Almost Sure Properties of Robust Regression Estimates

by

Raymond J. Carroll*

and

David Ruppert**

We consider Huber's Proposal 2 for robust regression estimates in the general linear model. The estimates are first shown to be strongly consistent. We then develop an almost sure expansion of these estimates, approximating them (to order $o(n^{-\frac{1}{2}})$) by a weighted sum of bounded random variables. The approximation is sufficiently strong to permit construction of sequential fixed-width confidence regions for the regression parameter.

AMS 1970 Subject Classifications: Primary 62G35; Secondary 62E20, 62J05

Key Words and Phrases: Regression, robustness, linear model, almost sure properties, consistency, M-estimates, Fixed-width confidence intervals.

*This work was supported by the Air Force Office of Scientific Research under Contract AFOSR-75-2796.

**This work was supported by the National Science Foundation Grant NSF MCS78-01240.

1. Introduction

Consider the general linear model

$$Y_i = \underline{c}_i \, \underline{\beta}_0 + \varepsilon_i , \tag{1.1}$$

where $\{\underline{c}_i\}$ is a sequence of $(1\times p)$ vectors of fixed constants, $\underline{\beta}_0 (p\times 1)$ is the regression parameter, and $\varepsilon_1, \varepsilon_2, \ldots$ are independent and identically distributed (i.i.d.) errors. Least squares estimation of $\underline{\beta}_0$ is known to be sensitive to outliers (Andrews (1974), Carroll (1979) give empirical demonstrations of this fact) and inefficient if the error distribution is heavier-tailed than the normal distribution (Huber (1973), (1977)). For this reason Huber (1973) (1977) has proposed a class of competitors to least squares called M-estimates. For given functions ψ, χ, Huber's Proposal 2 involves solving the simultaneous equations

$$n^{-1} \sum_{i=1}^{n} \psi((Y_i - \underline{c}_i \, \underline{\beta})/\sigma)\underline{c}_i = 0 \tag{1.2}$$

$$n^{-1} \sum_{i=1}^{n} \chi(Y_i - \underline{c}_i \, \underline{\beta})/\sigma) = \xi = E_{\Phi} \, \chi(z) , \tag{1.3}$$

where the expectation is taken under the standard normal distribution function. The original Proposal 2 (Huber (1973)) choses

$$\chi(u) = \psi^2(u) , \tag{1.4}$$

a choice that, for convenience, we make throughout.

There are two classes of functions ψ commonly used (Andrews, et al (1972)). The first are bounded and monotone non-decreasing, the prototype of which is Huber's

$$\psi(u) = \max(-k, \min(u,k)) . \tag{1.5}$$

Often, better robustness properties are obtained by using the solution to (1.2)-(1.3) with ψ given by (1.5) as an initial estimate, and then performing one step of Newton's algorithm using a function ψ which redescends to zero, such as Hampel's

$$(1.6) \qquad \psi(u) = -\psi(-u) = \begin{cases} u & 0 < u \leq a \\ a & a < u \leq b \\ a\left(\frac{u-b}{c-b}\right) & b < u \leq c \\ 0 & u > c\ . \end{cases}$$

The asymptotic properties of such one-step estimates can be established (Bickel (1975), (Carroll (1977)) *if one knows the asymptotic properties of the estimates based on (1.5)*.

Recent Monte-Carlo (Huber (1973), Gross (1977)) and empirical (Andrews (1974), Carroll (1979)) studies have established the superiority of M-estimates to least squares estimates. However, only limited theoretical work is available. Huber's (1973) methods may be used to show that there is a sequence of solutions to (1.2)-(1.3) which is asymptotically normally distributed when ψ has two bounded continuous derivatives (not satisfied for (1.5)), but while his results are truly remarkable in letting $p \to \infty$, he does not show that *all* solutions must be asymptotically normal.

Maronna and Yohai (1978) consider monotone functions ψ and treat the design $\{c_i\}$ as an i.i.d. sequence of random variables, which enables them to use Glivenko-Cantelli results. They show that all solutions are strongly consistent and asymptotically normal.

We consider the almost sure properties of M-estimates in the usual case that the design $\{\underline{c}_i\}$ is a sequence of constants. When ψ is monotone, we show in Section 2 that all solutions to (1.2)-(1.3) are strongly consistent.

In Section 3 we present a result which is much stronger than asymptotic normality and yields insight into the asymptotic behavior of M-estimates. The specific result generalize work of Carroll (1978a) to show that, if a sequence of solutions is strongly consistent, even when ψ is not monotone (as in (1.6)) the solutions can be approximated by a weighted sum of bounded i.i.d. random variables, with remainder term of order $o(n^{-\frac{1}{2}})$ almost surely. This approximation can be motivated as follows. If F is symmetric, if $n^{\frac{1}{2}-\delta}\,||\hat{\underline{\beta}}_n-\underline{\beta}_0|| \to 0$ (a.s.) and $n^{\frac{1}{2}-\delta}\,|\hat{\sigma}_n-\sigma_0| \to 0$ (a.s) for all $\delta > 0$, and if ψ has two continuous bounded derivatives (not true for (1.5)-(1.6)), then Taylor expansions can be used to show that if $\sum_n = n^{-1}\sum_{i=1}^{n} \underline{c}_i'\,\underline{c}_i$, then

$$(1.7)\quad (E\,\psi'(\varepsilon_1/\sigma_0))(\hat{\underline{\beta}}_n - \underline{\beta}_0)/\sigma_0 = \textstyle\sum_n^{-1}\, n^{-1}\sum_{i=1}^{n} \underline{c}_i\,\psi(\varepsilon_i/\sigma_0) + o(n^{-\frac{1}{2}}) \text{ (a.s.)} .$$

