High Breakdown Point and High Efficiency Robust Estimates for Regression

by

Victor J. Yohai

TECHNICAL REPORT No. 66

August 1985

Department of Statistics
University of Washington
Seattle, Washington 98195
High Breakdown-point and high efficiency robust estimates for regression

by Victor J. Yohai

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

SUMMARY

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 normal distribution, (ii) they are qualitatively robust and (iii) their breakdown-points 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 qualitatively robust and with high breakdown-point but not necessarily efficient. In the second stage a M-estimate of the errors scale is computed using residuals based on the initial estimate. Finally in the third stage a 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 is proved. A convergent iterative numerical algorithm is given. Finally, the asymptotic biases under contamination of optimal bounded influence estimates and MM-estimates are compared.

Key words and phrases: linear model, robust estimation, high breakdown-point, high efficiency.
AMS 1980 subject classification:
Primary 62P35 Secondary 62J05
Short Title: Robust estimates for regression.
1. Introduction. Consider the usual regression model with random carriers, i.e., we observe $z_i = (y_i', x_i')$, $1 \leq i \leq n$, i.i.d. $(p+1)$-dimensional vectors, where $y_i \in \mathbb{R}$, $x_i \in \mathbb{R}^p$ and denotes transpose, satisfying

\begin{equation}
(1.1) \quad y_i = \varrho_o' x_i + u_i, \quad 1 \leq i \leq n,
\end{equation}

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

\begin{equation}
(1.2) \quad H_0(z) = G_0(x)F_0(y-\varrho_o'x).
\end{equation}

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

\begin{equation}
(1.3) \quad S(\varrho) = \sum_{i=1}^{n} r_i^2(\varrho),
\end{equation}

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

\begin{equation}
(1.4) \quad r_i(\varrho) = y_i - \varrho_o' x_i.
\end{equation}

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

However it is very well known that the LS-estimates is very sensitive to small deviations from the regression model with normal errors.
(i) The performance of the LS-estimates, as measured by the trace of its asymptotic covariance matrix may be very much impaired when the observations satisfy the regression model given by (1.1) and (1.2) with $F_0$ closed to a normal distribution but with heavier tails, e.g., when $F_0$ belongs to the $\epsilon$-gross error contamination neighborhood $V_{\epsilon,\sigma}$ of the $N(0,\sigma^2)$ distribution ($N(\mu,\sigma^2)$ stands for the normal distribution with mean $\mu$ and variance $\sigma^2$) given by

$$V_{\epsilon,\sigma} = \{ F : F(u) = (1-%)\phi(u/\sigma)+\epsilon F^*(u) \},$$

where $\phi$ stands for the $N(0,1)$ distribution function and $F^*$ is an arbitrary symmetric distribution. If $F^*$ is chosen with dispersion larger than $\sigma$, e.g. $F^*(u) = \phi(u/\tau)$ with $\tau > \sigma$, then the LS-estimate may have very low efficiency, even when $\epsilon$ is very small.

(ii) A small fraction of bad observations (outliers), even one observation, may have a very large effect on the value of the estimate.

If we assume as a central model that the distribution of the $z_i$'s is $H_0$ given by (1.2) with $F_0$ normal, a robust estimate should have the following properties:

(a) It should be highly efficient when all the observations satisfy a linear model, i.e., if $H$ is given by (1.2), and $F_0$ is normal.

(b) The performance of the estimate should be only slightly impaired when $H_0$ is given by (1.2) with $F_0$ is in a small
neighborhood of a normal distribution, e.g., in $V_{\epsilon, \sigma^2}$.

(c) A small fraction of outliers should have only a small effect on the estimate.

(d) A somewhat large number of outliers should not spoil the estimate completely.

Condition (a) may be assessed by the asymptotic efficiency of the estimate (the ratio between the trace of its asymptotic covariance matrix and the trace of the Rao-Cramer bound matrix) under a regression model with normal errors, condition (b) by the maximum of the trace of the asymptotic covariance matrix of the estimate in the neighborhood $V_{\epsilon, \sigma}$ (Huber's (1964) minimax criterion), (c) by Hampel's (1971) concept of qualitative robustness and (d) by Hampel's (1971), concept of breakdown-point. Donoho (1982) and Donoho and Huber (1983) give a finite-sample version of this last concept which we will use here.

Motivated by robustness considerations, Huber (1973) proposes the class of maximum likelihood estimates (M-estimates) for the regression parameters. These estimates are defined by the minimization of

$$S(\theta) = \sum_{i=1}^{n} \rho(x_i(\theta)),$$

where $\rho(u)$ is a symmetric function, non decreasing in $|u|$ strictly increasing at $u = 0$, and which asymptotically increases slower than $u^2$. Differentiating $S(\theta)$ we get that the M-estimates satisfy the following system of equations
\[ \sum_{i=1}^{n} \psi(r_i(\theta)) \xi_i = 0, \]  

where \( \psi = \rho' \). Since the estimates defined by minimizing (1.6) or by solving (1.7) are not scale equivariant, Huber proposes to estimate simultaneously \( \theta_o \) and the scale error \( \sigma_o \) by the following system of equations

\[ \sum_{i=1}^{n} \psi(r_i(\theta)/\sigma) \xi_i = 0, \]

\[ (1/n) \sum_{i=1}^{n} \chi(r_i(\theta)/\sigma) = b, \]

where \( \chi \) has the same properties as \( \rho \). The constant \( b \) may be defined by

\[ b = E_\theta(\chi(u)). \]

Moreover, in order to get robustness properties \( \psi \) and \( \chi \) should be bounded.

Another possibility for getting scale equivariance is to obtain separately a scale error estimate \( s_n \) and then estimate \( \theta_o \) by the value which minimizes

\[ S(\theta) = \sum_{i=1}^{n} \rho(r_i(\theta)/s_n). \]

Huber's results (1964, 1973), give the optimal M-estimates using as criterion the minimization of the maximum of the trace of the asymptotic covariance matrix when \( F \) belongs to \( V_{\epsilon, \sigma} \). The optimal \( \psi \)-function belongs to the family given by
\[ \psi_{H,k}(u) = \text{sgn}(u) \max(|u|, k), \]

where \( k = k_0(\varepsilon) \) and \( k_0 \) is decreasing.

If \( k \) is properly chosen, these estimates satisfy criterion (a). However any M-estimate with monotone \( \psi \)-function is not qualitatively robust and its breakdown-point is 0. The bad behaviour of these estimates occurs in the presence of high leverage outliers. In order to avoid the large influence of the high leverage observations, Mallows (1976) proposed to replace (1.8) by

\[ \sum_{i=1}^{n} w(x_i) \frac{\psi(r_i(\hat{\theta})/\sigma)}{r_i(\hat{\theta})} x_i = 0 \]

where the function \( w \) downweights high leverage points.

A more general class of estimates, generalized M-estimates (GM-estimates), is defined replacing (1.9) by

\[ \sum_{i=1}^{n} w(x_i, r_i(\hat{\theta})/\sigma) r_i(\hat{\theta}) x_i = 0. \]

These estimates were studied by Maronna, Bustos and Yohai (1979) and Maronna and Yohai (1981). Krasker (1980), Krasker and Welsch (1982), Ronchetti and Rousseeuw (1982) and Samarov (1983) give optimal choices of \( w \) using different criteria.

However these GM-estimates have two shortcomings

(i) Their breakdown-point tends to 0 when \( p \) tends to \( \infty \), see Maronna, Bustos and Yohai (1979).

(ii) They have very low efficiency in the presence of good high leverage observations, i.e., if the distribution \( H \) of
the $z_i$'s satisfies (1.2) but the distribution of the carriers $G_0$ has heavy tails, see Maronna, Bustos and Yohai (1979).

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 (1982) proposed the least median of squares (LMS) estimate which is defined by the minimization of

$$\text{median}(r_1^2(\theta), r_2^2(\theta), \ldots, r_n^2(\theta)).$$

Another class of high breakdown-point estimates, the scale estimates (S-estimates), was proposed by Rousseeuw and Yohai (1984). These estimates are defined by the minimization of an $M$-estimate of the residuals scale.

However all these estimates are highly inefficient when all the observations satisfy the regression model with normal errors, and therefore they do not satisfy criterion (a). Moreover, Siegel's RM-estimator is not affine equivariant.

The purpose of this paper is to present a new class of estimates, which we call MM-estimates, satisfying simultaneously criteria (a), (b), (c), and (d).

The MM-estimates are defined in three stages. In the first stage we compute any high breakdown-point estimate which may be inefficient. In the second stage we obtain residuals using the initial estimate and then we compute an $M$-estimate $s_n$ of their scale with breakdown point 0.5. In the third stage we compute an $M$-estimate of $\theta_0$ defined by the minimization of
(1.1) using a proper bounded \( \rho \)-function and the scale estimate \( s_n \) computed in stage 2.

