The Annals of Statistics
1979, Vol. 7, No. 2, 258–268

ASYMPTOTIC BEHAVIOR OF M-ESTIMATORS FOR THE LINEAR MODEL

BY VÍCTOR J. YOHAI AND RICARDO A. MARONNA

Universidad de Buenos Aires and Fundación Bariloche

This paper deals with M-estimators for the linear model $y_i = \mathbf{x}_i'\boldsymbol{\theta} + u_i 1 \leq i \leq n$, where the $\mathbf{x}_i$ are fixed p-dimensional vectors, and the u_i are i.i.d. random variables with distribution F. The estimators considered are solutions $\hat{\boldsymbol{\theta}}$ of the equation $\Sigma_{j=1}^n \psi(y_j - \mathbf{x}_j'\hat{\boldsymbol{\theta}})\mathbf{x}_j = \mathbf{0}$ for some function ψ. Let $\mathbf{X}$ be the matrix whose ith row is $\mathbf{x}_i'$. Then it is proved that $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is bounded in probability assuming that ψ satisfies a set of conditions which include ψ to be monotone and X to have full rank. This implies that a sufficient condition for consistency is that the smallest eigenvalue of $\mathbf{X}'\mathbf{X}$ tends to infinity. For the case in which $p = p_n \to \infty$ it is proved that $p^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is bounded in probability, assuming that $p\varepsilon \to 0$ where $\varepsilon = \max_{1 \leq i \leq n}(\mathbf{x}_i'\mathbf{X}'\mathbf{X}\mathbf{x}_i)$.

The asymptotic normality of these estimators is proved for both the cases of p fixed and $p \to \infty$. The proof of the former is an easy consequence of a result of Bickel on one-step M-estimators. In the case of $p \to \infty$ we assume that ψ has a bounded derivative and that $p^{3/2}\varepsilon \to 0$. This improves an analogous result by Huber, who requires $p^2\varepsilon \to 0$.

1. Introduction. Consider the general linear model

$$\mathbf{y}_n = \mathbf{X}_n\boldsymbol{\theta} + \mathbf{u}_n, \tag{1.1}$$

where $\mathbf{X}_n$ is a given $n \times p$-matrix; $\boldsymbol{\theta}$ is the unknown p-dimensional vector of regression coefficients; $\mathbf{u}_n$ is the vector with components $(u_1, \ldots, u_n)$, where the u_i's are independent identically distributed random variables with distribution function F; and $\mathbf{y}_n = (y_{1n}, \ldots, y_{nn})$ is the vector of observations. Let $\mathbf{x}_{in}' \in R^p$ be the ith row of $\mathbf{X}_n$, with elements $x_{jin}(j = 1, \ldots, p;\ i = 1, \ldots, n)$.

The classical estimator of $\boldsymbol{\theta}$ is the least-squares estimator, defined as the vector $\hat{\boldsymbol{\theta}}_n$ which minimizes $|\mathbf{y}_n - \mathbf{X}_n\hat{\boldsymbol{\theta}}_n|$ (here $|\cdot|$ denotes Euclidean norm). It is well known that if

$$E_F(u_i) = 0 \quad \text{and} \quad E(u_i^2) < \infty, \tag{1.2}$$

then

$$(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})'(\mathbf{X}_n'\mathbf{X}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = O_p(1), \tag{1.3}$$

(where $O_p(1)$ means "bounded in probability"). As a consequence of (1.3), a sufficient condition for the consistency of $\hat{\boldsymbol{\theta}}_n$ is that

$$\lambda_1(\mathbf{X}_n'\mathbf{X}_n) \to \infty, \tag{1.4}$$

where $\lambda_1(\mathbf{A})$ denotes the smallest eigenvalue of the matrix $\mathbf{A}$.

Received November 1976; revised September 1977.
AMS 1970 *subject classifications.* Primary 62G35; secondary 62J05.
Key words and phrases. Robust estimation, linear model, consistency, asymptotic normality.

Eicker (1963) proved that if $\hat{\boldsymbol{\theta}}$ is consistent under any F which satisfies (1.2) then (1.4) must hold.

When F has heavy tails, $\hat{\boldsymbol{\theta}}$ is not efficient, and may not even be consistent. Hence, other "robust" estimators of $\boldsymbol{\theta}$ have been proposed. In particular Relles (1968), Huber (1973) and Yohai (1972) have studied the family of Huber M-estimators, defined as solutions of

$$\Sigma_{i=1}^{n}\rho(y_{in} - \mathbf{x}_{in}'\hat{\boldsymbol{\theta}}) = \text{minimum}, \tag{1.5}$$

where the function ρ is chosen conveniently. If ρ is convex with derivative ψ, then (1.5) is equivalent to

$$\Sigma_{i=1}^{n}\psi(y_{in} - \mathbf{x}_{in}'\hat{\boldsymbol{\theta}})\mathbf{x}_{in} = \mathbf{0}. \tag{1.6}$$

These estimators are not scale equivariant, i.e., they do not satisfy $\hat{\boldsymbol{\theta}}(ay_1, \ldots, ay_n) = a\hat{\boldsymbol{\theta}}(y_1, \ldots, y_n)$. Scale equivariant estimators may be obtained as solutions of

$$\Sigma_{i=1}^{n}\psi_{\hat{\sigma}}(y_{in} - \mathbf{x}_{in}'\hat{\boldsymbol{\theta}})\mathbf{x}_{in} = \mathbf{0}, \tag{1.7}$$

where

$$\psi_{\sigma}(x) = \psi(x/\sigma),$$

and $\hat{\sigma}(y_1, \ldots, y_n)$ is an estimator of a scale paramater σ satisfying

$$\hat{\sigma}(ay_1, \ldots, ay_n) = |a|\hat{\sigma}(y_1, \ldots, y_n). \tag{1.8}$$

