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

$$(1.1) \qquad y_i = \underline{c}_i \underline{\beta}_0 + \gamma_0 z_i \qquad (i = 1,2,\ldots) ,$$

where the $\{z_i\}$ are i.i.d. with distribution function symmetric about zero. The parameter γ_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 $\{\underline{c}_i\}$ 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 $\underline{\beta} = (\mu\ \alpha\ \underline{\beta}_*)$, where μ is the intercept and α is the treatment effect. An individual design point will take the form $(1\ \pm 1\ \underline{c}_*)$, 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,γ_0); particular choices of γ_n are "Proposal 2" (Huber (1977)) or "MAD" (Andrews et al. (1972)). Define $\hat{\underline{\beta}}_n$ as the solution to

$$(1.2) \qquad 0 = \sum_{i=1}^{n} \underline{c}_i' g_1(|\underline{c}_i S_n \underline{c}_i'|^{\frac{1}{2}})\psi(g_2(|\underline{c}_i S_n \underline{c}_i'|^{\frac{1}{2}} (y_i - \underline{c}_i \underline{\beta})/\gamma_n) .$$

Particular choices of g_1, g_2, and ψ give important special cases, e.g.,

Least Squares	$g_1(x) = g_2(x) = 1$,	$\psi(x) = x$
Classical M	$g_1(x) = g_2(x) = 1$,	$\psi(x)$ bounded
BIR	$\|xg_1(x)\|$ bounded ,	$\psi(x)$ bounded .

Special cases of BIR estimates which are particularly noteworthy include:

Mallows (unpublished) and Maronna and Yohai (1978)	$g_2(x) = 1$,	$S_n = I$
Schweppe (unpublished)	$g_1(x)g_2(x) = 1$,	$S_n = I$
Hampel and Krasker (unpublished)	$g_1(x)g_2(x) = 1$,	$g_1(x) = 1/x$.

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 ψ 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 = constant, but at a cost of notational complication.

Define the process $\{G_N(\cdot)\}$ by

$$(2.1) \qquad G_N(s) = 0 \qquad\qquad 0 \le s < K_N/N$$

$$= [Ns](\hat{\underline{\beta}}_{[Ns]} - \underline{\beta}_0)(\gamma_0 N^{\frac{1}{2}})^{-1} \qquad K_N/N \le s \le 1 .$$

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

$$(2.2) \qquad \sup_{0\le s\le 1} [Ns]N^{-1} |\Sigma_1 G_N(s) - N^{-\frac{1}{2}} \sum_{i=1}^{[Ns]} \underline{c}_i' \, g_1(\underline{c}_i)\psi(g_2(\underline{c}_i)z_i)| \overset{P}{\to} 0 .$$

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

$$(2.3) \qquad \sup_{0\le s\le 1} |\Sigma_1 G_N(s) - N^{-\frac{1}{2}} \sum_{i=1}^{[Ns]} \underline{c}_i' \, g_1(\underline{c}_i)\psi(g_2(\underline{c}_i)z_i)| \overset{P}{\to} 0 .$$

Clearly, Theorem 1 gives a weak convergence result for the process $[Ns]G_N(s)/N$ 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 $(\underline{c},\underline{z})$ be distributed as $(\underline{c}_1, \underline{z}_1)$.

A1. ψ is bounded, odd, absolutely continuous, and Lipschitz of order one.

A2. $|\underline{c}|g_1(\underline{c}) \le M_0$, $g_1(\underline{c}) > 0$ almost surely.

A3. $E|\underline{c}|g_2(\underline{c}) \le M_0$, $g_2(\underline{c}) \ge 1$ almost surely.

A4. There exists $A_0 > 0$ such that for every $|\underline{t}| \le A_0$,

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

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

$$|E\underline{c}'\,g_1(\underline{c})\psi(g_2(\underline{c})(1+r)(z-\underline{c}\,\underline{t})) + \Sigma_1\underline{t}| = o(|t|+|r|) .$$

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

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

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

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

B2. $P\{|z| \le \varepsilon\} > 0$ for all $\varepsilon > 0$.

B3. The minimal eigenvalue λ_{min} of the matrix

$$Eg_1(\underline{c})g_2(\underline{c})\underline{c}'\,\underline{c}I\{g_2(\underline{c})\le a\,,\ g_2(\underline{c})|\underline{c}|\le b\}$$

is positive.