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 fitting" property, which was introduced by Rousseeuw (1982). In Section 3 we study qualitative robustness, in Section 4 consistency and in Section 5 asymptotic normality. In Section 6 we give a numerical algorithm for computing the MM-estimates. In Section 7 we compare the asymptotical biases of MM-estimates and optimal GM-estimates under contamination when the \( \phi_i \)'s are multivariate normal. Section 8 contains an Appendix with all the proofs.
2. MM-estimates. Huber (1981) defines the scale M-estimates as follows: let \( \rho \) be a real function satisfying the following assumptions:

(A) (i) \( \rho(0) = 0 \), (ii) \( \rho(-u) = \rho(u) \), (iii) \( 0 \leq u \leq v \) implies \( \rho(u) \leq \rho(v) \), (iv) \( \rho \) is continuous, (v) Let \( a = \sup \rho(u) \), then \( 0 < a < \infty \), (vi) If \( \rho(u) < a \) and \( 0 \leq u \leq v \), then \( \rho(u) < \rho(v) \).

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

\[
(2.1) \quad \frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{u_i}{s}\right) = b,
\]

where \( b \) may be defined by \( E_{\phi}(\rho(u)) = b \).

It is easy to show that if

\[
c(\mathbf{u}) = \#\{1 \leq i \leq n, u_i = 0\}/n < 1 - (b/a),
\]

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

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

Stage 1: Take an estimate \( T_{0,n} \) of \( \theta_0 \) with high breakdown-point, possible 0.5 and qualitatively robust. See Remark 2.4 for the selection of this initial estimate.

Stage 2: Compute the residuals

\[
(2.2) \quad r_i(T_{0,n}) = y_i - T_{0,n}' x_i \quad 1 \leq i \leq n
\]

and compute the M-scale \( s_n = s(\mathbf{r}(T_{0,n})) \) defined by (2.1) using a function \( \rho_0 \) satisfying assumption (A) and using a
constant b such that

\[(2.3) \quad b/a = .5,\]

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 \(\rho_1\) be another function satisfying assumption (A) and such that

\[(2.4) \quad \rho_1(u) < \rho_0(u),\]

\[(2.5) \quad \sup \rho_1(u) = \sup \rho_0(u) = a.\]

Let \(\psi_1 = \rho_1\), then the MM-estimate \(T_{1,n}\) is defined as any solution of

\[(2.6) \quad \sum_{i=1}^{n} \psi_1(r_i(\theta)/s_n) x_i = 0,\]

which satisfies

\[(2.7) \quad S(T_{1,n}) < S(T_{0,n}),\]

where

\[(2.8) \quad S(\theta) = \sum_{i=1}^{n} \rho_1(r_i(\theta)/s_n)\]

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 \(S(\theta)\) exists. It is clear that
this absolute minimum should satisfy (2.6) and (2.7). However
any other value of $\theta$ 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 $T_{o,n}$ is affine equivariant, i.e., if

$$T_{o,n}((y_1 + \theta'x_1, x_1), (y_2 + \theta'x_2, x_2), \ldots, (y_n + \theta'x_n, x_n))$$

$$= T_{o,n}((y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)),$$

and if $T_{1,n}$ is defined as the absolute minimum of $S(\theta)$, then $T_{1,n}$ will be equivariant too.

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

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

Let $Z_n = (z_1, z_2, \ldots, z_n)$ be any sample of size $n$, and let $T = (T_n)_{n \geq p}$ ($T_n$ is the estimate corresponding to a sample of size $n$) be a sequence of estimates. Let

$$b(m, T, Z_n) = \sup_{Z_{n+m}} |T_{m+n}(Z_{n+m} \cup W_m) - T_n(Z_n)|,$$

where the supremum is taken over all the samples $W_m$ of size
\( m \), \( Z_n \cup W_m \) denotes the sample of size \( n+m \) which contains the observations of both samples and \( \| \cdot \| \) denotes Euclidean norm. The breakdown-point of \( T \) at the sample \( Z_n \) is defined by

\[
\varepsilon_n^*(T, Z_n) = \min \left( \frac{m}{m+n} : b(m, T, Z_n) = \infty \right).
\]

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

Let

\[(2.10) \quad c_n = \max_{\theta \in \mathbb{R}^p} \frac{\# \{i: 1 \leq i \leq n \text{ and } \theta' x_i = 0 \}}{n}. \]

Then, if any set of \( p \) carriers is linearly independent, we have \( c_n = (p-1)/n \). Let \( \{T_n \}_{n>p} \) be the initial sequence of estimates, and \( \{T_1, n_{n>p} \} \) the corresponding MM-estimate. The following Theorem implies that if \( \varepsilon_n^*(T_0, Z_n) \) is asymptotically 0.5 and \( c = (p-1)/n \), then \( \varepsilon_n^*(T_1, Z_n) \) is asymptotically 0.5 too.

**Theorem 2.1.** Suppose that \( \rho_0 \) and \( \rho_1 \) satisfy assumption (A), that (2.3), (2.4) and (2.5) hold and \( c_n > 0.5 \). Then if \( T_0 = \{T_0, n_{n>p} \} \) is any sequence of estimates which satisfies (2.7), we have

\[
\varepsilon_n^*(T_1, Z_n) > \min(\varepsilon_n^*(T_0, Z_n), (1-2c_n)/(2-2c_n)).
\]

**Remark 2.4.** A possible choice for \( T_0 \) is Siegel's RM-estimate whose breakdown-point is asymptotically 0.5. Another
estimate with asymptotical breakdown-point equal 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 \( \ell \) which fits exactly. Then we have \( N = \binom{n}{p} \) estimates, \( \hat{\ell}_1, \hat{\ell}_2, \ldots, \hat{\ell}_N \) of \( \ell_0 \). For each of these estimates \( \hat{\ell}_i \), we compute the residuals \( r_{ij} = y_j - \hat{\ell}_i' x_j \), \( i < j < n \) and \( \hat{\sigma}_i = \text{median} \{ r_{ij} \mid 1 < j < n \} \). Then \( T_n \) is defined as the value \( \hat{\ell}_i \) which corresponds to the minimum \( \hat{\sigma}_i \). 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 (1982), and called here "exact fitting" property (EFP), is the following: An estimate \( T_n \) has the EFP if given any sample of size \( n \), \( (y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n) \), for which there exists \( \ell \) such that \( \# \{ i : y_i = \ell' x_i \} > n/2 \), then \( \# \{ i : y_i = T_n' x_i \} > n/2 \) too.

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

**Theorem 2.2.** Assume that \( \rho_0 \) and \( \rho_1 \) satisfy (A). Suppose that \( T_{0,n} \) has the EFP and let \( T_{1,n} \) be any estimate
satisfying (2.7). Then $T_{1,n}$ has the EFP too.

Remark 2.5. The RM-, LSM-, finite LMS-, and S-estimates have the EFP. Therefore if we take any of these estimates as $T_{0,n}$, the MM-estimate $T_{1,n}$ will also have the EFP.
3. Qualitative robustness. Let \( z_1 = (y_1, x_1), z_2 = (y_2, x_2), \ldots, z_n = (y_n, x_n), \ldots \) be i.i.d. observations. Hampel (1971) gives the following definition of qualitative robustness.

**Definition.** A sequence of estimates \( \{T_n\}_{n \geq p} \) of \( \theta_0 \) is qualitatively robust at \( H_0 \), if given \( \epsilon > 0 \), there exists \( \delta > 0 \) such that \( \pi_{p+1}(H_0, H_0) < \delta \) implies \( \pi_p(L(T_n, H), L(T_n, H_0)) < \epsilon \) for all \( n \geq p \). \( L(T, H) \) denotes the law of \( T \) under \( H \) and \( \pi_j \) the Prohorov metric between distributions in \( R^j \).

Suppose that \( T_n = T(H_n) \), where \( T \) is a functional on a subset of distributions on \( R^{p+1} \) and \( H_n \) is the empirical distribution of \( x_1, x_2, \ldots, x_n \). Then Hampel (1971) proves that the following two conditions are sufficient for qualitative robustness at \( H_0 \).

(a) \( T \) is continuous at \( H_0 \) as a functional defined in the space of distributions in \( R^{p+1} \) endowed with the metric \( \pi_{p+1} \) and with values in \( R^p \).

(b) \( T_n \) is continuous as a function defined in \( R^{(p+1)n} \) and with values in \( R^p \).

We are able to prove that under general regularity conditions the MM-estimates satisfy (a), but we are not able to prove (b). However as observed by Papantoni-Kazakos and Gray (1979), (a) is a sufficient condition for the following asymptotic version of the qualitative robustness concept.

**Definition.** \( \{T_n\}_{n \geq p} \) is asymptotically qualitatively robust at \( H_0 \) if given \( \epsilon > 0 \) there exists \( \delta > 0 \) and \( n_0 \) such that
\[ \pi_{p+1}(H^*, H) < \delta \text{ and } n > n_0 \text{ implies } \pi_p(L(T_n, H^*), L(T_n, H)) > \varepsilon. \]

In order to prove the asymptotical qualitative robustness of the MM-estimates we need some additional assumptions.