It is the purpose of Section 3 to verify (1.7) under reasonable conditions. One can grasp the importance of (1.7) by noting that it implies that, except for a negligible remainder term, the normalized M-estimate is a normalized least squares estimate based on observations $\underline{c}_i\,\underline{\beta}_0 + \psi(\varepsilon_i/\sigma_0)$; the boundedness of the "errors" $\psi(\varepsilon_i/\sigma_0)$ is the essential reason for the robustness of M-estimates to outliers in the responses or heavier-tailed distributions. Note also that (1.7) shows quite clearly that the present version of M-estimates is not robust against outliers in the design; Maronna and Yohai (1978) have suggested weighting (1.2), replacing $\underline{c}_i$ by $\underline{c}_i\,w(\underline{c}_i)$, but here much work remains to be done.

Finally, the expansion (1.7) and its relationship with least squares yields as simple consequences two classes of results:

(i) The one-step estimates mentioned above are strongly consistent, asymptotically normal, and representable as least sqaures estimates with bounded "errors" to order $o(n^{-\frac{1}{2}})$.

(ii) Robust sequential fixed-width confidence bounds for the regression parameter $\underline{\beta}_0$ can be constructed and analyzed exactly as in Gleser (1965); only a change of notation is necessary.

We assume in Section 2 that ψ is monotone and bounded. We assume throughout that there exists η_0, σ_0 with

$$E\ \psi((\varepsilon_1 - \eta_0)/\sigma_0) = 0$$

$$E\ \psi^2((\varepsilon_1 - \eta_0)/\sigma_0) = \xi\ .$$

If ψ is given by (1.5), then this assumption is met if for all η, $P(\varepsilon_1 = \eta) < 1 - \xi/k^2$; this can be seen by a simple modification of an argument by Huber (1964, p. 97). By including an intercept parameter in the problem, we can reparameterize so that $\underline{\beta}_0 = 0$, $\sigma_0 = 1$, $E\ \psi(Y_1) = 0$ and $E\ \psi^2(Y_1) = \xi$, which we do throughout the paper. Thus, in proving consistency for example, we will attempt to show that all solutions $(\hat{\underline{\beta}}_n, \hat{\sigma}_n)$ converge almost surely to $(\underline{0},1)$.

2. Strong Consistency

Throughout this section we will make the following assumptions.

(2.1) The function ψ is continuous, odd, nondecreasing, and Lipschitz of order one. (For notational convenience, we take the Lipschitz constant equal to one).

(2.2) For some a, K_1, $K_2 > 0$,

$$\psi(x) = K_1 \qquad \text{for all } |x| \geq K_2$$

$$|\psi(x)| \geq a|x| \qquad \text{if} \qquad |x| \leq K_2\ .$$

(2.3) If λ_n is the minimum eigenvalue of

$$n^{-1} \sum_{i=1}^{n} \underline{c}_i' \underline{c}_i\ (1 + ||\underline{c}_i||)^{-1}\ ,$$

then $\liminf \lambda_n = \lambda_\infty > 0$.

(2.4) If λ_n^* is the maximum eigenvalue of

$$\Sigma_n = n^{-1} \sum_{i=1}^{n} \underline{c}_i' \underline{c}_i ,$$

then $\limsup \lambda_n^* = \lambda_\infty^* < \infty$.

(2.5) $Y_1, Y_2, \ldots$ are i.i.d. with $E|Y_1| < \infty$.

(2.6) Let $G_n(\underline{\beta},\sigma) = (G_{n1}(\underline{\beta},\sigma),\ G_{n2}(\underline{\beta},\sigma))$, where

$$G_{n1}(\underline{\beta},\sigma) = n^{-1} \sum_{i=1}^{n} E\ \psi((Y_1 - \underline{c}_i\ \underline{\beta})/\sigma)(\underline{c}_i\ \underline{\beta}/\sigma)$$

$$G_{n2}(\underline{\beta},\sigma) = n^{-1} \sum_{i=1}^{n} E\ \psi^2((Y_1 - \underline{c}_i\ \underline{\beta})/\sigma) - \xi\ .$$

Then for any compact subset C of $\{(\underline{\beta},\sigma):\ \sigma > 0,\ \sigma \neq 1,\ \underline{\beta} \neq \underline{0}\}$,

$$\liminf_{n\to\infty}\ \inf_{C}\ ||G_n(\underline{\beta},\sigma)|| > 0\ .$$

The assumptions on the design ((2.3)-(2.4)) are considerably stronger than what is needed for strong consistency of least squares estimates (see Lai, Robbins and Wei (1978)), but the latter requires $E|Y_1|^2$, which is of course stronger than our (2.5).

Limit theorem for regression often require that $n^{-1}\ \Sigma\ \underline{c}_i'\ \underline{c}_i \to \Sigma$ (positive definite); under this assumption (2.3) is implied by rather weak conditions (cf. Proposition 2 below).

