WEAK CONVERGENCE OF BOUNDED INFLUENCE REGRESSION ESTIMATES WITH APPLICATIONS

by

Raymond J. Carroll and David Ruppert

University of North Carolina

Abstract

We obtain weak convergence results for bounded influence regression M-estimates and apply the results to sequential clinical trials, with special reference to repeated significance tests in the two-sample problem with covariates.

Key Words and Phrases: Regression, Robustness, Sequential Analysis, Repeated Significance Tests, M-Estimates, Weak Convergence.

AMS 1970 Subject Classifications: Primary 62E20; Secondary 62F10, 62G35.

Raymond J. Carroll's research was supported by the Air Force Office of Scientific Research under Grant AFOSR-80-0080. David Ruppert's research was supported by the National Science Foundation under Grant MCS78-01240.
1. **Introduction.**

Our primary concern is the comparison of two treatments in a clinical setting, although our results are quite general and this problem is only one application. To fix the discussion, in our major example we assume that patients are recruited sequentially and assigned at random into one of the two groups. Important covariable information is recorded at the time of recruitment. The response variable is numerical (e.g., systolic blood pressure, cholesterol level) and is observed either instantaneously or at a fixed time (from recruitment) in the future. We assume that there is either a maximum sample size or a target sample size \( N \) (see Section 3).

In most experiments of this type, the investigator will want to study the data periodically or as it is collected and (for ethical reasons if no other) stop if obvious differences are observed. In this instance a sequential analysis is called for, possibly based on repeated significance testing (RST) (see for example Siegmund (1977, 1979)). The usual methods for such tests would be based on least squares estimates and F-tests for a linear model. Such an analysis has two potentially serious difficulties:

(i) if there are outliers or if the error distribution is heavier-tailed than the normal distribution, the estimates will be inefficient and the tests will have low power;

(ii) particular design points (which may be "outliers in the design") can be extremely influential, even though the observed response is not an outlier; this could cause difficulties in interpretation at the very least.
To overcome the first problem, Huber (1977) and others have suggested the "classical" M-estimates, which give bounded influence to outliers in the response but unfortunately give unbounded influence to outliers in the design. Recently, there have been a number of proposals to handle problems (i) and (ii) jointly (Maronna and Yohai (1978)); these are the "bounded influence regression" estimates (BIR).

Our model takes the general form

\[ y_i = c_i \beta_0 + \gamma_0 z_i \quad (i = 1, 2, \ldots) , \]

where the \( \{z_i\} \) are i.i.d. with distribution function symmetric about zero. The parameter \( \gamma_0 \) is a scale factor to be explicitly defined later. For our purposes it is convenient and not unrealistic to assume that the design vectors \( \{c_i\} \) are also i.i.d. and independent of the errors; modifications of this assumption are discussed in Section 5.

In the two-sample problem, we will write \( \beta = (\mu \alpha \beta_n) \), where \( \mu \) is the intercept and \( \alpha \) is the treatment effect. An individual design point will take the form (1 \( \pm 1 \ c_n \)), where the sign is determined by the treatment group.

All the estimates mentioned so far can be generated in the following manner. Let \( (S_n(p \times p), \gamma_n) \) be estimates of scale which converge to \( (S_0, \gamma_0) \); particular choices of \( \gamma_n \) are "Proposal 2" (Huber (1977)) or "MAD" (Andrews et al. (1972)). Define \( \hat{\beta}_n \) as the solution to

\[ 0 = \frac{1}{n} \sum_{i=1}^{n} c_i g_1 \left( \left| c_i S_n c_i' \right|^{1/2} \right) \psi \left( \left| c_i S_n c_i' \right|^{1/2} (y_i - c_i \hat{\beta}_n) / \gamma_n \right) . \]
Particular choices of $g_1$, $g_2$, and $\psi$ give important special cases, e.g.,

**Least Squares**

\[ g_1(x) = g_2(x) = 1, \quad \psi(x) = x \]

**Classical M**

\[ g_1(x) = g_2(x) = 1, \quad \psi(x) \text{ bounded} \]

**BIR**

\[ |xg_1(x)| \text{ bounded}, \quad \psi(x) \text{ bounded}. \]

Special cases of BIR estimates which are particularly noteworthy include:

Mallows (unpublished) and Maronna and Yohai (1978)

\[ g_2(x) = 1, \quad S_n = I \]

Schweppe (unpublished)

\[ g_1(x)g_2(x) = 1, \quad S_n = I \]

Hampel and Krasker (unpublished)

\[ g_1(x)g_2(x) = 1, \quad g_1(x) = 1/x. \]

In Section 2 we state the major weak convergence results, apply them to the two-sample problem in Section 3, and prove them in Section 4. Extensions are discussed in Section 5.

2. **Major convergence results.**

Except where noted, we discuss only the special case that $S_n \equiv S_0$ (and we take $S_0 = I$ with no loss). Our two results depend on whether or not $\psi$ is monotone. If so, our technique is based on Yohai and Maronna (1979), while if not it is based on Carroll (1978). Since there are two results, we have chosen to list groups of possible assumptions.