B4. $\sup\{[Ns]|\gamma_{[Ns]} - \gamma_0|N^{-\frac{1}{2}}:\ 0 \le s \le 1\} = O_p(1)$.

B5. For every $\varepsilon > 0$, there exists Q, N_1 such that $N \geq N_1$ implies

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

C1. In addition to A1, ψ has two bounded continuous derivatives except on a finite set $\mathcal{B}$.

C2. ψ is constant outside an interval.

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

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

$$P\{|\hat{\underline{\beta}}_n - \underline{\beta}_0| > \varepsilon \text{ for some } K_N \leq n \leq N\} \to 0 .$$

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

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

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 ψ is monotone, B2-B5 are specially adapted versions of assumptions in Yohai-Maronna. In an unpublished manuscript, they have also shown C4-C5 when $g_2(\underline{c}) \equiv 1$, ψ is monotone, and scale is estimated by Proposal 2.

3. Two-sample problem.

For the two-sample problem, the treatment effect is $\underline{a}' \underline{\beta}_0 = \alpha$ where $\underline{a}' = (0\ 1\ 0 \ldots 0)$. 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,

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

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 γ_0, Σ_1, and Σ_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_N observations. Reject H_0 in favor of H_1 if either of the following obtains:

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

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

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

The constant A_∞ can be estimated. One reasonable estimate is

$$A_N(s) = \gamma_{[Ns]}\ \underline{a}'\ \Sigma_{1N}(s)^{-1}\ \Sigma_{2N}(s)\ \Sigma_{1N}(s)^{-1}\ \underline{a}$$

where

$$(3.1) \qquad \Sigma_{1N}(s) = [Ns]^{-1} \sum_{i=1}^{[Ns]} \underline{c}_i'\ \underline{c}_i g_1(\underline{c}_i) g_2(\underline{c}_i) \cdot \psi'(g_2(\underline{c}_i)(y_i - \underline{c}_i \hat{\underline{\beta}}_{[Ns]})/\gamma_{[Ns]})$$

and

$$(3.2) \qquad \Sigma_{2N}(s) = [Ns]^{-1} \sum_{i=1}^{[Ns]} \underline{c}_i'\ \underline{c}_i \{g_1(\underline{c}_i) \cdot \psi(g_2(\underline{c}_i)(y_i - \underline{c}_i \hat{\underline{\beta}}_{[Ns]})/\gamma_{[Ns]})\}^2 .$$

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

<u>Lemma 3.1</u>. *Under the assumptions of Theorem 1,* C4 *holds if we also assume that for all* $L > 0$, $\sigma > 0$ *and* $|\underline{\theta}| = 1$,

$$(3.3) \qquad Eg_1(\underline{c})(\underline{c}\,\underline{\theta})\psi(\sigma g_2(\underline{c})(z - L\underline{c}\,\underline{\theta})) < 0 .$$

(We will analyze the power of these tests for contiguous alternatives so the term $\underline{\alpha}'\ \underline{\beta}_0$ below will depend upon N, though this will not be explicit in the notation.)

<u>Theorem 3</u>. *Suppose that* $N^{\frac{1}{2}}\ \underline{a}'\ \underline{\beta}_0 \to \eta$ *(finite), that* C4 *and* C5 *hold, and that either* (2.2) *or* (2.3) *holds. Suppose further that* ψ *satisfies* A1-A7 *and*

(3.4) $$\lim_{\varepsilon\to 0} E|\underline{c}|g_2(\underline{c}) \sup_{|r|\le\varepsilon,\ |t|\le\varepsilon} |\psi'(g_2(\underline{c})(1+r)(z-\underline{c}\,\underline{t})) - \psi'(g_2(\underline{c})z| = 0 .$$

Then the power of the RST converges to

(3.5) $$P\{|W(s) + s\eta_*| > C(s) \text{ for some } 0 \le s \le 1 \text{ or } |W(1) + \eta_*| > D\} ,$$

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

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

$$\gamma_0^2\ \underline{a}'\ \Sigma_1^{-1}\ \Sigma_2\ \Sigma_1^{-1}\ \underline{a} .$$

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

$$C_N(s) = (2[Ns]^2\ N^{-1}(\exp(a/[Ns])-1))^{\frac{1}{2}}$$

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: $K'\ \underline{\beta} = 0$ by rejecting for large values of

$$(K'\ \underline{\hat{\beta}}_{[Ns]})'\ [K'(\Sigma_{1N}(s)^{-1}\ \Sigma_{2N}(s)\ \Sigma_{1N}(s)^{-1})K]^{-1}\ K'\ \underline{\hat{\beta}}_{[Ns]} .$$

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 $\gamma_n = \sigma_n^{-1}$.