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

(D) \( P_{F_0} (\theta'X = 0) < 0.5 \) for all \( \theta \in \mathbb{R}^p \).

Suppose that the initial sequence \( \{T_{0,n}\}_{n \geq p} \) is given by a functional \( T_0 \), i.e., \( T_{0,n} = T_0(H_n) \), where \( H_n \) is the empirical distribution of \( z_1, \ldots, z_n \). The following Theorem proves that the functional which defines the MM-estimate is continuous.

**Theorem 3.1.** Let \( H_0 \) be given by (1.2). Suppose that \( T_0 \) is continuous at \( H_0 \) and \( T_0(H_0) = \theta_0 \). Assume also that \( \rho_0 \) and \( \rho_1 \) satisfy (A), (2.3), (2.4) and (2.5) holds, \( F \) satisfies (C) and \( G \) satisfies (D). Then the functional \( T_1 \) corresponding to the sequence of MM-estimates \( \{T_{1,n}\}_{n \geq p} \) is continuous at \( H_0 \) too.
4. Consistency. Theorems 4.1 and 4.2 establish the consistency of the scale estimate $s_n$ defined in stage 2 and of the MM-sequence of estimates $\{T_{1,n}\}_{n \geq p}$ of $\theta_0$.

**Theorem 4.1.** Let $(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)$ be i.i.d. observations with distribution $H_0$ given by (1.2). Assume that $\rho_0$ satisfies (A) and $\{T_{0,n}\}_{n \geq p}$ is a sequence of estimates which is strongly consistent for $\theta_0$. Then $s_n$ is strongly consistent for $\sigma_0$ defined by

\begin{equation}
E_{F_0}(\rho_0(u/\sigma_0)) = b.
\end{equation}

**Theorem 4.2.** Let $(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)$ be i.i.d. observations with distribution $H_0$ given by (1.2). Assume that $\rho_0$ and $\rho_1$ satisfy (A), that (2.3), (2.4) and (2.5) hold, $F_0$ satisfies (C) and $G_0$ satisfies (D). Assume also that the sequence $\{T_{0,n}\}_{n \geq p}$ is strongly consistent for $\theta_0$ then any other sequence $\{T_{1,n}\}_{n \geq p}$ which satisfies (2.7) is strongly consistent too.
5. 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 $x_i$'s. We are going to give here a proof of asymptotic normality for the MM-estimates which requires only second moments on the carriers. We need some additional assumptions.

(B) $\rho_1$ is odd, twice continuously differentiable and there exists $m$ such that $|u| \geq m$ implies $\rho_1(u) = a$

(E) $G_o$ has second moments and

\begin{equation}
V = E_{G_o}(x_i x_i')
\end{equation}

is non singular.

The following Theorem gives the asymptotical normality of M-estimates with scale estimated separately, which include as special case the MM-estimates.

**Theorem 5.1.** Let $z_1, z_2, \ldots, z_n$ be i.i.d. with distribution $H_o$ given by (1.2). Assume $\rho_1$ satisfies (B) and $G$ satisfies (E). Let $s_n$ be an estimate of the error scale which converges strongly to $\theta_o$. Let $T_n$ be a sequence of estimates which satisfies (2.6) and which is strongly consistent to the true value $\theta_o$. Then

\begin{equation}
n^{1/2}(T_n - \theta_o) \overset{d}{\rightarrow} N(0, \sigma_o^2 [A(\psi, F_o)/B(\psi, F_o)] V^{-1}),
\end{equation}

where $d$ denotes convergence in distribution,

\begin{equation}
A(\psi, F) = E_F(\psi^2(u/\sigma_o))
\end{equation}
and
\begin{equation}
(5.4) \quad B(\psi, F) = E_F(\psi'(u/\sigma_0)).
\end{equation}

**Remark 5.1.** Let $\rho_0$ and $\rho_1$ be as in Remark 2.3 where $\rho(u/k)$ is equivalent to $u^2$ for large $k$. For example, let $\rho$ be given by
\begin{equation}
(5.5) \quad \rho_B(u) = \begin{cases} 
\frac{u^2}{2} - \frac{u^4}{2} + \frac{u^6}{6} & \text{if } |u| \leq 1 \\
\frac{1}{6} & \text{if } |u| > 1,
\end{cases}
\end{equation}
which corresponds to the bisquare psi-function
\begin{equation}
(5.6) \quad \psi_B(u) = \begin{cases} 
\frac{u(1-u^2)^2}{6} & \text{if } |u| \leq 1 \\
0 & \text{if } |u| > 1.
\end{cases}
\end{equation}

Suppose also that the MM-estimate is computed using $s_n$ defined in stage 2 with $b = E_F(\rho(u/k_0))$, where $u$ is $N(0,1)$. Then if the $u_i$'s are $N(0, \sigma^2)$, we have $\sigma^2 = 1$ and according to Theorem 5.1, the asymptotic variance of the MM-estimate depends only on $k_1$. Therefore we can choose $k_1$ so that the MM-estimate be highly efficient without affecting its breakdown-point which depends only on the choice of $k_0$. When $\rho = \rho_B$ the value $k_0$ which 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.

**Remark 5.2.** According to Theorem 5.1, 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).
6. Computer algorithm. Here we propose a computer 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 \( z_i = (y_i, x_i) \), \( 1 \leq i \leq n \), be a sample of size \( n \) and suppose that we have already computed the initial estimate \( T_{0,n} \) and the scale estimate \( s_n \) defined in stage 2. For each \( \xi \in \mathbb{R}^p \) define the weights

\[
(6.1) \quad w_i(\xi) = \psi_1(r_i(\xi)/s_n)/(r_i(\xi)/s_n).
\]

Define also

\[
(6.2) \quad g(\xi) = (1/s_n^2) \sum_{i=1}^{n} w_i(\xi) r_i(\xi)x_i = (1/s_n^2) \sum_{i=1}^{n} \psi_1(r_i(\xi)/s_n)x_i
\]

and

\[
(6.3) \quad M(\xi) = (1/s_n^2) \sum_{i=1}^{n} w_i(\xi)x_ix_i^T.
\]

It is easy to show that \( -g(\xi) \) is the gradient of \( S(\xi) \). The recursion step of the IWLS is defined as follows. If \( \xi(j) \) is the value of the estimate in the \( j \)-th step then \( \xi(j+1) \) is defined by

\[
(6.4) \quad \xi(j+1) = \xi(j) + \Delta(\xi(j)),
\]

where

\[
(6.5) \quad \Delta(\xi) = M^{-1}(\xi)g(\xi).
\]
Using this recursion we can not guarantee that $S(t^{(j+1)}_k) < S(t^{(j)}_k)$ and therefore if $T_{1,j}$ is computed as a limit of the sequence $t^{(j)}_k$, (2.7) may not hold. We propose the following modification. Take $0 < \delta < 1$, then since $-g(t)$ is the gradient of $S(t)$, we can find an integer $k$ such that

\[(6.6) \quad S(t^{(j)}_k + \delta(t^{(j)}_k/2^k)) < S(t^{(j)}_k) - \delta(t^{(j)}_k/2^k)g(t^{(j)}_k).\]

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

\[(6.7) \quad t^{(j+1)}_k = t^{(j)}_k + (1/2)k^{(j)}_{2,j} \delta(t^{(j)}_k)\]

starting with $t^{(0)} = T_{0,n}$. Clearly we have now $S(t^{(j+1)}_k) < S(t^{(j)}_k)$. The following Theorem shows that any limit point of the sequence $t^{(j)}_k$ satisfies (2.6) and (2.7), and therefore is a MM-estimate.

**Theorem 6.1.** Suppose $\rho_0$ and $\rho_1$ satisfy (A), (2.3), (2.4) and (2.5) hold, $\psi_1$ continuous, $\lim_{u \to 0} \psi_1(u)/u > 0$, $u \neq 0$ and $\rho_1(u) < a$ implies $\psi_1(u) > 0$ and finally $c_n < 0.5$. Then if $t^{(j)}_k$ is defined by (6.7) with $t^{(0)} = T_{0,n}$, we have

(i) The sequence $t^{(j)}_k$ is bounded

(ii) Any limit point of $t^{(j)}_k$ satisfies (2.6) and (2.7)

(iii) If $t_0$ and $t_1$ are two limit points of $t^{(j)}_k$, we have $S(t_0) = S(t_1)$. 

