The Annals of Statistics
1987, Vol. 15, No. 20, 642–656

HIGH BREAKDOWN-POINT AND HIGH EFFICIENCY ROBUST ESTIMATES FOR REGRESSION

BY VICTOR J. YOHAI

Universidad de Buenos Aires and C.E.M.A.

A class of robust estimates for the linear model is introduced. These estimates, called MM-estimates, have simultaneously the following properties: (i) they are highly efficient when the errors have a normal distribution and (ii) their breakdown-point is 0.5. The MM-estimates are defined by a three-stage procedure. In the first stage an initial regression estimate is computed which is consistent robust and with high breakdown-point but not necessarily efficient. In the second stage an M-estimate of the errors scale is computed using residuals based on the initial estimate. Finally, in the third stage an M-estimate of the regression parameters based on a proper redescending psi-function is computed. Consistency and asymptotical normality of the MM-estimates assuming random carriers are proved. A convergent iterative numerical algorithm is given. Finally, the asymptotic biases under contamination of optimal bounded influence estimates and MM-estimates are compared.

1. Introduction. Consider the usual regression model with random carriers, i.e., we observe $\mathbf{z}_i = (y_i, \mathbf{x}_i)$, $1 \le i \le n$, i.i.d. $(p+1)$-dimensional vectors, where $y_i \in R$, $\mathbf{x}_i \in R^p$ and the prime denotes transpose, satisfying

$$y_i = \theta_0' \mathbf{x}_i + u_i, \qquad 1 \le i \le n, \tag{1.1}$$

where $\theta \in R^p$ is the vector of the regression parameters and u_i is independent of $\mathbf{x}_i$. Let $G_0(\mathbf{x})$ be the distribution of the carrier $\mathbf{x}_i$ and $F_0(u)$ the distribution of the error u_i. Then the distribution of $\mathbf{z}_i$ is given by

$$H_0(\mathbf{z}) = G_0(\mathbf{x}) F_0(y - \theta_0' \mathbf{x}). \tag{1.2}$$

The least-squares estimate (LS estimate) is defined by the value $\hat{\theta}_{\mathrm{LS}}$ which minimizes

$$S(\theta) = \sum_{i=1}^{n} r_i^2(\theta), \tag{1.3}$$

where the residuals $r_i(\theta)$ are defined by

$$r_i(\theta) = y_i - \theta' \mathbf{x}_i. \tag{1.4}$$

When F_0 is normal, $\hat{\theta}_{\mathrm{LS}}$ corresponds to the maximum likelihood estimate. In this case $\hat{\theta}_{\mathrm{LS}}$ is efficient since its covariance matrix attains the Rao–Cramér bound matrix.

However, it is well known that the LS-estimator is not robust: A small fraction of outliers, even one outlier may have a large effect on the estimate.

Received July 1985; revised May 1986.
AMS 1980 *subject classifications.* Primary 62F35; secondary 62J05.
Key words and phrases. Linear model, robust estimation, high breakdown-point, high efficiency.

The degree of robustness of an estimate in the presence of outliers may be measured by the concept of breakdown-point which was introduced by Hampel (1971). Donoho (1982) and Donoho and Huber (1983) gave a finite sample version of this concept which will be used here. The finite sample breakdown-point measures the maximum fraction of outliers which a given sample may contain without spoiling the estimate completely.

Unfortunately, many of the proposals for robust estimation in regression fail to have high breakdown-point. The M-estimates with monotone psi-function introduced by Huber (1973) have breakdown-point 0. Maronna, Bustos and Yohai (1979) showed that GM-estimates have a breakdown-point tending to 0 when p increases. This class contains the optimal bounded influence estimates obtained by Krasker (1980) and Krasker and Welsch (1982). It also contains the optimal estimates with bounded change-of-variance function derived by Ronchetti and Rousseeuw (1985).

In the last years several estimates with high breakdown-point, i.e., 0.5, were proposed. Siegel (1982) proposed the repeated median (RM) estimate. Rousseeuw (1984) proposed the least median of squares (LMS) and the least trimmed squares (LTS) which are defined by the minimization of the median or the trimmed mean of the squares of the residuals, respectively. Rousseeuw and Yohai (1984) proposed a class of estimates based on the minimization of a robust M-estimate of the residual scale (S-estimates).

However all these estimates are highly inefficient when all the observations satisfy the regression model with normal errors. Moreover Siegel's RM-estimator is not affine equivariant. See Remark 2.2 for the definition of affine equivariance.

Rousseeuw (1984) aiming at reconciling high breakdown-point and high efficiency, proposed to use a high breakdown-point estimate followed by a one-step M-estimate or a one-step reweighted least squares. It seems quite plausible that this type of procedure keeps the breakdown-point high and improves the efficiency of the initial estimate. However, the exact breakdown-point of this type of procedure is not known and therefore there is no guarantee that the breakdown-point of the initial estimate is kept unchanged. Moreover the one-step reweighted least-squares estimate, contrary to what happens with the one-step M-estimate, does not have the same asymptotic efficiency as the fully iterated estimate. In fact, its asymptotic efficiency depends on the initial estimate and may be quite difficult to compute.

The purpose of this paper is to present a new class of estimates, which we call MM-estimates, having simultaneously (i) high breakdown-point and (ii) high efficiency under normal errors.

In Section 2 we define the MM-estimates and establish that they have high breakdown-point. We also give another robustness property of the MM-estimates, called here the "exact fit" property, which was introduced by Rousseeuw (1984). In Section 3 we study consistency and in Section 4 asymptotic normality. In Section 5 we give a numerical algorithm for computing the MM-estimates. In Section 6 we compare the asymptotical biases of MM-estimates and optimal GM-estimates under contamination when the z_i's are multivariate normal. The last section is an Appendix with the proofs.

2. MM-estimates. Huber (1981) defines the scale M-estimates as follows. Let ρ be a real function satisfying the following assumptions.

(A1) (i) $\rho(0) = 0$; (ii) $\rho(-u) = \rho(u)$; (iii) $0 \le u \le v$ implies $\rho(u) \le \rho(v)$; (iv) ρ is continuous; (v) let $a = \sup \rho(u)$, then $0 < a < \infty$; (vi) if $\rho(u) < a$ and $0 \le u < v$, then $\rho(u) < \rho(v)$.

Given a sample of size n, $\mathbf{u} = (u_1, u_2, \dots, u_n)$, the scale estimate $s(\mathbf{u})$ is defined as the value of s which is the solution of

$$\boxed{(1/n) \sum_{i=1}^{n} \rho(u_i/s) = b,} \tag{2.1}$$

where b may be defined by $E_\phi(\rho(u)) = b$, and where ϕ stands for the standard normal distribution.

It is easy to show that if

$$c(\mathbf{u}) = \#\{i: 1 \le i \le n,\ u_i = 0\}/n < 1 - (b/a),$$

then (2.1) has a unique solution and this solution is different from 0. If $c(u) \ge 1 - (b/a)$, we define $s(u) = 0$.