In the applications we are going to assume that if $N$ is the maximum sample size, then one takes at least $K_N$ observations before stopping, where $K_N \to \infty$, $K_N/N \to 0$. We make use of this convention in this section. The proofs and assumptions can be modified to the case $K_N = \text{constant}$, but at a cost of notational complication.
Define the process \( \{G_N(\cdot)\} \) by

\[ G_N(s) = 0 \quad \text{for} \quad 0 \leq s < K_N/N \]

\[ = [N^s] \left[ \frac{\hat{\beta}}{[N^s]} - \frac{\beta}{E_0} \right] (Y_0, N^2)^{-1} \quad \text{for} \quad K_N/N \leq s \leq 1. \]

**Theorem 1.** Assume A1-A7 and B1-B5 below. Then

\[ \sup_{0 \leq s \leq 1} [N^s] N^{-1} \left| \sum_{i=1}^{[N^s]} c_i g_1(c_i) \psi(g_2(c_i)z) \right| \xrightarrow{p} 0. \]

**Theorem 2.** Assume A1-A7, B4 and C1-C5. Then

\[ \sup_{0 \leq s \leq 1} \left| \sum_{i=1}^{[N^s]} c_i g_1(c_i) \psi(g_2(c_i)z) \right| \xrightarrow{p} 0. \]

Clearly, Theorem 1 gives a weak convergence result for the process \([N^s] G_N(s)/N\) in \(D^P[0,1]\), while Theorem 2 considers the more interesting process \(G_N(s)\). In the next section these results and their applications are discussed.

Let \((c_1, z)\) be distributed as \((c_1, z_1)\).

A1. \(\psi\) is bounded, odd, absolutely continuous, and Lipschitz of order one.

A2. \(|c| g_1(c) \leq M_0, g_1(c) > 0\) almost surely.

A3. \(E|c| g_2(c) \leq M_0, g_2(c) \geq 1\) almost surely.
A4. There exists \( A_0 > 0 \) such that for every \( |t| \leq A_0 \),

\[
\sup_{|q|, |h| \leq A_0} h^{-2} E[\psi((1+q+h)g_2(c)(z-c\tau)) - \psi((1+q)g_2(c)(z-c\tau))]^2 < \infty.
\]

A5. \( \Sigma_1 = Ec'c'g_1(c)g_2(c)\psi'(g_2(c)z) \) is of full rank, and as \( |t| \to 0 \), \( r \to 0 \),

\[
|Ec'g_1(c)\psi(g_2(c)(1+r)(z-c\tau)) + \Sigma_1 t| = o(|t|+|r|).
\]

A6. \( \Sigma_2 = E[c'c\{g_1(c)\psi(g_2(c)z)\}]^2 \) is of full rank.

A7. \( \lim_{\varepsilon \to 0} \sup_{|q|, |t| \leq \varepsilon} E[\psi((1+q)g_2(c)(z-c\tau)) - \psi(g_2(c)z)]^2 = 0 \).

(This actually follows from A1 - A4, but it is convenient to state it separately.)

B1. \( \psi \) is monotone nondecreasing and nonconstant, and if \( D(u,z) = (\psi(u+z) - \psi(u))/z \), then \( |u| \leq a, |z| \leq b \) imply \( D(u,z) \geq d > 0 \).

B2. \( P(|z| \leq \varepsilon) > 0 \) for all \( \varepsilon > 0 \).

B3. The minimal eigenvalue \( \lambda_{\min} \) of the matrix

\[
Eg_1(c)g_2(c)c'c\{g_2(c) \leq a, g_2(c) |c| \leq b} \]

is positive.

B4. \( \sup\{[N_s]|\gamma_{[N_s]} - \gamma_0|N^{-1/2} : 0 \leq s \leq 1\} = o_p(1) \).
B5. For every $\varepsilon > 0$, there exists $Q, N_1$ such that $N \geq N_1$ implies

$$P\left\{ Q^{-1} \leq \gamma_n \leq Q \text{ for all } K_N \leq n \leq N \right\} \geq 1 - \varepsilon .$$

C1. In addition to A1, $\psi$ has two bounded continuous derivatives except on a finite set $E$.

C2. $\psi$ is constant outside an interval.

C3. The distribution function of $g_2(c)z$ is continuous at all $b \in B$.

C4. For every $\varepsilon > 0$ as $N \to \infty$,

$$P\left\{ |\hat{B}_n - B_0| > \varepsilon \text{ for some } K_N \leq n \leq N \right\} \to 0 .$$

C5. For every $\varepsilon > 0$ as $N \to \infty$,

$$P\left\{ |\gamma_n - \gamma_0| > \varepsilon \text{ for some } K_N \leq n \leq N \right\} \to 0 .$$

Remark 2.1. A1 holds for the common M-estimators. B1 is as in Yohai and Maronna (1979), while C1-C3 are needed by Carroll (1978).

Remark 2.2. A2-A3 are not unusual for BIR estimates. A4, A5, and A7 are smoothness conditions corresponding to those of Bickel (1975).

Remark 2.3. When $\psi$ is monotone, B2-B5 are specially adapted versions of assumptions in Yohai-Maronna. In an unpublished manuscript, they have also shown C4-C5 when $g_2(c) \equiv 1$, $\psi$ is monotone, and scale is estimated by Proposal 2.
3. Two-sample problem.