-22-
7. **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 models is subject to an infinitesimal contamination. Suppose that the distribution of \( z_i = (y_i, x_i) \), \( H_0 \), is given by (1.2) and that the sequence of estimates \( \{T_n \}_{n \in \mathbb{P}} \) is defined by a functional \( T \) applied to the empirical distribution i.e., \( T = T(H_n) \) which is consistent for \( \theta_0 \), i.e., \( T(H_0) = \theta_0 \). If \( H_0 \) is subject to a contamination of size \( \epsilon \), with the distribution \( \delta_{y, x} \) concentrated at the point \( (y, x) \), the asymptotic bias of the estimate given by the functional \( T \) is

\[
\mathcal{B}(T, H_0, \epsilon, y, x) = T((1-\epsilon)H_0 + \epsilon \delta_{y, x}) - T(H_0),
\]

and the influence curve is defined by

\[
\mathcal{IC}(T, H_0, y, x) = \lim_{\epsilon \to 0} \frac{\mathcal{B}(T, H_0, \epsilon, y, x)}{\epsilon}.
\]

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

\[
\mathcal{IC}(T, H_0, y, x) = \psi_1(y - \hat{\theta}_0, x) \sigma_0^2(B(\psi_1, F_0)y)^{-1},
\]

where \( V \) is given in (5.1) and \( B(\psi, F) \) in (5.4).

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 \( \epsilon \)-influence curve by
\[ I_{e} (\mathcal{T}, H_{0}, y, x) = b (\mathcal{T}, H_{0}, \varepsilon, y, x) / \varepsilon. \]

It may be proved, by arguments similar to those used in Theorem 2.1, that if \( \mathcal{T} \) is an MM-estimate, then \( \mathcal{T}_e((1-\varepsilon)H_{0} + \varepsilon H_{*}) \) is bounded in \( H_{*} \) for any \( \varepsilon < 0.5 \), and consequently \( I_{e} (\mathcal{T}, H_{0}, y, x) \) is bounded in \( (y, x) \) too.

In this section we compare the \( \varepsilon \)-influence curves of MM-estimates and the optimal Krasker and Welsch (1982) bounded influence estimates (KW-estimates) when \( P_{0} \) and \( G_{0} \) are normals.

The MM-estimate considered here is based in the bisquare rho-function given by (5.5); \( \rho_{B}(u) = \rho_{B}(u/k_{1}) \), \( i = 0, 1 \). The values of \( k_{0} \) and \( k_{1} \) are those given in Remark 5.1 and correspond to an estimate with breakdown-point 0.5 and efficiency 0.95 for normal errors. We denote this MM-estimate by \( \mathcal{T}_{1} \).

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

\[ \sum_{i=1}^{n} \psi_{H,k}(\{(y_{i} - \theta^\prime x_{i})/s_{n}\} \mid x_{i}^{\prime} x_{i}^{\prime}) (x_{i}^{\prime} x_{i}^{\prime}) = 0, \]

where \( |x_{i}^{\prime} x_{i}^{\prime}| = (x_{i}^{\prime} x_{i}^{\prime})^{1/2} \), \( \Sigma \) is the covariance matrix of \( x_{i} \) (supposed known here) and \( \psi_{H,k} \) is the Huber psi-function given by (1.11).

The values of the constant \( k \) is chosen so that the asymptotical efficiency of this estimate, which we denote by \( \mathcal{T}_{2} \), be 0.95 when the distribution \( H_{0} \) of \( z = (y, x) \) is multivariate normal and may be found in Maronna, Bustos and Yohai (1979). This estimate has the property of minimizing the in-
variant gross error sensitivity (see Krasker and Welsch (1982)) defined by

$$\gamma^*(\mathcal{T}) = (\sup_{y, x} \text{IC}(\mathcal{T}, \mathcal{H}_0, y, x) \tilde{V}^{-1} \text{IC}(\mathcal{T}, \mathcal{H}_0, y, x))^{1/2},$$

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 $\theta_0 = 0$, $\Sigma = I$ and therefore $y = u$ has distribution $N(0,1)$.

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

$$\gamma^*_\epsilon(\mathcal{T}) = (\sup_{y, x} \text{IC}_\epsilon(\mathcal{T}, \mathcal{H}_0, y, x) \tilde{V}^{-1} \text{IC}(\mathcal{T}, \mathcal{H}_0, y, x))^{1/2}$$

for $\epsilon = 0.1, 0.15$ and $0.2$, $p = 1, 2, 3, 5$ and 10 and $\mathcal{T}$ equal to $T_1$ and $T_2$.

We observe that $\gamma^*_\epsilon(T_2) < \gamma^*_\epsilon(T_1)$ for $p = 10$ if $\epsilon \geq 0.1$, for $p = 3$ if $\epsilon \geq 0.15$ and for $p = 1$ if $\epsilon \geq 0.20$. This shows that the infinitesimal gross error sensitivity may not be enough to compare the bias robustness of the two estimates even for small $\epsilon$. This table also shows that the MM-estimate $T_2$ may be better than the optimal bounded influence estimate $T_1$ in terms of $\gamma^*_\epsilon$ especially for large $p$.

**Table 1**

-25-
Table 1

ε - gross error sensitivities

<table>
<thead>
<tr>
<th>p</th>
<th>ε = 0.10</th>
<th></th>
<th>ε = 0.15</th>
<th></th>
<th>ε = 0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T₁</td>
<td>T₂</td>
<td>T₁</td>
<td>T₂</td>
<td>T₁</td>
</tr>
<tr>
<td>1</td>
<td>8.7</td>
<td>3.8</td>
<td>8.1</td>
<td>5.2</td>
<td>8.3</td>
</tr>
<tr>
<td>2</td>
<td>8.7</td>
<td>4.7</td>
<td>8.1</td>
<td>7.4</td>
<td>8.3</td>
</tr>
<tr>
<td>3</td>
<td>8.7</td>
<td>5.2</td>
<td>8.1</td>
<td>9.0</td>
<td>8.3</td>
</tr>
<tr>
<td>5</td>
<td>8.7</td>
<td>6.5</td>
<td>8.1</td>
<td></td>
<td>8.3</td>
</tr>
<tr>
<td>10</td>
<td>8.7</td>
<td>9.5</td>
<td>8.1</td>
<td></td>
<td>8.3</td>
</tr>
</tbody>
</table>
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 depending 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). The data are shown in Table 2 and the spurious observations are marked with *. 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.

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

**TABLE 2**
Table 2

Data of example
(Number of calls in ten of millions)

<table>
<thead>
<tr>
<th>YEAR</th>
<th>NUMBER OF CALLS</th>
<th>YEAR</th>
<th>NUMBER OF CALLS</th>
<th>YEAR</th>
<th>NUMBER OF CALLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.44</td>
<td>58</td>
<td>1.06</td>
<td>66</td>
<td>14.20*</td>
</tr>
<tr>
<td>51</td>
<td>0.47</td>
<td>59</td>
<td>1.20</td>
<td>67</td>
<td>15.90*</td>
</tr>
<tr>
<td>52</td>
<td>0.47</td>
<td>60</td>
<td>1.35</td>
<td>68</td>
<td>18.20*</td>
</tr>
<tr>
<td>53</td>
<td>0.59</td>
<td>61</td>
<td>1.49</td>
<td>69</td>
<td>21.20*</td>
</tr>
<tr>
<td>54</td>
<td>0.66</td>
<td>62</td>
<td>1.61</td>
<td>70</td>
<td>4.30</td>
</tr>
<tr>
<td>55</td>
<td>0.73</td>
<td>63</td>
<td>2.12</td>
<td>71</td>
<td>2.40</td>
</tr>
<tr>
<td>56</td>
<td>0.81</td>
<td>64</td>
<td>11.90*</td>
<td>72</td>
<td>2.70</td>
</tr>
<tr>
<td>57</td>
<td>0.88</td>
<td>65</td>
<td>12.40*</td>
<td>73</td>
<td>2.90</td>
</tr>
</tbody>
</table>
8. Appendix. Before proving Theorem 2.1 we will prove the following Lemma.

Lemma 2.1. Let $Z = \{z_1, z_2, \ldots, z_n\}$ 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 $\epsilon < (1-2c_n)/(2-2c_n)$ and $k_o$, there exists $k_1$ such that $m/n + m < \epsilon$ and $s_{m+n} \leq k_o$ imply

$$\inf_{|\hat{\theta}|} \sum_{i=1}^{m+n} \rho_1(x_i(\theta)/s_{m+n}) \geq \sum_{i=1}^{m+n} \rho_1(x_i(T_o, m+n)/s_{m+n})$$

for all samples $Z_n \cup W_m$ with $\# W_m = m$ and where $\| \|$ denotes Euclidean norm.