Relles (1968) proved the consistency of these estimators when ψ belongs to the Huber family

$$\psi(x, k) = \min(|x|, k)\ \text{sgn}(x), \tag{1.9}$$

and Yohai (1972) proved the consistency for ψ monotone. Both Relles and Yohai assumed a Lindeberg-type condition for $\mathbf{X}$, i.e.,

$$\lim_{n\to\infty}\max_{j\leqslant p}\max_{i\leqslant n}x_{jin}^2/\Sigma_{l=1}^{n}x_{jln}^2 = 0. \tag{1.10}$$

In Theorem 2.1 of this paper it is proved that (1.3) is satisfied when $\hat{\boldsymbol{\theta}}$ is any solution of (1.6), assuming that ψ is monotone, but without any assumption of $\mathbf{X}_n$, except that its rank be p. Then (1.4) is a sufficient condition for consistency. It is clear that (1.4) may not be weakened, since, as said before, it is necessary for the consistency of the least-squares estimator.

In Theorem 2.2 we treat the case in which p may tend to infinity with n. Using a strong Lindeberg condition, it is proved that $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'(\mathbf{X}_n'\mathbf{X}_n)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})/p = O_p(1)$.

Results similar to Theorems 2.1 and 2.2 are stated without proof for the solutions of (1.7).

In Section 3 we study the asymptotic normality of M-estimators, for both the cases of fixed p, and of $p \to \infty$. In the former, normality is obtained as a corollary of Theorem 4.1 of Bickel (1975). In the latter, we improve a result of Huber (1973).

It is important to remark that (1.5) holds with probability one if $\{\mathbf{x}_{1n}, \ldots, \mathbf{x}_{nn}\}$ is a sample from a p-variate distribution that is not concentrated on any subspace.

On the contrary, the probability that a sample satisfies (1.10) may be less than one; e.g., if $p = 1$ and $|\mathbf{x}_i|$ has a stable distribution with characteristic $\alpha < 1$.

2. Consistency. The following conditions will be assumed about ψ and F:

(A1) ψ is nondecreasing;

(A2) there exist positive numbers b, c and d such that

$$D(u, z) \geqslant d \quad \text{if} \quad |u| \leqslant c \quad \text{and} \quad |z| \leqslant b, \tag{2.1}$$

where

$$D(u, z) = (\psi(u + z) - \psi(u))/z \tag{2.2}$$

and c satisfies

$$q = F(c) - F(-c) > 0. \tag{2.3}$$

(A3) $E_F(\psi^2(u)) = v < \infty$;

(A4) $E_F(\psi(u)) = 0$;

(A5) $\mathbf{X}_n'\mathbf{X}_n$ is nonsingular for all $n \geqslant n_0$.

From now on, the subscript n will be generally omitted. Let $\hat{\boldsymbol{\theta}}$ be any solution of (1.6), and put

$$\boldsymbol{\theta}^* = \mathbf{M}\boldsymbol{\theta},\ \hat{\boldsymbol{\theta}}^* = \mathbf{M}\hat{\boldsymbol{\theta}},\ \mathbf{z}_i = (\mathbf{M}')^{-1}\mathbf{x}_i, \tag{2.4}$$

where $\mathbf{M}$ is any $p \times p$ matrix such that $\mathbf{M}'\mathbf{M} = \mathbf{X}'\mathbf{X}$. Then

$$\Sigma_{i=1}^n \mathbf{z}_i\mathbf{z}_i' = \mathbf{I}, \tag{2.5}$$

$$\Sigma_{i=1}^n |\mathbf{z}_i|^2 = p, \tag{2.6}$$

and $\hat{\boldsymbol{\theta}}^*$ is a solution of

$$\Sigma_{i=1}^n \psi(y_i - \mathbf{z}_i'\hat{\boldsymbol{\theta}}^*)\mathbf{z}_i = \mathbf{0}. \tag{2.7}$$

The main results of this section are the following:

THEOREM 2.1. *Assume* (A1) *through* (A5). *If p is fixed, then*

(i) $|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*|$ *is bounded in probability.*

(ii) *If* $\lambda_1(\mathbf{X}'\mathbf{X}) \to \infty$, *then* $\hat{\boldsymbol{\theta}}$ *is a consistent estimator of* $\boldsymbol{\theta}$.

COROLLARY. *Let* $\{\mathbf{x}_i\}$ *be a sequence of* i.i.d. *p-dimensional random vectors such that* $P(\mathbf{x}_i'\boldsymbol{\theta} = 0) < 1$ *for all* $\boldsymbol{\theta} \in R^p$; *and* $\{u_i\}$ *a sequence of* i.i.d *random variables with distribution function F, such that* $\{\mathbf{x}_i\}$ *and* $\{u_i\}$ *are independent. Put* $y_i = \mathbf{x}_i'\boldsymbol{\theta} + u_i$ *(this is equivalent to stating that the conditional distribution of y_i given x_i is* $F(y - \mathbf{x}_i'\boldsymbol{\theta})$*). Assume* (A1) *through* (A5). *Then the estimator* $\hat{\boldsymbol{\theta}}$ *defined in* (1.6) *is consistent.*

THEOREM 2.2. *Assume* (A1) *through* (A5). *Suppose that* $p = p_n$ *depends on n and*

$$\lim_{n\to\infty} p_n \max_{i\leqslant n} |\mathbf{z}_i|^2 = 0. \tag{2.8}$$

Then $p_n^{-\frac{1}{2}}|\hat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}^*| = O_p(1)$.