For the two-sample problem, the treatment effect is $a' \hat{\beta}_0 = \alpha$ where $a' = (0 \ 1 \ 0 \ ... \ 0)$. We consider the situation in which we are allowed to take at most $N$ observations and let the process start by taking at least $K_N$ observations. The usual RST methodology for testing $H_0: \alpha = 0$ vs. $H_1: \alpha \neq 0$ follows this form: for each $K_N \leq n \leq N$, one computes the usual $F$- or $t$-statistic based on $n$ observations and compares it to a cutoff point (possibly depending on $n,N$). $H_1$ is chosen and experimentation stopped if for any $K_N \leq n < N$ the test statistic is "large"; at the final observation ($n = N$), a new cutoff point is defined.

Note that in the previous section we have shown that if $W$ is Brownian motion,

$$[Ns]N^{-1} a' G_N(s)(\Sigma_1^{-1} \Sigma_2^{-1} a)^{-1/2} \Rightarrow sW(s)$$

where "$\Rightarrow$" denotes weak convergence in $D[0,1]$. Let $\{C_N(\cdot)\}$ be a sequence of nonnegative functions in $D[0,1]$ converging (uniformly) to a continuous function $C(\cdot)$ and let $D$ be a positive number. If $\gamma_0$, $\Sigma_1$, and $\Sigma_2$ were all known, then a general class of RST's which fits into the framework outlined above would take the following form:

Take an initial sample of $K$ observations. Reject $H_0$ in favor of $H_1$ if either of the following obtains:

(i) $[Ns]N^{-1/2} |a' \hat{\beta} |_{[Ns]} > C_N(s)A_\infty$ for some $K_N/N \leq s \leq 1$;

(ii) $N^{1/2} |a' \hat{\beta}_N | > DA_\infty$

(where $A_\infty = \gamma_0 a' \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1} a$).

The constant $A_\infty$ can be estimated. One reasonable estimate is
\[ A_N(s) = \gamma_{[Ns]} a' \Sigma_{1N}(s)^{-1} \Sigma_{2N}(s) \Sigma_{1N}(s)^{-1} a \]

where

\[ \Sigma_{1N}(s) = [Ns]^{-1} \sum_{i=1}^{[Ns]} c_i c_i g_1(c_i) g_2(c_i) \]

\[ \cdot \psi'(g_2(c_i)(y_i - c_i \hat{\theta}_{[Ns]}/\gamma_{[Ns]})^2 \]

and

\[ \Sigma_{2N}(s) = [Ns]^{-1} \sum_{i=1}^{[Ns]} c_i c_i g_1(c_i) \]

\[ \cdot \psi(g_2(c_i)(y_i - c_i \hat{\theta}_{[Ns]}/\gamma_{[Ns]}))^2 \]

The procedure we propose is the same as above but with \( A_N(s) \) replacing \( A_\infty \). Before analyzing the power of this test, we need the following:

**Lemma 3.1.** Under the assumptions of Theorem 1, C4 holds if we also assume that for all \( L > 0 \), \( \sigma > 0 \) and \( |\theta| = 1 \),

\[ \text{Eg}_1(c)(c \theta)\psi(\text{dg}_2(c)(z-Lc \theta)) < 0 \]

(We will analyze the power of these tests for contiguous alternatives so the term \( a' \beta_0 \) below will depend upon \( N \), though this will not be explicit in the notation.)

**Theorem 3.** Suppose that \( N^k a' \beta_0 + \eta \) (finite), that C4 and C5 hold, and that either (2.2) or (2.3) holds. Suppose further that \( \psi \) satisfies A1-A7 and
(3.4) \[
\lim_{\varepsilon \to 0} E \left| g_2(c) \right| \sup_{|r| \leq \varepsilon, |t| \leq \varepsilon} \left| \psi'(g_2(c)(1+r)(z-c)t) - \psi'(g_2(c)z) \right| = 0 .
\]

Then the power of the RST converges to

(3.5) \[
P\left\{ |W(s) + s\eta_*| > C(s) \text{ for some } 0 \leq s \leq 1 \text{ or } |W(1) + \eta_*| > D \right\},
\]

where \( \eta_* = \eta(\gamma_0^2 a', \Sigma_1^{-1} \Sigma_2^{-1} \Sigma_1^{-1} a)^{-1/2} \).

Remark 3.1. The quantity (3.5) can be computed from Anderson (1960). In comparing different tests, the relevant quantity is

\[
\gamma_0^2 a' \Sigma_1^{-1} \Sigma_2^{-1} \Sigma_1^{-1} a .
\]

Remark 3.2. Siegmund (1979) has considered RST's with (essentially)

\[
\tilde{C}_N(s) = (2[Ns]^2 N^{-1}(\exp(a/[Ns])-1))^{1/2}
\]

and has derived approximations in the normal case which are better than (3.5).

Remark 3.3. The approximations (2.2) and (2.3) can be used to construct analogues to classical F-tests (see Schrader and Hettmansperger (1980) for another approach). One could test \( H_0: K' \beta = 0 \) by rejecting for large values of

\[
(K' \beta \tilde{\beta}_{[Ns]}')' \left[ K' (\Sigma_1 N(s))^{-1} \Sigma_2 N(s) \Sigma_1 N(s)^{-1} K' \right]^{-1} K' \beta \tilde{\beta}_{[Ns]} .
\]
4. **Proofs.**

The proofs of Theorems 1-3 are broken into a number of steps. In Lemmas 4.1 - 4.4 we make the assumptions of Theorem 1. Also define $\gamma = \sigma^{-1}$ and $\gamma_n = \sigma^{-1}_n$.