Proof. by definition of $c_n$ we have for all $\theta$

$$\#\{i : 1 < i < n, \|z_i\| > 0\}/n > 1-c_n .$$

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

$$(8.1) \quad \inf_{|\hat{\theta}|=\delta} \#\{i : 1 < i < n, \|z_i\| > \delta \}/n > 1-c_n^* ,$$

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

$$(8.2) \quad a_o n/(n+m) > (1-\epsilon)a_o > a/(2-2c_n^*) .$$

By (2.5) there exists $k_2$ such that $\rho_1(k_2) = a_o$ and let
\[ k_1 = \left( \max_{1 \leq i \leq n} |y_i| + k_0 k_2 \right)/\delta. \]

Therefore, using the monotonicity of \( \rho_1 \), (8.1) and (8.2) we have \( m/(m+n) < \epsilon \) implies
\[
\inf_{|\theta| > k_1, s_{m+n} > k_0} \sum_{k=1}^{m+n} \rho_1(\rho_i(\theta)/s_{m+n}) \\
> \inf_{|\theta| = 1} \sum_{i=1}^{n} \rho_1((|y_i| - k_1(\theta') x_i)) / k_0 \\
> n(1 - c_n^*) \rho_1(k_2) = n(1 - c_n^*) a_0 > (m+n)a/2.
\]

On the other hand by (2.1), (2.4), and (2.5) we have
\[
\sum_{i=1}^{m+n} \rho_1(\rho_i(T_{0, m+n})/s_{m+n}) \leq \sum_{i=1}^{m+n} \rho_0(\rho_i(T_{0, m+n})/s_{m+n}) \leq (m+n)a/2.
\]

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 \( \epsilon < \min(\epsilon^*(T_0, Z_n), 0.5) \), there exists \( k_0 \) such that for any sample \( Z_n \cup W_m \), where \( \#W_m = m \) and \( m/(n+m) < \epsilon \), we have \( s_{m+n}(Z_n \cup W_m) < k_0 \).

Since \( \epsilon < \epsilon^*(T_0, Z_n) \), there exists \( k_1 \) such that
\[
T_{0, m+n}(Z_n \cup W_m) < k_1 \quad \forall W_m.
\]

Therefore there exists \( k_2 \) such that
\[
\sup_{1 \leq i \leq n} \rho_1(T_{0, m+n}(Z_n \cup W_m)) < k_2. \quad 1 \leq i \leq n
\]

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

Then using (8.3) we have

\[
\begin{align*}
&\left(\frac{1}{n+m}\right) \sum_{i=1}^{m+n} \rho_0(r_i(T_{0,m+n}(Z_n \cup W_m))/k_0) \\
&\leq \left(\frac{1}{n+m}\right) \sum_{i=1}^{n} \rho_0(r_i(T_{0,m+n}(Z_n \cup W_m))/k_0) \\
&+ \left(\frac{1}{n+m}\right) \sum_{i=n+1}^{n+m} \rho_0(r_i(T_{0,m+n}(Z_n \cup W_m))/k_0) \\
&\leq \left(\frac{n}{n+m}\right) \rho_0(k_2/k_0) + \left(\frac{m}{n+m}\right)a < \rho_0(\delta) + \varepsilon a \\
&= \varepsilon a + \gamma < b.
\end{align*}
\]

Therefore \( s_{m+n}(Z_n \cup W_m) \leq k_0 \). This proves the Theorem.

**Proof of Theorem 2.2:** Let \( Z_n = \{z_1 = (y_1, x_1), \ldots, z_n = (y_n, x_n)\} \)

be a sample and suppose that there exists \( \hat{\vartheta}^* \) such that

\[
\#\{i : 1 < i < n, y_i - \hat{\vartheta}^* x_i = 0\} > n/2.
\]

Since \( T_{0,n} \) have the EFP we have

\[
\#\{i : 1 < i < n, y_i - T_{0,n} x_i = 0\} > n/2,
\]

and therefore \( s_n = 0 \). Then for any \( \vartheta \) we have

\[
S(\vartheta) = \sum_{i=1}^{n} \rho_1(y_i - \vartheta' x_i/s_n) = a\#\{i : 1 < i < n, y_i - \vartheta' x_i \neq 0\}.
\]

Then \( S(T_{0,n}) < an/2 \) and therefore (2.7) implies \( S(T_{1,n}) \)

\( < an/2 \) too. This implies \( \#\{i : 1 < i < n, y - T_{1,n} x_i = 0\} > n/2 \).
and therefore $T_{1,n}$ has the EFP too.

From now on we will assume without loss of generality that $\theta_0 = 0$. Before proving Theorem 3.1 we need to prove Lemmas 3.1, 3.2, 3.3 and 3.4.

**Lemma 3.1.** Let $\rho$ be a function satisfying (A) and let $F$ be a distribution function satisfying (C), then $g(\lambda) = E_F(\rho(u - \lambda))$ has a unique minimum at $\lambda = 0$.

**Proof.** It is easy to see that for any $\lambda \neq 0$ the distribution function $R_{\lambda}$ of $|u - \lambda|$ satisfies: (i) $R_{\lambda}(u) < R_0(u)$ for all $u > 0$ and (ii) there exists $\delta > 0$ such that $R_{\lambda}(u) < R_0(u)$ for $0 \leq u \leq \delta$. Then since $\rho$ is non decreasing in $|u|$ and strictly increasing in a neighborhood of 0, the Lemma follows.

Lemma 3.2 shows the Fisher consistency of the M-estimates based in a $\rho$-function which satisfies (A), when the error distribution $F$ satisfies (C).

**Lemma 3.2.** Let $\rho$ be a function satisfying (A) and $(y, x)$ a $(p+1)$-dimensional random vector with distribution $H_0$ given by (1.2). Suppose that $F_0$ satisfies (C) and that for all $\theta \neq \theta_0$ $P_{G_0}(\theta'x = 0) < 1$. Define for $\xi \in \mathbb{R}^p$, $g^*(\xi) = E_{H_0}(\rho(y - \theta'x))$. Then $g$ has a unique minimum at $\xi = \xi_0$.

**Proof.** We have $g^*(\xi) = E_{H_0}(g(\theta'x))$. Then since for $\theta \neq \theta_0$ we have $P_{G_0}(\theta'x = 0) < 1$, Lemma 3.2 follows from Lemma 3.1.

**Lemma 3.3.** Suppose $T_0(H)$ is continuous at $H_0$ and
$T_0(H_0) = \emptyset$, where $H_0$ is given by (1.2). Assume also that $\rho_0$ satisfies (A), and $F_0$ satisfies (C). Then $s(H)$ defined by

$$E_H(\rho_0(y-T_0(H),s(H))) = b$$

is continuous at $H_0$.

**Proof.** Given $\epsilon > 0$, assumption (A) implies

$$E_{H_0}(\rho_0(y-T_0(H_0),x)/(\sigma_0 + \epsilon)) < b, \quad E_{H_0}(\rho_0(y-T_0(H_0),x)/(\sigma_0 - \epsilon)) > b.$$

Define

$$q_1(y,x,\gamma) = \sup \{ \rho_0((y-x)/(\sigma_0 + \epsilon)) : |\theta - T_0(H_0)| < \gamma \},$$

$$q_2(y,x,\gamma) = \sup \{ \rho_0((y-x)/(\sigma_0 + \epsilon)) : |\theta - T_0(H_0)| < \gamma \}.$$

Then, by the Lebesgue dominated convergence Theorem, there exists $\gamma_0 > 0$ and $c > 0$ such that

$$E_{H_0}(q_1(y,x,\gamma_0)) < b - c, \quad E_{H_0}(q_1(y,x,\gamma_0)) > b + c.$$

Since $q_i$, $i = 1, 2$, are uniformly continuous and bounded we can find $\delta_1$ such that $\pi_{p+1}(H,H_0) < \delta_1$ implies

$$E_H(q_1(y,x,\gamma_0)) < b - c/2, \quad E_H(q_1(y,x,\gamma_0)) > b + c/2.$$

Let $\delta_2$ be such that $\pi_{p+1}(H,H_0) < \delta_2$ implies $|T_0(H) - T_0(H_0)| < \gamma_0$. Take $\delta = \min(\delta_1, \delta_2)$, then if $\pi_{p+1}(H,H_0) < \delta$ we have
\[ E_H(\rho_0((y - T_0(H)'x)/(\sigma_0+\epsilon))) \leq b - \eta/2, \]
\[ E_H(\rho_0((y - T_0(H)'x)/(\sigma_0-\epsilon))) \leq b + \eta/2. \]

and therefore \( \sigma_0 - \epsilon \leq s(H) \leq \sigma_0 + \epsilon \). Then the Lemma is proved.

**Lemma 3.4.** Assume \( P_{G_0}(|\theta'x| > 0) > \lambda \) for all \( \theta \), then there exists \( \varphi > 0, \delta > 0, \gamma > 0 \) and a finite number of compact sets \( C_1, C_2, \ldots, C_s \) included in \( \mathbb{R}^p \) such that

(i) \( \bigcup_{i=1}^s C_i \subset C = \{ \theta : \theta \in \mathbb{R}^p, |\theta| = 1 \} \).

(ii) \( \pi_p(G, G_0) \leq \epsilon \) implies \( P_{G_0}(\inf_{\theta \in C_i} |\theta'x| \geq \varphi) > \lambda \).