We begin by proving some auxiliary results.

LEMMA 1. *Assume* (A3), (A4) *and* (A5). *Let* $J_n = \{1, \ldots, n\}$. *Then for any* $k > 0$ *and any set* $J \subseteq J_n$

$$P\big(|\Sigma_{j \in J}\psi(u_j)\mathbf{z}_j| \geqslant k\big) \leqslant pv/k^2.$$

PROOF. It is straightforward to prove that

$$E|\Sigma_{j\in J}\psi(u_j)\mathbf{z}_j|^2 = v\Sigma_{j\in J}|\mathbf{z}_j|^2;$$

then (2.6) and the Markov inequality yield the desired result.

Define now the matrices

$$\mathbf{D}(\varepsilon) = d\Sigma_{j=1}^n \mathbf{z}_j\mathbf{z}_j' I(|u_j| \leqslant c) I(|\mathbf{z}_j| \leqslant \varepsilon), \tag{2.9}$$

and

$$\mathbf{D}_0(\varepsilon) = E\mathbf{D}(\varepsilon) = dq\Sigma_{j=1}^n \mathbf{z}_j\mathbf{z}_j' I(|\mathbf{z}_j| \leqslant \varepsilon), \tag{2.10}$$

where q and d are defined in (A2), and $I(\cdot)$ is the indicator function of a set.

LEMMA 2. *Assume* (A5). *Define for any matrix* $\mathbf{A} = \{a_{ij}\}$, $\|\mathbf{A}\|^2 = \Sigma_j\Sigma_i a_{ij}^2$. *Then for any* $\delta > 0$ *we have*

$$P(\|\mathbf{D}(\varepsilon) - \mathbf{D}_0(\varepsilon)\| \geqslant \delta) \leqslant r^2\varepsilon^2 p/\delta^2,$$

where $r^2 = d^2 q(1-q)$.

PROOF. The result follows applying (2.6) and the Markov inequality to

$$E\big(\|\mathbf{D}(\varepsilon) - \mathbf{D}_0(\varepsilon)\|^2\big) = r^2\Sigma_{j=1}^n|\mathbf{z}_j|^4 I(|\mathbf{z}_j| \leqslant \varepsilon)$$
$$\leqslant r^2\varepsilon^2\Sigma_{j=1}^n|\mathbf{z}_j|^2 I(|\mathbf{z}_j| \leqslant \varepsilon).$$

Let k_0 be any positive number such that

$$k_0 < \psi(\infty) \quad \text{and} \quad \psi(-\infty) < -k_0; \tag{2.11}$$

then we have the following lemma.

LEMMA 3. *Let* J_0 *be any set such that* $J_0 \subseteq J_n$ *and let* m *be the number of its elements. Let* η *be a given positive number, and put* $T = \{\mathbf{t} \in R^n : \eta \leqslant |t_j| \leqslant 1, j = 1, \ldots, n\}$. *Assume* ($A_1$). *Then, given* $\delta > 0$, *there exists* $L = L(\eta, \delta, m)$ (*which does not depend on* n) *such that*

$$P\big(\sup_{J \subseteq J_0, \mathbf{t} \in T}\Sigma_{j\in J}[\psi(u_j - Lt_j)t_j + k_0|t_j|] \geqslant 0\big) \leqslant \delta. \tag{2.12}$$

PROOF. Assumption (A1) implies that, if $|t| \geqslant \eta$ and $L > 0$, then for all u

$$\psi(u - Lt)\operatorname{sgn} t \leqslant \min[\psi(u - L\eta), -\psi(u + L\eta)].$$

Besides, with probability one,

$$\lim_{L\to\infty}\psi(u - L\eta) < -k_0, \lim_{L\to\infty}(-\psi(u + L\eta)) < -k_0.$$

Hence there exists L such that, for any j,

$$P\big(\sup_{|t|\geqslant\eta}\psi(u_j - Lt)\operatorname{sgn} t + k_0 \geqslant 0\big) \leqslant \delta/m;$$

hence the left-hand term in (2.12) is less than

$$P\Big(\bigcup_{j\in J_0}\big[\sup_{\eta\leqslant|t_j|\leqslant 1}\psi(u_j - Lt_j)t_j + k_0|t_j|\big] \geqslant 0\Big) \leqslant \delta.$$

PROOF OF THEOREM 2.1. There is no loss of generality in assuming $\boldsymbol{\theta}^* = \mathbf{0}$. To simplify notation, define for $\boldsymbol{\theta} \in R^p$ and $L > 0$

$$U(\boldsymbol{\theta}, L) = \Sigma_{j=1}^n \psi(u_j - L\mathbf{z}_j'\boldsymbol{\theta})(\mathbf{z}_j'\boldsymbol{\theta}). \tag{2.13}$$

Then by (A1), since $U(\hat{\boldsymbol{\theta}}^*/|\hat{\boldsymbol{\theta}}^*|, |\hat{\boldsymbol{\theta}}^*|) = 0$:

$$P(|\hat{\boldsymbol{\theta}}^*| > L) \leqslant P(\sup_{|\boldsymbol{\theta}|=1} U(\boldsymbol{\theta}, L) > 0),$$

and hence it suffices to prove that, given δ, there exists L such that

$$\limsup_{n\to\infty} P(\sup_{|\boldsymbol{\theta}|=1} U(\boldsymbol{\theta}, L) \geqslant 0) < \delta. \tag{2.14}$$