Then the MM-estimate is defined in three stages as follows.

STAGE 1. Take an estimate $\mathbf{T}_{0,n}$ of $\boldsymbol{\theta}_0$ with high breakdown-point, possibly 0.5. See Remark 2.4 for the selection of this initial estimate.

STAGE 2. Compute the residuals

$$r_i(\mathbf{T}_{0,n}) = y_i - \mathbf{T}'_{0,n}\mathbf{x}_i, \qquad 1 \le i \le n, \tag{2.2}$$

and compute the M-scale $s_n = s(\mathbf{r}(\mathbf{T}_{0,n}))$ defined by (2.1), using a function ρ_0 satisfying assumption (A1) and using a constant b such that

$$b/a = 0.5, \tag{2.3}$$

where $a = \max \rho_0(u)$. As Huber (1981) proves, (2.3) implies that this scale estimate has breakdown-point equal to 0.5.

STAGE 3. Let ρ_1 be another function satisfying assumption (A1) and such that

$$\rho_1(u) \le \rho_0(u), \tag{2.4}$$

$$\sup \rho_1(u) = \sup \rho_0(u) = a. \tag{2.5}$$

Let $\psi_1 = \rho_1'$. Then the MM-estimate $\mathbf{T}_{1,n}$ is defined as any solution of

$$\sum_{i=1}^{n} \psi_1(r_i(\boldsymbol{\theta})/s_n)\mathbf{x}_i = 0, \tag{2.6}$$

which verifies

$$S(\mathbf{T}_{1,n}) \le S(\mathbf{T}_{0,n}), \tag{2.7}$$

where

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} \rho_1(r_i(\boldsymbol{\theta})/s_n) \tag{2.8}$$

and where $\rho_1(0/0)$ is defined as 0.

REMARK 2.1. Lemma 2.1, proved in the Appendix, implies that the absolute minimum of $S(\boldsymbol{\theta})$ exists. It is clear that this absolute minimum should satisfy (2.6) and (2.7). However any other value of $\boldsymbol{\theta}$ which satisfies (2.6) and (2.7), e.g., a local minimum, will be also a MM-estimate with high breakdown-point and with high efficiency under a regression model with normal errors.

REMARK 2.2. If $\mathbf{T}_{0,n}$ is regression equivariant, i.e., if

$$\begin{aligned}\mathbf{T}_{0,n}&((y_1 + \boldsymbol{\theta}'\mathbf{x}_1, \mathbf{x}_1), (y_2 + \boldsymbol{\theta}'\mathbf{x}_2, \mathbf{x}_2), \dots, (y_n + \boldsymbol{\theta}'\mathbf{x}_n, \mathbf{x}_n))\\ &= \mathbf{T}_{0,n}((y_1, \mathbf{x}_1), (y_2, \mathbf{x}_2), \dots, (y_n, \mathbf{x}_n)) + \boldsymbol{\theta}\end{aligned}$$

and if $\mathbf{T}_{1,n}$ is defined as the absolute minimum of $S(\boldsymbol{\theta})$, then $\mathbf{T}_{1,n}$ will be equivariant too. A similar statement can be made when $\mathbf{T}_{0,n}$ is affine equivariant, i.e., if

$$\mathbf{T}_{0,n}((y_1, A\mathbf{x}_1), \dots, (y_1, A\mathbf{x}_n)) = A^{-1}\mathbf{T}_{0,n}((y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n))$$

for any nonsingular matrix A.

REMARK 2.3. One way of choosing ρ_0 and ρ_1 satisfying (A1), (2.4) and (2.5) is as follows. Let ρ be a function satisfying (A1), and let $0 < k_0 < k_1$. Let $\rho_0(u) = \rho(u/k_0)$ and $\rho_1(u) = \rho(u/k_1)$. The value k_0 should be chosen such that (2.3) holds. The choice of k_1 will determine the asymptotic efficiency of the estimate.

Donoho (1982) and Donoho and Huber (1983) give the following finite sample version of Hampel's breakdown-point concept.

Let $\mathbf{Z}_n = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n)$ be any sample of size n, and let $\mathbf{T} = \{\mathbf{T}_n\}_{n \geq p}$ ($\mathbf{T}_n$ is the estimate corresponding to a sample of size n) be a sequence of estimates. Let

$$b(m, \mathbf{T}, \mathbf{Z}_n) = \sup |\mathbf{T}_{m+n}(\mathbf{Z}_n \cup \mathbf{W}_m) - \mathbf{T}_n(\mathbf{Z}_n)|,$$

where the supremum is taken over all the samples $\mathbf{W}_m$ of size m, $\mathbf{Z}_n \cup \mathbf{W}_m$ denotes the sample of size $n + m$ which contains the observations of both samples and $|\ |$ denotes Euclidean norm. The breakdown-point of $\mathbf{T}$ at the sample $\mathbf{Z}_n$ is defined by

$$\varepsilon_n^*(\mathbf{T}, \mathbf{Z}_n) = \min\{m/(m+n) \colon b(m, \mathbf{T}, \mathbf{Z}_n) = \infty\}.$$

We can interpret ε_n^* as the maximum fraction of outliers that we can add to the original sample without spoiling the estimate completely.

Let

$$c_n = \max_{\boldsymbol{\theta} \in R^p} \#\{i \colon 1 \leq i \leq n \text{ and } \boldsymbol{\theta}'\mathbf{x}_i = 0\}/n. \tag{2.9}$$

Then, if any set of p carriers is linearly independent, we have $c_n = (p-1)/n$. Let $\mathbf{T}_0 = \{\mathbf{T}_{0,n}\}_{n \geq p}$ be the initial sequence of estimates, and $\mathbf{T}_1 = \{\mathbf{T}_{1,n}\}_{n \geq p}$ the corresponding MM-estimate. The following theorem implies that if $\varepsilon_n^*(\mathbf{T}_0, Z_n)$ is asymptotically 0.5 and $c_n = (p-1)/n$, then $\varepsilon_n^*(T_1, Z_n)$ is asymptotically 0.5 too.

