Statistics & Probability Letters 14 (1992) 413–419
North-Holland

11 August 1992

# On the optimality of S-estimators

Ola Hössjer
*Uppsala University, Uppsala, Sweden*

Received June 1991

*Abstract*: The efficiency of an S-estimator in the linear model is maximized under a constraint on the breakdown point and the form of the optimal score function is derived. Numerical calculations for normally distributed errors are performed, and the maximal possible efficiency is then 0.33 when the breakdown point equals 0.5.

*AMS 1991 Subject Classifications*: 62F35, 62J05.

*Keywords*: Asymptotic efficiency; breakdown point; robust linear regression; S-estimator.

## 1. Introduction

Consider the linear model

$$y_i = \boldsymbol{x}_i' \boldsymbol{\theta}_0 + e_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where the carriers or explanatory variables $\boldsymbol{x}_i$ and the unknown regression parameter $\boldsymbol{\theta}_0$ are $p$-dimensional column vectors and the error terms $e_i$ are assumed to be independent and identically distributed (i.i.d.) with a common distribution function $F(e/\sigma)$, where $\sigma$ represents the unknown scale parameter.

The breakdown point $\varepsilon^*$ is a nonlocal criterion for developing estimators robust against leverage points, and it is loosely speaking the smallest fraction of contamination of the data vectors $(y_i, \boldsymbol{x}_i)$ that the estimator can tolerate before it collapses. It was first introduced by Hodges (1967) in the location setting and more generally by Hampel (1971). Later, Donoho and Huber (1983) introduced a more data-oriented definition. Many of the commonly used estimators have breakdown point zero, i.e. least squares (LS), M-estimators based on a non-decreasing $\psi$-function (Huber, 1973), R-estimators based on a non-decreasing score generating function (Jurečková, 1971, Jaeckel, 1972, for linear rank statistics and Van Eeden, 1972, Hettmansperger and McKean, 1983, for signed rank statistics). The first example of an estimator with the maximal value $\varepsilon^* = 0.5$, independent of $p$ was given by Siegel (1982) — the repeated median (RM). This estimator has the disadvantage of not being affinely equivariant. Affinely equivariant estimators with $\varepsilon^* = 0.5$ have then been introduced, the least median of squares (LMS) estimator (Rousseeuw, 1984), the least trimmed squares (LTS) estimator (Rousseeuw, 1985), S-estimators (Rousseeuw and Yohai, 1984), MM-estimators (Yohai, 1987), $\tau$-estimators (Yohai and Zamar, 1988) and R-estimators (based on signed rank statistics) with nonmonotone score functions (Hössjer, 1991). All these estimators except LMS have the optimal rate of convergence $n^{-1/2}$. (For LMS the rate of convergence is $n^{-1/3}$.) Furthermore, for normally distributed errors, the MM- and $\tau$-estimators may be

*Correspondence to*: Ola Hössjer, Department of Mathematics, Uppsala University, Thunbergsvägen 3, S-752 38 Uppsala, Sweden

This paper was written under support by the Swedish Board for Technical Development, under contract 712-89-1073.

chosen with efficiency arbitrarily close to one. A one-step M-estimator, which inherits the breakdown point of the preliminary estimator, has been considered by Jurečková and Portnoy (1987). Even this estimator may be chosen with efficiency arbitrarily close to one. The book by Rousseeuw and Leroy (1987) gives a good overview of robust regression estimators with high breakdown point.

In this paper we consider S-estimators, and answer the following question: Given a certain number $\varepsilon$ $(0 < \varepsilon \leqslant 0.5)$, what is the optimal choice of S-estimator under the constraint that the breakdown point is at least $\varepsilon$. It turns out that this optimality problem can be formulated as a standard variational problem, with the breakdown point condition formulated as an integral side condition. The resulting score function has a rather unexpected form and has not (up to the knowledge of the author) been used before.