**Lemma 4.1.** Define $\xi_1(r) = c_i g_1(c_i) \psi(g_2(c_i)) z_1(1+r)$. Then the process

$$H_N(r,s) = N^{-\frac{1}{2}} \sum_{i=1}^{[Ns]} \xi_1(r)$$

on $D([-A_0, A_0] \times [0,1])$ converges weakly to a Gaussian process which is almost surely continuous.

**Proof.** The finite dimensional distributions converge by A1 and A2. By A1, A2, A4 and the remark following Theorem 3 of Bickel and Wichura (1971, p. 1665), $H_N$ is tight (in the notation of their equation (3), $\gamma_1 = \gamma_2 = 2$, $\beta_1 = \beta_2 = 1$). Define the continuous process

$$H_N^*(r,s) = H_N(r,s) + (Ns - [Ns]) \xi_{[Ns]+1}(r) N^{-\frac{1}{2}}$$

$$= H_N(r,s) + o_p(1)$$

uniformly in $(r,s)$, from which the lemma follows.

**Lemma 4.2.** Define

$$V_N^*(r,s,t) = [Ns] N^{-3/2} \sum_{i=1}^{[Ns]} c_i g_1(c_i) \psi(g_2(c_i)) (z_i - c_i t N^{1/2})^{1/2} (1+r N^{1/2}) ([Ns]) (1+r N^{1/2}) ([Ns])$$

$$- V_N^*(r,s,t) = V_N^*(r,s,t) - V_N^*(0,s,0) - EV_N^*(r,s,t).$$

Then for all $\epsilon > 0, M > 0$ there exists $B = B(\epsilon, M)$ such that
\[ P\{ |V_N(r,s,t)| > \varepsilon \text{ for some } |r| \leq M, |t| \leq M, BN^{-1/2} s \leq 1 \} < \varepsilon . \]

**Proof.** Fix \( \varepsilon > 0, M > 0 \). For any \( B > 0, \delta > 0, A2-A3 \) and the fact that \( \psi \) is Lipschitz show

\[
\sup\{|V_N(r,s,t) - V_N(r,s,0)| : BN^{-1/2} s \leq \delta, |r| \leq M, |t| \leq M\} = O_p(\delta) .
\]

From Lemma 4.1 we can choose \( B \) large and \( \delta \) small so that

\[
P\{ \sup\{|V_N(r,s,0)| : BN^{-1/2} s \leq \delta, |r| \leq M, |s| > \varepsilon\} < \varepsilon/2 .
\]

Hence it suffices to prove the lemma when \( \delta \leq s \leq 1 \) for arbitrary \( \delta \); since \( M \) is arbitrary we need only prove the lemma for the process generated by

\[
V_N^*(r,s,t) = N^{-1/3} \sum_{i=1}^{[Ns]} \frac{c g_1(c_i) \psi(g_2(c_i)(z_i - c_i t N^{-1/2}))(1 + r N^{-1/2})}{\epsilon_i} .
\]

With this new process \( V_N^* \) we need only prove tightness since (\( A_2 \) and \( A_7 \)) \( \text{Var}(V_N^*(r,s,t)) \to 0 \). For any \( t_* \), by \( A1-A3 \)

\[
\sup\{|V_N^*(r,s,t) - V_N^*(r,s,t_*)| : 0 \leq s \leq 1, |r| \leq M, |t - t_*| \leq \eta\} = O_p(\eta) .
\]

This means we need only prove tightness of \( V_N^*(\cdot, \cdot, t_*) \), which follows from \( A4 \) as in the proof of Lemma 4.1. \( \Box \)

**Lemma 4.3.** For every \( M > 0 \),

\[
\sup\{n^2 |\bar{\beta}_n| N^{-3/2} : K_n \leq n \leq MN^{1/2}\} \overset{P}{\to} 0 .
\]
Proof. Fix $\epsilon, M > 0$. From B3,

$$\sup\{P(|\theta - 0| = 0): |\theta| = 1\} < 1,$$

so that for every $D > 0$ there exists $\eta(D)$ with

$$(4.1) \quad \sup\{P(0 < |\theta - 0| < \eta(D)): |\theta| = 1\} < D/2.$$ 

Choose $Q > 1$ to satisfy the inequality of B5 and $P(|z| > Q) < \epsilon$. Define

$$R_n(\theta, L) = n^{-1} \sum_{i=1}^{n} g_1(c_i - \theta) \psi(g_2(c_i)(z_i - Lc_i - \theta)) \theta_n.$$