THEOREM 2.1. *Suppose that ρ_0 and ρ_1 satisfy assumption* (A1), *that* (2.3), (2.4) *and* (2.5) *hold and $c_n < 0.5$. Then if $\mathbf{T}_0 = \{\mathbf{T}_{0,n}\}_{n \geq p}$ is any sequence of estimates which satisfies* (2.7), *we have*

$$\varepsilon_n^*(\mathbf{T}_1, \mathbf{Z}_n) \geq \min(\varepsilon_n^*(\mathbf{T}_0, \mathbf{Z}_n), (1-2c_n)/(2-2c_n)).$$

REMARK 2.4. A possible choice for $\mathbf{T}_0$ is Siegel's RM-estimate whose breakdown-point is asymptotically 0.5. Another estimate with asymptotical breakdown-point equal to 0.5, but which is affine equivariant was proposed by Leroy and Rousseeuw (1984). This estimate may be considered as a finite variant of Rousseeuw's LMS-estimate, and is defined as follows. For each set of p observations of the sample we compute the value of $\boldsymbol{\theta}$ which fits exactly. Then we have $N = \binom{n}{p}$ estimates, $\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \ldots, \hat{\boldsymbol{\theta}}_N$ of $\boldsymbol{\theta}_0$. For each of these estimates $\hat{\boldsymbol{\theta}}_i$, we compute the residuals $r_{ij} = y_j - \hat{\boldsymbol{\theta}}_i'\mathbf{x}_j$, $i \leq j \leq n$, and $\sigma_i^2 = \operatorname{med}\{r_{ij}^2, 1 \leq j \leq n\}$. Then $\mathbf{T}_n$ is defined as the value $\hat{\boldsymbol{\theta}}_i$ which corresponds to the minimum $\hat{\sigma}_i^2$. We call this estimate finite LMS. If p is large the finite LMS-estimate may be computationally very expensive. Then Leroy and Rousseeuw (1984) propose to use only a sample of all the possible sets of p observations drawn out of the n observations. In this case we can only guarantee breakdown-point 0.5 with some probability which depends on the size of this sample.

Another important robustness property used by Rousseeuw (1984), and called here "exact fit" property (EFP), is the following: An estimate $\mathbf{T}_n$ has the EFP if given any sample of size n, $(y_1, \mathbf{x}_1), (y_2, \mathbf{x}_2), \ldots, (y_n, \mathbf{x}_n)$, for which there exists $\boldsymbol{\theta}$ such that $\#\{i: y_i = \boldsymbol{\theta}'\mathbf{x}_i\} > n/2$, then $\#\{i: y_i = \mathbf{T}_n'\mathbf{x}_i\} > n/2$ too.

The following theorem shows that the MM-estimate inherits the EFP from the initial estimate.

THEOREM 2.2. *Assume that ρ_0 and ρ_1 satisfy* (A1). *Suppose $\mathbf{T}_{0,n}$ has the EFP and let $\mathbf{T}_{1,n}$ be any estimate satisfying* (2.7). *Then $\mathbf{T}_{1,n}$ has the EFP too.*

REMARK 2.5. The RM-, LMS-, finite LMS-, and S-estimates have the EFP. Therefore if we take any of these estimates as $\mathbf{T}_{0,n}$, the MM-estimate $\mathbf{T}_{1,n}$, will also have the EFP.

3. Consistency. In order to prove the consistency of the MM-estimates we need the following additional assumptions.

(A2) The function $g(a) = E_{F_0}(\rho_1((u-a)/\sigma_0))$, where σ_0 is defined by

$$E_{F_0}(\rho_0(u/\sigma_0)) = b, \tag{3.1}$$

has a unique minimum at $a = 0$.

(A3) $P_{G_0}(\boldsymbol{\theta}'\mathbf{x} = 0) < 0.5$ for all $\boldsymbol{\theta} \in R^p$.

If ρ_1 satisfies (A1), then a sufficient condition for (A2), see Lemma 3.1. of Yohai (1985), is given by:

(A2*) The error distribution F_0 has density f_0 with the following properties: (i) f_0 is even, (ii) $f_0(u)$ is monotone nonincreasing in $|u|$, and (iii) $f_0(u)$ is strictly decreasing in $|u|$ in a neighborhood of 0.

Theorems 3.1 and 3.2 establish the consistency of the scale estimate s_n defined in stage 2 and of the MM-sequence of estimates $\{\mathbf{T}_{1,n}\}_{n\geq p}$ of $\boldsymbol{\theta}_0$. The proofs are omitted and may be found in Yohai (1985).

THEOREM 3.1. *Let* $(y_1, \mathbf{x}_1), (y_2, \mathbf{x}_2), \ldots, (y_n, \mathbf{x}_n)$ *be i.i.d. observations with distribution* H_0 *given by* (1.2). *Assume that* ρ_0 *satisfies* (A1) *and* $\{\mathbf{T}_{0,n}\}_{n\geq p}$ *is a sequence of estimates which is strongly consistent for* $\boldsymbol{\theta}_0$. *Then* s_n *is strongly consistent for* σ_0, *where* σ_0 *is defined by* (2.1).

THEOREM 3.2. *Let* $(y_1, \mathbf{x}_1), (y_2, \mathbf{x}_2), \ldots, (y_n, \mathbf{x}_n)$ *be i.i.d. observations with distribution* H_0 *given by* (1.2). *Assume that* ρ_0 *and* ρ_1 *satisfy* (A1) *and that* (A2), (A3), (2.3), (2.4) *and* (2.5) *holds. Assume also that the sequence* $\{\mathbf{T}_{0,n}\}_{n\geq p}$ *is strongly consistent for* $\boldsymbol{\theta}_0$. *Then any other sequence* $\{\mathbf{T}_{1,n}\}_{n\geq p}$ *which satisfies* (2.7) *is strongly consistent too.*

REMARK 3.1. Suppose that (A2) does not hold but $g(a)$ has a unique minimum at $a_0 \neq 0$, and suppose also that the regression model has a constant term, i.e., $\mathbf{x}_{i,p} = 1$ for all i. Then in this case Theorem 3.2 shows that the first $p-1$ coordinates of $\mathbf{T}_{1,n}$ are strongly consistent with their true values and the last coordinate has an asymptotic bias equal to a_0.

4. Asymptotic normality. Asymptotic theory of M-estimates with random carriers can be obtained from Theorem 4.1 in Maronna and Yohai (1981). However, when applied to M-estimates, this theorem requires fourth moments on the $\mathbf{x}_i$'s. We are going to give here sufficient conditions for the asymptotic normality of the MM-estimates which require only second moments on the carriers. We need some additional assumptions.

(A4) ρ_1 is even, twice continuously differentiable and there exists m such that $|u| \geq m$ implies $\rho_1(u) = a$.

(A5) G_0 has second moments and

$$\mathbf{V} = E_{G_0}(\mathbf{x}_i\mathbf{x}_i') \tag{4.1}$$

is nonsingular.

The following theorem gives the asymptotical normality of M-estimates with scale estimated separately, with include as a special case the MM-estimates. The proof is omitted here and can be found in Yohai (1985).