Lemma 4.1. *Define* $\xi_i(r) = \underline{c}_i g_1(\underline{c}_i)\psi(g_2(\underline{c}_i)z_i(1+r))$. *Then the process*

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

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*

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

$$V_N(r,s,\underline{t}) = V_N^*(r,s,\underline{t}) - V_N^*(0,s,0) - EV_N^*(r,s,\underline{t}) .$$

Then for all $\varepsilon>0$, $M>0$ *there exists* $B = B(\varepsilon,M)$ *such that*

$$P\{|V_N(r,s,\underline{t})| > \varepsilon \textit{ for some } |r|\le M\,,\ |t|\le M\,,\ BN^{-\frac{1}{2}}\le s\le 1\} < \varepsilon\ .$$

Proof. Fix $\varepsilon>0$, $M>0$. For any $B>0$, $\delta>0$, A2-A3 and the fact that ψ is Lipschitz show

$$\sup\{|V_N(r,s,\underline{t}) - V_N(r,s,\underline{0})|:\ BN^{-\frac{1}{2}}\le s\le\delta\,,\ |r|\le M\,,\ |t|\le M\} = O_p(\delta)\ .$$

From Lemma 4.1 we can choose B large and δ small so that

$$P\{\{\sup|V_N(r,s,\underline{0})|:\ BN^{-\frac{1}{2}}\le s\le\delta\,,\ |r|\le M\} > \varepsilon\} < \varepsilon/2\ .$$

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

$$V_N^*(r,s,\underline{t}) = N^{-\frac{1}{2}}\sum_{i=1}^{[Ns]} \underline{c}_i g_1(\underline{c}_i)\psi(g_2(\underline{c}_i)(z_i-\underline{c}_i\underline{t}N^{-\frac{1}{2}})(1+rN^{-\frac{1}{2}}))\ .$$

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

$$\sup\{|V_N^*(r,s,\underline{t}) - V_N^*(r,s,\underline{t}_*)|:\ 0\le s\le 1\,,\ |r|\le M\,,\ |t-\underline{t}_*|\le\eta\} = O_p(\eta)\ .$$

This means we need only prove tightness of $V_N^*(\cdot,\cdot,\underline{t}_*)$, which follows from A4 as in the proof of Lemma 4.1. □

Lemma 4.3. *For every* $M>0$,

$$\sup\{n^2|\underline{\hat\beta}_n|N^{-3/2}:\ K_N\le n\le MN^{1/2}\}\xrightarrow{p} 0\ .$$

Proof. Fix $\varepsilon, M > 0$. From B3,

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

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

$$\sup\{P(0<|\underline{c}\,\underline{\theta}|<\eta(D)):\ |\underline{\theta}| = 1\} < D/2 . \tag{4.1}$$

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

$$R_n(\underline{\theta},L) = n^{-1} \sum_{i=1}^{n} g_1(\underline{c}_i)(\underline{c}_i\,\underline{\theta})\psi(g_2(\underline{c}_i)(z_i - L\underline{c}_i\,\underline{\theta})\sigma_n) .$$

As in the proof of Theorem 2.2 of Yohai and Maronna (1979), $n^2|\hat{\underline{\beta}}_n|N^{-3/2} > \varepsilon$ for some $K_N \le n \le MN^{\frac{1}{2}}$ implies $|\hat{\underline{\beta}}_n| > \varepsilon N^{\frac{1}{2}}/M^2$ so that $R_n(\underline{\theta}, \varepsilon N^{\frac{1}{2}}/M^2) > 0$ for some $|\underline{\theta}| = 1$. Since ψ is monotone and skew symmetric, with probability at least $1-2\varepsilon$ by B5,