We believe that the techniques used in this paper may be extended to S-estimators of multivariate location and covariance matrices, since the breakdown point condition may then be formulated in a similar fashion, cf. Rousseeuw (1985), Davies (1987), Lopuhaä (1989) and Lopuhaä and Rousseeuw (1991).

In Section 2 we formulate the problem and in Section 3 – 4 the optimal score functions are derived. Finally, in Section 5 the maximal possible efficiencies are calculated for normally distributed (or Gaussian) errors.

## 2. Definitions and regularity conditions

The S-estimators are defined by minimizing a robust measure $s$ of scale of the residuals $r_i = r_i(\boldsymbol{\theta}) = y_i - \boldsymbol{x}_i'\boldsymbol{\theta}$, i.e. the estimate of $\boldsymbol{\theta}$ is defined as

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} s(r_1(\boldsymbol{\theta}), \ldots, r_n(\boldsymbol{\theta})).$$

Usually (and in this paper), $s$ is an M-estimator of scale defined by a function $\rho$, which satisfies:

(A1) $\rho$ is symmetric, non-decreasing on $(0, \infty)$ and absolutely continuous with bounded derivative $\psi$.
(A2) $\rho(0) = 0$, and there exists a $c > 0$ such that $\rho$ is constant on $[c, \infty)$.

The scale estimate is then defined implicitly by means of

$$\frac{1}{n}\sum_{i=1}^{n} \rho(r_i/s) = K, \tag{2.1}$$

with $K = E_F(\rho)$. For the error distribution $F$ we assume that:

(B1) $F$ has a symmetric, unimodal and absolutely continuous density $f$ with $f'$ bounded.

Finally, we assume that:

(AB1) $f$ and $-\rho$ have a common point of decrease in $(0, \infty)$.

Asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ was proved by Rousseeuw and Yohai (1984) and further considered by Davies (1990). Assume for simplicity that $\sigma = 1$. After a possible reparametrization, we may also suppose that

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' = \boldsymbol{I}_p, \tag{2.2}$$

for each $n$, with $I_p$ the $p \times p$ identity matrix. Under assumptions (A1)–(A2), (B1) [1] and (AB1) and some additional assumptions on the carriers (assumptions (D1) and (D3) in Davies, 1990), it then holds that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{\mathrm{d}} \mathrm{N}(0, V(\psi, F))\boldsymbol{I}_p, \tag{2.3}$$

where the asymptotic variance $V(\psi, F)$ is given by

$$V(\psi, F) = \frac{\int \psi(x)^2 f(x)\,\mathrm{d}x}{\left(\int \psi(x)\Lambda(x)f(x)\,\mathrm{d}x\right)^2} = \frac{A(\psi, F)}{B(\psi, F)^2}. \tag{2.4}$$

In other words, the asymptotic variance is the same as for an M-estimator with score function $\psi$.

Given a sample $\boldsymbol{Z}_n = (y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$, Donoho and Huber (1983) define the finite sample breakdown point (with replacement) $\varepsilon_n^*(\hat{\boldsymbol{\theta}}_n, \boldsymbol{Z}_n)$ as the smallest fraction of data points $(y_i, \boldsymbol{x}_i)$ that after being replaced may change $\hat{\boldsymbol{\theta}}_n$ $(\boldsymbol{Z}_n)$ arbitrarily much. Restricting ourselves to samples in general position (cf. Rousseeuw, 1984), the asymptotic breakdown point is then defined as

$$\varepsilon^* = \lim_{n \to \infty} \varepsilon_n^*\left(\hat{\boldsymbol{\theta}}_n, \boldsymbol{Z}_n\right) \tag{2.5}$$

whenever the limit exists and is independent of the sequence of data points. The asymptotic breakdown point of the S-estimator is given by (Rousseeuw and Yohai, 1984)

$$\varepsilon^*(\psi) = \min\left(\frac{E_F(\rho)}{\rho(c)}, 1 - \frac{E_F(\rho)}{\rho(c)}\right). \tag{2.6}$$