THEOREM 4.1. *Let $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$ be i.i.d. with distribution H_0 given by* (1.2). *Assume ρ_1 satisfies* (A4) *and G_0 satisfies* (A5). *Let s_n be an estimate of the error scale which converges strongly to σ_0. Let $\mathbf{T}_n$ be a sequence of estimates which satisfies* (2.6) *and which is strongly consistent to the true value $\boldsymbol{\theta}_0$. Then*

$$n^{1/2}(\mathbf{T}_n - \boldsymbol{\theta}_0) \to_d N\big(\mathbf{0}, \sigma_0^2\big[A(\psi_1, F_0)/B^2(\psi_1, F_0)\big]\mathbf{V}^{-1}\big), \tag{4.2}$$

where $\to_d$ denotes convergence in distribution,

$$A(\psi, F) = E_F\big(\psi^2(u/\sigma_0)\big) \tag{4.3}$$

and

$$B(\psi, F) = E_F\big(\psi'(u/\sigma_0)\big). \tag{4.4}$$

REMARK 4.1. Let ρ_0 and ρ_1 be as in Remark 2.3, where $\rho(u/k)$ is equivalent to u^2 for large k. For example, let ρ be given by

$$\rho_B(u) = \begin{cases} u^2/2 - u^4/2 + u^6/6 & \text{if } |u| \le 1, \\ 1/6 & \text{if } |u| > 1, \end{cases} \tag{4.5}$$

which corresponds to the bisquare psi-function

$$\psi_B(u) = \begin{cases} u(1 - u^2)^2 & \text{if } |u| \le 1, \\ 0 & \text{if } |u| > 1. \end{cases} \tag{4.6}$$

Suppose also that the MM-estimate is computed using s_n defined in stage 2 with $b = E_\Phi(\rho(u/k_0)$. Then if the u_i's are $N(0,1)$, we have $\sigma_0^2 = 1$ and according to Theorem 4.1, the asymptotic variance of the MM-estimate depends only on k_1. Therefore we can choose k_1 so that the MM-estimate can be highly efficient without affecting its breakdown-point, which depends only on the choice of k_0. When $\rho = \rho_B$ the value k_0 that makes (2.3) holds is 1.56, the corresponding $b = 0.0833$ and the value k_1 which gives efficiency 0.95 for normal errors is 4.68. However it is worth noting that even if this estimate has breakdown-point 0.5, it will be less robust, i.e., more sensitive to outliers, than the estimates corresponding to a smaller value of k_1.

REMARK 4.2. According to Theorem 4.1, the asymptotic efficiency of the MM-estimates with respect to the LS-estimate is independent of the distribution G_0 of the carriers. This represents a clear advantage over the GM-estimates whose efficiency may be seriously affected by high leverage observations; see Maronna, Bustos and Yohai (1979).

5. Computing algorithm. Here we propose a computing algorithm for the MM-estimate which is a modified version of the iterated weighted least-squares (IWLS) algorithm used for computing M-estimates [see Huber (1981), Chapter 7]. Let $\mathbf{z}_i = (y_i, \mathbf{x}_i)$, $1 \le i \le n$, be a sample of size n and suppose that we have already computed the initial estimate $\mathbf{T}_{0,n}$ and the scale estimate s_n defined in stage 2. For each $\mathbf{t} \in R^p$ define the weights

$$w_i(\mathbf{t}) = \psi_1\big(r_i(\mathbf{t})/s_n\big)/\big(r_i(\mathbf{t})/s_n\big). \tag{5.1}$$

Also define

$$\text{(5.2)}\qquad \mathbf{g}(\mathbf{t}) = \left(1/s_n^2\right)\sum_{i=1}^{n} w_i(\mathbf{t})r_i(\mathbf{t})\mathbf{x}_i = (1/s_n)\sum_{i=1}^{n}\psi_1\left(r_i(\mathbf{t})/s_n\right)\mathbf{x}_i$$

and

$$\text{(5.3)}\qquad \mathbf{M}(\mathbf{t}) = \left(1/s_n^2\right)\sum_{i=1}^{n} w_i(\mathbf{t})\mathbf{x}_i\mathbf{x}_i'.$$

It is easy to show that $-\mathbf{g}(\mathbf{t})$ is the gradient of $S(\mathbf{t})$. The recursion step of the IWLS is defined as follows. If $\mathbf{t}^{(j)}$ is the value of the estimate in the jth step, then $\mathbf{t}^{(j+1)}$ is defined by

$$\text{(5.4)}\qquad \mathbf{t}^{(j+1)} = \mathbf{t}^{(j)} + \Delta(\mathbf{t}^{(j)}),$$

where

$$\text{(5.5)}\qquad \Delta(\mathbf{t}) = \mathbf{M}^{-1}(\mathbf{t})\mathbf{g}(\mathbf{t}).$$

Using this recursion we cannot guarantee that $S(\mathbf{t}^{(j+1)}) \le S(\mathbf{t}^{(j)})$ and, therefore, if $\mathbf{T}_{1,j}$ is computed as a limit of the sequence $\mathbf{t}^{(j)}$, (2.7) may not hold. We propose the following modification. Take $0 < \delta < 1$; then since $-\mathbf{g}(\mathbf{t})$ is the gradient of $S(\mathbf{t})$, we can find an integer k such that

$$\text{(5.6)}\qquad S\left(\mathbf{t}^{(j)} + \Delta(\mathbf{t}^{(j)})/2^k\right) \le S(\mathbf{t}^{(j)}) - \delta\left(\Delta(\mathbf{t}^{(j)})/2^k\right)'g(\mathbf{t}^{(j)}).$$

Let $k_{1,j}$ be the minimum of such k's and let $k_{2,j}$ be the value of k, $0 \le k \le k_{1,j}$, which gives the minimum of $S(\mathbf{t}^{(j)} + \Delta(\mathbf{t}^{(j)})/2^k)$. Then define the recursion step by

$$\text{(5.7)}\qquad \mathbf{t}^{(j+1)} = \mathbf{t}^{(j)} + \left(1/2^{k_{2,j}}\right)\Delta(\mathbf{t}^{(j)})$$

starting with $\mathbf{t}^{(0)} = \mathbf{T}_{0,n}$. Clearly we now have $S(\mathbf{t}^{(j+1)}) \le S(\mathbf{t}^{(j)})$. The following theorem shows that any limit point of the sequence $\mathbf{t}^{(j)}$ satisfies (2.6) and (2.7), and therefore is an MM-estimate.

THEOREM 5.1. *Suppose ρ_0 and ρ_1 satisfy* (A1), (2.3), (2.4) *and* (2.5) *hold*, ψ_1 *is continuous*, $\lim_{u\to 0}\psi_1(u)/u > 0$, $u \ne 0$, *and* $\rho_1(u) < a$ *implies* $\psi_1(u) > 0$ *and finally* $c_n < 0.5$. *Then if* $\mathbf{t}^{(j)}$ *is defined by* (5.7) *with* $\mathbf{t}^{(0)} = \mathbf{T}_{0,n}$,

(i) *the sequence* $\mathbf{t}^{(j)}$ *is bounded;*
(ii) *any limit point of* $\mathbf{t}^{(j)}$ *satisfies* (2.6) *and* (2.7);
(iii) *if* $\mathbf{t}_0$ *and* $\mathbf{t}_1$ *are two limit points of* $\mathbf{t}^{(j)}$, *we have* $S(\mathbf{t}_0) = S(\mathbf{t}_1)$.