$$\begin{aligned}
R_n(\underline{\theta}, \varepsilon N^{\frac{1}{2}}/M^2) &= n^{-1} \sum_{i=1}^{n} g_1(\underline{c}_i)|\underline{c}_i\,\underline{\theta}|\psi(\sigma_n g_2(\underline{c}_i)(\mathrm{sign}(\underline{c}_i\,\underline{\theta})z_i - \varepsilon N^{\frac{1}{2}}|\underline{c}_i\,\underline{\theta}|M^{-2})) \\
&\le n^{-1} \sum_{i=1}^{n} g_1(\underline{c}_i)|\underline{c}_i\,\underline{\theta}|\psi(\sigma_n g_2(\underline{c}_i)(|z_i| - \varepsilon N^{\frac{1}{2}}|\underline{c}_i\,\underline{\theta}|M^{-2})) \\
&\le M_0 \sup|\psi| n^{-1} \sum_{i=1}^{n} I\{|z_i| > Q\} \\
&\quad + n^{-1} \sum_{i=1}^{n} g_1(\underline{c}_i)|\underline{c}_i\,\underline{\theta}|\psi(Q^2 g_2(\underline{c}_i)(1-\varepsilon N^{\frac{1}{2}}|\underline{c}_i\,\underline{\theta}|/M^2Q^3)) \\
&\quad \cdot I\{|z_i| \le Q\} \\
&= R_{n1} + R_{n2} .
\end{aligned}$$

By the strong law of large numbers (SLLN),

$$\limsup_{n\to\infty} R_{n1} \le \varepsilon M_0 \sup|\psi| \quad \text{(a.s.)} \ .$$

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

$$\text{(4.2)} \qquad \limsup\{R_{n2}: |\underline{\theta}| = 1,\ K_N \le n \le N\}$$

$$\le \sup_{|\theta|=1} Eg_1(\underline{c})|\underline{c}\,\underline{\theta}|\psi(g_2(\underline{c})Q^2(1-L|\underline{c}\,\underline{\theta}|))P(|z_i| \le Q) \quad \text{(a.s.)} \ .$$

Since ψ is monotone nondecreasing and nonconstant, $\liminf_{x\to-\infty} \psi(x) < 0$. Therefore from (4.1), $g_1(\underline{c}) > 0$, (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_∞ is independent of ε. One completes the proof by choosing ε_* sufficiently small. □

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

$$P\{n|\hat{\underline{\beta}}_n|N^{-\frac{1}{2}} > L \text{ for some } MN^{\frac{1}{2}} \le n \le N\} < 4\varepsilon \ .$$

Proof. Choose Q as in B5 and define

$$R_N(\underline{\theta},L,n) = N^{-\frac{1}{2}} \sum_{i=1}^{n} g_1(\underline{c}_i)(\underline{c}_i\,\underline{\theta})\psi(g_2(\underline{c}_i)(z_i - L\underline{c}_i\,\underline{\theta})\sigma_n) \ .$$

Let $L = M/Q$. As in Theorem 2.2 of Yohai and Maronna (1979), $n|\hat{\underline{\beta}}_n|N^{-\frac{1}{2}} > L$ implies $R_N(\underline{\theta}, LN^{\frac{1}{2}}/n, n) > 0$ for some $|\underline{\theta}| = 1$. When $Q^{-1} \le \sigma_n \le Q$ (with probability at least $1-\varepsilon$ from B5) uniformly in $|\underline{\theta}| = 1$, then by B1,

$$R_N(\underline{\theta}\,, LN^{\frac{1}{2}}/n\,, n) \le A_1(n,N,\underline{\theta}) - A_2(n,\underline{\theta})$$

$$= N^{-\frac{1}{2}} \sum_{i=1}^{n} g_1(\underline{c}_i)(\underline{c}_i\,\underline{\theta})\psi(g_2(\underline{c}_i)z_i\sigma_n)$$

$$- dMQ^{-2}n^{-1} \sum_{i=1}^{n} g_1(\underline{c}_i)g_2(\underline{c}_i)(\underline{c}_i\,\underline{\theta})^2$$

$$\cdot\, I\{|z_i| \le Q^{-1}\}I\{g_2(\underline{c}_i)\le a\,,\ g_2(\underline{c}_i)|\underline{c}_i|\le b\}\ .$$

By the SLLN and B3,

$$\limsup_{n\to\infty} A_2(n,\underline{\theta}) \ge dM\lambda_{min}\, P\{|z| \le Q^{-1}\}\ (a.s.)\ .$$

Therefore, it suffices to show that

$$\sup_{\substack{|\underline{\theta}|=1 \\ MN^{\frac{1}{2}}\le n\le N}} |A_1(n,N,\theta)| = O_p(1)$$

for some M.