Condition (2.6) is used to insure eventual uniqueness of the solutions and corresponds to Huber's (1967) condition (B-3) (in fact, in the location -scale problem it is exactly his (B-3)). If we know that F is symmetric then it is easy to write various reasonable conditions which imply (2.6). For example, consider

(2.7) F is symmetric and, for each fixed $t > 0$, $F(s,t) = E\ \psi(Y_1-s)/t)(s/t)$ has a unique zero at $s = 0$.

Proposition 1. (2.7) implies (2.6).

Note that

$$\sum(\underline{c}_i \underline{\beta})^2 = \underline{\beta}' \sum_n \underline{\beta} \leq \lambda_n^* ||\underline{\beta}||^2 .$$

Consider the expressions

$$n^{-1} \sum_{i=1}^{n} \psi((Y_i - \underline{c}_i \underline{\beta})/\sigma)\underline{c}_i \underline{\beta}/\sigma = 0 \tag{2.8}$$

$$n^{-1} \sum_{i=1}^{n} \psi^2((Y_i - \underline{c}_i \underline{\beta})/\sigma) = \xi . \tag{2.9}$$

Huber's Proposal 2 M-estimates satisfy (2.8) and (2.9).

Lemma 1. If (2.1)-(2.5) hold then, for some $M > 0$,

$$\Pr\{\text{there exists } N \text{ such that } n \geq N, (2.8) \text{ and } (2.9) \text{ imply } ||\beta|| \leq M, \bar{\sigma} \leq M\} = 1 .$$

Theorem 1. If (2.1)-(2.6) hold then for all $\varepsilon > 0$,

$$\Pr\{\text{there exists } N \text{ such that } n \geq N, (2.8) \text{ and } (2.9) \text{ imply } ||\beta|| \leq \varepsilon, |\sigma-\sigma_0| \leq \varepsilon\} = 1 ,$$

i.e., M-estimates of regression are strongly consistent.

Proof are contained in Appendix A.

3. An Almost Sure Approximation

In this section we do not assume that ψ is monotone, but we do insist that the following assumption holds:

(3.1) The sequence $\{\underline{\beta}_n, \sigma_n\}$ of solutions to (1.2) and (1.3) is strongly consistent, converging to $(\underline{0}, 1)$ (without loss of generality).

Additionally, we assume

(3.2) The function ψ is odd, bounded, continuous, and constant outside a finite interval. Further, ψ is twice boundedly and continuously differentiable except possibly at a finite number of points $a_1,\ldots,a_k$ (note that ψ need not be monotone).

(3.3) $Y_1, Y_2, \ldots$ are i.i.d. (F).

(3.4) F is Lipschitz in neighborhoods of $a_1,\ldots,a_k$.

(3.5) $E\,\psi(Y_1) = 0$, $E\,\psi^2(Y_1) = \xi$, but $E\,\psi'(Y_1) \neq 0$, $E\,Y_1\,\psi(Y_1)\psi'(Y_1) \neq 0$.

(3.6) For all $\delta > 0$, $n^{-(\frac{1}{2}-\delta)} \max_{1\le i\le n} ||\underline{c}_i|| \to 0$.

(3.7) There is a $\delta_* > 0$ for which

$$\limsup_{n\to\infty} n^{-1} \sum_{i=1}^{n} ||\underline{c}_i||^{2(1+\delta_*)} < \infty$$

(3.8) $\sum_n = n^{-1} \sum_{i=1}^{n} \underline{c}_i\,\underline{c}_i' \to \sum$ (positive definite).

(3.9) If $\underline{c}_n^* = n^{-1} \sum_{i=1}^{n} \underline{c}_i$, then the following matrix is non-singular for sufficiently large n:

$$A_n = \begin{bmatrix} E\,\psi'(Y_1) I_p & \underline{c}_n^{*\prime}\, E\,Y_1\,\psi'(Y_1) \\ 2\,E\psi(Y_1)\psi'(Y_1)\underline{c}_n^* & -2\,E\,Y_1\,\psi(Y_1)\psi'(Y_1) \end{bmatrix} .$$

Note that (3.9) holds if F is symmetric (since $E\,Y_1\,\psi'(Y_1) = 0$). If the design is centered and has an intercept term so that $\underline{c}_n^* \to (1\ 0\ \ldots\ 0)$, one shows that A_n is eventually nonsingular if

$$E\ \psi'(Y_1)E\ Y_1\ \psi(Y_1)\psi'(Y_1) + E\ Y_1\ \psi'(Y_1)E\ \psi(Y_1)\psi'(Y_1) \neq 0\ ;$$

this condition is also required by Maronna and Yohai (1978) in proving asymptotic normality for their situation ($\underline{c}_i$ random).

Theorem 2. If (3.1)-(3.9) hold, then

$$(3.10)\qquad (\Sigma_n\ E\psi'(Y_1))\hat{\underline{\beta}}_n = n^{-1}\sum_{i=1}^{n} \underline{c}_i(\psi(Y_i) + (\hat{\sigma}_n^{-1} - 1)E\ Y_1\ \psi'(Y_1)) + H_n\ ,$$

$$(3.11)\qquad -2(E\ Y_1\ \psi(Y_1)\psi'(Y_1))(\hat{\sigma}_n^{-1} - 1) =$$

$$n^{-1}\sum_{i=1}^{n} \{\psi^2(Y_i) - \xi - 2(E\ \psi(Y_1)\psi'(Y_1))\underline{c}_i\ \hat{\beta}_n\} + G_n\ ,$$

where $n^{\frac{1}{2}} H_n \to 0$ (a.s.), $n^{\frac{1}{2}} G_n \to 0$ (a.s.)

