The Annals of Statistics
1990, Vol. 18, No. 4, 1651–1675

THE ASYMPTOTICS OF S-ESTIMATORS IN THE LINEAR REGRESSION MODEL

BY LAURIE DAVIES

University of Essen

We consider the consistency and weak convergence of S-estimators in the linear regression model. Sufficient conditions for consistency with varying dimension are given which are sufficiently weak to cover, for example, polynomial trends and i.i.d. carriers. A weak convergence theorem for the Hampel–Rousseeuw least median of squares estimator is obtained, and it is shown under rather general conditions that the correct norming factor is $n^{1/3}$. Finally, the asymptotic normality of S-estimators with a smooth ρ-function is obtained again under weak conditions on the carriers.

1. Introduction. It is well known that the least squares estimator in the linear regression model is not robust: one errant observation can give rise to arbitrarily poor estimates of the coefficients.

Hampel (1975) proposed an estimator based on minimizing the median absolute deviation of the residuals and gave its breakdown point as $\frac{1}{2}$, the highest possible value for affine equivariant estimates. His idea was taken up by Rousseeuw (1984) who developed it into a practical method for estimating the parameters and identifying possible outliers in the linear regression model [see Rousseeuw and Leroy (1987)]. Rousseeuw (1984) showed that the estimate, now known as the least median of squares, was affine equivariant and determined its finite sample breakdown point. He also considered the weak convergence of the estimator. A generalization of least median of squares was given by Rousseeuw and Yohai (1984) who introduced a new class of estimators, the S-estimators. For a discussion of the problem of robust estimation in the linear regression model we refer to Hampel, Ronchetti, Rousseeuw and Stahel (1986), Huber (1981), Rousseeuw (1984), Rousseeuw and Yohai (1984) and Rousseeuw and Leroy (1987).

Rousseeuw and Yohai (1984) proved consistency and asymptotic normality (with a norming factor of $n^{1/2}$) for a restricted class of S-estimators. They assumed, however, that the carriers, the x-variables, were i.i.d. and this excludes many cases of practical interest such as polynomial trends. The theory for the Hampel–Rousseeuw least median of squares is even less satisfactory. Its behaviour is related to that of the shorth. There exist heuristic arguments showing that the correct norming factor for the shorth and also for middle of the shortest half is $n^{1/3}$ and these arguments also identify the limiting distribution [see Andrews, Bickel, Hampel, Huber, Rogers and Tukey (1972) and Shorack and Wellner (1986)]. Similar arguments were used by

Received March 1988; revised October 1989.
AMS 1980 *subject classifications*. Primary 62J05; secondary 62F12, 62F35.
Key words and phrases. S-estimators, linear regression, least median of squares, consistency, weak convergence, asymptotic normality.

Rousseeuw (1984) who considered the asymptotic behaviour of the least median of squares for i.i.d. carriers. He also obtained an $n^{-1/3}$ rate of convergence and identified the limiting distribution.

In this paper we show that the heuristic arguments give the correct answer for middle of the shortest half and we also extend the result to least median of squares. Kim and Pollard (1989) have independently given a correct proof of the asymptotic behaviour of the least median of squares estimator. It is one example of their general approach to $n^{-1/3}$ rates of convergence. However, they also consider i.i.d. carriers. The first version of the present paper was written without the benefit of Kim and Pollard. Subsequent versions were written with the benefit of Kim and Pollard and this has resulted in a considerable improvement as anyone who was unfortunate enough to read the first version, in particular the referees, can confirm. The improvements concern mostly a shortening of the proofs but one substantial improvement in Theorem 3 resulted.

In spite of its poor rate of convergence, least median of squares would seem to be the most appropriate estimator of those currently available for obtaining robust estimates in the sense of high finite sample breakdown point. The reason for this is that S-estimators can only be calculated by brute force, usually using some version of the random search method. In practice this is most easily implemented for the least median of squares estimates.

This paper is organized as follows. In Section 2 we consider the question of consistency for a sequence of regression models. Section 3 is devoted to the asymptotic behaviour of least median of squares and Section 4 to the asymptotic behaviour of S-estimators with a smooth ρ-function (definition below).

In the following we shall not distinguish between sets and their indicator functions. The expectation operator will be denoted by $\mathbb{E}$ and the variance operator by $\mathbb{V}$.

2. Consistency. We consider a sequence of regression models

$$y_i(n) = x_i(n)^T\beta(n) + \varepsilon_i(n), \qquad 1 \le i \le n,$$

where $y(n) \in \mathbb{R}^n$, $x_i(n) \in \mathbb{R}^{k_n}$, $1 \le i \le n$, $\beta(n) \in \mathbb{R}^{k_n}$ and the $\varepsilon_i(n)$, $1 \le i \le n$, are independently and identically distributed random variables. We shall assume that the random variables $\varepsilon_i(n)$ are defined on a common complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Elements of Ω will be denoted by ω. We write

$$\mathbb{P}(\varepsilon_1(n) \le u) = F(u/\sigma), \qquad u \in \mathbb{R},$$

where $\sigma \in \mathbb{R}^+ = (0, \infty)$ is a scale parameter and $F\colon \mathbb{R} \to [0,1]$ a nondegenerate distribution function.

Throughout we consider fixed carriers $x_i(n)$, $1 \le i \le n$. The case of random carriers $x_i(n)$ which are independent of the errors $\varepsilon_i(n)$ may be covered by conditioning on the $x_i(n)$ and then checking whether the assumptions placed on the carriers hold almost surely.

To define S-estimators we introduce a function ρ: $\mathbb{R} \to [0,1]$ with the following properties:

R1. (a) $\rho(0) = 1$.
(b) $\rho(u) = \rho(-u)$, $u \in \mathbb{R}$.
(c) ρ: $\mathbb{R}_+ \to [0,1]$ is nonincreasing, continuous at 0 and continuous on the left.
(d) For some $e > 0$, $\rho(u) > 0$ if $|u| < e$ and $\rho(u) = 0$ if $|u| > e$.

For any $y \in \mathbb{R}$ and $s \in \mathbb{R}_+$, we define

$$R(y,s) = \int \rho\left(\frac{u-y}{s}\right) dF(u) \quad \{s > 0\}. \tag{1}$$

Given ε, $0 < \varepsilon < 1$, we choose, as is always possible, a function ρ with the above properties such that

$$R(0,1) = 1 - \varepsilon. \tag{2}$$

This ε has nothing to do with the errors $\varepsilon_i(n)$. It is used as it is standard notation for the breakdown point of an estimator. No confusion should arise.

The S-estimator $(\hat{b}(n), \hat{s}(n))$ of $(\beta(n), \sigma)$ is now defined to be a solution of the following problem: Choose $b \in \mathbb{R}^{k_n}$ and $s \in \mathbb{R}_+$ so as to minimize s subject to

$$\frac{1}{n}\sum_{1}^{n} \rho\left(\frac{y_i(n) - x_i(n)^T b}{s}\right) \geq 1 - \varepsilon. \tag{3}$$

We denote this problem by $\hat{\mathscr{B}}_n$.

Although we are not explicitly concerned with this, the breakdown point of the S-estimator defined above is $\varepsilon^* = \min(\varepsilon, 1-\varepsilon)$. Having chosen ε the statistician may ensure the Fisher consistency of the estimates by choosing a function ρ as above. This may, for example, be the function

$$\rho(u) = \left(1 - \frac{u^2}{e^2}\right)^2 \qquad \{|u| \leq e\}$$

where the cutoff point e is adjusted to guarantee $R(0,1) = 1 - \varepsilon$. We assume that F is known, the standard choice being the standard normal distribution.

THEOREM 1. *If* R1 *holds, F is continuous and $k_n < [(1-\varepsilon)n]$, then with probability one, $\hat{\mathscr{B}}_n$ has at least one solution $(\hat{b}(n), \hat{s}(n))$ and $\hat{s}(n) > 0$. Furthermore, $\hat{b}(n)$ and $\hat{s}(n)$ can be chosen to be random variables.*

PROOF. For $\alpha \in \mathbb{R}^n$, $b \in \mathbb{R}^{k_n}$ and $s \in \mathbb{R}_+$, we write

$$g(\alpha, b, s) = \frac{1}{n}\sum_{1}^{n} \rho\left(\frac{\alpha_i - x_i(n)^T b}{s}\right) \qquad \{s > 0\},$$

and $s(\alpha, b) = \inf\{s: g(\alpha, b, s) \geq 1 - \varepsilon\}$.

As $\lim_{s\to\infty} g(\alpha, b, s) = 1$ and $\lim_{s\to 0} g(\alpha, b, s) = 0$ because of R1, it follows that $s(\alpha, b)$ is well-defined and finite. Further, as $g(\alpha, b, s)$ is, for fixed α and b, a right-continuous nondecreasing function of s, it follows that $s(\alpha, b)$ is measurable with respect to the Borel σ-algebra on $\mathbb{R}^{n+k_n}$. In particular, the class of functions

$$\mathscr{S} = \{s(\cdot, b) : b \in \mathbb{R}^{k_n}\}$$

is permissible [see Appendix C of Pollard (1984)].

For fixed α we define $s(\alpha) = \inf_b s(\alpha, b)$.