6. Bias under contamination. The influence curve was introduced by Hampel (1974) to measure the degree of bias robustness of an estimate when the distribution of a central nominal model is subject to an infinitesimal contamination. Suppose that the distribution of $\mathbf{z}_i = (y_i, \mathbf{x}_i)$, H_0, is given by (1.2) and that the sequence of estimates $\{\mathbf{T}_n\}_{n\ge p}$ is defined by a functional $\mathbf{T}$ applied to the empirical distribution, i.e., $\mathbf{T}_n = \mathbf{T}(H_n)$ which is consistent for $\boldsymbol{\theta}_0$, i.e., $\mathbf{T}(H_0) = \boldsymbol{\theta}_0$. If H_0 is subject to a contamination of size ε, with the distribution $\delta_{y,\mathbf{x}}$ concentrated at the point $(y, \mathbf{x})$, the asymptotic bias of the estimate given by the functional $\mathbf{T}$ is

$$\mathbf{b}(\mathbf{T}, H_0, \varepsilon, y, \mathbf{x}) = \mathbf{T}\big((1-\varepsilon)H_0 + \varepsilon\delta_{y,\mathbf{x}}\big) - \mathbf{T}(H_0)$$

and the influence curve is defined by

$$\mathbf{IC}(\mathbf{T}, H_0, y, \mathbf{x}) = \lim_{\varepsilon \to 0} \mathbf{b}(\mathbf{T}, H_0, \varepsilon, y, \mathbf{x})/\varepsilon.$$

Since the MM-estimate is an M-estimate, its influence curve [see Krasker (1980)] is given by

$$\mathbf{IC}(\mathbf{T}, H_0, y, \mathbf{x}) = \psi_1(y - \boldsymbol{\theta}_0'\mathbf{x})\mathbf{x}\sigma_0^2\big(B(\psi_1, F_0)\mathbf{V}\big)^{-1},$$

where $\mathbf{V}$ is given in (4.1) and $B(\psi, F)$ in (4.4).

Therefore, the influence curve of the MM-estimates is not bounded. However, for practical purposes it is more meaningful to consider the case of small but positive contamination size. For this purpose we define the ε-influence curve by

$$\mathbf{IC}_\varepsilon(\mathbf{T}, H_0, y, \mathbf{x}) = \mathbf{b}(\mathbf{T}, H_0, \varepsilon, y, \mathbf{x})/\varepsilon.$$

It may be proved, by arguments similar to those used in Theorem 2.1, that if $\mathbf{T}$ is an MM-estimate, then $\mathbf{T}((1-\varepsilon)H_0 + \varepsilon H^*)$ is bounded in H^* for any $\varepsilon < 0.5$, and consequently $\mathbf{IC}_\varepsilon(\mathbf{T}, H_0, y, \mathbf{x})$ is bounded in $(y, \mathbf{x})$ too.

In this section we compare the ε-influence curves of MM-estimates and the optimal Krasker and Welsch (1982) bounded influence estimates (KW-estimates) when F_0 and G_0 are normals.

The MM-estimate considered here is based in the bisquare rho-function given by (4.5): $\rho_i(u) = \rho_B(u/k_i)$, $i = 0, 1$. The values of k_0 and k_1 are those given in Remark 4.1. and correspond to an estimate with breakdown-point 0.5 and efficiency 0.95 for normal errors. As initial estimate $\mathbf{T}_0$ we use the S-estimate with rho-function ρ_0. Rousseeuw and Yohai (1984) proved that this estimate has breakdown-point 0.5. We denote this MM-estimate by $\mathbf{T}_1$.

The KW-estimate belongs to the class of GM-estimates and is defined as a solution of

$$\sum_{i=1}^{n} \psi_{H,k}\big(((y_i - \boldsymbol{\theta}'\mathbf{x}_0)/s_n) \cdot |\mathbf{x}_i|_{\boldsymbol{\Sigma}}\big)(\mathbf{x}_i/|\mathbf{x}_i|_{\boldsymbol{\Sigma}}) = \mathbf{0},$$

where $|\mathbf{x}|_{\boldsymbol{\Sigma}} = (\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x})^{1/2}$, $\boldsymbol{\Sigma}$ is the covariance matrix of $\mathbf{x}_i$ (supposed known here) and $\psi_{H,k}$ is the Huber psi-function given by $\psi_{H,k}(u) = \mathrm{sign}(u)\max(|u|, k)$; s_n is estimated simultaneously using the same equation as in stage 2 of the MM-estimate $\mathbf{T}_1$, i.e., using the equation of an M-estimate of scale based on ρ_0.

The value of the constant k is chosen so that the asymptotical efficiency of this estimate, which we denote by $\mathbf{T}_2$, be 0.95 when the distribution H_0 of $\mathbf{z} = (y, \mathbf{x})$ is multivariate normal and may be found in Maronna, Bustos and Yohai (1979). This estimate has the property of minimizing the invariant gross error sensitivity [see Krasker and Welsch (1982)] defined by

$$\gamma^*(\mathbf{T}) = \Big(\sup_{y,\mathbf{x}} \mathbf{IC}(\mathbf{T}, H_0, y, \mathbf{x})'\mathbf{V}^{-1}\mathbf{IC}(\mathbf{T}, H_0, y, \mathbf{x})\Big)^{1/2},$$

TABLE 1
Gross error sensitivities ε

	$\varepsilon = 0.10$		$\varepsilon = 0.15$		$\varepsilon = 0.20$	
p	T_1	T_2	T_1	T_2	T_1	T_2
1	8.7	3.8	8.1	5.2	8.3	18.4
2	8.7	4.7	8.1	7.4	8.3	∞
3	8.7	5.2	8.1	9.0	8.3	∞
5	8.7	6.5	8.1	15.3	8.3	∞
10	8.7	9.5	8.1	∞	8.3	∞

subject to the restriction that the trace of its asymptotical covariance matrix be less than $1.05 = 1/0.95$ times the trace of the covariance matrix of the LS-estimate.

Without loss of generality we may assume that $\boldsymbol{\theta}_0 = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathbf{I}$ and, therefore, $y = u$ has distribution $N(0,1)$.

In Table 1 we show the values of a positive-ε version of the gross error sensitivity given by

$$(6.1) \qquad \gamma_\varepsilon^*(\mathbf{T}) = \left(\sup_{y,\mathbf{x}} \mathbf{IC}_\varepsilon(\mathbf{T}, H_0, y, \mathbf{x})'\mathbf{V}^{-1}\mathbf{IC}_\varepsilon(\mathbf{T}, H_0, y, \mathbf{x})\right)^{1/2}$$

for $\varepsilon = 0.1, 0.15$ and 0.2, $p = 1, 2, 3, 5, 10$ and $\mathbf{T}$ equal to $\mathbf{T}_1$ and $\mathbf{T}_2$.