Corollary 2. If (3.1)-(3.9) hold and F is symmetric, then

$$(3.12)\qquad (\Sigma_n\ E\ \psi'(Y_1))\hat{\underline{\beta}}_n = n^{-1}\sum_{i=1}^{n} \underline{c}_i\ \psi(Y_i) + H_n$$

$$(3.13)\qquad -2(E\ Y_1\ \psi(Y_1)\psi'(Y_1))(\hat{\sigma}_n^{-1} - 1) = n^{-1}\sum_{i=1}^{n} (\psi^2(Y_i) - \xi) + G_n\ ,$$

where $n^{\frac{1}{2}} H_n \to 0$ (a.s.), $n^{\frac{1}{2}} G_n \to 0$ (a.s.).

All proofs are given in Appendix B.

Remark. In terms of the original model $Y_i = \underline{c}_i\ \underline{\beta}_0 + \varepsilon_i$, $E\ \psi(z_1/\sigma_0) = 0$, $E\ \psi^2(z_1/\sigma_0) = \xi$, Theorem 2 and its Corollary may be written by substituting ε_i/σ_0 for Y_i, $(\hat{\underline{\beta}}_n - \underline{\beta}_0)/\sigma_0$ for $\hat{\underline{\beta}}_n$ and $(\hat{\sigma}_n - \sigma_0)/\sigma_0$ for $\hat{\sigma}_n - 1$.

Remark. If the design includes an intercept and is centered so that $n^{-1}\sum_{i=1}^{n} \underline{c}_i \to (1\ 0\ \ldots\ 0)$, one can show that the common estimate of the variance covariance matrix of $\hat{\underline{\beta}}_n$

$$\frac{\hat{\sigma}_n^2\, n^{-1} \sum_{i=1}^{n} \psi^2((Y_i - \underline{c}_i\, \hat{\beta}_n)/\hat{\sigma}_n)}{\{n^{-1} \sum_{i=1}^{n} \psi'((Y_i - \underline{c}_i\, \hat{\beta}_n)/\hat{\sigma}_n)\}^2}\, \Sigma_n^{-1}$$

is a consistent estimate of variance of all the terms of $\hat{\beta}_n$ if F is symmetric, while it inconsistently estimates only the variance of the intercept term if F is asymmetric and E $Y_1\ \psi'(Y_1) \neq 0$ (see Carroll (1978b) for empirical demonstrations of this result).

4. Relationship between the assumptions

It is clear that (3.7) is guaranteed in the design is bounded, i.e., $\sup||\underline{c}_i|| < \infty$. In addition, the following proposition shows that (2.3) is not a very strong condition.

Proposition 2. (3.7)-(3.8) imply (2.3).

Proof: Choose M > 1 so large that

$$\lambda_\infty/2 - 2K_0/(M-1)^\delta = \varepsilon_M > 0,$$

where $\limsup n^{-1} \sum_{i=1}^{n} ||\underline{c}_i||^{2+\delta} = K_0 < \infty$. By (3.7)-(3.8), for sufficiently large n

$$\lambda_{min}(\Sigma_n) > \lambda_\infty/2$$

$$n^{-1} \sum_{i=1}^{n} ||\underline{c}_i||^{2+\delta} < 2K_0\ ,$$

where $\lambda_{min}(A)$ is the minimum eigenvalue of A. For fixed $\underline{x} \in R^p$,

$$n^{-1} \sum_{i=1}^{n} (\underline{c}_i\ \underline{x})^2 (1 + ||\underline{c}_i||)^{-1}$$

$$= n^{-1} \sum_{i=1}^{n} (\underline{c}_i\ \underline{x})^2 (1/M + (1 + ||\underline{c}_i||)^{-1} - 1/M)$$

$$\geq ||x||^2\, \lambda_\infty/(2M) - n^{-1} \sum_{i=1}^{n} (\underline{c}_i\ \underline{x})^2\, I(1 + ||\underline{c}_i|| > M)/M$$

$$\geq ||x||^2 \lambda_\infty/(2M) - n^{-1} \sum_{i=1}^{n} \{(\underline{c}_i \underline{x})^2/M\}\{||\underline{c}_i||^\delta/(M-1)^\delta\}$$

$$\geq (||x||^2/M)(\lambda_\infty/2 - 2K/(M-1)^\delta) = ||x||^2 \epsilon_M/M , \text{ say with } \epsilon_M > 0.$$

Thus, since $\underline{x}$ was arbitrary

$$\lambda_{min}(n^{-1} \sum_{i=1}^{n} \underline{c}_i' \underline{c}_i (1 + ||\underline{c}_i||)^{-1}) \geq \varepsilon_M/M . \quad \square$$

Appendix A

Proposition A1. Assumptions (2.1) and (2.2) imply that for some $K_3, K_4 > 0$ and for all z,

(A.1) $$\sup_{-\infty<\mu<\infty} |\psi(z-\mu)\mu - \psi(-\mu)\mu| \leq K_3(K_4 + |z|) .$$