Let $(s_m(\alpha))_1^\infty$ be a decreasing sequence with $\lim_{m\to\infty} s_m(\alpha) = s(\alpha)$ and for each m, let $b_m(\alpha)$ be such that $g(\alpha, b_m(\alpha), s_m(\alpha)) \geq 1 - \varepsilon$. Consider now a fixed m. There exist an n_0, $[n(1-\varepsilon)] \leq n_0 \leq n$ and real numbers γ_{i_j}, $1 \leq j \leq n_0$, with $|\gamma_{i_j}| \leq e$, $1 \leq j \leq n_0$, $|\alpha_i - x_i(n)^T b_m(\alpha)| > e$ for all $i \notin \{i_1, \dots, i_{n_0}\}$ and

$$\alpha_{i_j} - x_{i_j}(n)^T b_m(\alpha) = \gamma_{i_j}, \qquad 1 \leq j \leq n_0. \tag{4}$$

We denote by $b_m^*(\alpha)$ that b which satisfies (4) and which minimizes $\|b\|$. It is seen that $s_m^*(\alpha) = s(\alpha, b_m^*(\alpha)) \leq s_m(\alpha)$ and $\sup_m \|b_m^*(\alpha)\| < \infty$. Thus there exists a subsequence $(m')_1^\infty$ such that $\lim_{m'\to\infty}(b_{m'}^*(\alpha), s_{m'}^*(\alpha)) = (b'(\alpha), s(\alpha))$. As $\lim_{u'\to u} \sup \rho(u') \leq \rho(u)$, it follows that $g(\alpha, b'(\alpha), s(\alpha)) \geq 1 - \varepsilon$ and hence $(b'(\alpha), s(\alpha))$ is a solution of $\hat{\mathscr{B}}_n$.

We now prove that measurable versions can be chosen. Let $\varepsilon(n, \omega)^T = (\varepsilon_1(n, \omega), \dots, \varepsilon_n(n, \omega))$. Then on setting $\alpha = \varepsilon(n, \omega)$ and using the fact that the class $\mathscr{S}$ above is permissible, it follows that $\hat{s}(n, \omega) = \inf_b s(\varepsilon(n, \omega), b)$ is measurable [Appendix C of Pollard (1984)]. Consider the set $\{(\omega, b): s(\varepsilon(n, \omega), b) = \hat{s}(n, \omega)\}$. The projection of this set onto Ω is Ω itself as we have a solution of $\hat{\mathscr{B}}_n$ for each ω. It now follows from the measurable cross-section theorem in Appendix C of Pollard (1984) that there exists a measurable function $\hat{b}(n, \omega)$ such that $s(\varepsilon(n, \omega), \hat{b}(n, \omega)) = \hat{s}(n, \omega)$ for almost all ω. This proves the theorem. □

The S-estimator $(\hat{b}(n), \hat{s}(n))$ is affine equivariant and consequently, without loss of generality, we may and shall assume that $\beta(n) = 0$ and $\sigma = 1$.

We now turn to the corresponding theoretical problem $\mathscr{B}_n$ which is the following: choose $b(n)$ in $\mathbb{R}^{k_n}$ and $s(n)$ in $\mathbb{R}_+$ so as to minimize $s(n)$ subject to

$$\frac{1}{n} \sum_1^n R\big(x_i(n)^T b(n), s(n)\big) \geq 1 - \varepsilon. \tag{5}$$

To obtain a unique solution we introduce the following conditions.

F1. (a) F has a bounded density f.
(b) $f(u) = f(-u)$, $u \in \mathbb{R}$.
(c) $f: \mathbb{R}_+ \to \mathbb{R}_+$ is nonincreasing.

FR1. $f\colon \mathbb{R}_+ \to \mathbb{R}_+$ and $\rho\colon \mathbb{R}_+ \to \mathbb{R}_+$ have a common point of decrease.

LEMMA 1. *Suppose* F1 *and* R1 *hold. Then*

(i) $R\colon \mathbb{R} \times \mathbb{R}_+ \to [0,1]$ *is continuous.*

If in addition, FR1 *holds, then*

(ii) $R(u,r) \le R(0,r)$ *for* $u \in \mathbb{R}$, $r \in \mathbb{R}_+$.
(iii) $\sup_{|u|>\eta} R(u,1) < R(0,1)$ *for all* $\eta > 0$.
(iv) $R(0,r') < R(0,1) < R(0,r'')$ *if* $0 \le r' < 1 < r''$.

PROOF. (i) may be proved, for example, by using dominated convergence and the fact that ρ has at most countably many discontinuities.

To prove (ii)–(iv) we note that

$$\int \left(\rho\left(\frac{y-u}{s}\right) - \rho\left(\frac{y}{s}\right)\right)(f(y-u) - f(y))\,dy \ge 0,$$

for all u with strict inequality if $s = 1$ and $u \ne 0$. On multiplying out and using the symmetry about zero, the claims of the lemma follow. □

THEOREM 2. *If* F1, R1 *and* FR1 *hold, then* $\mathscr{B}_n$ *has the unique solution*

$$(b(n), s(n)) = (0,1).$$

PROOF. This follows from Lemma (ii) and (iii) on noting that each summand in (5) is individually maximized by setting $b = 0$. □

We now show that the S-estimator $(\hat{b}(n), \hat{s}(n))$ tends in probability to $(0,1)$.

To do this and in the following, we make extensive use of results for empirical processes as to be found, for example, in Pollard (1984). Most of these results are formulated for i.i.d. random variables and as we consider fixed but otherwise arbitrary carriers $x_i(n)$, they are not directly applicable to the empirical processes occurring in this paper. Although the necessary adjustments are straightforward, we give an explicit proof in one particular case, Lemma 2 below, to demonstrate the nature of the adjustments.

In the following, we shall denote positive constants whose actual value is of no importance by $c_1, \dots, c_{18}$.

LEMMA 2. *Let*

$$Z_n(b,s) = \frac{1}{n}\sum_{1}^{n} \left(\left\{|\varepsilon_i(n) + x_i(n)^T b| \le s\right\} - R\left(x_i(n)^T b, s\right)\right).$$

Then there exists a constant c_1 *such that*

$$\mathbb{P}\left(\sup_{b,s}|Z_n(b,s)| \ge 8\eta\right) \le c_1 n^{k_n+2}\exp\left(-\tfrac{1}{2}n\eta^2\right),$$

for all $\eta > 0$.

PROOF. We write $Z_n(b,s)$ in the form

$$Z_n(b,s) = \frac{1}{n}\sum_1^n (E_{i,n}(b,s) - \mathbb{P}(E_{i,n}(b,s))),$$

where $E_{i,n}(b,s)$ denotes the indicator function of the set $\{|\varepsilon_i(n) + x_i(n)^T b| \le s\}$. We have $\mathbb{E}(Z_n(b,s)) = 0$ and $\mathbb{V}(Z_n(b,s)) \le 1/(4n)$. From the symmetrization lemma of Pollard [(1984), page 14] we obtain

$$\mathbb{P}\Big(\sup_{b,s}|Z_n(b,s)| \ge 8\eta\Big) \le 2\mathbb{P}\Big(\sup_{b,s}|Z_n(b,s) - Z_n'(b,s)| \le 4\eta\Big),$$

for all $n \ge 1/(8\eta^2)$, where $Z_n'(b,s) = (1/n)\Sigma_1^n(E_{i,n}'(b,s) - \mathbb{P}(E_{i,n}'(b,s)))$, $E_{i,n}'(b,s) = \{|\varepsilon_i'(n) + x_i(n)^T b| \le s\}$ and the $\varepsilon_i'(n)$, $1 \le i \le n$, have the same distribution as (but are independent of) the $\varepsilon_i(n)$, $1 \le i \le n$.

Let δ_i, $1 \le i \le n$, be independent of the $\varepsilon_i(n)$ and $\varepsilon_i'(n)$, $1 \le i \le n$, and satisfy $\mathbb{P}(\delta_i = 1) = \mathbb{P}(\delta_i = -1) = \frac{1}{2}$ and define $Z_n^0(b,s) = (1/n)\Sigma_1^n \delta_i E_{i,n}(b,s)$. Then as in Pollard [(1984), page 15], we obtain

$$\mathbb{P}\Big(\sup_{b,s}|Z_n(b,s)| \ge 8\eta\Big) \le 4\mathbb{P}\Big(\sup_{b,s}|Z_n^0(b,s)| \ge 2\eta\Big),$$

for all $n \ge 1/(8\eta^2)$. For fixed $(\varepsilon_i(n), x_i(n))$, $1 \le i \le n$, consider now $Z_n^0(b,s) = (1/n)\Sigma_1^n \delta_i E_{i,n}(b,s)$.