We have for $|\underline{\theta}| = 1$ and $MN^{\frac{1}{2}} \le n \le N$

$$|A_1(n,N,\underline{\theta})| \le \sup_{MN^{\frac{1}{2}}\le n\le N} |N^{-\frac{1}{2}} \sum_{i=1}^{n} g_1(c_i)c_i\ \psi(g_2(c_i)z_i\sigma_n)|\ .$$

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

$$P\{\sup\{[Ns]\,|\gamma_{[Ns]} - \gamma_0|N^{-\frac{1}{2}} : 0 \le s \le 1\} \ge M_*\} < \varepsilon\ .$$

Hence

$$P\{\sup\{n|\gamma_n - \gamma_0|N^{-\frac{1}{2}} : MN^{\frac{1}{2}} \le n \le N\} \ge M_*\} < \varepsilon$$

or

$$P\{\sup\{|\gamma_n - \gamma_0| : MN^{\frac{1}{2}} \le n \le N\} \le M_* N^{\frac{1}{2}} n^{-1}\} \ge 1 - \varepsilon .$$

Therefore for M sufficiently large

$$P\{\sup\{|\gamma_n - \gamma_0| : MN^{\frac{1}{2}} \le n \le N\} \le \gamma_0/2\} \ge 1 - \varepsilon ;$$

hence for some M_{**} not depending on M

$$P\{|\sigma_n - \sigma_0| \le M_{**} \text{ for all } MN^{\frac{1}{2}} \le n \le N\} \ge 1 - \varepsilon .$$

Thus, since $\sigma_0 = 1$,

$$\sup_{\substack{|\underline{\theta}|=1 \\ MN^{\frac{1}{2}} \le n \le N}} |A_1(n,N,\underline{\theta})| \le \sup_{\substack{0 \le s \le 1 \\ 0 \le |r| \le M_{**}}} \left|N^{-\frac{1}{2}} \sum_{i=1}^{[Ns]} g_1(c_i)c_i \, \psi(g_2(c_i)z_i(1+r))\right| = O_p(1)$$

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 $K_N/N \le s \le \delta$, (2.2) is $O_p(\delta)$. Since δ is arbitrary, it suffices to prove (2.2) when the sup is taken over $a \le s \le 1$ for any $a > 0$. In Lemma 4.2 set $\underline{t} = \sigma_0 [Ns](\hat{\underline{\beta}}_{[Ns]} - \underline{\beta}_0)N^{-\frac{1}{2}}$, $r = [Ns](\hat{\sigma}_{[Ns]}/\sigma_0 - 1)N^{-\frac{1}{2}}$, both of which are $O_p(1)$ (by Lemma 4.4 and B4). Then use A5; since $a \le s \le 1$, 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* $|\sigma-\sigma_0| \le \varepsilon$, $|\underline{\beta}-\underline{\beta}_0| \le \varepsilon$

$$(4.3) \qquad \Big| n^{-1} \sum_{i=1}^{n} \underline{c}_i' \, g_1(\underline{c}_i)\{\psi(g_2(\underline{c}_i)(y_i - \underline{c}_i\underline{\beta})\sigma) - \psi(g_2(\underline{c}_i)z_i) + \sigma\underline{c}_i g_2(\underline{c}_i)\psi'(g_2(\underline{c}_i)z_i)(\underline{\beta}-\underline{\beta}_0) - g_2(\underline{c}_i)z_i \, \psi'(g_2(\underline{c}_i)z_i)(\sigma-\sigma_0)\}\Big| \le h(\varepsilon)(|\sigma-\sigma_0| + |\underline{\beta}-\underline{\beta}_0|) .$$

Proof. Set $\sigma_0=1$, $\underline{\beta}_0=0$ without loss. Define

$$M_2 = 2 + \max\{|K| , K\varepsilon\mathcal{B}\}$$

$$d(\varepsilon) = (2M_2 \, \varepsilon^{\frac{1}{2}})^{-1}$$

$$A(\varepsilon) = \{|\underline{c}|\, g_2(\underline{c}) \ge d(\varepsilon)\}$$

$$L(\varepsilon,\sigma,\underline{\beta}) = \{|g_2(\underline{c})\sigma(z-\underline{c}\,\underline{\beta}) - K| > 2\varepsilon d(\varepsilon) \text{ for all } K\varepsilon\mathcal{B}\} .$$