Proof: Suppose that $|z| < |\mu| - K_2$ (and hence $|\mu| \geq K_2$). Then (2.2) implies

$$|\psi(z-\mu) - \psi(-\mu)| = 0 .$$

Since $|\psi(x)| \leq K_1$, if $|z| \geq |\mu| - K_2$,

$$|\psi(z-\mu) - \psi(-\mu)| \; |\mu| \leq 2K_1(K_2 + |z|). \quad \square$$

Proposition A2. There exists $A_1 > 0$, $A_2 > 0$ such that for almost all w, there exists N(w) such that

(A.2) $$\left\{ \begin{array}{l} n \geq N(w), \ (2.8) \\ ||\beta|| \geq A_1 \end{array} \right\} \Rightarrow \sigma \geq A_2 \; ||\underline{\beta}||^{1/2} .$$

Proof: There exists $N_1(w)$ such that $n \geq N_1(w)$ implies

(A.3) $$n^{-1} \sum_{i=1}^{n} |Y_i| \leq 2 \, E|Y_1| .$$

From (A.1) and (A.3)

(A.4) $$\sup_{\sigma,\underline{\beta}} n^{-1} \sum_{i=1}^{n} |\psi((Y_i - \underline{c}_i\ \underline{\beta})/\sigma) - \psi(-\underline{c}_i\ \underline{\beta}/\sigma)|\ |\underline{c}_i\ \underline{\beta}/\sigma|$$

$$\leq 2K_3(K_4 + E|Y_1|/\sigma) .$$

From (2.2) and (2.3), there exists N_0 such that $n > N_0$ implies

(A.5) $$n^{-1} \sum_{i=1}^{n} \psi(-\underline{c}_i\ \underline{\beta}/\sigma)(\underline{c}_i\underline{\beta}/\sigma)$$

$$\leq -a\ n^{-1} \sum_{i=1}^{n} (\underline{c}_i\ \underline{\beta}/\sigma)^2\ I\{|\underline{c}_i\ \underline{\beta}/\sigma| \leq K_2\}$$

$$- K_1\ n^{-1} \sum_{i=1}^{n} |\underline{c}_i\ \underline{\beta}/\sigma|\ I\{|\underline{c}_i\ \underline{\beta}/\sigma| \geq K_2\}$$

$$\leq -\min(a,K_1)\ n^{-1} \sum_{i=1}^{n} (\underline{c}_i\ \underline{\beta}/\sigma)^2\ (1 + ||\underline{c}_i||)^{-1}\ (1 + ||\underline{\beta}||/\sigma)^{-1}$$

$$\leq -\ \min(a,K_1)\ \lambda_\infty\ ||(\underline{\beta}/\sigma)||^2\ (1 + ||\beta||/\sigma)^{-1} .$$

Thus from (A.4) and (A.5), there exists K_5,K_6 and K_7 for which $n > \max(N_0,N_1(w))$ implies

(A.6) $$n^{-1} \sum_{i=1}^{n} \psi((Y_i - \underline{c}_i\ \underline{\beta})/\sigma)\underline{c}_i\ \underline{\beta}/\sigma \leq K_5 + K_6/\sigma - K_7\ ||\underline{\beta}/\sigma||^2(1 + ||\underline{\beta}/\sigma||)^{-1}.$$

Now, (2.8), $\sigma \leq A_2\ ||\underline{\beta}||^{1/2}$, and the choice of sufficiently large A_1 would imply

$$K_5 + K_6/\sigma - K_7\ ||\underline{\beta}/\sigma||^2\ (1 + ||\underline{\beta}/\sigma||)^{-1} < 0 ,$$

a contradiction. □

Since ψ is Lipschitz and $\psi(x) = K$, for $|x| \geq K_2$ we can choose K_8 such that $\psi^2(x) \leq K_8\ |x|$. Define

$$C_1 = 2\ E|Y_1|$$

$$C_2 = (\xi/K_8 - K_1/M_0)/(2\ \lambda^*_\infty)^{\frac{1}{2}} ,$$

where M_0 is chosen so that $C_2 > 0$. Further fix δ so that if

$$\delta' = (2\ \lambda^*_\infty)^{\frac{1}{2}}\ \delta/(\lambda_\infty \min(a,K_1)) ,$$

then

$$\delta' < C_2(1 - \delta') .$$

Choose M_1 (by dominated convergence) and $N_3(w)$ so that if $n \geq N_3(w)$,

$$n^{-1} \sum_{i=1}^{n} \{\min(2K_1,\ |Y_i|/M_1)\}^2 \leq 2\ E\{\min(2K_1,\ |Y_1|/M_1)\}^2 \leq \delta^2 .$$

Proposition A3. There exists $N_4(w)$ for which

$$\left\{ \begin{array}{l} n \geq N_4(w) \\ (2.9) \\ \sigma > M_0 \end{array} \right\} \Rightarrow ||\underline{\beta}||\ \sigma > C_2 .$$

Proof: If $n \geq \max(N_1(w),\ N_3(w))$

(A.7) $$n^{-1} \sum_{i=1}^{n} \psi^2((Y_i - \underline{c}_i\ \underline{\beta})/\sigma) \leq K_8\ n^{-1} \sum_{i=1}^{n} (|Y_i| + |\underline{t}_i\ \underline{\beta}|)/\sigma$$

$$\leq (K_8/\sigma)(c_1 + (n^{-1} \sum_{i=1}^{n} |\underline{c}_i\ \underline{\beta}|^2)^{\frac{1}{2}}) \leq (K_8/\sigma)(c_1 + (2\ \lambda^*_\infty)^{\frac{1}{2}}\ ||\underline{\beta}||) .$$