From Hoeffding's inequality [Pollard (1984), page 192] we obtain

$$\mathbb{P}\big(|Z_n^0(b,s)| \ge 2\eta | (\varepsilon_i(n), x_i(n)), 1 \le i \le n\big) \le 2\exp(-2n\eta^2).$$

This implies

$$\mathbb{P}\Big(\sup_{b,s}|Z_n(b,s)| \ge 8\eta\Big) \le 8\exp(-2n\eta^2)\mathbb{E}(M_n),$$

where M_n denotes the number of different points $(E_{1,n}(b,s),\dots,E_{n,n}(b,s))$ as (b,s) ranges over $\mathbb{R}^{k_n} \times \mathbb{R}_+$. This is equal to the number of different subsets of the n point set $\{(\varepsilon_i(n), x_i(n))\colon 1 \le i \le n\}$ picked out by sets of the form $\{(\varepsilon, x)\colon |\varepsilon - x^T b| \le s\}$. For fixed (b,s) and $\alpha \in \mathbb{R}$, we define $g_{b,s,\alpha}(\varepsilon, x) = \alpha\varepsilon + x^T b + s$. This is a linear space of dimension $k_n + 2$ and hence the sets $\{(\varepsilon, x)\colon |\varepsilon - x^T b| \le s\}$ have polynomial discrimination of order at most $k_n + 2$ [Pollard (1984), Lemma 18, page 20 and Lemma 15, page 18]. It follows that $M_n \le c_2 n^{k_n+2}$ and this gives

$$\mathbb{P}\Big(\sup_{b,s}|Z_n(b,s)| \ge 8\eta\Big) \le c_3 n^{k_n+2}\exp(-2n\eta^2),$$

for all $n \ge 1/(8\eta^2)$. For $n \le 1/(8\eta^2)$, we have $\exp(-2n\eta^2) \ge \exp(-\frac{1}{4})$ and hence, by adjusting the constant c_3 if necessary, we obtain the claim of the lemma. □

To obtain consistency we must, however, place a weak condition on the carriers $x_i(n)$, $1 \le i \le n$. We denote the number of elements of the set $\{\cdots\}$

by $|\{\cdots\}|$. With this notation we define $\lambda_n(\alpha)$ for any α, $0 < \alpha < 1$, by

$$\lambda_n(\alpha) = \min_{\substack{\mathscr{S} \subset \{1,\dots,n\} \\ |\mathscr{S}| = [n\alpha]}} \left(\min_{\substack{\theta \\ \|\theta\|=1}} \left(\max_{i \in \mathscr{S}} |x_i(n)^T \theta| \right) \right).$$

Let $X(n)$ denote the $n \times k_n$ design matrix with ith row $x_i(n)^T$. Then a sufficient condition for the least squares estimate to be weakly consistent is that $X(n)^T X(n)$ be nonsingular and $\lim_{n\to\infty}(X(n)^T X(n))^{-1} = 0$. This condition is not sufficient for S-estimators as the following can occur.

EXAMPLE 1. We set $k_n = 1$, $x_i(n) = 0$, $1 \le i \le n-3$, $x_{n-2}(n) = x_{n-1}(n) = n^{-2}$, $x_n(n) = n$. The least squares estimate is consistent because of $x_n(n)$ but an S-estimator may declare $x_n(n)$ to be an outlier and concentrate on the $x_i(n)$, $1 \le i \le n-1$ and these carriers are badly conditioned.

The function $\lambda_n(\alpha)$ measures in some sense the worst possible conditioning of any $[n\alpha]$ subset of the carriers.

THEOREM 3. *Suppose*

$$k_n < [n(1-\varepsilon)], \qquad \lim_{n\to\infty} \frac{k_n \log n}{n} = 0$$

and that F1, R1 *and* FR1 *hold. Then*

$$\operatorname*{plim}_{n\to\infty} \left(\lambda_n(\alpha)\|\hat{b}(n)\|, \hat{s}(n)\right) = (0,1),$$

for each α, $0 < \alpha < 1$.

PROOF. First, we show that

$$(6) \quad \operatorname*{plim}_{n\to\infty} \sup_{\substack{b \in \mathbb{R}^{k_n} \\ 0 \le s \le 2}} \left| \frac{1}{n} \sum_1^n \left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b}{s} \right) - R\left(x_i(n)^T b, s\right) \right) \right| = 0.$$

Because of R1, ρ may be uniformly approximated by functions of the form $\sum_1^m a_j \{|u| \le b_j\}$. It is therefore sufficient to consider the case $\rho(u) = \{|u| \le 1\}$.
Let

$$Z(b,s) = \frac{1}{n} \sum_1^n \left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b}{s} \right) - R\left(x_i(n)^T b, s\right) \right),$$

as in Lemma 2.

From Lemma 2 and the assumption $\lim_{n\to\infty}(k_n \log n)/n = 0$, we may deduce

$$\operatorname*{plim}_{n\to\infty} \sup_{\substack{b \in \mathbb{R}^{k_n} \\ 0 \le s \le 2}} \left| \frac{1}{n} \sum_1^n \left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b}{s} \right) - R\left(x_i(n)^T b, s\right) \right) \right| = 0,$$

for $\rho(u) = \{|u| \le 1\}$ and this implies (6).

The law of large numbers shows that for any $\eta > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\sum_{1}^{n} \rho\left(\frac{\varepsilon_i(n)}{1+\eta}\right) > 1 - \varepsilon\right) = 1,$$

where we have used Lemma 1(iii). This implies $\lim_{n\to\infty} \mathbb{P}(\hat{s}(n) \le 1 + \eta) = 1$, for each $\eta > 0$. We may therefore conclude

$$\operatorname*{plim}_{n\to\infty} \left|\frac{1}{n}\sum_{1}^{n} \left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T\hat{b}(n)}{s(n)}\right) - R\big(x_i(n)^T\hat{b}(n), \hat{s}(n)\big)\right)\right| = 0,$$

and hence, as $R(0, \hat{s}(n)) \ge R(u, \hat{s}(n))$ for all u [Lemma 1(ii)],

$$\lim_{n\to\infty} \mathbb{P}(R(0, \hat{s}(n)) \ge 1 - \varepsilon - \eta) = 1,$$

for all $\eta > 0$. This, together with Lemma 1(iii) and $\lim_{n\to\infty} \mathbb{P}(\hat{s}(n) \le 1 + \eta) = 1$ for all $\eta > 0$, implies $\operatorname{plim}_{n\to\infty} \hat{s}(n) = 1$.

For any $\eta > 0$, let N be the number of $x_i(n)$ with $|x_i(n)^T\hat{b}(n)| \ge \eta$. From Lemma 1(ii) we may conclude that for any α, $0 < \alpha < 1$,

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(N \ge (1-\alpha)n) \\
&\quad \le \lim_{n\to\infty} \mathbb{P}(\alpha R(0, \hat{s}(n)) + (1-\alpha)(R(0, \hat{s}(n)) - \delta) \ge 1 - \varepsilon) \\
&\quad \le \lim_{n\to\infty} \mathbb{P}(R(0, \hat{s}(n)) \ge 1 - \varepsilon + \delta(1-\alpha)),
\end{aligned}$$

for some $\delta > 0$. Thus $\lim_{n\to\infty} \mathbb{P}(N \ge (1-\alpha)n) = 0$ for each α, $0 < \alpha < 1$, and from the definition of $\lambda_n(\alpha)$, we obtain $\operatorname{plim}_{n\to\infty}(\lambda_n(\alpha)\|\hat{b}(n)\|) = 0$, proving the theorem. □

EXAMPLE 2. We suppose that $k_n = k \ge 1$ and that the $x_i(n)$, $1 \le i \le n$, are independently and identically distributed. If the $x_i(n)$, $1 \le i \le n$, satisfy

$$\sup_{\theta, \|\theta\|=1} \mathbb{P}\big(x_i(n)^T\theta = 0\big) < 1,$$

then there exists an $\eta > 0$ such that

$$\sup_{\theta, \|\theta\|=1} \mathbb{P}\big(|x_i(n)^T\theta| < \eta\big) \le 1 - \eta.$$

The sets $B(\theta) = \{x\colon |x^T\theta| < \eta\}$ have polynomial discrimination [Pollard (1984), page 17] and consequently [Pollard (1984), Theorem 14, page 18]

$$\operatorname*{plim}_{n\to\infty} \sup_{\theta, \|\theta\|=1} \left|\frac{1}{n}\sum_{1}^{n} \Big(\{|x_i(n)^T\theta| < \eta\} - \mathbb{P}\big(|x_i(n)^T\theta| < \eta\big)\Big)\right| = 0.$$

If α, $0 < \alpha < 1$, is such that $\alpha > 1 - \eta$, it follows that for any θ, $\|\theta\| = 1$, and any subset $\mathscr{S}$ of $\{1, \dots, n\}$ of size $[n\alpha]$, there exists an $i \in \mathscr{S}$ with $|x_i(n)^T\theta| \ge \eta$. Thus for all sufficiently large n, we have $\lambda_n(\alpha) \ge \eta > 0$. If the $\varepsilon_i(n)$ have a common density function strictly decreasing on $[0, \infty)$, for example the normal distribution, then the conditions of Theorem 3 are satisfied and we have $p\lim_{n\to\infty}(\hat{b}(n), \hat{s}(n)) = (0,1)$. This generalizes Theorem 2 of Rousseeuw and Yohai (1984).