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

$$2\alpha/(1+\varepsilon) - \varepsilon d(\varepsilon)/(1-\varepsilon) \ge \alpha$$

and

$$2\alpha\varepsilon + \varepsilon d(\varepsilon)(1+\varepsilon) \le 2\varepsilon d(\varepsilon) .$$

Now suppose that $|\sigma - 1| \le \varepsilon$, $|\underline{\beta}| \le \varepsilon$, $(z,\underline{c}) \in L(\varepsilon,\sigma,\underline{\beta})$ and $\underline{c} \notin A(\varepsilon)$. Then if $g_2(\underline{c})z \ge 2\alpha$, it follows that

$$g_2(\underline{c})\sigma z \geq 2\alpha/(1+\varepsilon) \geq \alpha$$

and

$$g^2(\underline{c})\sigma(z - \underline{c}\,\underline{\beta}) \geq 2\alpha/(1+\varepsilon) - |g_2(\underline{c})\underline{c}\,\underline{\beta}|/(1-\varepsilon) \geq \alpha$$

so that $\psi(g_2(\underline{c})\sigma(z - \underline{c}\,\underline{\beta})) = \psi(g_2(\underline{c})z)$; similarly $g_2(\underline{c})z \leq -2\alpha$ implies that $\psi(g_2(\underline{c})\sigma(z - \underline{c}\,\underline{\beta})) = \psi(g_2(\underline{c})z)$. If $|g_2(\underline{c})z| < 2\alpha$, then

$$|g_2(\underline{c})(z - \sigma(z - \underline{c}\,\underline{\beta}))| \leq |g_2(\underline{c})z(\sigma-1)| + |g_2(\underline{c})\sigma\underline{c}\,\underline{\beta}|$$
$$\leq 2\alpha\varepsilon + \varepsilon d(\varepsilon)(1+\varepsilon) \leq 2\varepsilon d(\varepsilon) ,$$

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

If $|\sigma - 1| > \varepsilon$, $|\underline{\beta}| > \varepsilon$, $(z,\underline{c}) \notin L(\varepsilon,\sigma,\underline{\beta})$, or $\underline{c} \in A(\varepsilon)$, then we can use the fact that ψ 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 ψ and

$$A_{n1} = n^{-1} \sum_{i=1}^{n} \{g_2(\underline{c}_i)(|\sigma - 1| + |\underline{c}_i||\underline{\beta}|)\}^2 \, I\{|\underline{c}_i|g_2(\underline{c}_i) \leq d(\varepsilon)\}$$

$$A_{n2} = n^{-1} \sum_{i=1}^{n} g_2(\underline{c}_i)(|\sigma - 1| + |\underline{c}_i||\underline{\beta}|) I\{(z_i, \underline{c}_i) \notin L(\varepsilon,\sigma,\underline{\beta})\} \cdot \{|\underline{c}_i|g_2(\underline{c}_i) \leq d(\varepsilon)\}$$

$$A_{n3} = n^{-1} \sum_{i=1}^{n} g_2(\underline{c}_i)(|\sigma - 1| + |\underline{c}_i||\underline{\beta}|) I\{|\underline{c}_i|g_2(\underline{c}_i) \geq d(\varepsilon)\}$$

$$A_{n4} = n^{-1} \sum_{i=1}^{n} |\psi(g_2(\underline{c}_i)\sigma(z_i - \underline{c}_i\,\underline{\beta})) - \psi(g_2(\underline{c}_i)z_i)| I\{|\underline{c}_i|g_2(\underline{c}_i) \geq d(\varepsilon)\} .$$

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 ψ is Lipschitz and A3 holds, we need only prove the lemma when $\underline{\beta} = 0$. Since ψ is constant off a compact set, there is a constant M_1 with

$$A_{n4}(\underline{\beta}=0) \le |\sigma - 1| n^{-1} \sum_{i=1}^{n} g_2(\underline{c}_i)|z_i| I\{|\underline{c}_i| g_2(\underline{c}_i) \ge d(\varepsilon) , g_2(\underline{c}_i)|z_i| \le M_1\} ,$$