Take k_1 such that

$$pv/k_1^2 \leqslant \delta/4. \tag{2.15}$$

Using (A1) and (A4) we can find a positive number k_0 satisfying (2.11). Define

$$\delta_2 = k_0/32k_1. \tag{2.16}$$

Let $\{V_1, \ldots, V_M\}$ be a covering of the spherical surface $\{\boldsymbol{\theta} : |\boldsymbol{\theta}| = 1\}$ with balls of diameter less that δ_2. Let δ_1 be such that

$$\delta_1 \leqslant qd/2, \tag{2.17}$$

$$M(32/k_0)^2(v/qd)\delta_1 \leqslant \delta/8. \tag{2.18}$$

Define L_1 and ε such that

$$k_1 < L_1\delta_1/2, \tag{2.19}$$

$$4\varepsilon^2 r^2 p/\delta_1^2 \leqslant \delta/4, \tag{2.20}$$

(where r^2 is defined in Lemma 2) and

$$\varepsilon L_1 < b. \tag{2.21}$$

Finally, define the sets

$$S = \{\boldsymbol{\theta};\ |\boldsymbol{\theta}| = 1,\ \boldsymbol{\theta}'\mathbf{D}_0(\varepsilon)\boldsymbol{\theta} \geqslant \delta_1\}, \tag{2.22}$$

$$S' = \{\boldsymbol{\theta};\ |\boldsymbol{\theta}| = 1,\ \boldsymbol{\theta}'\mathbf{D}_0(\varepsilon)\boldsymbol{\theta} < \delta_1\}. \tag{2.23}$$

It will be proved in the first place that

$$\limsup_{n\to\infty} P(\sup_{\boldsymbol{\theta}\in S} U(\boldsymbol{\theta}, L_1) \geqslant 0) \leqslant \delta/2; \tag{2.24}$$

then it will be proved that there exists L_2 such that

$$\limsup_{n\to\infty} P(\sup_{\boldsymbol{\theta}\in S'} U(\boldsymbol{\theta}, L_2) \geqslant 0) \leqslant \delta/2. \tag{2.25}$$

From this will follow (2.14), putting $L = \max(L_1, L_2)$.

To prove (2.24) decompose U as

$$U(\boldsymbol{\theta}, L_1) = \mathbf{g}'\boldsymbol{\theta} - h(\boldsymbol{\theta}),$$

where

$$g = \Sigma_{j=1}^{n}\psi(u_j)\mathbf{z}_j,$$

and

$$h(\boldsymbol{\theta}) = L_1\Sigma_{j=1}^{n}D(u_j, -L_1\mathbf{z}_j'\boldsymbol{\theta})(\mathbf{z}_j'\boldsymbol{\theta})^2,$$

where D is defined in (A2). Then by (2.1), (2.9) and (2.21)

$$h(\boldsymbol{\theta}) \geqslant L_1\boldsymbol{\theta}'\mathbf{D}(\varepsilon)\boldsymbol{\theta}\forall\boldsymbol{\theta} \quad \text{such that} \quad \|\boldsymbol{\theta}\| = 1.$$

Besides, by Lemma 2 and (2.20)

$$P(\|\mathbf{D}(\varepsilon) - \mathbf{D}_0(\varepsilon)\| \geqslant \delta_1/2) \leqslant \delta/4;$$

hence by (2.22), with probability larger than $1 - \delta/4$,

$$\begin{aligned}\inf_{\boldsymbol{\theta}\in S}\boldsymbol{\theta}'\mathbf{D}(\varepsilon)\boldsymbol{\theta} &= \inf_{\boldsymbol{\theta}\in S}\left[\boldsymbol{\theta}'\mathbf{D}_0(\varepsilon)\boldsymbol{\theta} + \boldsymbol{\theta}'(\mathbf{D}(\varepsilon) - \mathbf{D}_0(\varepsilon))\boldsymbol{\theta}\right] \\ &\geqslant \delta_1 - \|\mathbf{D}(\varepsilon) - \mathbf{D}_0(\varepsilon)\| \geqslant \delta_1/2.\end{aligned}$$

Hence by (2.19)

$$P(\inf_{\boldsymbol{\theta}\in S}h(\boldsymbol{\theta}) \geqslant k_1) \geqslant 1 - \delta/4, \tag{2.26}$$

and finally (2.26), (2.15) and Lemma 1 yield

$$P(\sup_{\boldsymbol{\theta}\in S}U(\boldsymbol{\theta}, L_1) \geqslant 0) \leqslant P(|\mathbf{g}| \geqslant \inf_{\boldsymbol{\theta}\in S}h(\boldsymbol{\theta})) \leqslant \delta/2.$$

This proves (2.24). Now we shall prove (2.25).

Put

$$J = \{j : j \leqslant n, |\mathbf{z}_j| > \varepsilon\}.$$

Then (2.6) implies that

$$\operatorname{card}(J) \leqslant m = : p/\varepsilon^2, \tag{2.27}$$

where "card" is the cardinal of the set.