EXAMPLE 3. Consider $k_n = 2$ and a linear trend $x_i(n)^T = (1, i)$, $1 \le i \le n$. We redefine the $x_i(n)$ and set $\tilde{x}_i(n)^T = (1, i/n)$, $1 \le i \le n$. Then for any θ with $\|\theta\| = 1$ and any subset $\mathscr{S}$ of $\{1, \dots, n\}$ of size $[n\alpha]$, we have

$$\max_{i\in\mathscr{S}} |\tilde{x}_i(n)^T\theta| = \max_{i\in\mathscr{S}} \left|\theta_1 + \frac{i\theta_2}{n}\right| \ge \frac{\alpha}{5}.$$

This gives $\tilde{\lambda}_n(\alpha) \ge \alpha/5$. If now the density function f is strictly decreasing on $[0,\infty)$, we obtain from Theorem 3 $p\lim_{n\to\infty} \tilde{b}(n) = 0$ or, in terms of the original S-estimator $\hat{b}(n)$,

$$\operatorname*{plim}_{n\to\infty}\left(\hat{b}_1(n), n\hat{b}_2(n)\right) = 0.$$

This result may be extended to a polynomial trend of any fixed order.

3. Least median of squares. In this section we restrict ourselves to a fixed dimension k. We assume as before that $\beta(n) \equiv 0$ and $\sigma = 1$.

We shall suppose

$$\sum_1^n x_i(n)x_i(n)^T = nI_k, \tag{7}$$

which implies

$$\sum_1^n \|x_i(n)\|^2 = nk. \tag{8}$$

This represents no loss of generality as it may be obtained by an appropriate transformation of the $x_i(n)$, $1 \le i \le n$, and $\beta(n)$.

The Hampel–Rousseeuw least median of squares estimator may be defined as that b which minimizes the median of the squared residuals $(y_i - x_i(n)^T b)^2$. In our notation this is equivalent to minimizing s subject to

$$\frac{1}{n}\sum_1^n \rho\left(\frac{y_i - x_i(n)^T b}{s}\right) \ge \frac{1}{n}\left(\left[\frac{n}{2}\right] + 1\right),$$

where ρ is the indicator function of some interval $[-e, e]$.

We shall consider the slightly more general case of minimizing s subject to

$$\frac{1}{n}\sum_1^n \rho\left(\frac{y_i - x_i(n)^T b}{s}\right) \ge 1 - \varepsilon,$$

and, abusing the meaning of the word median, still refer to this as the least median of squares estimator.

The conditions of Theorem 3 will be assumed to hold and, in addition, $\liminf \lambda_n(\alpha) > 0$ for some α, $0 < \alpha < 1$. It then follows that the least median of squares estimators $\hat{b}(n)$ and $\hat{s}(n)$ are consistent.

We first show that $\liminf \lambda_n(\alpha) > 0$ for some α, $0 < \alpha < 1$, is equivalent to the following assumption:

D1. There exist positive numbers η_1, η_2 and n_0 such that

$$\sum_1^n x_i(n) x_i(n)^T \{\|x_i(n)\| < \eta_1\} - n\eta_2 I_k$$

is positive definite for all $n \geq n_0$.

LEMMA 3. *Suppose* (7) *holds. Then* $\liminf \lambda_n(\alpha) > 0$ *for some* α, $0 < \alpha < 1$, *if and only if* D1 *holds.*

PROOF. Suppose D1 holds and for any α, $0 < \alpha < 1$. Let $\mathscr{S}_n(\alpha)$ be a subset of $\{1, \dots, n\}$ of size $[n\alpha]$ such that

$$\lambda_n(\alpha) = \min_{\|\theta\|=1} \left(\max_{i \in \mathscr{S}_n(\alpha)} |x_i(n)^T \theta| \right).$$

Then as

$$\sum_{i \in \mathscr{S}_n(\alpha)} |x_i(n)^T \theta|^2 \leq n\alpha \left(\max_{i \in \mathscr{S}_n(\alpha)} |x_i(n)^T \theta|^2 \right),$$

we have

$$\begin{aligned} n\alpha\lambda_n(\alpha)^2 &\geq \min_{\|\theta\|=1} \left(\sum_{\substack{i \in \mathscr{S}_n(\alpha) \\ \|x_i(n)\| < \eta_1}} |x_i(n)^T \theta|^2 \right) \\ &= \min_{\|\theta\|=1} \left(\sum_{\|x_i(n)\| < \eta_1} |x_i(n)^T \theta|^2 - \sum_{\substack{i \notin \mathscr{S}_n(\alpha) \\ \|x_i(n)\| < \eta_1}} |x_i(n)^T \theta|^2 \right) \\ &\geq n\eta_2 - (n - [n\alpha])\eta_1^2, \end{aligned}$$

where we have used D1. If α is such that $(1-\alpha)\eta_1^2 < \frac{1}{2}\eta_2$, then we obtain $\liminf \lambda_n(\alpha) > 0$.

In the other direction let α, $0 < \alpha < 1$, be such that $\liminf \lambda_n(\alpha) = \lambda > 0$. From (8) it follows that $|\{i\colon \|x_i(n)\| \geq \eta\}| \leq nk^2/\eta^2$. Thus there exist η_1 and α_1, $0 < \alpha < \alpha_1 < 1$, such that $|\mathscr{I}_n| \geq n\alpha_1$, where $\mathscr{I}_n = \{i\colon \|x_i(n)\| \leq \eta_1\}$.

Let θ' with $\|\theta'\| = 1$ be such that

$$\frac{1}{2} \sum_{i \in \mathscr{I}_n} |x_i(n)^T \theta'|^2 = \min_{\|\theta\|=1} \frac{1}{2} \sum_{i \in \mathscr{I}_n} |x_i(n)^T \theta|^2,$$

and let $\mathscr{S}_n(\alpha)$ consist of those i's in $\mathscr{I}_n$ giving rise to the smallest $[n\alpha]$ of the $|x_i(n)^T \theta'|^2$.

The definition of $\lambda_n(\alpha)$ gives

$$\lambda_n(\alpha) \le \max_{i \in \mathscr{G}_n(\alpha)} |x_i(n)^T \theta'|$$

and hence $|x_i(n)^T\theta'| \ge \lambda_n(\alpha)$, for all $i \in \mathscr{I}_n \setminus \mathscr{G}_n(\alpha)$. This implies

$$\min_{\|\theta\|=1} \frac{1}{n} \sum_{i \in \mathscr{I}_n} |x_i(n)^T\theta|^2 \ge (\alpha_1 - \alpha)\lambda_n(\alpha)^2$$

and hence the desired result. □

In order to obtain weak convergence results we require a maximal inequality as given, for example, in Kim and Pollard (1989) and taken from Pollard (1989). We state it here using our notation.

Maximal inequality. Let $\mathscr{H}$ be a manageable class of functions with an envelope H for which $\mathbb{E}(H^2) < \infty$. Suppose $0 \in \mathscr{H}$. Then there exists a function J, not depending on n, such that

(i) $$\sqrt{n}\,\mathbb{E}\Big(\sup_{\mathscr{H}} |\mathbb{P}_n h - \mathbb{E}(h)|\Big) \le \mathbb{E}\Big(\sqrt{\mathbb{P}_n(H^2)}\, J\Big(\sup_{\mathscr{H}} \mathbb{P}_n(h^2)/\mathbb{P}_n(H^2)\Big)\Big) \le J(1)\sqrt{\mathbb{E}(H^2)}\,.$$

(ii) $$n\mathbb{E}\Big(\sup_{\mathscr{H}} |\mathbb{P}_n h - \mathbb{E}(h)|^2\Big) \le \mathbb{E}\Big(\mathbb{P}_n(H^2) J\Big(\sup_{\mathscr{H}} \mathbb{P}_n(h^2)/\mathbb{P}_n(H^2)\Big)^2\Big) \le J(1)\mathbb{E}(H^2).$$

The function J is continuous and increasing with $J(0) = 0$ and $J(1) < \infty$.

The above inequalities are stated for i.i.d. random variables and are therefore not directly applicable to the present situation. Relevant inequalities may be obtained as follows. The functions h are assumed to be defined on $\mathbb{R}^{1+k}$ and $\mathbb{P}_n$ is taken to be the empirical measure obtained by placing a mass of $1/n$ at each of the points $(\varepsilon_i(n), x_i(n))$, $1 \le i \le n$. This gives

$$\mathbb{P}_n(h) = \frac{1}{n}\sum_1^n h(\varepsilon_i(n), x_i(n)).$$

The expression $\mathbb{E}(h)$ must be replaced by

$$\frac{1}{n}\sum_1^n \mathbb{E}(h(\varepsilon_i(n), x_i(n))).$$

The $x_i(n)$, $1 \le i \le n$, may be interpreted as being fixed or, alternatively, the $(\varepsilon_i(n), x_i(n))$, $1 \le i \le n$, may be interpreted as being independent random variables, the $\varepsilon_i(n)$ having a common distribution and the $x_i(n)$ being degenerate random variables.

In order to obtain the maximal inequalities given above, the process

$$Z_n(h) = \frac{1}{n}\sum_1^n h(\varepsilon_i(n), x_i(n)) - \frac{1}{n}\sum_1^n \mathbb{E}(h(\varepsilon_i(n), x_i(n)))$$

is replaced by the symmetrized process

$$Z_n^0(h) = \frac{1}{n}\sum_1^n \pm h(\varepsilon_i(n), x_i(n)),$$

just as in the proof of Lemma 2. This step makes use only of the independence of the random variables. It is not necessary for them to be identically distributed. The remainder of the proof is independent of any distributional properties of the $(\varepsilon_i(n), x_i(n))$, $1 \le i \le n$. It requires only that the class $\mathscr{H}$ of functions $h: \mathbb{R}^{1+k} \to \mathbb{R}$ be manageable.