Integration by parts shows that the condition $E_F(\rho)/\rho(c) = \alpha$, $0 < \alpha < 1$, may be replaced by

$$\int_0^\infty \psi(x) F_\alpha(x)\,\mathrm{d}x = 0, \tag{2.7}$$

with

$$F_\alpha(x) = F(x) + \tfrac{1}{2}\alpha - 1. \tag{2.8}$$

We will need the following additional regularity conditions on $F$:

(B2) $f(x) > 0$ and $f'(x) < 0$ for all $x > 0$, the function $h_{\alpha,k}(x) = f'(x) + kF_\alpha(x)$ continuous, has a unique zero $c(\alpha, k)$ for all $k > 0$, $0 < \alpha < 1$, and $h_{\alpha,k}(x) \gtreqless 0$ as $x \gtreqless c(\alpha, k)$.

(B3) $\int_0^\infty \Lambda(x)\,\mathrm{d}x = \infty$, where $\Lambda(x) = -f'(x)/f(x)$ is the likelihood score function.

Note that (AB1) is contained in (A1) and (B2), so in the sequel we may neglect this condition. The optimization problem may be formulated as:

(P) Given an $F$ satisfying (B1)–(B3) and $\varepsilon$, $0 < \varepsilon \leqslant 0.5$, minimize $V(\psi, F)$ with respect to $\psi$ among all $\psi$-functions satisfying (A1)–(A2), and $\varepsilon^*(\psi) \geqslant \varepsilon$.

Since $\psi$ (or $\rho$) may be multiplied with a scalar $a$ without affecting the S-estimator or $V(\psi, F)$, we may replace each $\psi$ with $a(\psi)\psi$ so that $B(a(\psi)\psi, F)$ equals a fixed value $C$ (which may be chosen to make the calculations convenient). We may thus reformulate our minimization problem as:

[1] Actually, the boundedness requirement on $f'$ in (B1) can be weakened to condition (R3) in Davies (1990).

(P′) Given an $F$ satisfying (B1)–(B3), and $\varepsilon$, $0 < \varepsilon \leqslant 0.5$, minimize $A(\psi, F)$ with respect to $\psi$ among all $\psi$-functions satisfying (A1)–(A2), $B(\psi, F) = C$ and (2.7) for some $\alpha$, $\varepsilon \leqslant \alpha \leqslant 1 - \varepsilon$.

## 3. Two preliminary lemmas

An attempt to solve (P′) by the method of Lagrange multipliers, with $B(\psi, F) = C$ and (2.7) as integral side conditions (and remembering that (A1) implies that $\psi(x) \geqslant 0$ as $x > 0$), suggests that odd score functions of the form [2]

$$\psi_{\alpha,k}(x) = \left( \Lambda(x) - k\frac{F_\alpha(x)}{f(x)} \right)_+, \quad x > 0, \tag{3.1}$$

may solve (P′). Indeed, Theorem 1 gives an affirmative answer. Note that the functions $\psi_{\alpha,k}$ satisfy (A1)–(A2). We start with a preliminary lemma, for which we need the quantity $c_\alpha = F^{-1}(1 - \frac{1}{2}\alpha)$.

**Lemma 1.** *Under assumptions* (B1)–(B2), *there exists to each* $k > 0$ *and* $0 < \alpha < 1$ *a unique* $c = c(\alpha, k)$ *such that* $\Lambda(x) \lessgtr kF_\alpha(x)/f(x)$ *whenever* $x > 0$ *and* $x \gtrless c(\alpha, k)$. *Moreover, with* $\alpha$ *fixed,* $c(\alpha, k)$ *is a continuous and strictly decreasing function of* $k$, *with* $c(\alpha, k) \to \infty$ *as* $k \to 0+$ *and* $c(\alpha, k) \to c_\alpha +$ *as* $k \to \infty$.