Let us now define the set

$$I = I(\eta, \boldsymbol{\theta}) = \{i : i \leqslant n, |\mathbf{z}'\boldsymbol{\theta}| > \eta\}.$$

It will be proved that there exists $\eta \in (0, 1)$ independent of n, such that, for all $\boldsymbol{\theta} \in S'$

$$\Sigma_{i\in I\cap J}|\mathbf{z}_j'\boldsymbol{\theta}| \geqslant \tfrac{1}{4}, \tag{2.28}$$

and

$$P(\sup_{\boldsymbol{\theta}\in S'}\eta\Sigma_{i\in I\cap J}|\psi(u_j)| \geqslant k_0/16) \leqslant \delta/16. \tag{2.29}$$

In effect, if $\eta \in (0, 1)$ and $\boldsymbol{\theta} \in S'$, (2.17) implies

$$qd/2 \geqslant \delta_1 > \boldsymbol{\theta}'\mathbf{D}_0(\varepsilon)\boldsymbol{\theta} = dq\Sigma_{i\in J^c}(\mathbf{z}_i'\boldsymbol{\theta})^2;$$

since $\Sigma_{i=1}^{n}(\mathbf{z}_i'\boldsymbol{\theta})^2 = 1$, this yields

$$1/2 \leqslant \Sigma_{i\in J}(\mathbf{z}_i'\boldsymbol{\theta})^2 = \Sigma_{i\in I\cap J}(\mathbf{z}_i'\boldsymbol{\theta})^2 + \Sigma_{i\in J-I}(\mathbf{z}_i'\boldsymbol{\theta})^2. \tag{2.30}$$

But by (2.27)

$$\Sigma_{i \in J - I}(\mathbf{z}_i'\boldsymbol{\theta})^2 \leqslant \eta^2 \Sigma_{i \in J} 1 \leqslant \eta^2 m;$$

hence, since $|\mathbf{z}_i'\boldsymbol{\theta}| \leqslant 1$, (2.30) yields

$$1/2 - \eta^2 m \leqslant \Sigma_{i \in I \cap J} |\mathbf{z}_i'\boldsymbol{\theta}|,$$

so that for η sufficiently small (independently of n), (2.28) holds; (2.29) results from $\operatorname{card}(J) \leqslant m$.

Let now $L = L(\eta, \delta/4, m)$ be the number obtained in Lemma 3; where η is obtained in (2.28)-(2.29), and m in (2.27). Then, by Lemma 3,

$$P(\sup_{\boldsymbol{\theta} \in S'}[\Sigma_{i \in I \cap J}\psi(u_i - L_2\mathbf{z}_i'\boldsymbol{\theta})(\mathbf{z}_i'\boldsymbol{\theta}) + k_0 \Sigma_{i \in I \cap J}|\mathbf{z}_i'\boldsymbol{\theta}|] \leqslant 0) \geqslant 1 - \delta/4. \tag{2.31}$$

Decompose the sum in $U(\boldsymbol{\theta}, L_2)$ into those terms corresponding respectively to $I \cap J$, $J - I$, and J^c. Then we have, applying (A1) in the last two terms:

$$U(\boldsymbol{\theta}, L_2) \leqslant s(\boldsymbol{\theta}) + t(\boldsymbol{\theta}) + u(\boldsymbol{\theta}), \tag{2.32}$$

where

$$s(\boldsymbol{\theta}) = \Sigma_{i \in I \cap J}\psi(u_i - L_2\mathbf{z}_i'\boldsymbol{\theta})(\mathbf{z}_i'\boldsymbol{\theta}),$$
$$t(\boldsymbol{\theta}) = \eta \Sigma_{i \in J - I}|\psi(u_i)|,$$
$$u(\boldsymbol{\theta}) = \Sigma_{i \in J^c}\psi(u_i)\mathbf{z}_i'\boldsymbol{\theta}.$$

Then (2.29) implies

$$P(\sup_{\boldsymbol{\theta} \in S'} t(\boldsymbol{\theta}) > k_0/16) \leqslant \delta/16. \tag{2.33}$$

By (2.28) and (2.31):

$$P(\sup_{\boldsymbol{\theta} \in S'} s(\boldsymbol{\theta}) \leqslant -k_0/4) \geqslant 1 - \delta/4. \tag{2.34}$$

Finally, it will be proved that

$$P(\sup_{\boldsymbol{\theta} \in S'} u(\boldsymbol{\theta}) > k_0/16) \leqslant 3/8\delta. \tag{2.35}$$

Let $W_{j'}(j' = 1, \ldots, M' \leqslant M)$ be the nonvoid intersections $V_j \cap S' \neq \varnothing$. Let $\boldsymbol{\theta}_j$ be any element of W_j, so that $\boldsymbol{\theta} \in W_j$ implies $|\boldsymbol{\theta} - \boldsymbol{\theta}_j| < \delta_2$.