We shall have reason to apply the maximal inequality to subclasses $\mathscr{H}'$ of some class $\mathscr{H}$. It will be then of importance that the same function J can still be used independently of the subclass. In Kim and Pollard (1989) and Pollard (1989) such classes of functions are called uniformly manageable. This will, in particular, be the case if the graphs of the functions h,

$$G_h = \{(\varepsilon, x, t): 0 \le t \le h(\varepsilon, x) \text{ or } h(\varepsilon, x) \le t \le 0\},$$

have polynomial discrimination. This will be the case in most of our applications, the others being covered by the fact that if $\mathscr{H}$ is uniformly manageable, as is

$$\tilde{\mathscr{H}} = \{h_1 - h_2: h_1, h_2 \in \mathscr{H}\}$$

[see Pollard (1989) and Kim and Pollard (1989)].

We now consider the class

$$\mathscr{H} = \{h_{b,s}: b \in \mathbb{R}^k, 0 \le s \le 2\},$$

where

$$h_{b,s}(\varepsilon, x) = \{|\varepsilon - x^T b| \le s\}.$$

The family of graphs of these functions has polynomial discrimination and the class $\mathscr{H}$ is therefore uniformly manageable.

THEOREM 4. *Suppose* F1, F2, FR1 *and* D1 *hold and that f is continuous at e. Then*

$$\sqrt{n}\,(\hat{s}(n) - 1) \Rightarrow N\left(0, \frac{\varepsilon(1-\varepsilon)}{4 f(e)^2}\right).$$

PROOF. We consider

$$Z_n(b,s) = \frac{1}{\sqrt{n}} \sum_1^n \left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b}{s} \right) - \rho\left(\frac{\varepsilon_i(n)}{s} \right) \right.$$

$$\left. - \Big(R\big(x_i(n)^T b, s\big) - R(0,s) \Big) \right).$$

The maximal inequality as given above with $H \equiv 1$ implies

$$\mathbb{E}\Big(\sup_{\|b\|<\delta,\,|s|<2} |Z_n(b,s)| \Big)$$

$$\leq 2\mathbb{E}\left(J\left(\sup_{\|b\|<\delta,\,|s|\leq 2} \frac{1}{n} \sum_1^n \left| \rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b}{s} \right) - \rho\left(\frac{\varepsilon_i(n)}{s} \right) \right| \right) \right).$$

As the

$$\left| \rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b}{s} \right) - \rho\left(\frac{\varepsilon_i(n)}{s} \right) \right|$$

are indicator functions of sets with polynomial discrimination, it follows, as in the proof of Lemma 2, that

$$\operatorname*{plim}_{n\to\infty} \sup_{\|b\|<\delta,\,|s|\leq 2} \left| \frac{1}{n} \sum_1^n \left| \rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b}{s} \right) - \rho\left(\frac{\varepsilon_i(n)}{s} \right) \right| \right.$$

$$\left. - \mathbb{E}\left(\left| \rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b}{s} \right) - \rho\left(\frac{\varepsilon_i(n)}{s} \right) \right| \right) \right| = 0.$$

This together with

$$\frac{1}{n} \sum_1^n \mathbb{E}\left(\left| \rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b}{s} \right) - \rho\left(\frac{\varepsilon_i(n)}{s} \right) \right| \right) \leq c_4 \frac{1}{n} \sum_1^n \|x_i(n)\| \|b\| \leq c_5 \|b\|$$

yields

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \mathbb{E}\Big(\sup_{\|b\|<\delta,\,|s|<2} |Z_n(b,s)| \Big) = 0. \tag{9}$$

It follows from D1, Lemma 3 and Theorem 3 that $\operatorname{plim}_{n\to\infty}\hat{b}(n) = 0$ and hence

$$
\begin{aligned}
1-\varepsilon &\le \frac{1}{n}\sum_{1}^{n}\rho\left(\frac{\varepsilon_i(n)-x_i(n)^T\hat{b}(n)}{\hat{s}(n)}\right)\\
&= \frac{1}{n}\sum_{1}^{n}\rho\left(\frac{\varepsilon_i(n)}{\hat{s}(n)}\right)+\frac{1}{n}\sum_{1}^{n}\Big(R\big(x_i(n)^T\hat{b}(n),\hat{s}(n)\big)\Big)\\
&\quad - R(0,\hat{s}(n)) + o_p\left(\frac{1}{\sqrt{n}}\right)\\
&\le \frac{1}{n}\sum_{1}^{n}\rho\left(\frac{\varepsilon_i(n)}{\hat{s}(n)}\right)+o_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
\tag{10}
$$

by Lemma 1.

We have

$$
\frac{1}{\sqrt{n}}\sum_{1}^{n}\rho\left(\frac{\varepsilon_i(n)}{s}\right)-(1-\varepsilon)=\frac{1}{\sqrt{n}}\sum_{1}^{n}\big(\rho(\varepsilon_i(n))-R(0,1)\big)+o_p(1),
\tag{11}
$$

for small $|s-1|$. This follows directly from the weak convergence of the empirical distribution function of the $(\varepsilon_i(n))_1^n$ to a continuous Brownian bridge. Alternatively, on writing

$$
Z_n'(s)=\frac{1}{\sqrt{n}}\sum_{1}^{n}\left(\rho\left(\frac{\varepsilon_i(n)}{s}\right)-\rho(\varepsilon_i(n))-R(0,s)+R(0,1)\right),
$$

an argument similar to the one leading to (9) will yield (11). As $\operatorname{plim}_{n\to\infty}\hat{s}(n) = 1$, we may conclude after some manipulation

$$
2f(e)(\hat{s}(n)-1)\ge -\frac{1}{n}\sum_{1}^{n}\big(\rho(\varepsilon_i(n))-(1-\varepsilon)\big)+o_p\left(\frac{1}{\sqrt{n}}\right).
\tag{12}
$$

In the other direction we have for small $s-1$,

$$
\begin{aligned}
\frac{1}{n}\sum_{1}^{n}\rho\left(\frac{\varepsilon_i(n)}{s}\right) &= (1-\varepsilon)+2f(e)(s-1)(1+o(1))\\
&\quad+\frac{1}{n}\sum_{1}^{n}\big(\rho(\varepsilon_i(n))-(1-\varepsilon)\big)+o_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

by (11).

If we define $s_n(\gamma)$ by

$$
2f(e)(s_n(\gamma)-1)=-\frac{1}{n}\sum_{1}^{n}\big(\rho(\varepsilon_i(n))-(1-\varepsilon)\big)+\frac{\gamma}{\sqrt{n}},
$$

for $\gamma > 0$, it follows that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\sum_1^n \rho\left(\frac{\varepsilon_i(n)}{s_n(\gamma)}\right) > 1-\varepsilon\right) = 1.$$

The pair $(0, s_n(\gamma))$ satisfies (3) with high probability and hence it follows from the minimality of $\hat{s}(n)$ that $\lim \mathbb{P}(\hat{s}_n \le s_n(\gamma)) = 1$, for all $\gamma > 0$. Thus

$$2f(e)(\hat{s}(n)-1) \le -\frac{1}{n}\sum_1^n \left(\rho(\varepsilon_i(n)) - (1-\varepsilon)\right) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

which together with (10) and

$$\frac{1}{\sqrt{n}}\sum_1^n \left(\rho(\varepsilon_i(n)) - (1-\varepsilon)\right) \Rightarrow N(0, \varepsilon(1-\varepsilon))$$

prove the theorem. □

A weak convergence result for $\hat{s}(n)$ as a function of ε is given by Grübel (1988) for the case $k = 1$ and $x_i(n) \equiv 1$, $1 \le i \le n$.

We now turn to the behaviour of $\hat{b}(n)$.

LEMMA 4. *Suppose that* (7) *holds. Then for each* $\eta > 0$, *there exist random variables* $(M_n)_1^\infty$ *of order* $O_p(1)$ *such that*

$$\sup_{\|b\|<1,\,0\le s\le 2}\left|\frac{1}{n}\sum_1^n\left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b}{s}\right) - \rho\left(\frac{\varepsilon_i(n)}{s}\right) - \left(R\left(x_i(n)^T b, s\right) - R(0,1)\right)\right)\right|$$
$$\le \eta\left(\|b\|^2 + |s-1|^2\right) + n^{-2/3}M_n^2,$$

for all (b, s) *satisfying* $\|b\| < 1$, $|s-1| < 1$.