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

$$A_{n2} \le (|\sigma - 1| + |\underline{\beta}|) n^{-1} \sum_{i=1}^{n} g_2(\underline{c}_i)|\underline{c}_i| I\{|g_2(\underline{c}_i)z_i - K| \le M_2\, \varepsilon d(\varepsilon)$$

$$\text{for some } K \varepsilon \mathcal{B}\} ,$$

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

<u>Proof of Theorem 2</u>. Fix $\varepsilon_1, \varepsilon_2 > 0$. Using the fact that Σ_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 $N \ge M_1$ implies

$$P\{|\hat{\underline{\beta}}_n - \beta_0| + |\sigma_n - \sigma_0| > \varepsilon_1 \text{ for some } K_N \le n \le N\} \le \varepsilon_1 \tag{4.4}$$

$$P\{N^{\frac{1}{2}}\{\frac{n}{N}|\sigma_n - \sigma_0| + |\Sigma_1^{-1} N^{-1} \sum_{i=1}^{n} \underline{c}_i' \, g_1(\underline{c}_i)\psi(g_2(\underline{c}_i)z_i)|\} > M_2 \text{ for some } K_N \le n \le N\} \le \varepsilon_2 \tag{4.5}$$

$$(4.6)\quad P\{|\Sigma_1^{-1} n^{-1} \sum_{i=1}^{n} \underline{c}'_i g_1(\underline{c}_i)\{\psi(g_2(\underline{c}_i)\sigma(y_i - \underline{c}_i \underline{\beta})) - \psi(g_2(\underline{c}_i)z_i)$$

$$+ \sigma\underline{c}_i g_2(\underline{c}_i)\psi'(g_2(\underline{c}_i)z_i)(\underline{\beta} - \underline{\beta}_0)$$

$$- g_2(\underline{c}_i)z_i\ \psi'(g_2(\underline{c}_i)z_i)(\sigma-\sigma_0)\}| > 4h(\varepsilon_1)(|\sigma-\sigma_0| + |\underline{\beta} - \underline{\beta}_0|)$$

$$\text{for some } K_N \le n \le N \text{ and some } |\sigma-\sigma_0| \le \varepsilon_1,\ |\underline{\beta} - \underline{\beta}_0| \le \varepsilon_1\} \le \varepsilon_1 .$$

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

$$(4.7)\qquad n|\hat{\underline{\beta}}_n - \underline{\beta}_0|N^{-\frac{1}{2}} \le 4h(\varepsilon_1)n|\hat{\underline{\beta}}_n - \underline{\beta}_0|N^{-\frac{1}{2}} + 2M_2 .$$

For (4.7) to occur we must have (for ε_1 small) $n|\hat{\underline{\beta}}_n - \underline{\beta}_0|N^{-\frac{1}{2}} \le 4M_2$, proving that

$$(4.8)\qquad \sup\{n|\hat{\underline{\beta}}_n - \underline{\beta}_0|N^{-\frac{1}{2}}:\ K_N \le n \le N\} = O_p(1) .$$

Placing $n(\hat{\underline{\beta}} - \underline{\beta}_0)N^{-\frac{1}{2}}$ into (4.6) with $\sigma=\sigma_n$, $\underline{\beta}=\beta_n$ completes the proof. □

<u>Proof of Lemma 3.1</u>. This follows by applying the technique of Lemma 4.3, B5, Lemma 1 of Yohai (1974), and (3.3). □

<u>Remark</u>. A1, A3, and B1-B2 imply (3.4) by dominated convergence.

<u>Proof of Theorem 3</u>. We will first show that for $j = 1,2$

$$(4.9)\qquad \sup\{|\Sigma_{jN}(s) - \Sigma_j|:\ K_N/N \le s \le 1\} \overset{p}{\to} 0 .$$

First consider $j = 1$. From A2, C4 and C5, with probability at least $1 - \varepsilon/2$,

$$(4.10) \qquad \left|\Sigma_{1N}(s) - [Ns]^{-1} \sum_{i=1}^{[Ns]} \underline{c}'_i \, \underline{c}_i g_1(\underline{c}_i) g_2(\underline{c}_i) \psi'(g_2(\underline{c}_i) z_i)\right|$$

$$\le [Ns]^{-1} \sum_{i=1}^{[Ns]} |\underline{c}_i| g_2(\underline{c}_i)$$

$$\sup_{|r| \le \varepsilon, \, |\underline{t}| \le \varepsilon} |\psi'(g_2(\underline{c}_i)(1+r)(z_i - \underline{c}_i \underline{t}))$$

$$- \psi'(g_2(\underline{c}_i) z_i)| \, .$$

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 Σ_1 and Σ_2 are of full rank, the power becomes

$$P\{[Ns]^2 N^{-3/2} |\underline{a}' \, \hat{\underline{\beta}}_{[Ns]}| > [Ns]N^{-1} A_N(s) C_N(s)$$

$$\text{for some } K_N/N \le s \le 1 \text{ or } N^{\frac{1}{2}} |\underline{a}' \, \hat{\underline{\beta}}_N| > A_N(1)D\} \, ,$$