Define the event

$$B = \{|\Sigma_{i \in J^c}\psi(u_i)z_i| < k_1\}, \tag{2.36}$$

so that Lemma 1 and (2.15) imply that $P(B^c) \leqslant \delta/4$. Hence

$$P(\sup_{\boldsymbol{\theta} \in S'} u(\boldsymbol{\theta}) > k_0/16) \leqslant \delta/4 + \Sigma_{j=1}^{M'} P\Big[B \bigcap \big(\sup_{\boldsymbol{\theta} \in W_j} u(\boldsymbol{\theta}) \geqslant k_0/16\big)\Big]. \tag{2.37}$$

In B we have for each j

$$\sup_{\boldsymbol{\theta} \in W_j}|u(\boldsymbol{\theta})| \leqslant |\Sigma_{i \in J^c}\psi(u_i)\mathbf{z}_i'\boldsymbol{\theta}_j| + \sup_{\boldsymbol{\theta} \in W_j}|\Sigma_{i \in J^c}\psi(u_i)\mathbf{z}_i'(\boldsymbol{\theta} - \boldsymbol{\theta}_j)|$$
$$\leqslant |\Sigma_{i \in J^c}\psi(u_i)\mathbf{z}_i'\boldsymbol{\theta}_j| + k_1\delta_2;$$

applying (2.16), the Chebychev inequality, and then (2.18) and (2.23) we have

$$P\Big[B \bigcap \big(\sup_{\theta \in W_j} |\Sigma_{i \in J^c} \psi(u_i) \mathbf{z}_i' \theta_j|\big) \geqslant k_0/16\Big]$$
$$\leqslant P\big(|\Sigma_{i \in J^c} \psi(u_i) \mathbf{z}_i' \theta_j| \geqslant k_0/32\big)$$
$$\leqslant (32/k_0)^2 v\big(\theta_j' \mathbf{D}_0(\varepsilon)\theta_j\big)/dq \leqslant \delta/8M.$$

Hence the expression in (2.37) is less than $\delta/4 + M\delta/8M = 3\delta/8$. Finally, (2.25) results from (2.32), (2.33), (2.34) and (2.35). This finishes the proof of part (i) of Theorem 2.1.

The proof of part (ii) is a trivial consequence of part (i).

PROOF OF THE COROLLARY. It is easy to prove that the fact that $P(\mathbf{x}_i'\theta = 0) < 1$ for all θ implies that $\lambda_1(\mathbf{X}'\mathbf{X}) \to \infty$ with probability one. Then part (ii) of the theorem implies that for all $\delta > 0$:

$$\lim_{n \to \infty} P\big(|\hat{\theta} - \theta| > \delta | \mathbf{x}_1, \ldots, \mathbf{x}_n\big) = 0 \qquad \text{a.s.,}$$

which implies the consistency of $\hat{\theta}$.

PROOF OF THEOREM 2.2. Put

$$\varepsilon = \varepsilon_n = \max_{i \leqslant n} |\mathbf{z}_i|^2, \tag{2.38}$$

so that by (2.8)

$$p_n \varepsilon_n \to 0. \tag{2.39}$$

Then note that Lemma 2 implies that

$$E\,\|\mathbf{D}(\varepsilon^{\frac{1}{2}}) - \mathbf{D}_0(\varepsilon^{\frac{1}{2}})\| \to 0, \tag{2.40}$$

where the matrices $\mathbf{D}$ and $\mathbf{D}_0$ are defined in (2.9) and (2.10). Note that

$$\mathbf{D}_0(\varepsilon^{\frac{1}{2}}) = dq\mathbf{I}. \tag{2.41}$$

Now, to prove the thesis, recall that, as in (2.13),

$$P\big(p^{-\frac{1}{2}}|\hat{\theta}^*| \geqslant L\big) \leqslant P\big(\sup_{|\theta|=1} p^{-\frac{1}{2}} U(\theta, p^{\frac{1}{2}}L) \geqslant 0\big). \tag{2.42}$$

Now decompose

$$p^{-\frac{1}{2}} U\big(\theta, p^{\frac{1}{2}}L\big) = \Sigma_{i=1}^n \psi(u_i)\mathbf{z}_i'\theta p^{-\frac{1}{2}} - L\Sigma_{i=1}^n D\big(u_i, -Lp^{\frac{1}{2}}\mathbf{z}_i'\theta\big)(\mathbf{z}_i'\theta)^2. \tag{2.43}$$

Put $A_n = \sup_{|\theta|=1} |\Sigma_{i=1}^n \psi(u_i)\mathbf{z}_i'\theta p^{-\frac{1}{2}}|$. Then by Lemma 1 it is possible to find L such that for all n

$$P(A_n \geqslant Ldq/2) \leqslant \frac{\delta}{2}. \tag{2.44}$$

By (2.8) there exists n_0 such that $p^{\frac{1}{2}}\varepsilon^{\frac{1}{2}}L \leqslant b$ for all $n \geqslant n_0$. Then by (A2) we have for all $n \geqslant n_0$

$$B_n = \inf_{|\theta|=1} \Sigma_{i=1}^n D\big(u_i, -Lp^{\frac{1}{2}}\mathbf{z}_i'\theta\big)(\mathbf{z}_i'\theta)^2$$
$$\geqslant \inf_{|\theta|=1} \theta'\mathbf{D}(\varepsilon^{\frac{1}{2}})\theta \geqslant qd - \|\mathbf{D}(\varepsilon^{\frac{1}{2}}) - \mathbf{D}_0(\varepsilon^{\frac{1}{2}})\|.$$

Then, by (2.8) and Lemma 2 there exists $n_1 \geqslant n_0$ such that

$$P(B_n < dq/2) \leqslant \frac{\delta}{2} \forall n \geqslant n_1. \tag{2.45}$$

Then by (2.42), (2.43), (2.44) and (2.45) we have for all $n \geqslant n_1$

$$P\big(p^{-\frac{1}{2}}|\hat{\boldsymbol{\theta}}^*| \geqslant L\big) \leqslant P(A_n \geqslant Ldq/2) + P(B_n < dq/2) \leqslant \delta.$$

This finishes the proof.