PROOF. This is a modified version of Lemma 4.1 of Kim and Pollard (1989). For $\|b\| + |s-1| < \delta$, we note that

$$\left|\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b}{s}\right) - \rho\left(\frac{\varepsilon_i(n)}{s}\right)\right| \le \sum\left\{\left|\frac{\varepsilon_i(n)}{1\pm\delta} - \pm e\right| \le \|x_i(n)\|\delta\right\},$$

where the sum extends over all choices of $\pm$. From F1 it follows that

$$\frac{1}{n}\sum_1^n \mathbb{E}\left(\left\{\left|\frac{\varepsilon_i(n)}{1\pm\delta} - \pm e\right| \le \|x_i(n)\|\delta\right\}\right) \le \frac{c_6}{n}\sum_1^n (1 + \|x_i(n)\|)\delta \le c_7\delta$$

and hence the condition on the envelope is fulfilled. □

THEOREM 5. *Suppose that, in addition to the assumptions of Theorem 4, f has a continuous first derivative at e.*

Then

$$\|\hat{b}(n)\| = O_p(n^{-1/3}).$$

PROOF. We first note that because of D1, $\operatorname{plim}_{n\to\infty}(\hat{b}(n), \hat{s}(n)) = (0, 1)$. As $\hat{b}(n)$ maximizes

$$\frac{1}{n}\sum_1^n \rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b}{\hat{s}(n)}\right),$$

we have

$$\begin{aligned}
0 \le \frac{1}{n}\sum_1^n &\left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b(n)}{\hat{s}(n)}\right) - \rho\left(\frac{\varepsilon_i(n)}{\hat{s}(n)}\right)\right.\\
&\qquad \left. -\left(R\left(x_i(n)^T\hat{b}(n), \hat{s}(n)\right) - R(0, \hat{s}(n))\right)\right)\\
&+\frac{1}{n}\sum_1^n \left(R\left(x_i(n)^T\hat{b}(n), \hat{s}(n)\right) - R(0, \hat{s}(n))\right)\\
&\le \eta\left(\|\hat{b}(n)\|^2 + |\hat{s}(n) - 1|^2\right) + O_p(n^{-2/3})\\
&\quad + \frac{1}{n}\sum_{\substack{1\\ \|x_i(n)\|\le\eta_1}}^n \left(R\left(x_i(n)^T\hat{b}(n), \hat{s}(n)\right) - R(0, \hat{s}(n))\right),
\end{aligned}$$

where we have used Lemmas 1 and 4.

As $\operatorname{plim}_{n\to\infty}\hat{b}(n) = 0$ and, by Theorem 4, $|\hat{s}(n) - 1| = O_p(n^{-1/2})$, it follows that

$$\begin{aligned}
0 &\le \eta\|\hat{b}(n)\|^2 + O_p(n^{-2/3}) + \frac{1}{n}\sum_{\|x_i(n)\|\le\eta_1} \|x_i(n)^T\hat{b}(n)\|^2 f^{(1)}(e)\left(1 + o_p(1)\right)\\
&\le \eta\|\hat{b}(n)\|^2 + O_p(n^{-2/3}) + \eta_2\|\hat{b}(n)\|^2 f^{(1)}(e)\left(1 + o_p(1)\right),
\end{aligned}$$

by the assumptions of the theorem. As η may be taken to be small and $f^{(1)}(e) < 0$ we obtain the claim of the theorem. □

Under slightly more restrictive conditions we can prove the weak convergence of $n^{1/3}\hat{b}(n)$.

We introduce the following condition:

D2. $$\lim_{n\to\infty}\frac{1}{n}\sum_1^n \|x_i(n)\|^2\{\|x_i(n)\| > \delta n^{1/3}\} = 0 \quad \text{for all } \delta > 0.$$

THEOREM 6. *Suppose that in addition to the conditions of Theorem* 5, D2 *holds and*

$$\lim_{n\to\infty} \frac{1}{n} \sum_1^n |x_i(n)^T b| = L(b), \qquad b \in \mathbb{R}^k. \tag{13}$$

Then

$$n^{1/3}\hat{b}(n) \Rightarrow \tau_{\max},$$

where $\tau_{\max}$ *is the uniquely defined* arg max *of*

$$2f^{(1)}(e)\|b\|^2 + \mathcal{G}(b)$$

and $\mathcal{G}(b)$ *is a zero-mean continuous Gaussian process with incremental variance*

$$\mathbb{E}\big((\mathcal{G}(b) - \mathcal{G}(b'))^2\big) = L(b - b').$$

PROOF. We first note that

$$0 < \eta\|b\| \le L(b) \le \sqrt{k}\,\|b\|,$$

for all $b \neq 0$. The last inequality follows from (8) and the second is a consequence of D1.

We define

$$\mathcal{G}_n(b,s) = n^{-1/3} \sum_1^n \left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b n^{-1/3}}{1 + sn^{-1/2}} \right) - \rho\left(\frac{\varepsilon_i(n)}{1 + sn^{-1/2}} \right) \right.$$
$$\left. - \left(R\big(x_i(n)^T b n^{-1/3}, 1 + sn^{-1/2}\big) - R(0, 1 + sn^{-1/2}) \right) \right).$$

Then, using D2 and the assumption (13) of the theorem, a Taylor expansion gives

$$\lim_{n\to\infty} \mathbb{E}\big((\mathcal{G}_n(b,s) - \mathcal{G}_n(b',s'))^2\big) = 2f(e)L(b - b').$$

The Lindeberg conditions for the central limit theorem hold and it follows that the finite dimensional distributions of $\mathcal{G}_n(b,s)$ converge to those of $\mathcal{G}(b)$.

To show that $\mathcal{G}_n$ converges weakly to $\mathcal{G}$, it remains to check the stochastic equicontinuity condition [see Pollard (1984), pages 139–140]. This, however, follows along the lines of Lemma 4.6 of Kim and Pollard (1989) using the envelope of Lemma 2. We therefore have $\mathcal{G}_n \Rightarrow \mathcal{G}$ and hence

$$\frac{1}{n^{1/3}} \sum_1^n \rho\left(\frac{\varepsilon_i - x_i(n)^T b n^{-1/3}}{1 + sn^{-1/2}} \right) \Rightarrow 2f^{(1)}(e)\|b\|^2 + \mathcal{G}(b), \tag{14}$$

where we have used D2 and (7).

As $f^{(1)}(e) < 0$, $\mathbb{E}(|\mathcal{G}(b) - \mathcal{G}(b')|^2) \le \sqrt{k}\,\|b - b'\|$ and $\mathbb{E}(|\mathcal{G}(b) - \mathcal{G}(b')|^2) = 0$, if and only if $b = b'$, it follows from Lemma 2.6 of Kim and Pollard (1989) that, with probability one, $2f^{(1)}(e)\|b\|^2 + \mathcal{G}(b)$ takes its maximum at only point $\tau_{\max}$.

From the definition of $(\hat{b}(n), \hat{s}(n))$, it follows that $\hat{b}(n)$ maximizes the left-hand side of (14) with $sn^{-1/2} = \hat{s}(n)$ and as $\hat{b}(n) = O_p(n^{-1/3})$, by Theorem 5 we may conclude that $n^{1/3}\hat{b}(n) \Rightarrow \tau_{\max}$. □

For an explanation of the $n^{-1/3}$ rate of convergence for discontinuous ρ we refer to Kim and Pollard (1989).

EXAMPLE 4. We consider Example 2 and suppose that

$$\mathbb{E}\left(x_i(n)x_i(n)^T\right) = \Sigma,$$

where Σ is nonsingular. Then the conditions of Theorem 6 are seen to hold for the transformed variables $\Sigma^{1/2}b$ and $\Sigma^{-1/2}x_i(n)$.

If

$$L(b) = \mathbb{E}\left(|x_i(n)^T\Sigma^{-1/2}b|\right),$$

then

$$\hat{b}(n) \Rightarrow \Sigma^{-1/2}\tau_{\max},$$

where $\tau_{\max}$ is defined as in Theorem 6.

EXAMPLE 5. As in Example 3, we suppose that

$$x_i(n)^T = (p_0(i/n), \dots, p_{k-1}(i/n)),$$

where p_r is a polynomial of degree r, $0 \le r \le k-1$, and

$$\int_0^1 p_r(n)p_s(n)\,dn = \{r = s\}. \tag{15}$$

If we replace the carriers $x_i(n)$ by $\Sigma_n^{-1/2}x_i(n)$, where

$$\Sigma_n = \frac{1}{n}\sum_1^n x_i(n)x_i(n)^T,$$

then (7) holds. Furthermore, as $\lim_{n\to\infty}\Sigma_n = I_k$, by (15) we obtain

$$\lim_{n\to\infty}\frac{1}{n}\sum_1^n |x_i(n)^T\Sigma_n^{-1/2}b| = \lim_{n\to\infty}\frac{1}{n}\sum_1^n |x_i(n)^T b| = \int_0^1\left|\sum_{j=0}^{k-1} b_j p_j(u)\right| du.$$

We may therefore deduce that $n^{1/3}\hat{b}(n) \Rightarrow \tau_{\max}$, where $\tau_{\max}$ is as in the theorem with

$$L(b) = \int_0^1\left|\sum_0^{k-1} b_j p_j(u)\right| du.$$

4. Smooth ρ. We now consider the case of smooth ρ. We shall continue to assume that (7) and hence also (8) holds.

We write

$$A(\rho) = \int \rho(u)^2 \, dF(u) - (1-\varepsilon)^2 \tag{16}$$

and make the following further assumptions concerning ρ:

R2. ρ is absolutely continuous with a bounded Lebesgue density ψ.