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 $E|\underline{c}|^2 < \infty$. 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 $\{\underline{c}_i\}$ 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_{n3}, A_{n4} 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 $E\underline{c}\, g_1(\underline{c})g_2(\underline{c})z\psi'(g_2(\underline{c})z) = 0$. For Section 3, the symmetry is unnecessary if $g_j(\underline{c}) = g_j(|\underline{c}|)$ $(j = 1,2)$ and the patients are randomly (with probability ½) assigned to each group, for then

$$E\underline{a}'\ \underline{c}\, g_1(\underline{c})g_2(\underline{c})z\psi'(g_2(\underline{c})z) = 0 \ .$$

(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

$$\underline{\beta}_N \approx \gamma_0\, N^{-1} \sum_{i=1}^{N} \Sigma_1^{-1}\ \underline{c}_i'\ g_1(\underline{c}_i)\psi(g_2(\underline{c}_i)z_i) \ .$$

Since the $\{\underline{c}_i\}$ are i.i.d., when γ_0 is estimated in an appropriately smooth fashion simultaneously with $\underline{\beta}$ (as in Huber's Proposal 2), the BIR estimates are minimum contrast estimates of a vector parameter. If ψ has a derivative then Berry-Esseen results and Edgeworth expansions are available (e.g., Pfanzagl (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

Anderson, T.W. (1960). A modification of the sequential probability ratio test to reduce the sample size. *Annal of Math. Statist. 31, 1*, 165-197.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., and Tukey, J.W. (1972). *Robust Estimates of Location: Survey and Advances*. Cambridge: Princeton University Press.

Bhattacharya, R.N. and Ghosh, J.K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist. 6*, 434-451.

Bickel, P.J. (1974). Edgeworth expansions in nonparametric statistics. *Ann. Statist. 2*, 1-20.

Bickel, P.J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc. 70*, 428-434.

Bickel, P.J. and Wichura, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist. 12*, 1656-1670.

Billingsley, P. (1968). *Convergence of Probability Measures*. New York: John Wiley and Sons, Inc.

Carroll, R.J. (1978). On almost sure expansions for M-estimates. *Ann. Statist. 6*, 314-318.

Huber, P.J. (1977). *Robust Statistical Procedures*. SIAM, Philadelphia.

Maronna, R.A. and Yohai, V.J. (1978). Robust M-estimators for regression with contaminated independent variables. Unpublished manuscript.

Pfanzagl, J. (1973). The accuracy of the normal approximation for estimates of vector parameters. *Z. Wahr. und Verw. Gebiete* 25, 171-198.

Schrader, R.M. and Hettmansperger, T.P. (1980). Robust analysis of variance based upon a likelihood ratio criterion. *Biometrika 67*, 93 and 102.

Siegmund, D. (1977). Repeated significance tests for a normal mean. *Biometrika 64*, 177-189.

Siegmund, D. (1979). Sequential χ^2 and F tests and the related confidence intervals. *Stanford University Technical Report No. 134*.

Yohai, V.J. (1974). Robust estimation in the linear model. *Ann. Statist. 2*, 562-567.

Yohai, V.J. and Maronna, R.A. (1979). Asymptotic behavior of M-estimators for the linear model. *Ann. Statist. 7*, 258-268.

REPORT DOCUMENTATION PAGE

UNCLASSIFIED

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Weak Convergence of Bounded Influence Regression Estimates with Applications

5. TYPE OF REPORT & PERIOD COVERED
TECHNICAL

6. PERFORMING ORG. REPORT NUMBER
Mimeo Series No. 1285

7. AUTHOR(s)
Raymond J. Carroll and David Ruppert

8. CONTRACT OR GRANT NUMBER(s)
AFOSR Grant AFOSR-80-0080
NSF Grant MCS78-01240

9. PERFORMING ORGANIZATION NAME AND ADDRESS

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC 20332

12. REPORT DATE
April 1980

13. NUMBER OF PAGES
23

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release -- Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

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

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

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.