As in the proof of Theorem 2.2 of Yohai and Maronna (1979), $n^2|\hat{\theta}_n| N^{-3/2} > \epsilon$

for some $K_N \leq n \leq MN^{3/2}$ implies $|\hat{\theta}_n| > \epsilon N^3/M^2$ so that $R_n(\theta, \epsilon N^3/M^2) > 0$

for some $|\theta| = 1$. Since $\psi$ is monotone and skew symmetric, with

probability at least $1 - 2\epsilon$ by B5,

$$R_n(\theta, \epsilon N^3/M^2) = n^{-1} \sum_{i=1}^{n} g_1(c_i) |c_i - \theta| \psi(\sigma_n g_2(c_i)(\text{sign}(c_i - \theta)z_i - \epsilon N^3/c_i \theta | M^{-2})).$$

$$\leq n^{-1} \sum_{i=1}^{n} g_1(c_i) |c_i - \theta| \psi(\sigma_n g_2(c_i)(|z_i| - \epsilon N^3/c_i \theta | M^{-2})).$$

$$\leq M_0 \sup |\psi| n^{-1} \sum_{i=1}^{n} I\{|z_i| > Q\}$$

$$+ n^{-1} \sum_{i=1}^{n} g_1(c_i) |c_i - \theta| \psi(Q^2 g_2(c_i)(1 - \epsilon N^3/c_i \theta | M^2 Q^3))$$

$$\cdot I\{|z_i| \leq Q\}$$

$$= R_{n1} + R_{n2}.$$ 

By the strong law of large numbers (SLLN),
\[
\limsup_{n \to \infty} R_{n1} \leq \varepsilon M_0 \sup |\psi| \quad (a.s.)
\]

Since \(\psi\) is monotone, applying Lemma 1 of Yohai (1974) shows that for every \(L > 1\) (since \(N^{1/2}/QM^2 \to \infty\))

\[
(4.2) \quad \limsup_{n \to \infty} \sup_{|\theta| = 1, K_n \leq n \leq N} \quad \leq \sup_{|\theta| = 1} E g_1(c) c \theta |\psi(g_2(c)Q^2(1-L|c \theta|))P(z_i \leq Q) \quad (a.s.)
\]

Since \(\psi\) is monotone nondecreasing and nonconstant, \(\liminf_{x \to -\infty} \psi(x) < 0\).

Therefore from (4.1), \(g_1(c) > 0\), (A2), dominated convergence, and Fatou's Lemma, one can choose \(L\) and \(Q\) sufficiently large so that the right side of (4.2) is at most \(A_\infty < 0\) and \(A_\infty\) is independent of \(\varepsilon\). One completes the proof by choosing \(\varepsilon_\star\) sufficiently small.

\[\square\]

**Lemma 4.4.** For every \(\varepsilon > 0\) there exists \(M=M(\varepsilon) > 0\), \(L(\varepsilon)\) such that for \(N\) sufficiently large

\[
P\{n \mid \hat{\theta}_n \mid N^{1/4} > L \text{ for some } \lambda N^{1/4} \leq n \leq N\} < 4 \varepsilon.
\]

**Proof.** Choose \(Q\) as in B5 and define

\[
R_N(\theta, L, n) = N^{1/2} \sum_{i=1}^{n} g_1(c_i)(c_i \theta) \psi(g_2(c_i)(z_i - Lc_i \theta)\sigma_n).
\]

Let \(L = M/Q\). As in Theorem 2.2 of Yohai and Maronna (1979),

\[
n \mid \hat{\theta}_n \mid N^{1/4} > L \implies R_N(\theta, LN^{1/4}/n, n) > 0 \text{ for some } |\theta| = 1. \quad \text{When } Q^{-1} \leq \sigma_n \leq Q \text{ (with probability at least } 1-\varepsilon \text{ from B5) uniformly in } |\theta| = 1, \text{ then by B1},
\]
\[ R_N(\theta, LN^{-1/2}/n, n) \leq A_1(n, N, \theta) - A_2(n, \theta) \]

\[ = N^{-1/2} \sum_{i=1}^{n} g_1(c_{i \theta})(c_{i \theta}) \psi(g_2(c_{i \theta})z_i \sigma_n) \]

\[ - dMQ^{-2} n^{-1} \sum_{i=1}^{n} g_1(c_{i \theta}) g_2(c_{i \theta}) (c_{i \theta})^2 \]

\[ \cdot I\{|z_i| \leq Q^{-1}\} I\{g_2(c_{i \theta}) \leq a, g_2(c_{i \theta}) |c_{i \theta}| \leq b\}. \]

By the SLLN and B3,

\[ \limsup_{n \to \infty} A_2(n, \theta) \geq dM \min P\{|z| \leq Q^{-1}\} (a.s.). \]

Therefore, it suffices to show that

\[ \sup_{|\theta| = 1} |A_1(n, N, \theta)| = O_p(1) \]

\[ \text{for some } M. \]

We have for \(|\theta| = 1\) and \(MN^{-1/2} \leq n \leq N\)

\[ |A_1(n, N, \theta)| \leq \sup_{|\theta| = 1} \sum_{i=1}^{n} g_1(c_{i \theta}) (c_{i \theta}) \psi(g_2(c_{i \theta})z_i \sigma_n). \]

Assume \(M > 1\). Then from B4 there exists \(M_*\) independent of \(M\) (we may take \(M_* < M\)) so that

\[ P\{\sup_{\{Ns\}} |Y_{Ns} - Y_0| N^{-1/2} : 0 \leq s \leq 1 \geq M_*\} < \epsilon. \]

Hence
\[ p\{\sup\{n | \gamma_n - \gamma_0 | N^{-1/2} : MN^{1/2} \leq n \leq N \} \geq M_* \} < \varepsilon \]

or

\[ p\{\sup\{|\gamma_n - \gamma_0| : MN^{1/2} \leq n \leq N \} \leq (M_* N^{1/2} n^{-1}) \} \geq 1 - \varepsilon. \]