R3. $$\lim_{(v,s)\to 0} \frac{R(v,1+s) - R(0,1+s)}{v^2} = -B(\rho) < 0.$$

From R2, it follows that ρ satisfies a uniform Lipschitz condition of order 1. Furthermore, if F1 holds then for small s we have

$$\left| R(0,1+s) - R(0,1) + s\int u\psi(u)\, dF(u) \right|$$
$$\leq c_8 |s| \int \left(\frac{1}{|su|} \int_{0 \leq |t| \leq 2|su|} |\psi(u+t) - \psi(u)|\, dt \right) f(u)\, du,$$

which is $o(|s|)$ by Lebesgue's theorem and dominated convergence. It therefore follows from R2 and $R(0,s)$ is differentiable at $s = 1$ with derivative

$$C(\rho) = -\int u\psi(u) f(u)\, du. \tag{17}$$

A similar argument shows

$$\lim_{(v,s)\to 0} \frac{1}{|v|^2} \int \left(\rho\left(\frac{u+v}{1+s} \right) - \rho\left(\frac{u}{1+s} \right) - v\psi(u) \right)^2 f(u)\, du = 0. \tag{18}$$

The assumption R3 will hold if either ψ or f is absolutely continuous with a bounded Lebesgue derivative.

THEOREM 7. *Suppose that* R1, R2, F1, FR1 *and* D1 *hold. Then*

$$\sqrt{n}\,(\hat{s}(n) - 1) \Rightarrow N\left(0, \frac{A(\rho)}{C(\rho)^2}\right).$$

PROOF. As ρ is an even function and nonincreasing on $\mathbb{R}_+$ the graphs of the functions

$$h_{b,s}(\varepsilon_i, x_i) = \rho\left(\frac{\varepsilon_i - x_i^T b}{s} \right),$$

have polynomial discrimination [see, for example, Pollard (1984), page 29]. The class of functions $h_{b,s}$ is therefore uniformly manageable and hence so is the class of functions

$$\tilde{h}_{b,s} = h_{b,s} - h_{0,s}. \tag{19}$$

We may therefore apply the maximal inequality to the $\tilde{h}_{b,s}$ taking as the envelope

$$\tilde{H}_R(\varepsilon_i, x_i) = c_9\|x_i\|R, \tag{20}$$

for suitable c_9. This follows from R2 because of the uniform Lipschitz condition. The proof of Theorem 4 may now be repeated with only minor changes using the differentiability of $R(0, s)$ at $s = 1$. □

We now turn to the behaviour of $\hat{b}(n)$ and show first that $\hat{b}(n) = O_p(n^{-1/2})$.

LEMMA 5. *Suppose* R1, R2, R3, F1, FR1 *and* D1 *hold. Then*

$$\hat{b}(n) = O_p(n^{-1/2}).$$

PROOF. The proof is a modified version of Lemma 4.1 of Kim and Pollard (1989). We note first that D1 implies $\operatorname{plim}_{n\to\infty} \hat{b}(n) = 0$. Furthermore $|\hat{s}(n) - 1| = O_p(n^{-1/2})$ by Theorem 7.

We now consider the $\tilde{h}_{b,s}$ of (19) and their envelope $\tilde{H}_R$ given by (20). It follows that

$$\frac{1}{n}\sum_1^n \mathbb{E}\big(\tilde{H}_R(\varepsilon_i(n), x_i(n))\big) = CR^2.$$

One can now repeat the reasoning of Lemma 4.1 of Kim and Pollard (1989) using, in their notation, $PG_R^2 \le CR^2$ instead of $PG_R^2 \le CR$ and replacing the factor $n^{-1/3}$, where appropriate, by $n^{-1/2}$. This leads to the following: For all $\eta > 0$, there exists a sequence $(M_n)_1^\infty$ of $O_p(1)$ random variables such that

$$\left|\frac{1}{n}\sum_1^n \Big(\tilde{h}_{b,s}(\varepsilon_i(n), x_i(n)) - \mathbb{E}\big(\tilde{h}_{b,s}(\varepsilon_i(n), x_i(n))\big)\Big)\right| \le \eta\big(\|b\|^2 + |s-1|^2\big) + n^{-1}M_n^2,$$

for all (b, s) satisfying $\|b\| < 1$ and $|s - 1| < \frac{1}{2}$.

As $\hat{b}(n)$ maximizes

$$\frac{1}{n}\sum_1^n \rho\left(\frac{\varepsilon_i - x_i(n)^T b}{\hat{s}(n)}\right),$$

we obtain

$$\begin{aligned}
0 &\le \frac{1}{n}\sum_1^n \left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T\hat{b}(n)}{\hat{s}(n)}\right) - \rho\left(\frac{\varepsilon_i(n)}{\hat{s}(n)}\right)\right)\\
&\le \eta\big(\|\hat{b}(n)\|^2 + |\hat{s}(n) - 1|^2\big) + n^{-1}M_n^2\\
&\quad + \frac{1}{n}\sum_1^n \Big(R\big(x_i(n)^T\hat{b}(n), \hat{s}(n)\big) - R(0, \hat{s}(n))\Big)\\
&\le \eta\big(\|\hat{b}(n)\|^2 + |\hat{s}(n) - 1|^2\big) + n^{-1}M_n^2\\
&\quad + \frac{1}{n}\sum_{\|x_i(n)\| < \eta_1} \Big(R\big(x_i(n)^T\hat{b}(n), \hat{s}(n)\big) - R(0, \hat{s}(n))\Big),
\end{aligned}$$

by (ii) of Lemma 1. As $\operatorname{plim}_{n\to\infty} \hat{b}(n) = 0$ and $|\hat{s}(n) - 1| = O_p(n^{-1/2})$, it follows from (R3) and D1 that

$$0 \le \eta \|\hat{b}(n)\|^2 - c_{10}\|\hat{b}(n)\|^2 + O_p(n^{-1}),$$

with $c_{10} > 0$. On taking $\eta = \frac{1}{2}c_{10}$, we obtain the claim of the lemma. □

Corresponding to D2, we now introduce the assumption D3:

D3. $$\lim_{n\to\infty} \frac{1}{n} \Sigma \|x_i(n)\|^2 \{\|x_i(n)\| > \delta\sqrt{n}\} = 0, \quad \text{for all } \delta > 0.$$

If D3 holds, then there exists a sequence $(\delta(n))_1^\infty$ with $\lim_{n\to\infty} \delta(n) = 0$ and

$$\lim_{n\to\infty} \frac{1}{\delta(n)^2 n} \sum_1^n \|x_i(n)\|^2 \{\|x_i(n)\| \ge \delta(n)\sqrt{n}\} = 0. \tag{21}$$

In particular, this implies

$$\lim_{n\to\infty} \sum_{\{\|x_i(n)\| > \delta(n)\sqrt{n}\}} 1 = 0. \tag{22}$$

We can now prove

THEOREM 8. *Suppose* R1, R2, R3, F1, FR1, D1 *and* D3 *hold. Then*

$$\sqrt{n}\,\hat{b}(n) \Rightarrow N\left(0, I_k \frac{\int \psi^2\, dF}{4B(\rho)^2}\right).$$

PROOF. The conditions of Theorem 7 and Lemma 5 are satisfied so we have $\|\hat{b}(n)\| + |\hat{s}(n) - 1| = O_p(n^{-1/2})$. We consider the rescaled processes

$$Z_n(b, s) = \sum_1^n \left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b/\sqrt{n}}{1 + s/\sqrt{n}}\right) - \rho\left(\frac{\varepsilon_i(n)}{1 + s/\sqrt{n}}\right) - R\left(x_i(n)^T \frac{b}{\sqrt{n}}, 1 + \frac{s}{\sqrt{n}}\right) + R\left(0, 1 + \frac{s}{\sqrt{n}}\right)\right)$$

and prove that they converge weakly to a continuous Gaussian process Z_∞ to be identified below.

If $(\delta(n))_1^\infty$ is as in (21), then, because ρ is bounded, it is sufficient to consider the process

$$Z_n'(b, s) = \sum_1^{n}{}' \left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b/\sqrt{n}}{1 + s/\sqrt{n}}\right) - \rho\left(\frac{\varepsilon_i(n)}{1 + s/\sqrt{n}}\right) - R\left(x_i(n)^T \frac{b}{\sqrt{n}}, 1 + \frac{s}{\sqrt{n}}\right) + R\left(0, 1 + \frac{s}{\sqrt{n}}\right)\right),$$

where $\Sigma_1'^n$ denotes the sum over all i with $\|x_i(n)\| \le \delta(n)\sqrt{n}$.

The first step is to prove that the sequence $(Z'_n)_1^\infty$ is tight. To this end we define for (b, b', s, s') in $\mathbb{R}^{2k+2}$ and $|s - 1| < \frac{1}{2}$, $|s' - 1| < \frac{1}{2}$,

$$h^*_{b,b',s,s'} = \tilde{h}_{b,s} - \tilde{h}_{b',s'},$$

where $\tilde{h}$ is as given by (19). As the class $\tilde{h}_{b,s}$ is uniformly manageable so also is the class $h^*_{b,b',s,s'}$. In order to apply the maximal inequality we require an envelope for this latter class. From R2 and the fact that ψ has a compact support, it follows that

$$\sup_{\|b-b'\|+|s-s'|<R} |h^*_{b,b',s,s'}(\varepsilon_i, x_i)| \le c_{11} R(\|x_i\| + 1) =: H^*_R(\varepsilon_i, x_i),$$

as long as $|s - 1| < \frac{1}{2}$ and $|s' - 1| < \frac{1}{2}$.

