ni}(dw) \to_P 0 . \tag{4.64}$$

Therefore equations (3.15) are soluble for n large and the solutions are equivalent to order $1/b(n)$ to the appropriate type 1 estimates given by (3.10).

To obtain the results of the theorem for type 1′ estimates apply Lemma 4.6 to each row separately using as observations in the ith row, $X_{ij}^* = [\text{sgn } c_{ij}]X_j$, and putting $d_j = |c_{ij}|$. We then get,

$$\begin{aligned} \frac{1}{b(n)} \Bigg[\Bigg[& \int_0^1 [[Q_{ni}^*]^{-1}(w, \boldsymbol{\beta}^*) - F^{-1}(w)]\Lambda(dw) \\ & + \int_0^1 \frac{(Q_{ni}^*(F^{-1}(w), 0) - w)}{q(w)} \Lambda(dw)\Bigg] \sum_{j=1}^n |c_{ij}| \\ & + \sum_{k=1}^p \beta_k^* \sum_{j=1}^n c_{kj} c_{ij} \Bigg] \to_P 0 . \end{aligned} \tag{4.65}$$

Now, by the symmetry of Λ and F,

$$\begin{aligned} \int_0^1 & \frac{(Q_{ni}^*(F^{-1}(w), 0) - w)}{q(w)} \Lambda(dw)[\sum_{j=1}^n |c_{ij}|] \\ & = \int_0^1 \frac{(S_{ni}(F^{-1}(w), 0) - w \sum_{j=1}^n c_{ij})}{q(w)} \Lambda(dw) \end{aligned} \tag{4.66}$$

where

$$S_{ni}(s, \boldsymbol{\beta}^*) = \sum_{j=1}^n c_{ij} I_{[Y_j(\boldsymbol{\beta}^*) \leq s]} . \tag{4.67}$$

The rest of the argument for the 1′ and 2′ estimates is essentially a recapitulation of that for Theorems 2.1 and 2.2. Note that the symmetry of Λ is not required for $p = 1$ but becomes necessary to preserve the right covariances between rows of the random matrix defining the "primed" estimates. The proofs of the analogue of Theorem 2.3 employ a suitable analogue of Lemma 4.5. The arguments are straightforward and we leave them to the reader.

REFERENCES

[1] Adichie, J. (1967) Estimation of regression parameters based on rank tests. *Ann. Math. Statist.* **38** 894–904.

[2] Bennett. C. (1952). Asymptotic properties of ideal linear estimators. Ph. D. thesis, Univ. of Michigan.

[3] Bickel, P. J. (1965). On some robust estimates of location. *Ann. Math. Statist.* **36** 847–858.

[4] Bickel, P. J. (1967). Some contributions to the theory of order statistics. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 575–591. Univ. of California Press.

[5] Bickel, P. J. (1971). A note on approximate (M) estimates in the linear model. Submitted to *J. Amer. Statist. Assoc.*

[6] Billingsley, P. (1968). *Weak Convergence of Probability Measures.* Wiley, New York.

[7] BIRNBAUM, A. AND MIKÉ, V. (1970). Asymptotically robust estimates of location. *J. Amer. Statist. Assoc.* **65** 1265-1283.

[8] CHERNOFF, H., GASTWIRTH, J. and JOHNS, M. V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *Ann. Math. Statist.* 38 52-72.

[9] GASTWIRTH, J. and RUBIN, H. (1969). On robust linear estimators. *Ann. Math. Statist.* **40** 24-39.

[10] HÁJEK, J. (1971). Asymptotic minimaxity and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 175-194. Univ. of California Press.

[11] HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

[12] HODGES, J. L. JR. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 599-611.

[13] HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73-101.

[14] HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 221-233. Univ. of California Press.

[15] HUBER, P. J. (1970). Robust regression. (Mimeographed).

[16] JAECKEL, L. B. (1971). Robust estimation of location: Symmetry and asymmetric contamination. *Ann. Math. Statist.* **42** 1020-1034.

[17] JUNG, J. (1955). On linear estimates defined by a continuous weight function. *Ark. Math.* **3** 199-209.

[18] JUREČKOVÁ, J. (1969). Asymptotic linearity of a rank statistic in regression parameter. *Ann. Math. Statist.* **40** 1889-1901.

[19] KOUL, H. L. (1969). Asymptotic behavior of Wilcoxon type confidence regions. *Ann. Math. Statist.* **40** 1950-1979.

[20] KRAFT, C. and VAN EEDEN, C. (1970). Efficient linearized estimates based on ranks. *Proc. First Internat. Symp. Nonparam. Statist.* Cambridge Univ. Press. Cambridge, 267-273.

[21] KRAFT, C. and VAN EEDEN, C. (1970). Linearized rank estimates and signed rank estimates for the general linear hypothesis. *Ann. Math. Statist.* **43** 42-57.

[22] KRAFT, C. and VAN EEDEN, C. (1972). Asymptotic efficiences of quick methods of computing efficient estimates based on ranks. *J. Amer. Statist. Assoc.* **67** 199-202.

[23] MOORE, D. S. (1968). An elementary proof of asymptotic normality of linear functions of order statistics. *Ann. Math. Statist.* **39** 263-265.

[24] MOSTELLER, F. (1946). On some useful inefficient statistics. *Ann. Math. Statist.* **17** 377-407.

[25] PYKE, R. and SHORACK, G. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *Ann. Math. Statist.* **39** 755-771.

[26] RELLES, D. (1968). Robust regression by modified least squares. Ph. D. thesis, Yale Univ.

[27] RIESZ, F. and SZ-NAGY, B. (1955). *Functional Analysis*, (2nd ed.). Ungar, New York.

[28] SHORACK, G. (1969). Asymptotic normality of linear combinations of functions of order statistics. *Ann. Math. Statist.* **40** 2041-2050.

[29] STIGLER, S. M. (1969). Linear functions of order statistics. *Ann. Math. Statist.* **40** 770-788.

[30] TUKEY, J. W. and MCLAUGHLIN, G. (1963). Less vulnerable confidence and significance procedures for location based on a single sample. *Sankhyā Ser. A* **35** 331-352.

[31] VAN EEDEN, C. (1968). Nonparametric estimation. *Lecture Notes, Séminaire de Mathématiques Supérieures.* Univ. of Montréal.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720