**Proof.** By a standard compactness argument is enough to show that given \( \theta \in C \) there exist \( \hat{\theta} > 0, \delta > 0, \gamma > 0 \) and \( \varphi > 0 \) such that

\[ \pi_p(G, G_0) \leq \delta \implies P_{G_0}(\inf_{|\theta' - \theta| < \eta} |\theta'x| \geq \varphi \geq \lambda + \gamma). \]

Since \( P_{G_0}(|\theta'x| > 0) > \lambda \), we can find \( \gamma_1 > 0 \) and \( \varphi_1 > 0 \) such that

\[ P_{G_0}(|\theta'x| \geq \varphi_1) > \lambda + \gamma_1 \]

with \( \varphi_1 \) continuity point of the distribution of \( |\theta'x| \) under \( G_0 \). Then we can find \( \delta_1 \) such that

\[ \pi_p(G, G_0) \leq \delta_1 \implies P_{G}(|\theta'x| \geq \varphi_1) > \lambda + \gamma_1. \]

We can also find \( K \) and \( \delta_2 \) such that

\[ \pi_p(G, G_0) \leq \delta_2 \implies P_{G}(|x| > K) < \gamma_1/2. \]
Let \( \eta = \varphi / (2K) \) and let \( \delta = \min(\delta_1, \delta_2) \), then using (8.7) and (8.8) we have that \( \pi_p (G, G_0) < \delta \) implies

\[
P_G (\inf \| \varphi - \varphi \| < \eta \| \varphi \| ) > \varphi / 2) > P_G (\| \varphi \| > \varphi_1) - P_G (\| \varphi \| > \varphi_1 / 2)
\]

\[
> P_G (\| \varphi \| > \varphi_1) - P_G (\| \varphi \| > K) > \lambda + \gamma_1 \gamma_1 / 2 = \lambda + \gamma_1 / 2 .
\]

Then (8.5) holds with \( \phi = \varphi_1 / 2 \) and \( \gamma = \gamma_1 / 2 \).

**Lemma 3.5.** Under the assumptions of Theorem 3.1, there exists \( \delta > 0 \) and \( L \) such that \( \pi_{p+1} (H, H_0) < \delta \) implies

\[|T_\lambda (H)| < L\]

**Proof.** Since (2.3), (2.4) and (8.4) imply

\[E_H (\rho_1 (y - T_\mu (H)'y) / s(H)) < a / 2 ,\]

by Lemma 3.3 and (2.7) it is enough to prove that for any \( \sigma > 0 \) there exists \( \delta > 0 \), \( L \) and \( \gamma_1 \) such that

\[
(8.9) \quad \pi_{p+1} (H, H_0) < \delta \text{ implies } \inf \| \varphi \| > L E_H (\rho_1 (y - \varphi)'y) > a / 2 + \gamma_1 .
\]

Let \( \varphi, \delta_1, \gamma \) and \( C_1, C_2, \ldots, C_8 \) as in Lemma 3.4 with \( \lambda = 0.5 \). Since \( \lim_{|u| \to \infty} \rho_1 (u) = a \), for any fixed \( \sigma > 0 \) we can find \( \delta_2 > 0 \) and \( \gamma_1 > 0 \) and \( M \) such that

\[
(8.10) \quad \pi_{p+1} (H, H_0) < \delta_2 \text{ implies } (E_H (\rho_1 (|y - M| / \sigma)))^{1/2} (0.5 + \gamma)^{1/2} > a / 2 + \gamma_1 .
\]

-33-
Put $L = M/\varphi$ and $\delta = \min(\delta_1, \delta_2)$ and denote by $I(A)$ the indicator function of the set $A$, then according to Lemma 3.4 and the Schwarz inequality we have

$$\inf_{|Q| > L} E_H(\rho_1((y-Q'(X))/\sigma))$$

$$\geq \inf_{|Q| > L} E_H(\rho_1((|Y| - \varphi L)/\sigma)I(|(Q/L)'X| \geq \varphi))$$

$$\geq (E_H(\rho_1^2((|Y| - M)/\sigma)))^{1/2} \inf_{|Q| = 1} (P_H(|Q'(X)| \geq \varphi))^{1/2}.$$

Then (8.9) follows from Lemma 3.4 and (8.10). This proves Lemma 3.5.

Proof of Theorem 3.1. Put $C(\epsilon, L) = \{Q : Q \in \mathbb{R}^p, \epsilon \leq |Q| \leq L\}$. Then according to (2.7) and Lemmas 3.3 and 3.5 is enough to show that for any $\epsilon > 0$ and arbitrarily large $L$ there exist $\sigma_1 > \sigma_0$, $\delta > 0$ and $\eta > 0$ such that $\pi_{p+1}(H, E_0) \leq \delta$ implies

(8.11) $$\inf_{Q \in C(\epsilon, L)} E_H(\rho_1((y-Q'(X))/\sigma_1))$$

$$\geq E_{E_0}(\rho_1(u/\sigma_0)) + \eta$$

and

(8.12) $$E_H(\rho_1(y-T_0'(H)'X)/s(H)) \leq E_F(\rho_1(u/\sigma_0)) + \eta/2.$$

Using Lemma 3.2, the Lebesgue dominated convergence Theorem and a standard compacity argument we can find $\sigma_1 > \sigma_0$, $\gamma > 0$ and a finite number of compact sets $C_1, C_2, \ldots, C_s$ included in $\mathbb{R}^p$ such that
\[ E_{H_0}(\inf_{\theta \in C_1} \rho_1((y-\theta,x)/\sigma)) \geq E_{F_0}(\rho_1(u/\sigma_0)) + \gamma \]

and

\[ C(\epsilon,L) \subset \bigcup_{i=1}^{s} C_i. \]

Put \( q_i(y,x) = \inf_{\theta \in C_i} \rho_1((y-\theta,x)/\sigma), \) \( 1 \leq i \leq s. \) Since the \( q_i \)'s functions are uniformly continuously and bounded there exists \( \delta > 0 \) and \( \eta > 0 \) such that for all \( i, \) \( 1 \leq i \leq s, \) we have

\[ \tau_{p+1}(H,H_0) < \delta \quad \text{implies} \quad E_{H}(q_i(y,x)) \geq E_{F_0}(\rho_1(u/\sigma_0)) + \eta, \]

and this implies (8.11).

By the Lebesgue dominated convergence Theorem we can find \( \varphi > 0 \) such that

\[ E_{H_0}(\sup |g| \leq \varphi, |y-\sigma_0| \leq \rho_1((y-\theta,x)/\sigma)) \]

\[ \leq E_{F_0}(\rho_1(u/\sigma_0)) + (1/4)\eta. \]

Put \( q(y,x) = \sup |g| \leq \varphi, |y-\sigma_0| \leq \rho_1((y-\theta,x)/\sigma). \) Since \( q(y,x) \) is uniformly continuous and bounded, there exists \( \delta_1 > 0 \) such that \( \tau_{p+1}(H,H_0) < \delta_1 \) implies \( E_{H}(q(y,x)) \leq E_{F_0}(\rho_1(u/\sigma_0)) + \eta/2. \)

Since \( T_0(Q) = Q, \) \( T_0 \) is continuous at \( H_0 \) and by Lemma 3.3, \( s(H) \) is continuous at \( H_0, \) there exists \( \delta_1 > 0 \) such that \( \tau_{p+1}(H,H_0) < \delta_1 \) implies \( |T_0(H) - T_0(H_0)| < \varphi \) and \( |s(H) - s(H_0)| < \varphi. \) Then (8.12) holds with \( \delta = \min(\delta_1, \delta_2), \) and then Theorem 3.1 is proved.
Proof of Theorem 4.1. Take \( \epsilon > 0 \), then since \( \rho_o \) satisfies (A), we can find \( \delta > 0 \) such that

\[
\mathbb{E} \left( \inf \left| g \right| < \delta \rho_o \left( \frac{(y - \mathbf{y'}) \mathbf{x}}{(\sigma_o - \epsilon)} \right) \right) \geq b + \delta
\]

and

\[
\mathbb{E} \left( \sup \left| g \right| < \delta \rho_o \left( \frac{(y - \mathbf{y'}) \mathbf{x}}{(\sigma_o + \epsilon)} \right) \right) \leq b - \delta.
\]

Therefore by the law of the large numbers we have

\[
\lim_{n \to \infty} (1/n) \sum_{i=1}^{n} \inf \left| g \right| < \delta \rho_o \left( \frac{(y - \mathbf{y'}) \mathbf{x}}{(\sigma_o - \epsilon)} \right) \geq b + \delta \quad \text{a.s.,}
\]

and

\[
\lim_{n \to \infty} (1/n) \sum_{i=1}^{n} \sup \left| g \right| < \delta \rho_o \left( \frac{(y - \mathbf{y'}) \mathbf{x}}{(\sigma_o + \epsilon)} \right) \leq b - \delta \quad \text{a.s.}
\]

Then, since \( \lim_{n \to \infty} T_o, n = 0 \) a.s., we have

\[
\lim_{n \to \infty} (1/n) \sum_{i=1}^{n} \rho_o \left( \frac{(y - T_o, n \mathbf{x})}{(\sigma_o - \epsilon)} \right) \geq b + \delta \quad \text{a.s.}
\]

and

\[
\lim_{n \to \infty} (1/n) \sum_{i=1}^{n} \rho_o \left( \frac{(y - T_o, n \mathbf{x})}{(\sigma_o + \epsilon)} \right) \leq b - \delta \quad \text{a.s.}
\]

Therefore by the monotonicity of \( \rho_o \), with probability 1 there exists \( n_o \) such that for all \( n > n_o \) we have \( \sigma_o - \epsilon < s_n < \sigma_o + \epsilon \).