Thus, by (A.7), $\sigma > M_0$, (2.9), $n \geq N_4(w)$ imply

$$\xi/K_8 \leq (C_1 + (2\ \lambda^*_\infty)^{\frac{1}{2}}\ ||\underline{\beta}||)/\sigma \leq C_1/M_0 + (2\ \lambda^*_\infty)^{\frac{1}{2}}\ ||\beta||/\sigma ,$$

which gives $||\beta||/\sigma > C_2$. □

Proposition A4.

$$\left\{ \begin{array}{l} n \geq N_4(w) \\ (2.8) \\ \sigma > M_1 \end{array} \right\} \Rightarrow ||\underline{\beta}||/\sigma < C_2 .$$

Proof: By (2.1), (2.2), the Schwarz inequality and the definition of M_1,

$$\text{(A.8)} \qquad |n^{-1} \sum_{i=1}^{n} \{\psi((Y_i - \underline{c}_i\ \underline{\beta})/\sigma) - \psi(-c_i\ \underline{\beta}/\sigma)\}\ \underline{c}_i\ \underline{\beta}/\sigma|$$
$$\leq n^{-1} \sum_{i=1}^{n} \min(2K_1,\ |Y_i|/\sigma)\ |\underline{c}_i\ \underline{\beta}/\sigma| \leq \delta(2\lambda_\infty^*)^{\frac{1}{2}}\ ||\underline{\beta}/\sigma||\ .$$

Also note that (2.8) implies

$$\text{(A.9)} \qquad 0 = |\,n^{-1} \sum_{i=1}^{n} \psi((Y_i - \underline{c}_i\ \underline{\beta})/\sigma)\underline{c}_i\ \underline{\beta}/\sigma|$$
$$= |\,n^{-1} \sum_{i=1}^{n} \psi(-\underline{c}_i\ \underline{\beta}/\sigma)\underline{c}_i\ \underline{\beta}/\sigma|$$
$$-\,|\,n^{-1} \sum_{i=1}^{n} \{\psi((Y_i - \underline{c}_i\ \underline{\beta})/\sigma) - \psi(-\underline{c}_i\ \underline{\beta}/\sigma)\}\ \underline{c}_i\ \underline{\beta}/\sigma|\ .$$

Again note that, as in (A.5),

$$\text{(A.10)} \qquad |n^{-1} \sum_{i=1}^{n} \psi(-\underline{c}_i\ \underline{\beta}/\sigma)\underline{c}_i\ \underline{\beta}/\sigma| \geq \min(a,K_1)\ \lambda_\infty\ ||\underline{\beta}/\sigma||^2\ (1 + ||\underline{\beta}||/\sigma)^{-1}\ .$$

Hence (A.9), (A.10) imply

$$\text{(A.11)} \qquad 0 = |n^{-1} \sum_{i=1}^{n} \psi((Y_i - \underline{c}_i\ \underline{\beta})/\sigma)\ \underline{c}_i\ \underline{\beta}/\sigma|$$
$$\geq \min(a,K_1)\lambda_\infty\{||\underline{\beta}/\sigma||^2\ (1 + ||\underline{\beta}/\sigma||)^{-1} - \delta'\ ||\underline{\beta}/\sigma||\}\ .$$

This implies $||\beta/\sigma|| \leq \delta'\ (1-\delta')^{-1} < \varepsilon$. □

Proof of Lemma 1. Propositions A3 and A4 show that σ must eventually be bounded (with probability one). Proposition A2 shows that this implies $||\underline{\beta}||$ must be bounded. □

The proof of Theorem 1 is based on the consistency prove of Huber (1967, Part B).

Proposition A5. Fix $\varepsilon > 0$ and let C be a compact subset of $S = \{(\underline{\beta},\sigma): \sigma > 0\}$. Then one can define open neighborhoods $U(\underline{\beta},\sigma)$ of $(\underline{\beta},\sigma)$ so that

$$\text{(A.12)} \quad \Pr\left\{ \begin{array}{l} \text{there exists N such that } n > N \text{ implies} \\ \sup_{(\beta,\sigma)\in C} \ \sup_{(\underline{\beta}_1,\sigma_1)\in U(\underline{\beta},\sigma)} n^{-1} \sum_{i=1}^{n} \left| \begin{array}{l} \psi((Y_i - \underline{c}_i \underline{\beta})/\sigma)(\underline{c}_i \underline{\beta})/\sigma) \\ -\psi((Y_i - \underline{c}_i\underline{\beta})/\sigma_1)(\underline{c}_i \underline{\beta}_1)/\sigma_1) \end{array} \right| > \varepsilon \end{array} \right\} = 0.$$

A similar result holds for $\psi^2((Y_i - \underline{c}_i \underline{\beta})/\sigma)$.