Therefore for \( M \) sufficiently large

\[ p\{\sup\{|\gamma_n - \gamma_0| : MN^{1/2} \leq n \leq N \} \leq \gamma_0/2 \} \geq 1 - \varepsilon; \]

hence for some \( M_{**} \) not depending on \( M \)

\[ p\{|\sigma_n - \sigma_0| \leq M_{**} \text{ for all } MN^{1/2} \leq n \leq N \} \geq 1 - \varepsilon. \]

Thus, since \( \sigma_0 = 1 \),

\[ \sup_{|\theta| = 1} |A_1(n, N, \theta)| \leq \sup_{0 \leq s \leq 1} \sup_{N^{1/2} \leq n \leq N} |N^{-1/2} \sum_{i=1}^{[Ns]} g_1(c_i) c_i \psi(g_2(c_i) z_i (1 + r))| = O_p(1) \]

by Lemma 4.1.

Proof of Theorem 1. Fix \( \varepsilon > 0 \) and \( \delta > 0 \). Then Lemma 4.4 shows that when the sup in (2.2) is taken over \( K_N/N \leq s \leq \delta \), (2.2) is \( O_p(\delta) \).

Since \( \delta \) is arbitrary, it suffices to prove (2.2) when the sup is taken over \( a \leq s \leq 1 \) for any \( a > 0 \). In Lemma 4.2 set

\[ t = \sigma_0 [Ns] (\hat{\rho}_{[Ns]} - \rho_0) N^{-1/2}, r = [Ns] (\hat{\sigma}_{[Ns]} / \sigma_0 - 1) N^{-1/2}, \]

both of which are \( O_p(1) \) (by Lemma 4.4 and B4). Then use A5; since \( a \leq s \leq 1 \), the proof is complete.
Lemma 4.5. Assume A1-A7 and C1-C5. There is a function $h(x) \to 0$ as $x \to 0$ for which the following obtains: for every $\varepsilon > 0$ sufficiently small, almost surely as $n \to \infty$ and uniformly in $|\sigma - \sigma_0| \leq \varepsilon, |\beta - \beta_0| \leq \varepsilon$

\begin{align}
(4.3) \quad |n^{-1} \sum_{i=1}^{n} c_i \, g_1(c_i)(\psi(g_2(c_i))(y_i - c_i \beta)\sigma - \psi(g_2(c_i)z_i) + \sigma c_i g_2(c_i)\psi'(g_2(c_i)z_i)(\beta - \beta_0) - g_2(c_i)z_i \psi'(g_2(c_i)z_i)(\sigma - \sigma_0)|
\leq h(\varepsilon)(|\sigma - \sigma_0| + |\beta - \beta_0|).
\end{align}

Proof. Set $\sigma_0 = 1, \beta_0 = 0$ without loss. Define

\[ M_2 = 2 + \max\{|K|, KeB\} \]

\[ d(\varepsilon) = (2M_2 \varepsilon^{2.5})^{-1} \]

\[ A(\varepsilon) = \{ |c| g_2(c) \geq d(\varepsilon) \} \]

\[ L(\varepsilon, \sigma, \beta) = \{ |g_2(c)\sigma(z - c\beta) - K| > 2\varepsilon d(\varepsilon) \text{ for all } KeB \} \]

Using C2 choose $\alpha > 1$ such that $\psi$ is constant on $(-\infty, -\alpha]$ and constant on $[\alpha, \infty)$. Choose $\varepsilon$ is so small that

\[ 2\alpha/(1+\varepsilon) - \varepsilon d(\varepsilon)/(1-\varepsilon) \geq \alpha \]

and

\[ 2\alpha \varepsilon + \varepsilon d(\varepsilon)(1+\varepsilon) \leq 2\varepsilon d(\varepsilon) . \]

Now suppose that $|\sigma - 1| \leq \varepsilon, |\beta| \leq \varepsilon, (z, c) \in L(\varepsilon, \sigma, \beta)$ and $c \notin A(\varepsilon)$. Then if $g_2(c)z \geq 2\alpha$, it follows that
\[ g_2(\zeta) \sigma z \geq 2\alpha/(1+\varepsilon) \geq \alpha \]

and
\[ g^2(\zeta) \sigma (z - c\beta) \geq 2\alpha/(1+\varepsilon) - |g_2(\zeta) c\beta|/(1-\varepsilon) \geq \alpha \]

so that \( \psi(g_2(\zeta) \sigma (z - c\beta)) = \psi(g_2(\zeta) z) \); similarly \( g_2(c) z \leq -2\alpha \) implies that \( \psi(g_2(c) \sigma (z - c\beta)) = \psi(g_2(c) z) \). If \( |g_2(c) z| < 2\alpha \), then
\[
|g_2(c)(z - \sigma(z - c\beta))| \leq |g_2(c)z(\sigma - 1)| + |g_2(c)\sigma c\beta|
\leq 2\alpha \varepsilon + \varepsilon d(\varepsilon)(1+\varepsilon) \leq 2\varepsilon d(\varepsilon),
\]

so that \( g_2(c) z \) is in the same interval between points of \( \mathcal{B} \) as is \( g_2(c) \sigma (z - c\beta) \), so that we may use a Taylor expansion to bound the summands in (4.3).