**Proof.** Since $f$ is strictly positive according to (B2), it suffices for the first part of the lemma to show that a unique number $c(\alpha, k)$ exists, such that

$$-f(x)\left( \Lambda(x) - k\frac{F_\alpha(x)}{f(x)} \right) = h_{\alpha,k}(x) \gtrless 0 \quad \text{whenever } x \gtrless c(\alpha, k). \tag{3.2}$$

However, (3.2) is immediate from (B2). According to (B2), $f'$ is strictly negative on the positive real line, and hence

$$h_{\alpha,k}(x) < kF_\alpha(x) \leqslant 0 \quad \text{as } 0 < x \leqslant c_\alpha, \tag{3.3}$$

and consequently $c(\alpha, k) > c_\alpha$. Suppose now that $0 < k' < k$. Then

$$h_{\alpha,k'}(x) < h_{\alpha,k}(x) \leqslant 0 \quad \text{as } c_\alpha < x \leqslant c(\alpha, k). \tag{3.4}$$

It follows from (3.4) that $c(\alpha, k') > c(\alpha, k)$. The continuity of $c(\alpha, k)$ in $k$ follows easily from the uniqueness of $c(\alpha, k)$ and continuity of $h_{\alpha,k}(x)$ in $k$. Finally, the fact that for each $x > c_\alpha$, $h_{\alpha,k}(x) < 0$ when $k$ is small enough and $> 0$ when $k$ is large enough, implies the stipulated convergence of $c(\alpha, k)$ as $k \to 0+$ and $k \to \infty$ respectively. □

Condition (B2) is not very explicit in its form. The following condition is stronger (and hence implies Lemma 1 together with (B1)) and easier to check:

(B2′) $f$ is strictly positive, $f'$ is differentiable outside the origin, with $f''$ having at most a finite number of discontinuities, $f'(x) < 0$ when $x > 0$ and $\lim_{x \to \infty} f'(x) = 0$. Finally, for each $k > 0$ and $y > 0$, $\Lambda(y)^2 - \Lambda'(y) + k > 0$ implies $\Lambda(x)^2 - \Lambda'(x) + k > 0$ for all $x > y$.

[2] + denotes the positive part operator.

Condition (B2′) is satisfied for instance by the standard normal, Laplace, logistic, Cauchy and Huber's least favourable minimax distributions. Theorem 1 is therefore valid for all these distributions, since they satisfy (B1) and (B3) as well.

In order to show that (B2′) in fact implies (B2), we first note that (3.3) remains valid, and $\lim_{x\to\infty} h_{\alpha,k}(x) = \frac{1}{2}\alpha > 0$. Hence

$$c(\alpha, k) = \sup\{x > 0,\ h_{\alpha,k}(x) < 0\}$$

is well defined, and since $h_{\alpha,k}$ is continuous, $h_{\alpha,k}(c(\alpha, k)) = 0$. It remains to show $h_{\alpha,k}(x) > 0$ for all $x > c(\alpha, k)$. Observe that

$$h'_{\alpha,k}(x) = f''(x) + kf(x) = \big(\Lambda(x)^2 - \Lambda'(x) + k\big)f(x). \tag{3.5}$$

By the mean value theorem of calculus and (3.3), $h'_{\alpha,k}(y) > 0$ for some $0 < y < c(\alpha, k)$, and hence $\Lambda(y)^2 - \Lambda'(y) + k > 0$. The last assumption in (B2′) and (3.5) then imply $h'_{\alpha,k}(x) > 0$ for all $x > y$ and hence $h_{\alpha,k}(x) > 0$ for all $x > c(\alpha, k)$.

Our next step is to find for each $\alpha$ a corresponding $k = k(\alpha)$ such that (2.7) is satisfied and hence $\varepsilon^*(\psi_{\alpha,k(\alpha)}) = \min(\alpha, 1-\alpha)$. For this purpose we introduce the function

$$J_\alpha(k) = \int_0^{c(\alpha,k)} \psi_{\alpha,k}(x) F_\alpha(x)\, \mathrm{d}x \tag{3.6}$$

defined for $k > 0$.