Proof: First note that since ψ is Lipschitz and bounded and C is compact,

$$\text{(A.13)} \quad n^{-1} \sum_{i=1}^{n} |\psi((Y_i - \underline{c}_i \underline{\beta})/\sigma)(\underline{c}_i \underline{\beta}/\sigma) - \psi((Y_i - \underline{c}_i \underline{\beta}_1)/\sigma_1)(\underline{c}_i \underline{\beta}_1/\sigma_1)|$$

$$\leq n^{-1} \sum_{i=1}^{n} |\psi((Y_i - \underline{c}_i \underline{\beta}_1)/\sigma_1)| \; |\underline{c}_i \underline{\beta}_1 - \underline{c}_i \underline{\beta}|/\sigma_1$$

$$+ n^{-1} \sum_{i=1}^{n} |\psi((Y_i - \underline{c}_i \underline{\beta}_1)/\sigma_1)||\underline{c}_i \underline{\beta}|| 1/\sigma - 1/\sigma_1|$$

$$+ n^{-1} \sum_{i=1}^{n} |\psi((Y_i - \underline{c}_i \underline{\beta}_1)/\sigma_1) - \psi((Y_i - \underline{c}_i \underline{\beta})/\sigma_1)| \; |\underline{c}_i\underline{\beta}|/\sigma$$

$$+ n^{-1} \sum_{i=1}^{n} |\psi((Y_i - \underline{c}_i \underline{\beta})/\sigma_1) - \psi((Y_i - \underline{c}_i \underline{\beta})/\sigma)| \; |\underline{c}_i \underline{\beta}|/\sigma$$

$$\leq M \sum_{i=1}^{n} \{\underline{c}_i(\underline{\beta}_1 - \underline{\beta})/\sigma_1 + \underline{c}_i \underline{\beta} \; |1/\sigma - 1/\sigma_1|$$

$$+ |\underline{c}_i \underline{\beta}| \; |\underline{c}_i(\underline{\beta}_1 - \underline{\beta})|/\sigma\sigma_1 + |\underline{c}_i \underline{\beta}|^2 \; |1/\sigma - 1/\sigma_1|/\sigma\}$$

$$\leq M'\{||\underline{\beta}_1 - \underline{\beta}|| + |1/\sigma - 1/\sigma_1|\} ,$$

if the $U(\underline{\beta},\sigma)$ are chosen so that

$$\sup_{(\underline{\beta},\sigma)\in C} \ \sup_{(\beta_1,\sigma_1)\in U(\underline{\beta},\sigma)} 1/\sigma_1 < \infty .$$

This completes the proof. □

Proposition A6. The functions $\{G_n\}_{n=1}^{\infty}$ are equicontinuous on compact subsets of S.

Proof: Similar to that of Proposition A5.

Proof of Theorem 1. From Lemma 1, all solutions to (2.8), (2.9) are eventually confined to a compact set C. Let U be an open neighborhood of $(\underline{0},1)$. Then by (2.6), one can choose $\varepsilon > 0$ (depending on U) so that

$$\lim_{n\to\infty} \inf_{C/U} ||G_n(\underline{\beta},\sigma)|| \geq 5\varepsilon .$$

We discuss only the case

$$\lim_{n} \inf_{C/U} |G_{n1}(\underline{\beta},\sigma)| \geq 5\varepsilon ,$$

as the other case is quite similar.

Now for every $(\underline{\beta},\sigma) \in C/U$, let $U(\underline{\beta})$ be a nieghborhood of $\underline{\beta}$ for which (A.14) and the following hold:

$$\sup_{(\underline{\beta}',\sigma')\in U(\underline{\beta},\sigma)} |G_{n1}(\underline{\beta}',\sigma') - G_{n1}(\underline{\beta},\sigma)| \leq \varepsilon .$$

Select a finite subcover $U_s (s = 1,\ldots,k)$ with associated points $(\underline{\beta}_s,\sigma_s)$ $(s = 1,\ldots,k)$. Then

(A.14) $$\sup_{(\underline{\beta},\sigma)\in C/U} |n^{-1} \sum_{i=1}^{n} \{\psi((Y_i - \underline{c}_i \underline{\beta})/\sigma)(\underline{c}_i \beta/\sigma) - G_{n1}(\underline{\beta},\sigma)\}|$$
$$\leq \sup_{1\leq s\leq K} |n^{-1} \sum_{i=1}^{n} \{\psi((Y_i - \underline{c}_i \underline{\beta}_s)/\sigma_s)(\underline{c}_i \underline{\beta}_s/\sigma_s) - G_{n1}(\underline{\beta}_s,\sigma_s)\}|$$
$$+ 2\varepsilon .$$

Suppose we show that for any (β,σ),

(A.15) $$n^{-1} \sum_{i=1}^{n} \{\psi((Y_i - \underline{c}_i \underline{\beta})/\sigma) \underline{c}_i \underline{\beta}/\sigma - G_n(\underline{\beta},\sigma)\} \to 0 \quad \text{(a.s.)} .$$

Then with probability one we can choose n sufficiently large that the left hand side of (A.14) is bounded by 4ε. Since $|G_n(\underline{\beta},\sigma)| \geq 5\varepsilon$, this implies

$$\inf_{(\underline{\beta},\sigma)\in C/U} |n^{-1} \sum_{i=1}^{n} ((Y_i - \underline{c}_i \underline{\beta})/\sigma)(\underline{c}_i \underline{\beta}/\sigma)| \geq \varepsilon ,$$

and since U was arbitrary, the proof would be complete. Now (A.15) follows from Proposition B1 below with $a_{nK} = \underline{c}_K \underline{\beta}/n\sigma$.