If \( |\sigma - 1| > \varepsilon, |\beta| > \varepsilon, (z, c) \notin L(\varepsilon, \sigma, \beta) \), or \( c \in A(\varepsilon) \), then we can use the fact that \( \psi \) is Lipschitz and constant on \( (-\infty, -\alpha) \) and on \( [\alpha, \infty) \) to bound the summands. This suggests that we can bound (4.3) by
\[ C_0 \{ A_{n1} + A_{n2} + A_{n3} + A_{n4} \} \]
where \( C_0 \) depends only on \( \psi \) and

\[
A_{n1} = n^{-1} \sum_{i=1}^{n} \{ g_2(c_{1,i}) (|\sigma - 1| + |c_{1,i}| |\beta|) \}^2 \mathbb{I}\{|c_{1,i}| g_2(c_{1,i}) \leq d(\varepsilon)\}
\]

\[
A_{n2} = n^{-1} \sum_{i=1}^{n} g_2(c_{1,i}) (|\sigma - 1| + |c_{1,i}| |\beta|) \mathbb{I}\{(z_{1,i}, c_{1,i}) \notin L(\varepsilon, \sigma, \beta)\}
\]

\[
\cdot \{ |c_{1,i}| g_2(c_{1,i}) \leq d(\varepsilon) \}
\]

\[
A_{n3} = n^{-1} \sum_{i=1}^{n} g_2(c_{1,i}) (|\sigma - 1| + |c_{1,i}| |\beta|) \mathbb{I}\{|c_{1,i}| g_2(c_{1,i}) \geq d(\varepsilon)\}
\]

\[
A_{n4} = n^{-1} \sum_{i=1}^{n} |\psi(g_2(c_{1,i}) \sigma(z_{1,i} - c_{1,i}\beta)) - \psi(g_2(c_{1,i}) z_{1,i})| \mathbb{I}\{|c_{1,i}| g_2(c_{1,i}) \geq d(\varepsilon)\}.
\]
We must prove the bound of the lemma for each term. This is easy to do for $A_{n1}$ and $A_{n3}$ by applying the SLLN and using A3. For $A_{n4}$, since $\psi$ is Lipschitz and A3 holds, we need only prove the lemma when $\beta = 0$. Since $\psi$ is constant off a compact set, there is a constant $M_1$ with

$$A_{n4}(\beta = 0) \leq |\sigma - 1| n^{-1} \sum_{i=1}^{n} g_2(c_i) |z_i| I\{|g_2(c_i)\geq d(\epsilon), g_2(c_i)|z_i| \leq M_1\},$$

so from A3 the lemma holds for $A_{n4}$. Finally, for some constant $M_2$,

$$A_{n2} \leq (|\sigma - 1| + |\beta|) n^{-1} \sum_{i=1}^{n} g_2(c_i) |c_i| I\{|g_2(c_i)z_i - K| \leq M_2 \epsilon d(\epsilon)\}$$

for some $K \epsilon B$,

so that A3, C3 and the SLLN and dominated convergence complete the proof.

Proof of Theorem 2. Fix $\epsilon_1, \epsilon_2 > 0$. Using the fact that $\Sigma_1$ is of full rank, and applying A1, A2, B4, C4, C5 and Lemma 4.5, we see there exists $M_1, M_2$ such that $N \geq M_1$ implies

$$P\{|\widehat{\beta}_n - \beta_0| + |\sigma_n - \sigma_0| > \epsilon_1 \text{ for some } K_N \leq n \leq N\} \leq \epsilon_1$$

$$P\{|n^{-1} \sum_{i=1}^{n} c_i g_1(c_i) \psi(g_2(c_i)z_i)| > M_2 \text{ for some } K_N \leq n \leq N\} \leq \epsilon_2$$
\[ P\left[ \Sigma_{1}^{n-1} \sum_{i=1}^{n} c_{i} s_{1}(c_{i}) \{ \psi(g_{2}(c_{i})\sigma(y_{i} - c_{i} \beta)) - \psi(g_{2}(c_{i})z_{i}) + \sigma c_{i} g_{2}(c_{i}) \psi'(g_{2}(c_{i})z_{i})(\beta - \beta_0) \right. \]
\[ \left. - g_{2}(c_{i})z_{i} \psi'(g_{2}(c_{i})z_{i})(\sigma - \sigma_0) \right\} > 4h(\varepsilon_1)(|\sigma - \sigma_0| + |\beta - \beta_0|) \]

for some \( K_{n} \leq n \leq N \) and some \( |\sigma - \sigma_0| \leq \varepsilon_1, |\beta - \beta_0| \leq \varepsilon_1 \), \( \varepsilon_1 \).