**Lemma 2.** *Assuming* $0 < \alpha < 1$ *and that* (B1)–(B3) *hold,* $J_\alpha(k)$ *is a continuous, strictly decreasing function of* $k$ *with a unique zero* $k(\alpha)$.

**Proof.** When $0 < x < c_\alpha$, $F_\alpha(x) < 0$ and $\psi_{\alpha,k}(x)$ is a strictly increasing function of $k$, whereas $x > c_\alpha$ implies $F_\alpha(x) > 0$ and that $\psi_{\alpha,k}(x)$ is a non-increasing function of $k$. (Remember that $c(\alpha, k)$ is a (strictly) decreasing function of $k$ by Lemma 1.) Altogether this implies that $J_\alpha(k)$ is a strictly decreasing function of $k$. Since $c(\alpha, k)$ is a continuous function of $k$, so is $\psi_{\alpha,k}(x)$ for each $x > 0$ and bounded convergence then yields that even $J_\alpha(k)$ is a continuous function of $k$. It remains to investigate the sign of $J_\alpha(k)$. Let us rewrite (3.6) into the form

$$J_\alpha(k) = \int_0^{c(\alpha,k)} \Lambda(x) F_\alpha(x)\, \mathrm{d}x - k\int_0^{c(\alpha,k)} \frac{F_\alpha(x)^2}{f(x)}\, \mathrm{d}x = J_1 - J_2. \tag{3.7}$$

As $k \to \infty$, $c(\alpha, k) \to c_\alpha +$ by Lemma 1, $J_1 \to \int_0^{c_\alpha} \Lambda(x) F_\alpha(x)\, \mathrm{d}x$, $J_2 \to \infty$ and hence $J_\alpha(k) \to -\infty$. Let now $c > 0$ be arbitrary. Since $c(\alpha, k) \to \infty$ as $k \to 0+$ it follows that

$$\liminf_{k\to 0+} J_\alpha(k) \geqslant \int_0^c \Lambda(x) F_\alpha(x)\, \mathrm{d}x = I(c)$$

for each $c$, and hence

$$\liminf_{k\to 0+} J_\alpha(k) \geqslant \sup_{0<c<\infty} I(c) \geqslant \limsup_{c\to\infty} I(c) = \infty, \tag{3.8}$$

where the last equality follows from (B3). It follows finally from (3.8) that $J_\alpha(k) \to \infty$ as $k \to 0+$ and hence there exists a unique zero $k(\alpha)$. □

## 4. The main theorem

Let us introduce the short notation $\psi_\alpha(x) = \psi_{\alpha,k(\alpha)}(x)$ and $c(\alpha) = c(\alpha, k(\alpha))$. The main theorem states that the class of optimal S-estimators solving (P′) are to be found among $\{\psi_\alpha\}_{0<\alpha\leq 0.5}$.

**Theorem 1.** *Given* $\varepsilon$, $0 < \varepsilon \leq 0.5$, *the score functions* $\{a\psi_\varepsilon\}$, *with* $a > 0$ *arbitrary, solve the optimization problem* (P) *uniquely a.e.* (F).

**Proof.** It suffices to show that $\psi_\varepsilon$ solves (P′) uniquely a.e. (F), i.e. minimizes $A(\psi, F)$ uniquely a.e. (F) among all $\psi$ satisfying (A1)–(A2), $\varepsilon^*(\psi) \geq \varepsilon$ and $B(\psi, F) = B(\psi_\varepsilon) = C_\varepsilon$. For any such $\psi$, it follows by (2.6) that (2.7) holds for some $\alpha$, $\varepsilon \leq \alpha \leq 1 - \varepsilon$. We intend to show that