We therefore have

$$\frac{1}{n} \sum_1^n{}' \mathbb{E}\big(H^*_R(\varepsilon_i(n), x_i(n))\big) \le c_{12} R,$$

and the maximal inequality implies

$$\mathbb{E}\left(\sup_{\substack{\|b-b'\|+|s-s'|<R \\ \|s-1\|<\frac{1}{2}, |s'-1|<\frac{1}{2}}} \left| \frac{1}{n} \sum_1^n{}' \Big(h^*_{b,b',s,s'}(\varepsilon_i(n), x_i(n)) - \mathbb{E}\big(h^*_{b,b',s,s'}(\varepsilon_i(n), s_i(n))\big)\Big) \right| \right) \le c_{13} R.$$

If we now replace R by $R/\sqrt{n}$, we obtain

$$\sqrt{n}\, \mathbb{E}\left(\sup_{\|b-b'\|+|s-s'|<R} \left| \frac{1}{n}\big(Z'_n(b,s) - Z'_n(b',s')\big) \right| \right) \le c_{14} \frac{R}{\sqrt{n}},$$

which, on letting R tend to zero, shows that the sequence $(Z_n)_1^\infty$ is tight.

To complete the proof of the weak convergence, it is necessary to show that the finite dimensional distributions converge. For those i with $\|x_i(n)\| \le \delta(n)\sqrt{n}$, we have for fixed b and sufficiently large n,

$$\big|R\big(x_i(n)^T b/\sqrt{n}, 1 + s/\sqrt{n}\big) - R\big(0, 1 + s/\sqrt{n}\big)\big| \le c_{15} \|x_i(n)\|^2 \|b\|^2/n,$$

by R3. This is seen to imply, using (8),

$$\sum_1^n{}' \Big(R\big(x_i(n)^T b/\sqrt{n}, 1 + s/\sqrt{n}\big) - R\big(0, 1 + s/\sqrt{n}\big) - R\big(x_i(n)^T b'/\sqrt{n}, 1 + s/\sqrt{n}\big) + R\big(0, 1 + s'/\sqrt{n}\big)\Big)^2 = c_{16} \delta(n)^2 \|b\|^4,$$

and hence

$$\mathbb{E}\Big(\big(Z_n'(b,s) - Z_n'(b',s')\big)^2\Big)$$
$$= \sum_1^n{}' \mathbb{E}\left(\left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b/\sqrt{n}}{1+s/\sqrt{n}}\right) - \rho\left(\frac{\varepsilon_i(n)}{1+s/\sqrt{n}}\right)\right.\right.$$
$$\left.\left. -\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b'/\sqrt{n}}{1+s'/\sqrt{n}}\right) + \rho\left(\frac{\varepsilon_i(n)}{1+s'/\sqrt{n}}\right)\right)^2\right) + o(1).$$

Now

$$\mathbb{E}\left(\left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b/\sqrt{n}}{1+s/\sqrt{n}}\right) - \rho\left(\frac{\varepsilon_i(n)}{1+s/\sqrt{n}}\right)\right.\right.$$
$$\left.\left. -\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b'/\sqrt{n}}{1+s'/\sqrt{n}}\right) + \rho\left(\frac{\varepsilon_i(n)}{1+s'/\sqrt{n}}\right)\right)^2\right)$$
$$= \frac{|x_i(n)^T(b-b')|^2}{n}\int \psi^2\, dF + o\left(\frac{\|x_i(n)\|^2}{n}(\|b\|^2 + \|b'\|^2)\right),$$

where we have used (17). The

$$o\left(\frac{\|x_i(n)\|^2}{n}(\|b\|^2 + \|b'\|^2)\right)$$

holds uniformly in n for those i with $\|x_i(n)\| \le \delta(n)\sqrt{n}$. We therefore obtain

$$\lim_{n\to\infty} \mathbb{E}\Big(\big(Z_n'(b,s) - Z_n'(b',s')\big)^2\Big) = \|b-b'\|^2 \int \psi^2\, dF.$$

As

$$\left|\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b/\sqrt{n}}{1+s/\sqrt{n}}\right) - \rho\left(\frac{\varepsilon_i(n)}{1+s/\sqrt{n}}\right)\right.$$
$$\left. -\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b'/\sqrt{n}}{1+s'/\sqrt{n}}\right) + \rho\left(\frac{\varepsilon_i(n)}{1+s'/\sqrt{n}}\right)\right|$$
$$\le c_{17}\frac{\|x_i(n)\|(\|b\| + \|b'\|)}{\sqrt{n}} \le c_{18}\delta(n)(\|b\| + \|b'\|),$$

the Lindeberg conditions are fulfilled and it follows that

$$\big(Z_n'(b_1,s_1),\dots,Z_n'(b_k,s_k)\big) \Rightarrow N(0,\Sigma(1,\dots,k)),$$

with

$$\Sigma_{ij}(1,\dots,k) = b_i^T b_j \int \psi^2\, dF.$$

Putting all this together we see that the processes $(Z_n')_1^\infty$ and hence the processes $(Z_n)_1^\infty$ converge weakly to a Gaussian process Z_∞ of the form

$$Z_\infty(b,s) = b^T Z,$$

where Z is a Gaussian random variable with zero mean and covariance matrix $I_k \int \psi^2 \, dF$.

The limiting process Z_∞ is independent of s. This implies that for any K,

$$\max_{\|b\|<K, |s|<K} |Z_n(b,s) - Z_n(b,0)| = o_p(1),$$

and hence, for any sequence of $O_p(1)$ random variable s_n, the processes $(Z_n(\cdot, s_n)_1^\infty)$ converge weakly to the process $Z_\infty(\cdot)$. As $\sqrt{n}(\hat{s}(n) - 1) = O_p(1)$ this implies

$$Z_n\big(\cdot, \sqrt{n}(\hat{s}(n) - 1)\big) \Rightarrow Z_\infty(\cdot). \tag{23}$$

Now R3 implies

$$\begin{aligned}\sum_1^n{}' \mathbb{E}\left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b/\sqrt{n}}{1 + s/\sqrt{n}}\right) - \rho\left(\frac{\varepsilon_i(n)}{1 + s/\sqrt{n}}\right)\right) \\ = \sum_1^n{}' - \frac{|x_i(n)^T b|^2 B(\rho)(1 + o(1))}{n},\end{aligned}$$

which together with D3 yields

$$\lim_{n\to\infty} \sum_1^n \mathbb{E}\left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b}{1 + s/\sqrt{n}}\right) - \rho\left(\frac{\varepsilon_i(n)}{1 + s/\sqrt{n}}\right)\right) = -B(\rho)\|b\|^2.$$

This combined with (23) gives

$$\sum_1^n \left(\rho\left(\frac{\varepsilon_i(n) - x_i(n)^T b/\sqrt{n}}{\hat{s}(n)}\right) - \rho\left(\frac{\varepsilon_i(n)}{\hat{s}(n)}\right)\right) \Rightarrow b^T Z - B(\rho)\|b\|^2.$$

The left-hand side is maximized by $b = \sqrt{n}\,\hat{b}(n)$, which is $O_p(1)$ and the right-hand side by $b = Z/(2B(\rho))$. Using Theorem 2.7 of Kim and Pollard (1989) on the weak convergence of the arg max we obtain

$$\sqrt{n}\,\hat{b}(n) \Rightarrow \frac{Z}{2B(\rho)} = N\left(0, I_k \frac{\int \psi^2 \, dF}{4B(\rho)^2}\right),$$

proving the theorem. □

Theorems 7 and 8 give the asymptotic normality of $\hat{b}(n)$ and $\hat{s}(n)$ under less stringent conditions than those of Theorem 3 of Rousseeuw and Yohai (1984).

Acknowledgment. I would like to acknowledge the help of the two anonymous referees whose comments on the first two versions lead to considerable improvements.

REFERENCES

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H. and TUKEY, J. W. (1972). *Robust Estimates of Location*. Princeton Univ. Press.

DAVIES, P. L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292.

DUDLEY, R. M. (1987). Universal Donsker classes and metric entropy. *Ann. Probab.* **15** 1306–1326.

GRÜBEL, R. (1988). The length of the shorth. *Ann. Statist.* **16** 619–628.

HAMPEL, F. R. (1975). Beyond location parameters: Robust concepts and methods (with discussion). *Proc. 40th ISI Conf.* **46**, Book 1, 375–391.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

KIM, J. and POLLARD, D. (1989). Cube root asymptotics. Statist. Dept., Yale Univ. Unpublished.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

POLLARD, D. (1989). Asymptotics via empirical processes (with discussion). *Statist. Sci.* **4** 341–356.

ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880.

ROUSSEEUW, P. J. and LEROY, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.

ROUSSEEUW, P. J. and YOHAI, V. (1984). Robust regression by means of S-estimators. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 256–272. Springer, New York.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

FACHBEREICH MATHEMATIK
UNIVERSITY OF ESSEN
D–4300 ESSEN
WEST GERMANY