It is straightforward to show that for the both estimates $\mathbf{T}_1$ and $\mathbf{T}_2$, γ_ε^* is obtained by restricting the supremum in (6.1) to $\mathbf{x}$ varying in the direction of the first coordinate (or any other fixed direction). As a consequence of this, only the first coordinates of the estimates are biased, and this will simplify considerably the computation of these estimates. Moreover, in the case of $\mathbf{T}_1$ it is easy to see that γ_ε^* is independent of p, and therefore we will need to compute it only for $p = 1$. Then in order to compute $\gamma_\varepsilon^*(\mathbf{T}_1)$ in the case of $p = 1$, we compute $T_1((1-\varepsilon)H_0 + \varepsilon\delta_{y,x})$ numerically for each pair (y, x), and then using the routine ZXMIN of the IMSL (1982) library we find its maximum value. Computing $\gamma_\varepsilon^*(T_2)$ is easier since it may be shown that the maximum of $T_2((1-\varepsilon)H_0 + \varepsilon\delta_{y,x})$ is obtained when $y \to \infty$, $x_1 \to \infty$ and $y/x_1 \to \infty$. This property follows from the monotony of $\psi_{H,k}$.

We observe that $\gamma_\varepsilon^*(\mathbf{T}_2) > \gamma_\varepsilon^*(\mathbf{T}_1)$ for $p = 10$ if $\varepsilon \geq 0.1$, for $p = 3$ if $\varepsilon \geq 0.15$ and for $p = 1$ if $\varepsilon \geq 0.20$. This shows that the infinitesimal gross error sensitivity may not be enough to compare the bias robustness of the two estimates even for small ε. This table also shows that the MM-estimate $\mathbf{T}_1$ may be better than the optimal bounded influence estimate $\mathbf{T}_2$ in terms of γ_ε^*, especially for large p.

The following example taken from Rousseeuw and Yohai (1984) will illustrate the robustness of the MM-estimates in the presence of a large fraction of outliers. The dependent variable y is the annual number of international calls made from Belgium and the independent variable x is the year. These variables contain heavy contamination from 1964 to 1969 due to the fact that a different recording system was used (the total number of minutes was registered).

TABLE 2
Data for the example (number of calls in ten of millions)

Year	Number of calls	Year	Number of calls	Year	Number of calls
50	0.44	58	1.06	66	14.20*
51	0.47	59	1.20	67	15.90*
52	0.47	60	1.35	68	18.20*
53	0.59	61	1.49	69	21.20*
54	0.66	62	1.61	70	4.30
55	0.73	63	2.12	71	2.40
56	0.81	64	11.90*	72	2.70
57	0.88	65	12.40*	73	2.90

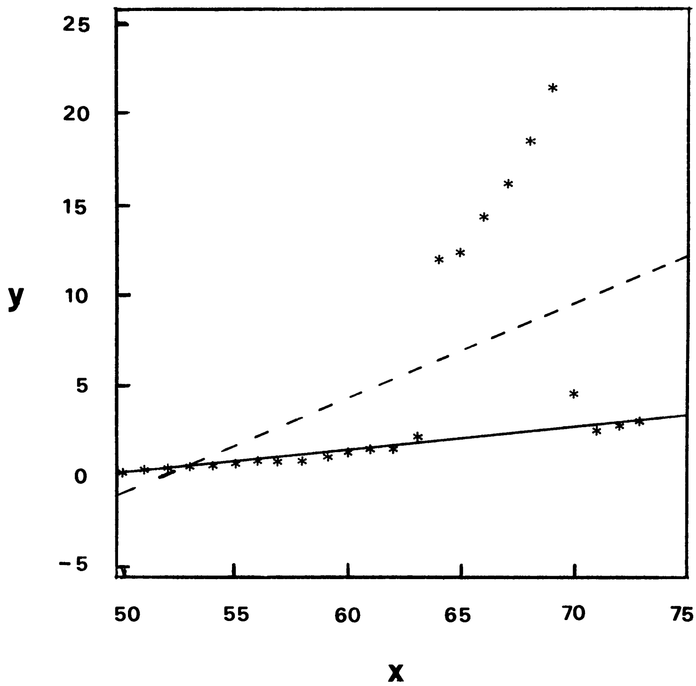

FIG. 1.

The data are shown in Table 2 and the spurious observations are marked with an asterisk. The LS-estimate gives $y = 0.504x - 26.01$ and corresponds to the dotted line in Figure 1. The KW-estimate gives a very similar result: $y = 0.489x - 25.16$. The MM-estimate gives $y = 0.11x - 5.24$ and is plotted in Figure 1 as a solid line. The initial estimator used for the MM-estimator was the finite variant of the LMS-estimator described in Remark 2.4.

We can observe in Figure 1 that contrary to what happens with the LS- and KW-estimates, the MM-estimate is not very much influenced for the outliers. The KW-estimate is not plotted but it is almost identical to the LS-estimate.

APPENDIX

Before proving Theorem 2.1 we will prove the following:

LEMMA 2.1. *Let* $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n)$ *be any sample of size* n *and let* c_n *be given by* (2.10). *Consider the same assumptions as in Theorem* 2.1. *Then given* $\varepsilon < (1 - 2c_n)/(2 - 2c_n)$ *and* k_0, *there exists* k_1 *such that* $m/n + m \le \varepsilon$ *and* $s_{m+n} \le k_0$ *imply*

$$\inf_{|\boldsymbol{\theta}| \ge k_1} \sum_{i=1}^{m+n} \rho_1(r_i(\boldsymbol{\theta})/s_{m+n}) > \sum_{i=1}^{m+n} \rho_1(r_i(\mathbf{T}_{0,\, m+n})/s_{m+n})$$

for all samples $\mathbf{Z}_n \cup \mathbf{W}_m$ *with* $\#\mathbf{W}_m = m$ *and where* $|\ |$ *denotes Euclidean norm.*

PROOF. By definition of c_n, we have for all $\boldsymbol{\theta}$

$$\#\{i{:}\ 1 \le i \le n,\ |\boldsymbol{\theta}'\mathbf{x}_i| > 0\}/n \ge 1 - c_n.$$

Take $c_n^* > c_n$ such that $\varepsilon < (1 - 2c_n^*)/(2 - 2c_n)$ too. Therefore, using a compacity argument we can find $\delta > 0$ such that

$$\inf_{|\boldsymbol{\theta}|=1} \#\{i{:}\ 1 \le i \le n,\ |\boldsymbol{\theta}'\mathbf{x}_i| > \delta\}/n \ge 1 - c_n^*. \tag{A.1}$$