REMARK. Suppose now that $\hat{\boldsymbol{\theta}}$ is a solution of (1.7) where $\hat{\sigma} \to \sigma$ in probability. Then Theorems 2.1 and 2.2 still hold if we replace (A1) . . . (A5) by (A′1) . . . (A′6) where

(A′1) (A1);
(A′2) ψ_σ satisfies (A2);
(A′3) there exists $\varepsilon > 0$ such that $E_F(\psi_\sigma^2((1+\varepsilon)u)) < \infty$;
(A′4) there exists $\varepsilon > 0$ such that for all λ such that $|\lambda| < \varepsilon$, we have $E_F(\psi_\sigma((1+\lambda)u)) = 0$;
(A′5) (A5);
(A′6) there exists $\varepsilon > 0$ such that

$$\sup_{|\lambda| \leqslant \varepsilon, |\lambda'| \leqslant \varepsilon} E_F\big([\psi_\sigma((1+\lambda)u) - \psi_\sigma((1+\lambda')u)]^2\big)/(\lambda - \lambda')^2 \leqslant \infty.$$

If $n^{\frac{1}{2}}(\hat{\sigma} - \sigma) = O_p(1)$, (A′4) may be replaced by $E_F(\psi_\sigma(u)) = 0$.

We do not give the proofs of these results here. They are based in tightness arguments similar to those used in Yohai (1972) and Bickel (1975).

3. Asymptotic normality. The asymptotic normality of the estimators will be proved for both the cases of p fixed and $p \to \infty$, respectively in Theorems 3.1 and 3.2. The proof for the former case is an easy consequence of a result of Bickel (1975) for one-step M-estimators. The asymptotic normality for the case $p \to \infty$ is proved assuming that ψ has a bounded third derivative and that $p^{\frac{3}{2}}\varepsilon \to 0$, where ε is defined in (2.38). This improves an analogous result by Huber (1973) who required $p^2\varepsilon \to 0$.

From now on, the notation of Section 2 will be kept, as well as the assumption that the "true parameter" is $\boldsymbol{\theta} = 0$.

THEOREM 3.1. *Assume that* ψ *and* F *satisfy* (A1) *through* (A5), *and the following conditions*:

(C1) $\int_{-\infty}^{\infty}[\psi(x+h) - \psi(x-h)]^2 dF(x) = o(1)$ *as* $h \to 0$

and

$\sup_{|q| \leqslant \varepsilon, |h| \leqslant \varepsilon}\{|h|^{-1}\int_{-\infty}^{\infty}[\psi(x+q+h) - \psi(x+q)]dF(x)\} < \infty$

for some $\varepsilon > 0$;

(C2) *There exists* $A(\psi, F)$ *such that*

$$\int_{-\infty}^{\infty}[\psi(x+h) - \psi(x)]dF(x) = hA(\psi, F) + o(h).$$

Assume also

(G)
$$\lim_{n\to\infty}\max_{i\leqslant n}|\mathbf{z}_i|^2 = 0.$$

Let $\hat{\boldsymbol{\theta}}^*$ *be defined as in* (2.4) *with* $\hat{\boldsymbol{\theta}}$ *given by* (1.6) *Then the distribution of* $(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)$ *tends to a multivariate normal with mean* $\mathbf{0}$ *and covariance matrix* $\tau^2\mathbf{I}$, *where*

(3.1)
$$\tau^2 = E_F\psi^2/A(\psi, F)^2.$$

THEOREM 3.2. *Assume that, in addition to* (A1) *through* (A5), ψ *and* F *satisfy the following conditions*:

(N1) ψ *is three times differentiable, and has a bounded third derivative*: $|\psi'''(u)| \leqslant c$;

(N2) $E_F|\psi'(u)|$ *and* $E_F(\psi''(u))^2$ *are finite*;

(N3) $E_F\psi''(u) = 0$.

Assume also that

(N4) $\lim_{n\to\infty} p_n^{\frac{3}{2}}\varepsilon_n = 0$,

where $\varepsilon_n = \max_{i\leqslant n}|\mathbf{z}_i|^2$.

Let $\mathbf{a}_n$ *be any sequence in* R^{p_n}, *with* $|\mathbf{a}_n| = 1$. *Then the distribution of* $\mathbf{a}_n'(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)$ *tends to* $N(0, \tau^2)$, *where* τ^2 *is defined in* (3.1) (*note that in this case*, $A(\psi, F) = E_F\psi'$).

PROOF OF THEOREM 3.1. Note that $\hat{\boldsymbol{\theta}}^*$ may be viewed as a "one-step" estimator with $\hat{\boldsymbol{\theta}}^*$ itself as the "initial estimator." By Theorem 2.1 the initial estimator has the order of convergence required by Theorem 4.1 of Bickel (1975) (see the remarks at the end of that paper). Hence our result follows easily from Bickel's theorem.