The following Lemmas will be necessary for the proof of Theorem 4.2.

Lemma 4.1. Under the assumptions of Theorem 4.2, there
exists \( L \) such that

\[
\lim_{n \to \infty} |T_{1,n}| \leq L \quad \text{a.s.}
\]

**Proof.** According to the definition of \( s_n \) (2.3) and (2.4) we have

\[
(1/n) \sum_{i=1}^{n} \rho_{1}( (y-T_{i}, n\times \xi)/s_n ) \leq a/2
\]

Then, since \( s_n \to \sigma_0 \) a.s., by (2.7), it would be enough to show that for any \( \sigma > 0 \) there exists \( L \) and \( \eta > 0 \) such that

\[
\lim_{n \to \infty} \inf |\sigma| > L (1/n) \sum_{i=1}^{n} \rho_{1}( (y-\theta \times \xi)/\sigma ) > a/2 + \eta \quad \text{a.s.}
\]

(8.13)

By the Lebesgue dominated convergence theorem, it is easy to show that for any \( \sigma > 0 \)

\[
\lim_{M \to \infty} E_{\mathcal{F}_0}^{\nu} (\rho_{1}( |y|-M / \sigma ) ) = a.
\]

(8.14)

By (D) and Lemma 3.4 there exist \( \nu > 0 \), \( \gamma > 0 \) and a finite number of sets \( C_1, C_2, \ldots, C_s \), included in \( \mathbb{R}^p \), such that

\[
\bigcup_{i=1}^{s} C_i \supset C = \{ \theta : \theta \in \mathbb{R}^p, |\theta| = 1 \}
\]

(8.15)

and

\[
\mathbb{P}_{\mathcal{F}_0} ( \inf_{\theta \in C_1} |\theta \times \xi| > \nu ) > 0.5 + \gamma.
\]

(8.16)

By (8.14) and (8.16) we can find \( M \) and \( \eta > 0 \) such that

\[
(0.5+\gamma) E_{\mathcal{F}_0}^{\nu} (\rho_{1}( |y|-M / \sigma )) > a/2 + \eta.
\]

(8.17)
Then from (8.16) and (8.17) we have

(8.18) \[ E_{H_0} \left( \inf_{q \in C_1} \mathbb{I}(|\hat{\theta}'x| > \varphi) \rho_1 \left( (|y| - M)/\sigma \right) \right) > a/2 + n. \]

Take \( L = M/\varphi \), then by (8.15) we have

\[ \inf_{|\varphi| > L} \left( (1/n) \sum_{i=1}^{n} \rho_1 \left( (|y_i| - (M/\varphi))/\sigma \right) I(|\varphi'x_i| > \varphi) \right) \]

\[ > \inf_{1 < i < s} \left( \inf_{q \in C_1} \rho_1 \left( (|y_i| - M)/\sigma \right) I(|\varphi'x_i| > \varphi) \right). \]

Then using the law of the larger numbers and (8.18) we get (8.13) and this proves the Lemma.

**Lemma 4.2.** Let \( g : \mathbb{R}^k \times \mathbb{R}^h \to \mathbb{R} \) continuous and let \( Q \) be a probability distribution on \( \mathbb{R}^k \) such that for some \( \delta > 0 \) we have

\[ E(\sup |\hat{\lambda}_m - \lambda_0| < \delta |g(z, \lambda)|) < \infty. \]

Let \( \hat{\lambda}_m \) be a sequence of estimates in \( \mathbb{R}^h \) such that

\[ \lim_{n \to \infty} \hat{\lambda}_m = \lambda_0 \quad \text{a.s..} \]

Then if \( z_1, z_2, \ldots, z_n \) are i.i.d. random variables in \( \mathbb{R}^k \) with distribution \( Q \), we have

\[ \lim_{n \to \infty} \left( (1/n) \sum_{i=1}^{n} g(z_i, \hat{\lambda}_m) \right) = E_Q(g(z, \lambda_0)). \quad \text{a.s.} \]

**Proof.** It is enough to show that for any \( \epsilon > 0 \) there exists \( \eta > 0 \) such that
(8.19) \[ \lim_{n \to \infty} \sup |\lambda - \lambda_0| \leq \eta (1/n) \sum_{i=1}^{n} g(z_i, \lambda) \leq E_Q (g(z, \lambda_0)) + \varepsilon \]
and

(8.20) \[ \lim_{n \to \infty} \inf |\lambda - \lambda_0| \leq \eta (1/n) \sum_{i=1}^{n} g(z_i, \lambda) \geq E_Q (g(z, \lambda_0)) - \varepsilon. \]

By the Lebesgue dominated convergence Theorem we can get \( \eta > 0 \) such that

\[ E(\sup |\lambda - \lambda_0| \leq \eta, g(z, \lambda)) \leq E_Q (g(z, \lambda_0)) + \varepsilon. \]

Then using the law of the large numbers we get (8.19). (8.20) is proved similarly.

Proof of Theorem 4.2. Put \( C(\varepsilon, L) = \{ \theta : \theta \in \mathbb{R}^p, \varepsilon < |\theta| < M \} \). According to Theorem 4.1, Lemma 4.1 and (2.7), it is enough to show that given any \( \varepsilon > 0 \) and arbitrarily large \( L \), there exist \( \gamma > 0 \) and \( \sigma_1 > \sigma_0 \) such that

(8.21) \[ \lim_{n \to \infty} \inf_{\theta \in C(\varepsilon, L)} \left( (1/n) \sum_{i=1}^{n} \rho_1 \left( \frac{(y_i - \theta, x_i)}{\sigma_1} \right) \right) \geq E_{F_0} (\rho_1(u/\sigma_0)) + \gamma \quad \text{a.s.} \]

and

(8.22) \[ \lim_{n \to \infty} \left( (1/n) \sum_{i=1}^{n} \rho_1 \left( \frac{(y_i - T_0, x_i)}{s_n} \right) \right) = E_{F_0} (\rho_1(u/\sigma_0)) \quad \text{a.s.} \]

By Lemma 3.2 and the Lebesgue dominated convergence Theo-
rem, using a standard compactness argument we can find \( \sigma_1 > \sigma_0 \), \( \gamma > 0 \) and a finite number of sets, \( C_1, C_2, \ldots, C_s \), such that

\[
E_{H_{\sigma_0}} (\inf_{g \in C_i} \phi_1 ((y-g'x)/\rho_1)) \geq E_{F_0} (\phi_1 (u/\sigma_0)) + \gamma \tag{8.23}
\]

and

\[
\cup_{i=1}^{s} C_i \supset C(\epsilon, \delta).
\]

By (8.24) we have

\[
\lim_{n \to \infty} \inf_{g \in C(\epsilon, \delta)} (1/n) \sum_{i=1}^{n} \phi_1 ((y_{i} - g'x_{i})/\sigma_1) > \inf_{1 \leq i \leq r} \lim_{n \to \infty} (1/n) \sum_{i=1}^{n} \inf_{g \in C_i} \phi_1 ((y_{i} - g'x_{i})/\sigma_1).
\]

Then by (8.23) and the law of the large numbers we get (8.21).

(8.22) follows from Lemma 4.2. This proves the Theorem.

In order to prove Theorem 5.1 we need the following Lemma.

Lemma 5.1. Under the assumptions of Theorem 5.1, we have

\[
p \lim (1/n^{1/2}) \sum_{i=1}^{n} \psi_1 (y_i/s_n) x_{ij} - \psi_1 (y_i/\sigma_0) x_{ij} = 0,
\]

where \( p \lim \) denotes limit in probability.

Proof. Put for \( 0 \leq t \leq 1 \)

\[
A_{n, j}(t) = (1/n^{1/2}) \sum_{i=1}^{n} \psi_1 (y_i/(0.5\sigma_0 + t\sigma_0)) x_{ij}.
\]

\( A_{n, j}(t) \) is an element of \( C \), the space of the continuous
functions defined in \([0,1]\). Since \(\lim_{n \to \infty} s_n = \sigma_o\) a.s., in order to prove the Lemma is enough to show that \(A_{n,j}(t),\ l \leq j \leq p,\) are tight. Then using Theorem 12.3 of Billingsley (1968), it will be enough to show the following two conditions

(i) \(A_{n,j}(0)\) is tight.