Since $1 - \varepsilon > 1/(2 - 2c_n^*)$, we can find $a_0 < a$ such that $m/(n + m) \le \varepsilon$ implies

$$a_0 n/(n + m) \ge (1 - \varepsilon) a_0 > a/(2 - 2c_n^*). \tag{A.2}$$

By (2.5) there exists k_2 such that $\rho_1(k_2) = a_0$ and let

$$k_1 = \Big(\max_{1 \le i \le n} |y_i| + k_0 k_2\Big)/\delta.$$

Therefore, using the monotonicity of ρ_1, (A.1) and (A.2) we see that $m/(m + n) \le \varepsilon$ implies

$$\begin{aligned}
&\inf_{|\boldsymbol{\theta}| \ge k_1,\, s_{m+n} \le k_0} \sum_{k=1}^{m+n} \rho_1(r_i(\boldsymbol{\theta})/s_{m+n}) \\
&\quad \ge \inf_{|\boldsymbol{\theta}|=1} \sum_{i=1}^{n} \rho_1\big((|y_i| - k_1|(\boldsymbol{\theta}'\mathbf{x}_i)|)/k_0\big) \\
&\quad \ge n(1 - c_n^*)\rho_1(k_2) = n(1 - c_n^*)a_0 > (n + m)a/2.
\end{aligned}$$

On the other hand by (2.1), (2.4), and (2.5) we have

$$\sum_{i=1}^{m+n} \rho_1(r_i(\mathbf{T}_{0,\, m+n})/s_{m+n}) \le \sum_{i=1}^{m+n} \rho_0(r_i(\mathbf{T}_{0,\, m+n})/s_{m+n}) \le (m + n)a/2.$$

Then the lemma follows. □

PROOF OF THEOREM 2.1. According to (2.7) and Lemma 2.1, it is enough to show that for any $\varepsilon < \min(\varepsilon_n^*(\mathbf{T}_0, \mathbf{Z}_n), 0.5)$, there exists k_0 such that for any sample $\mathbf{Z}_n \cup \mathbf{W}_m$, where $\#\mathbf{W}_m = m$ and $m/(n+m) \le \varepsilon$, we have $s_{m+n}(\mathbf{Z}_n \cup \mathbf{W}_m) \le k_0$.

Since $\varepsilon < \varepsilon_n^*(\mathbf{T}_0, \mathbf{Z}_n)$, there exists k_1 such that

$$\mathbf{T}_{0,\, m+n}(\mathbf{Z}_n \cup \mathbf{W}_m) \le k_1 \qquad \forall\, \mathbf{W}_m.$$

Therefore, there exists k_2 such that

$$\sup_{1 \le i \le n} r_i\big(\mathbf{T}_{0,\, m+n}(\mathbf{Z}_n \cup \mathbf{W}_m)\big) \le k_2, \qquad 1 \le i \le n. \tag{A.3}$$

Since $\varepsilon < 0.5$, by (2.3) we can find $\gamma > 0$ such that $\varepsilon a + \gamma < b$. Let δ be defined by $\rho_0(\delta) = \gamma$ and let $k_0 = k_2/\delta$. Then using (A.3) we have

$$\begin{aligned}
&(1/(n+m)) \sum_{i=1}^{m+n} \rho_0\big(r_i(\mathbf{T}_{0,\, m+n}(\mathbf{Z}_n \cup \mathbf{W}_m))/k_0\big) \\
&\quad \le (1/(n+m)) \sum_{i=1}^{n} \rho_0\big(r_i(\mathbf{T}_{0,\, m+n}(\mathbf{Z}_n \cup \mathbf{W}_m))/k_0\big) \\
&\qquad + (1/(n+m)) \sum_{i=n+1}^{n+m} \rho_0\big(r_i(\mathbf{T}_{0,\, m+n}(\mathbf{Z}_n \cup \mathbf{W}_m))/k_0\big) \\
&\quad \le (n/(n+m))\rho_0(k_2/k_0) + (m/(n+m))a \le \rho_0(\delta) + \varepsilon a \\
&\quad = \varepsilon a + \gamma < b.
\end{aligned}$$

Therefore, $s_{m+n}(\mathbf{Z}_n \cup \mathbf{W}_m) \le k_0$. This proves the theorem. □

PROOF OF THEOREM 2.2. Let $\mathbf{Z}_n = \{\mathbf{z}_1 = (y_1, \mathbf{x}_1), \dots, \mathbf{z}_n = (y_n, \mathbf{x}_n)\}$ be a sample and suppose that there exists $\boldsymbol{\theta}^*$ such that

$$\#\{i: 1 \le i \le n,\ y_i - \boldsymbol{\theta}^{*\prime}\mathbf{x}_i = 0\} > n/2.$$

Since $\mathbf{T}_{0,\, n}$ have the EFP we have

$$\#\{i: 1 \le i \le n,\ y_i - \mathbf{T}_{0,\, n}'\mathbf{x}_i = 0\} > n/2,$$

and, therefore, $s_n = 0$. Then for any $\boldsymbol{\theta}$ we have

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} \rho_1\big((y_i - \boldsymbol{\theta}'\mathbf{x}_i)/s_n\big) = a\#\{i: 1 \le i \le n,\ y_i - \boldsymbol{\theta}'\mathbf{x} \ne 0\}.$$

Then $S(\mathbf{T}_{0,\, n}) < an/2$ and, therefore, (2.7) implies $S(\mathbf{T}_{1,\, n}) < an/2$ too. This implies $\#\{i: 1 \le i \le n,\ y - \mathbf{T}_{1,\, n}'\mathbf{x}_i = 0\} > n/2$, and, therefore, $\mathbf{T}_{1,\, n}$ has the EFP too. □

PROOF OF THEOREM 5.1. Since $S(\mathbf{t}^{(i)}) \le S(\mathbf{T}_{0,\, n})$, (i) follows from Lemma 2.1.

In order to prove (ii), is enough to show that $\lim_{j \to \infty} \mathbf{g}(\mathbf{t}^{(j)}) = \mathbf{0}$. Suppose this is not true; then there exists a subsequence $\mathbf{t}^{(i_j)}$ such that $\lim_{j \to \infty} \mathbf{t}^{(i_j)} = \mathbf{t}^*$ with $\mathbf{g}(\mathbf{t}^*) \ne \mathbf{0}$. We will show that $\mathbf{M}(\mathbf{t}^*)$ is positive definite. Let $m = \#\{i: \rho_1(r_i(\mathbf{t}^*)/s_n) = a\}$. We will show that $m \le n/2$. We have

$$S(\mathbf{t}^*) = \sum_{i=1}^{n} \rho_1\big(r_i(\mathbf{t}^*)/s_n\big) \ge ma. \tag{A.4}$$

According to the definition of s_n, (2.3) and (2.4), we get $S(\mathbf{T}_{0,n}) \le na/2$. Then, since $S(\mathbf{t}^*) \le S(\mathbf{T}_{0,n})$, from (A.4) we get $m \le n/2$. Let $H = \{i\colon \rho_1(r_i(\mathbf{t}^*)/s_n) < a)\}$ and $w_0 = \inf_{i \in H} w_i(\mathbf{t}^*)$. Then $w_0 > 0$ and $\mathbf{M}(\mathbf{t}^*) \ge w_0 \Sigma_{i \in H} \mathbf{x}_i \mathbf{x}_i'$. Since $c_n < 0.5$ and $\#H \ge n/2$ we have rank $\{\mathbf{x}_i\colon i \in H\} = p$ and, therefore, by (A.5), we get that $\mathbf{M}(\mathbf{t}^*)$ is positive definite.