REMARK. If $\hat{\boldsymbol{\theta}}$ is a solution of (1.7) with $n^{\frac{1}{2}}(\hat{\sigma} - \sigma) = 0_p(1)$ and (A1)-(A5) are replaced by (A′1)-(A′6) and condition (G) holds, then Theorem 4.1 of Bickel gives also sufficient conditions for the convergence of $(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)$ to a normal distribution $N(\mathbf{0}, \tau^2\mathbf{I})$, where

$$\tau^2 = E_F\psi_\sigma^2/A(\psi_\sigma, F)^2.$$

PROOF OF THEOREM 3.2. Multiplying (2.7) by $\mathbf{a}$, and performing a second order Taylor expansion, we have

(3.2)
$$\begin{aligned} 0 &= \Sigma_{i=1}^n\psi\big(u_i - \mathbf{z}_i'\hat{\boldsymbol{\theta}}^*\big)\mathbf{z}_i'\mathbf{a} \\ &= w_1 - w_2 - \mathbf{w}_3'\hat{\boldsymbol{\theta}}^* + \tfrac{1}{2}\boldsymbol{\theta}^{*\prime}\mathbf{W}_4\boldsymbol{\theta}^* - \tfrac{1}{6}w_5, \end{aligned}$$

where

$$\begin{aligned} w_1 &= \Sigma_{i=1}^n\psi(u_i)\mathbf{z}_i'\mathbf{a}, \\ w_2 &= (E\psi')\Sigma_{i=1}^n\mathbf{a}'\mathbf{z}_i\mathbf{z}_i'\hat{\boldsymbol{\theta}}^*, \\ \mathbf{w}_3 &= \Sigma_{i=1}^n\big[\psi'(u_i) - E\psi'\big](\mathbf{a}'\mathbf{z}_i)\mathbf{z}_i, \\ \mathbf{W}_4 &= \Sigma_{i=1}^n\psi''(u_i)(\mathbf{a}'\mathbf{z}_i)\mathbf{z}_i\mathbf{z}_i', \\ w_5 &= \Sigma_{i=1}^n\psi'''\big(u_i + \eta_i\mathbf{z}_i'\hat{\boldsymbol{\theta}}^*\big)\big(\mathbf{z}_i'\hat{\boldsymbol{\theta}}^*\big)^3(\mathbf{a}'\mathbf{z}_i). \end{aligned}$$

with $|\eta_i| < 1$.

Note first that $Ew_1 = 0$ and Var $(w_1) = E\psi^2(u)$; since $\varepsilon \to 0$, the Lindeberg theorem yields

$$w_1 \text{ converges in law to } N(0, E\psi^2). \tag{3.3}$$

Besides, (2.5) implies that

$$w_2 = (E\psi')\mathbf{a}'\hat{\boldsymbol{\theta}}^*. \tag{3.4}$$

Hence, to prove the thesis, it suffices to show that the remaining three terms of (3.2) are $o_p(1)$. It will be first proved that

$$p^{\frac{1}{2}}|\mathbf{w}_3| = o_p(1). \tag{3.5}$$

In effect, $E\mathbf{w}_3 = \mathbf{0}$, and its covariance matrix is:

$$\mathrm{Cov}(\mathbf{w}_3) = \mathrm{Var}(\psi')\Sigma_{i=1}^n(\mathbf{a}'\mathbf{z}_i)^2\mathbf{z}_i\mathbf{z}_i';$$

hence, since $\Sigma_{i=1}^n(\mathbf{a}'\mathbf{z}_i)^2 = 1$

$$pE|\mathbf{w}_3|^2 = p\mathrm{Var}(\psi')\Sigma_{i=1}^n(\mathbf{a}'\mathbf{z}_i)^2|\mathbf{z}_i|^2 \leqslant \mathrm{Var}(\psi')\varepsilon p \to 0.$$

This proves (3.5). Since by Theorem 2.2, $p^{-\frac{1}{2}}|\hat{\boldsymbol{\theta}}^*| = O_p(1)$, this deals with the third term of (3.2).

Now it will be proved that

$$p\|\mathbf{W}_4\| = o_p(1). \tag{3.6}$$

In effect, since $\|\mathbf{W}_4\|^2 = \text{trace } (\mathbf{W}_4\mathbf{W}_4')$:

$$p^2E\|\mathbf{W}_4\|^2 \leqslant E(\psi'')^2\Sigma_{i=1}^n(\mathbf{a}'\mathbf{z}_i)^2|\mathbf{z}_i|^4p^2 \leqslant E(\psi'')^2\varepsilon^2p^2 \to 0.$$

Application of Theorem 2.2 finishes with the fourth term in (3.2). Finally, it will be proved that

$$w_5 = o_p(1). \tag{3.7}$$

In effect, the Cauchy-Schwarz inequality and (2.5) yield

$$|w_5| \leqslant c\Sigma_{i=1}^n(\mathbf{z}_i'\hat{\boldsymbol{\theta}}^*)^2|\hat{\boldsymbol{\theta}}^*||\mathbf{z}_i|^2 \leqslant c|\hat{\boldsymbol{\theta}}^*|\varepsilon\Sigma_{i=1}^n(\mathbf{z}_i'\hat{\boldsymbol{\theta}}^*)^2 = c|\hat{\boldsymbol{\theta}}^*|^3\varepsilon = |p^{-\frac{1}{2}}\hat{\boldsymbol{\theta}}^*|^3cp^{\frac{3}{2}}\varepsilon;$$

then Theorem 2.2 and (N4) imply 3.7, thus finishing the proof of the theorem.

REFERENCES

[1] BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–433.

[2] EICKER, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Ann. Math. Statist.* **34** 447–456.

[3] HUBER, P. (1973). Robust regression. *Ann. Statist.* **1** 799–821.

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

[5] YOHAI, V. J. (1972). Robust (M) estimates for the general linear model. *Notas de Matemática*, No. 23, Univ. Nacional de La Plata, Facultad de Ciencias Exactas y Naturales.

DEPARTAMENTO DE MATEMATICAS
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
UNIVERSIDAD NACIONAL DE BUENOS AIRES
PABELLON 1,
CIUDAD UNIVERSITARIA
BUENOS AIRES, ARGENTINA