Choosing \( \beta = \hat{\beta}_n, \sigma = \sigma_n \) we see that if \( N \geq M_1 \) the probability that the following event obtains for all \( K_{n} \leq n \leq N \) is at least \( 1 - 2(\varepsilon_1 + \varepsilon_2) \):

\[ n|\hat{\beta}_n - \beta_0|N^{-\frac{1}{2}} \leq 4h(\varepsilon_1)n|\hat{\beta}_n - \beta_0|N^{-\frac{1}{2}} + 2M_2. \]

For (4.7) to occur we must have (for \( \varepsilon_1 \) small) \( n|\hat{\beta}_n - \beta_0|N^{-\frac{1}{2}} \leq 4M_2 \), proving that

\[ \sup\{n|\hat{\beta}_n - \beta_0|N^{-\frac{1}{2}} : K_{n} \leq n \leq N\} = 0 \quad p(1). \]

Placing \( n(\hat{\beta} - \beta_0)N^{-\frac{1}{2}} \) into (4.6) with \( \sigma = \sigma_n, \beta = \beta_n \) completes the proof. \( \square \)

**Proof of Lemma 3.1.** This follows by applying the technique of Lemma 4.3, B5, Lemma 1 of Yohai (1974), and (3.3). \( \square \)

**Remark.** A1, A3, and B1-B2 imply (3.4) by dominated convergence.

**Proof of Theorem 3.** We will first show that for \( j = 1, 2 \)

\[ \sup\{|\Sigma_{jN}(s) - \Sigma_{j}| : K_{n}/N \leq s \leq 1\} \overset{P}{\to} 0. \]
First consider \( j = 1 \). From A2, C4 and C5, with probability at least \( 1 - \epsilon/2 \),

\[
(4.10) \quad \left| \Sigma_1 N(s) - \frac{[N s]}{[N s]} \right| \leq \frac{[N s]}{[N s]} \sum_{i=1}^{[N s]} c_i \frac{c_i}{g_1(c_i)g_2(c_i)\psi'(g_2(c_i)z_i)}
\]

\[
\leq \frac{[N s]}{[N s]} \sum_{i=1}^{[N s]} |c_i| g_2(c_i)
\]

\[
|r| \leq \varepsilon, |t| \leq \varepsilon
\]

\[
\sup_{|r| \leq \varepsilon, |t| \leq \varepsilon} \left| \psi'(g_2(c_i)(1+r)(z_i - c_i t)) - \psi'(g_2(c_i)z_i) \right|
\]

One applies (3.4) and the SLLN to prove (4.9); a similar proof using A7 and Lemma 1 of Yohai handles \( j = 2 \). Since \( \Sigma_1 \) and \( \Sigma_2 \) are of full rank, the power becomes

\[
P\left( \left| a - \frac{\hat{b}}{[N s]} \right| > \frac{[N s]}{[N s]} A_N(s)C_N(s) \right)
\]

for some \( K_N/N \leq s \leq 1 \) or \( N^{1/2} |a - \frac{\hat{b}}{N} | > A_N(1)D \),

from which (3.5) follows by Theorem 1.

5. Extensions.

(a) It is possible to extend the results to classical M-estimators. In Theorem 2 we need \( E|c|^2 < \infty \). For Theorem 1, A4 and A7 must be modified so that Lemma 4.1 holds.

(b) We have dealt with the case in which the design vectors \( \{c_i\} \) are i.i.d., both because it is often reasonable and because it simplifies the exposition. If these vectors are (i) constants or (ii) generated in Section 3 by blocking on each group of two patients, it is still possible
to apply our techniques. Detailed and reasonably simple conditions can be worked out, both because of the nature of BIR estimates and because our results depend in large part on easily generalized standard SLLN and weak convergence techniques. The major difficulties lie in (4.1) and (4.2) and the terms $A_{n^3}, A_{n^4}$ in Lemma 4.5.

(c) It is possible but extremely messy to extend the results to estimating $S_0$ by $S_n$.

(d) In Lemma 4.5, the major use of symmetry is that it assures $E_{c} g_1(c)g_2(c)z\psi'(g_2(c)z) = 0$. For Section 3, the symmetry is unnecessary if $g_j(c) = g_j(\lvert c \rvert)$ ($j = 1, 2$) and the patients are randomly (with probability $\frac{1}{2}$) assigned to each group, for then

$$E_a' c g_1(c)g_2(c)z\psi'(g_2(c)z) = 0 .$$

(e) The approximation (2.3), while only of first order, can in fixed sample problems give some information about the dependence of the design on the rate of convergence through Berry-Esseen results for

$$\bar{\beta}_N \approx \gamma_0 N^{-1} \sum_{i=1}^{N} \Sigma_i^{-1} c_i' g_1(c_i) \psi(g_2(c_i)z_i) .$$

Since the $\{c_i\}$ are i.i.d., when $\gamma_0$ is estimated in an appropriately smooth fashion simultaneously with $\bar{\beta}$ (as in Huber's Proposal 2), the BIR estimates are minimum contrast estimates of a vector parameter. If $\psi$ has a derivative then Berry-Esseen results and Edgeworth expansions are available (e.g., Pflanzagl (1973) and Bhattacharya and Ghosh (1978)). For the Huber function $\psi(x) = \max(-b, \min(x, b))$ when the design is bounded, it should be possible to prove a version of (3.8) in Bickel (1974) and obtain an Edgeworth expansion. (In the location case with scale not estimated, Jurecková (unpublished) has verified Bickel's (3.8).)
REFERENCES


Weak Convergence of Bounded Influence Regression Estimates with Applications

Raymond J. Carroll and David Ruppert

Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC 20332

April 1980

Regression, Robustness, Sequential Analysis, Repeated Significance Tests, M-Estimates, Weak Convergence

We obtain weak convergence results for bounded influence regression M-estimates and apply the results to sequential clinical trials, with special reference to repeated significance tests in the two-sample problem with covariates.