Let $\lambda_{1,i} = (1/2)^{k_{1,i}}$ and $\lambda_{2,i} = (1/2)^{k_{2,i}}$. We can assume without loss of generality that $\lim_{j\to\infty} \lambda_{1,i_j} = \lambda_1^*$. We will show that $\lambda_1^* > 0$. Since $\mathbf{g}(\mathbf{t})$ and $\mathbf{M}(\mathbf{t})$ are continuous, $\mathbf{g}(\mathbf{t}^*) \ne \mathbf{0}$ and $\mathbf{M}(\mathbf{t}^*)$ is positive definite, there exists $\varepsilon > 0$ such that $|\mathbf{t}_1 - \mathbf{t}^*| \le \varepsilon$ and $|\mathbf{t}_2 - \mathbf{t}^*| \le \varepsilon$ imply

$$\begin{aligned} &|\mathbf{g}(\mathbf{t}_1)'\mathbf{M}^{-1}(\mathbf{t}_2)\mathbf{g}(\mathbf{t}_2) - \mathbf{g}(\mathbf{t}_2)'\mathbf{M}^{-1}(\mathbf{t}_2)g(\mathbf{t}_2)|/\mathbf{g}(\mathbf{t}_2)'\mathbf{M}^{-1}(\mathbf{t}_2)\mathbf{g}(\mathbf{t}_2) \\ &\qquad \le 1 - \delta, \end{aligned} \tag{A.5}$$

$$|\mathbf{g}(\mathbf{t}_1)| \le 2|\mathbf{g}(\mathbf{t}^*)| \tag{A.6}$$

and

$$\mu_p(\mathbf{t}_1) \le 2\mu_p(\mathbf{t}^*), \tag{A.7}$$

where $\mu_p(\mathbf{t})$ is the maximum eigenvalue of $\mathbf{M}^{-1}(\mathbf{t})$.

Let j^* be such that $j > j^*$ implies $|\mathbf{t}^{(i_j)} - \mathbf{t}^*| \le \varepsilon/2$, then $\Delta(\mathbf{t}^{(i_j)}) \le 4\mu_p(\mathbf{t}^*)|\mathbf{g}(\mathbf{t}^*)|$ for $j > j^*$. Then $\lambda \le \varepsilon/(8\mu_p(\mathbf{t}^*)|\mathbf{g}(\mathbf{t}^*)|)$ and $j > j^*$ implies $|\lambda \Delta(\mathbf{t}^{(i_j)})| \le \varepsilon/2$ and, therefore,

$$S\big(\mathbf{t}^{(i_j)} + \lambda\Delta(\mathbf{t}^{(i_j)})\big) = S(\mathbf{t}^{(i_j)}) - \lambda\mathbf{g}(\mathbf{s})'\Delta(\mathbf{t}^{(i_j)}), \tag{A.8}$$

where $|\mathbf{s} - \mathbf{t}^*| \le \varepsilon$. Then using (A.5) and (A.8) we have

$$S\big(\mathbf{t}^{(i_j)} + \lambda\Delta(\mathbf{t}^{(i_j)})\big) \le S(\mathbf{t}^{(i_j)}) - \delta\mathbf{g}(\mathbf{t}^{(i_j)})'\big(\lambda\Delta(\mathbf{t}^{(i_j)})\big).$$

This implies $k_{1,i_j} \le \ln(8\mu_p(t^*)|g(\mathbf{t}^*)|/\varepsilon)/\ln 2 + 1$ and, therefore, $\lambda_1^* > 0$.
We can find j_1^* such that $j > j_1^*$ implies $\lambda_{1,i_j} > \lambda_1^*/2$ and

$$\mathbf{g}(\mathbf{t}^{(i_j)})'\mathbf{M}^{-1}(\mathbf{t}^{(i_j)})\mathbf{g}(\mathbf{t}^{(i_j)}) \ge \mathbf{g}(\mathbf{t}^*)'\mathbf{M}^{-1}(\mathbf{t}^*)g(\mathbf{t}^*)/2.$$

Then according to the definition of $\lambda_{2,i}$ it is easy to show that we have

$$S(\mathbf{t}^{(i_j+1)}) \le S(\mathbf{t}^{(i_j)}) - (1/4)\delta\lambda_1^* g(\mathbf{t}^*)'\mathbf{M}^{-1}(\mathbf{t}^*)\mathbf{g}(\mathbf{t}^*). \tag{A.9}$$

Since $S(\mathbf{t}) \ge 0$ for all $\mathbf{t}$ and $g(\mathbf{t}^*)'\mathbf{M}^{-1}(\mathbf{t}^*)g(\mathbf{t}^*) > 0$, (A.9) can not hold for all $j > j_1^*$. Therefore, $\lim_{j\to\infty} \mathbf{g}(\mathbf{t}^{(j)}) = \mathbf{0}$. This proves (ii).

(iii) follows immediately from the fact that $S(\mathbf{t}^{(i_j)})$ is monotone decreasing. □

REFERENCES

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Harvard Univ.

DONOHO, D. L. and HUBER, P. J. (1983). The notion of breakdown-point. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, Calif.

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

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

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

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

IMSL (1982). *IMSL Library Reference Manual*, 9th ed. International Mathematical and Statistical Libraries, Houston, Tex.

KRASKER, W. S. (1980). Estimation in linear regression models with disparate data points. *Econometrica* **48** 1333–1346.

KRASKER, W. S. and WELSCH, R. E. (1982). Efficient bounded influence regression estimation. *J. Amer. Statist. Assoc.* **77** 595–604.

LEROY, A. and ROUSSEEUW, P. (1984). PROGRESS: A program for robust regression. Technical Report CS00TW/201, Center for Statistics and Operations Research, VUB Brussels.

MARONNA, R., BUSTOS, O. H. and YOHAI, V. J. (1979). Bias and efficiency robustness of general M-estimators for regression with random carriers. In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds.). Springer, New York.

MARONNA, R. A. and YOHAI, V. J. (1981). Asymptotic behaviour of general M-estimates for regression and scale with random carriers. *Z. Wahrsch. verw. Gebiete* **58** 7–20.

RONCHETTI, E. and ROUSSEEUW, P. J. (1985). Change of variance sensitivities in regression analysis. *Z. Wahrsch. verw. Gebiete* **68** 503–519.

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

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

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

YOHAI, V. J. (1985). High breakdown-point and high efficiency robust estimates for regression. Technical Report No. 66, Dept. Statistics, Univ. Washington, Seattle.

AVENUE LIBERTADOR 5256
PISO 7
1426 BUENOS AIRES
ARGENTINA