(ii) There exists \(\gamma > 0, \alpha > 0\) and a monotone continuous function \(R\) on \([0,1]\), such that for any \(0 \leq t_1 < t_2 \leq 1\) and any \(\lambda > 0\) we have

\[P_{\mathcal{H}_0}(|A_{n,j}(t_2) - A_{n,j}(t_1)| > \lambda) \leq (1/\lambda^{\gamma})(R(t_2) - R(t_1))^\alpha.\]

(i) holds by the Central Limit Theorem.

In order to prove (ii), by the Chebyshev inequality, it would be enough to show that there exists a constant \(K\) such that

\[(8.25) \quad E_{\mathcal{H}_0}((A_{n,j}(t_2) - A_{n,j}(t_1))^2) \leq K(t_2 - t_1)^2.\]

We have

\[(8.26) \quad E_{\mathcal{H}_0}((A_{n,j}(t_2) - A_{n,j}(t_1))^2) \leq E_{G_{\sigma_0}^0}(x_{i,j}^2)E_{F_0}((\psi_1(y/(0.5\sigma_0 + t_2\sigma_1)) - \psi_1(y/(0.5\sigma_0 + t_1\sigma_1)))^2)\]

According to assumption (B), let \(k_o = \sup |\psi_1(u)|.\) Then, since \(|y| > m\) implies \(\psi_1(y) = 0,\) we have
(8.27) \[ E_F \left( (\psi(y/(0.5\sigma_0 + t_2\sigma_0)) - \psi(y/(0.5\sigma_0 + t_1\sigma_0))^2 \right) \]
\[ \leq K_0^2 \left( \frac{1}{0.5\sigma_0 + t_2\sigma_0} - \frac{1}{0.5\sigma_0 + t_1\sigma_0} \right)^2 E_F (y^2 I(|y| < 1.5m\sigma_0) \]
\[ \leq 30 K_0^2 m^2 (t_2 - t_1)^2. \]

(8.26) and (8.27) imply (8.25) and therefore the Lemma is proved.

**Proof of Theorem 5.1.** We know by Theorems 4.1 and 4.2 that
\[ \lim_{n \to \infty} s_n = \sigma_0 \ a.s. \text{ and } \lim_{n \to \infty} T_{1,n} = 0 \ a.s. \]. We also have by (2.6)
\[ \sum_{i=1}^{n} \psi_1 ((y_i - T_{1,n}/s_n)/s_n)x_i = 0. \]

Using the Mean Value Theorem we get

(8.26) \[ (n^{1/2} T_{1,n} = s_n W_n^{-1} a_n, \]

where
\[ a_n = (1/n^{1/2}) \sum_{i=1}^{n} \psi_1 (y_i/s_n)x_i, \]
\[ W_n = (1/n) \sum_{i=1}^{n} \psi'_1 ((y_i - \theta^*_n x_i)/s_n)x_i x_i', \]
and \[ \lim_{n \to \infty} \theta^*_n = 0 \ a.s. \]. Using Lemma 4.2, we have

(8.29) \[ \lim_{n \to \infty} W_n = B(\psi_1, F_0)Y, \ a.s. \]

and using Lemma 5.1 and the Central Limit Theorem we have
\begin{equation}
\mathbb{E}_n \xrightarrow{d} N(0, A(\psi_1, F_0)\psi).
\end{equation}

Theorem 5.1 follows from (8.28), (8.29) and (8.30).

**Proof of Theorem (6.1).** Since \(S(t^{(i)}) < S(T_{0,n})\), (i) follows from Lemma 2.1.

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

\begin{equation}
S(t^*) = \sum_{i=1}^{n} \rho_1(r_i(t^*)/s_n) \geq ma.
\end{equation}

According to the definition of \(s_n\), (2.3) and (2.4), we get

\(S(T_{0,n}) < na/2\).

Then, since \(S(t^*) < S(T_{0,n})\), from (8.30) we get \(m < n/2\). Let \(H = \{ i : \rho_1(r_i(t^*)/s_n) < a \}\) and \(w_o = \inf_{i \in H} w_i(t^*)\). Then \(w_o > 0\) and

\begin{equation}
M(t^*) \geq w_o \sum_{i \in H} x_i x_i^t.
\end{equation}

Since \(c_n < 0.5\) and \(#H > n/2\) we have rank \(\{ x_i : i \in H \} = p\) and therefore by (8.31), we get that \(M(t^*)\) 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 $\lim_{j \to \infty} \lambda_{1,i} = \lambda_i^*$. We will show that $\lambda_i^* > 0$. Since $g(t)$ and $M(t)$ are continuous, $g(t^*) \neq 0$ and $M(t^*)$ positive definite, there exists $\epsilon > 0$ such that $|t_1 - t^*| < \epsilon$ and $|t_2 - t^*| < \epsilon$ imply

$$
(8.32) \quad |g(t_1) \cdot M^{-1}(t_2) g(t_2) - \frac{g(t_2)^T M^{-1}(t_2) g(t_2)}{g(t_2)^T M^{-1}(t_2) g(t_2)}| < 1 - \delta,
$$

and

$$
(8.33) \quad |g(t_1)| < 2|g(t^*)|
$$

and

$$
(8.34) \quad \mu_p(t_1) < 2\mu_p(t^*),
$$

where $\mu_p(t)$ is the maximum eigenvalue of $M^{-1}(t)$. Let $j^*$ be such that $j > j^*$ implies $|t (^{(i_j)} - t^*| < \epsilon/2$, then $A(t)^{(i_j)} < 4\mu_p(t^*)|g(t^*)|$ for $j > j^*$. Then $\lambda < \epsilon/(8\mu_p(t^*)|g(t^*)|)$ and $j > j^*$ implies $|\lambda A(t)^{(i_j)}| < \epsilon/2$ and therefore

$$
(8.35) \quad S(t) = S(t)^{(i_j)} + \lambda A(t)^{(i_j)} = S(t)^{(i_j)} - \lambda g(s)^T A(t)^{(i_j)}
$$

where $|s - t^*| < \epsilon$. Therefore by (8.32) we have

$$
g(s)^T M^{-1}(t)^{(i_j)} g(t)^{(i_j)} > g(t)^{(i_j)} M^{-1}(t)^{(i_j)} g(t)^{(i_j)} - (1 - \delta) g(t)^{(i_j)} M^{-1}(t)^{(i_j)} g(t)^{(i_j)} = \delta g(t)^{(i_j)} M^{-1}(t)^{(i_j)} g(t)^{(i_j)}.
$$
Therefore, by (8.35) we have

\[ S(\xi_{i_j}^*(t)) + \lambda \Delta(\xi_{i_j}^*) < S(\xi_{i_j}^*) - \beta \delta \eta(\xi_{i_j}^*) \cdot (\lambda \Delta(\xi_{i_j}^*)). \]

This implies \( k_{1,i_j} < \ln(\beta \eta p(\xi^*) | g(\xi^*) | / \xi) / \ln 2 + 1 \), and therefore \( \lambda^* > 0 \).

We can find \( j^* \) such that \( j > j^* \) implies \( \lambda_{1,i_j} > \lambda^*_1 / 2 \) and

\[ g(\xi_{i_j}^*) \cdot M^{-1}(\xi_{i_j}^*) g(\xi_{i_j}^*) > g(\xi^*) \cdot M^{-1}(\xi^*) g(\xi^*) / 2. \]

Then according to the definition of \( \lambda_{2,i} \) we have

(8.36) \[ S(\xi_{i_j+1}^*) = S(\xi_{i_j}^*) + \lambda_{2,i} \Delta(\xi_{i_j}^*) < S(\xi_{i_j}^*) + \lambda_{1,i} \Delta(\xi_{i_j}^*) \]

\[ < S(\xi_{i_j}^*) - \delta \lambda_{1,i} \cdot g(\xi_{i_j}^*) \cdot \Delta(\xi_{i_j}^*) \]

\[ = S(\xi_{i_j}^*) - \delta \lambda_{1,i} \cdot g(\xi_{i_j}^*) \cdot M^{-1}(\xi_{i_j}^*) g(\xi_{i_j}^*) \]

\[ < S(\xi_{i_j}^*) - (1/4) \delta \lambda_{1,i} \cdot g(\xi^*) \cdot M^{-1}(\xi^*) g(\xi^*). \]

Since \( S(\xi) > 0 \) for all \( \xi \) and \( g(\xi^*) \cdot M^{-1}(\xi^*) g(\xi^*) > 0 \), (8.36) can not hold for all \( j > j^* \). Therefore \( \lim_{j \to \infty} g(\xi_{i_j}^*) = 0 \). This proves (ii).

(iii) follows immediately from the fact that \( S(\xi_{i1}) \) is monotone decreasing.
REFERENCES