$$A(\psi, F) \geq A(\psi_\varepsilon, F), \tag{4.1}$$

with equality iff $\psi = \psi_\varepsilon$ a.e. $(F)$. Because of (A2), there exists a $c > c(\varepsilon)$ (depending on $\psi$) such that $\psi(x) = 0$ when $|x| > c$. Consider then the integral

$$\begin{aligned} R_c(\psi, F) &= \int_0^c \left(\psi(x) - \Lambda(x) + k(\varepsilon)\frac{F_\varepsilon(x)}{f(x)}\right)^2 f(x)\,\mathrm{d}x \\ &= \tfrac{1}{2}A(\psi, F) - C_\varepsilon + 2k(\varepsilon)\int_0^c \psi(x)F_\varepsilon(x)\,\mathrm{d}x + \int_0^c \left(\Lambda(x) - k(\varepsilon)\frac{F_\varepsilon(x)}{f(x)}\right)^2 f(x)\,\mathrm{d}x. \end{aligned} \tag{4.2}$$

Since $\psi(x) \geq 0$ when $x > 0$ it follows that

$$\begin{aligned} R_c(\psi, F) &= \int_0^{c(\varepsilon)} \left(\psi(x) - \Lambda(x) + k(\varepsilon)\frac{F_\varepsilon(x)}{f(x)}\right)^2 f(x)\,\mathrm{d}x \\ &\quad + \int_{c(\varepsilon)}^c \left(\psi(x) - \Lambda(x) + k(\varepsilon)\frac{F_\varepsilon(x)}{f(x)}\right)^2 f(x)\,\mathrm{d}x \\ &\geq \int_{c(\varepsilon)}^c \left(\Lambda(x) - k(\varepsilon)\frac{F_\varepsilon(x)}{f(x)}\right)^2 f(x)\,\mathrm{d}x = R_c(\psi_\varepsilon, F), \end{aligned} \tag{4.3}$$

with equality iff $\psi = \psi_\varepsilon$ a.e. (F). Inserting (4.2) into (4.3) and observing that $\psi_\varepsilon$ satisfies (2.7) with $\varepsilon$ in place of $\alpha$ (by Lemma 2) yields

$$A(\psi, F) \geq A(\psi_\varepsilon, F) - 4k(\varepsilon)\int_0^c \psi(x)F_\varepsilon(x)\,\mathrm{d}x \geq A(\psi_\varepsilon, F). \tag{4.4}$$

The last inequality follows from (2.7) and the fact that $F_\varepsilon(x) \leq F_\alpha(x)$ for all $x$. □

## 5. A numerical example for Gaussian errors

Assume now that $F = \Phi$, the standard normal distribution. The optimal score functions are then odd functions of the form

$$\psi_\varepsilon(x) = \begin{cases} x - \sqrt{2\pi}\,k(\varepsilon)\left(\Phi(x) - 1 + \tfrac{1}{2}\varepsilon\right)\mathrm{e}^{x^2/2}, & 0 < x < c(\varepsilon), \\ 0, & x \geq c(\varepsilon). \end{cases} \tag{5.1}$$

Table 1 gives values of $c(\varepsilon)$, $k(\varepsilon)$ and the efficiency $e(\psi_\varepsilon, \Phi) = 1/V(\psi_\varepsilon, \Phi)$ when $\varepsilon$ is varied.

Table 1
Values of $e(\psi_\varepsilon, \Phi)$, $c(\varepsilon)$ and $k(\varepsilon)$ for the optimal S-estimators at the standard normal distribution

| $\varepsilon$ | $e$ | $c$ | $k$ |
|---|---|---|---|
| 0.50 | 0.329 | 1.613 | 0.891 |
| 0.45 | 0.414 | 1.767 | 0.794 |
| 0.40 | 0.505 | 1.928 | 0.693 |
| 0.35 | 0.601 | 2.098 | 0.590 |
| 0.30 | 0.697 | 2.283 | 0.485 |
| 0.25 | 0.790 | 2.491 | 0.376 |
| 0.20 | 0.874 | 2.738 | 0.266 |
| 0.15 | 0.944 | 3.055 | 0.155 |
| 0.10 | 0.987 | 3.544 | 0.053 |

Rousseeuw and Yohai (1984) calculate efficiencies and rejection points $c$ when $\psi$ equals Tukey's biweight function. The efficiencies in Table 1 are about 0.03–0.04 higher (0.042 when $\varepsilon = 0.5$). The rejection points $c(\varepsilon)$ are smaller for the estimator (6.1), except when $\varepsilon = 0.5$ ($c(0.5)$ is 0.066 larger). The advantage of higher efficiency should be weighted against the fact that the estimator based on Tukey's biweight function is easier to calculate.

## References


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

Davies, L. (1990), The asymptotics of S-estimators in the linear regression model, *Ann. Statist.* **18**, 1651–1675.

Donoho, D.L. and P.J. Huber (1983). The notion of breakdown point, in: P. Bickel, K. Doksum and J.L. Hodges Jr., eds., *A festschrift for Erich Lehmann* (Wadsworth, Belmont, CA).

Hampel, F.R. (1971), A general qualitative definition of robustness, *Ann. Math. Statist.* **42**, 1887–1896.

Hettmansperger, T.P. and J.W. McKean (1983), A geometric interpretation of inferences based on ranks in the linear model, *J. Amer. Statist. Assoc.* **78**, 885–893.

Hodges Jr., J.L. (1967), Efficiency in normal samples and tolerance of extreme values for some estimates of location, *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* (Univ. California Press, Berkeley, CA) pp. 163–186.

Huber, P.J. (1973), Robust regression: Asymptotics, conjectures and Monte Carlo, *Ann. Statist.* **1**, 799–821.

Hössjer, O. (1991), Rank-based estimates in the linear model with high breakdown point, U.U.D.M. Report 1991:5, Dept. of Math., Uppsala Univ. (Uppsala, Sweden).

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

Jurečková, J. (1971), Nonparametric estimate of regression coefficients, *Ann. Math. Statist.* **42**, 1328–1338.

Jurečková, J. and S. Portnoy (1987), Asymptotics for one-step M-estimators in regression with application to combining efficiency and high breakdown point, *Comm. Statist. Theory Methods* **16**, 2187–2199.

Lopuhaä, H.P. (1989), On the relation between S-estimators and M-estimators of multivariate location and covariance, *Ann. Statist.* **17**, 1662–1683.

Lopuhaä, H.P. and P. Rousseeuw (1991), Breakdown points of affine equivariant estimators of multivariate location and covariance matrices, *Ann. Statist.* **19**, 229–248.

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

Rousseeuw, P.J. (1985), Multivariate estimation with high breakdown point, in: W. Grossmann, G. Pflug, I. Vincze and W. Wertz, eds., *Mathematical Statistics and Applications, Vol. B* (Reidel, Dordrecht) pp. 283–297.

Rousseeuw, P.J. and A. Leroy (1987). *Robust Regression and Outlier Detection* (Wiley, New York).

Rousseeuw, P.J. and V.Y. Yohai (1984), Robust regression by means of S-estimators, in: J. Franke, W. Härdle and R.D. Martin, eds., *Robust and Nonlinear Time Series Analysis*, Lecture Notes in Statistics No. 26 (Springer, New York) pp. 256–272.

Siegel, A.F. (1982), Robust regression using repeated medians, *Biometrika* **69**, 242–244.

Van Eeden, C. (1972), An analogue, for signed rank statistics, of Jurečková's asymptotic linearity theorem for rank statistics, *Ann. Math. Statist.* **43**, 791–802.

Yohai, V. (1987), High breakdown-point and high efficiency robust estimates for regression, *Ann. Statist.* **15**, 642–656.

Yohai, V. and R. Zamar (1988), High breakdown-point estimates of regression by means of the minimization of an efficient scale, *J. Amer. Statist. Assoc.* **83**, 406–413.