Proof of Proposition 1. Suppose that

$$\liminf_{n\to\infty} \inf_{C} |G_{n1}(\underline{\beta},\sigma)| = 0 .$$

Then, since C is compact, Proposition A6 shows that there exists $(\underline{\beta},\sigma)$ in C with

$$\liminf_{n\to\infty} |G_{n1}(\underline{\beta},\sigma)| = 0 .$$

Now, ψ is odd so that (2.7) implies

$$G_{n1}(\underline{\beta},\sigma) \leq n^{-1} \sum_{i=1}^{n} E\, \psi((Y_1 - \underline{c}_i \beta)/\sigma)(\underline{c}_i \underline{\beta}/\sigma) I\{|\underline{c}_i \underline{\beta}| > \varepsilon\}$$

$$< E\, \psi((Y_1 - \varepsilon)/\sigma)(\varepsilon/\sigma)\, n^{-1} \sum_{i=1}^{n} I\{|\underline{c}_i \underline{\beta}| > \varepsilon\} ,$$

so it suffices to show that there exists $\varepsilon > 0$ for which

(A.19) $$\liminf_{n\to\infty} n^{-1} \sum_{i=1}^{n} I\{|\underline{c}_i \underline{\beta}| > \varepsilon\} > 0.$$

Let $\epsilon^* = \liminf n^{-1} \sum_{i=1}^{n} |\underline{c}_i \underline{\beta}|$. By (2.3)

$$\epsilon^* \geq \liminf n^{-1} \sum_{i=1}^{n} \frac{|\underline{c}_i \underline{\beta}|^2}{(1+||\underline{c}_i||)(1+||\underline{\beta}||)} \geq \lambda_\infty \frac{||\beta||}{1+||\beta||} > 0.$$

Since (2.4) implies that

$$n^{-1}\sum_{i=1}^{n}|\underline{c}_i\ \underline{\beta}| \leq \epsilon + n^{-1}\sum_{i=1}^{n}|\underline{c}_i\ \beta|\ I\{|\underline{c}_i\ \underline{\beta}| > \epsilon\}$$

$$\leq \epsilon + (n^{-1}\sum_{i=1}^{n}|\underline{c}_i\ \beta|^2)^{\frac{1}{2}}(n^{-1}\sum_{i=1}^{n}I\{|\underline{c}_i\ \underline{\beta}| > \epsilon\})^{\frac{1}{2}}$$

$$\leq \epsilon + ||\beta||(\lambda^*_\infty)^{\frac{1}{2}}(n^{-1}\sum_{i=1}^{n}I\{|\underline{c}_i\ \underline{\beta}| > \epsilon\})^{\frac{1}{2}},$$

if $\epsilon < \epsilon^*$, then (A.19) holds.

Appendix B

Proposition B1. Let $X_1, X_2, \ldots$ be i.i.d. bounded mean zero random variables. Let a_{nK} $(K = 1,\ldots,n)$ be a triangular array of constants with

$$|a_{nK}| \leq n^{-\alpha} \quad \text{for some } 0 < \alpha \leq 1$$

$$\sum_{K=1}^{n} a_{nK}^2 \leq n^{-\beta} \quad \text{for some } 0 < \beta.$$

Then

$$T_n = \sum_{K=1}^{n} a_{nK}\ X_K \to 0 \quad \text{(a.s.)} .$$

Proof: Theorem 4.1.3 of Stout (1974). □

Fix $\varepsilon > 0$ and define $A(\underline{c},\varepsilon) = \{2||\underline{c}|| \geq (M/\varepsilon)^{\frac{1}{2}}\}$, where

$$\limsup_n\ 4\ n^{-1}\sum_{i=1}^{n}||\underline{c}_i||^2 \leq M.$$

Then

$$\text{(B.1)} \qquad \limsup_n\ n^{-1}\sum_{i=1}^{n} I(A(\underline{c}_i,\varepsilon)) \leq 4\ n^{-1}\sum_{i=1}^{n}||\underline{c}_i||^2\ \varepsilon/M \leq \varepsilon.$$

Choose $d = d(\varepsilon) = (M/\varepsilon)^{\frac{1}{2}}$. Since $\varepsilon d = (M\varepsilon)^{\frac{1}{2}}$, for sufficiently small ε,

$$\text{(B.2)} \qquad \varepsilon(\max|a_j| + \varepsilon d) + (1 + \varepsilon)\ \varepsilon d/2 < \varepsilon d,$$

where $a_1, \ldots, a_K$ are as in (3.2). Now define

$$B(\varepsilon) = \bigcup_{j=1}^{K} [a_j - \varepsilon d, a_j + \varepsilon d] .$$

Proposition B2. The conditions $||\underline{\beta}|| \le \varepsilon$, $|\sigma^{-1}| \le \varepsilon$, $Y_i \notin B$, $\underline{c}_i \notin A(c_i, \varepsilon)$ imply that for some j,

$$\text{(B.3)} \qquad a_{j-1} < Y_i < a_j \qquad a_{j-1} < (Y_i - \underline{c}_i\,\underline{\beta})/\sigma < a_j$$

Proof: Note

$$|\sigma^{-1}(Y_i - \underline{c}_i\,\underline{\beta}) - Y_i|$$

$$\le \varepsilon|Y_i| + |1+\varepsilon|\ ||\underline{c}_i||\ \ ||\underline{\beta}||$$

$$\le \varepsilon(\max|a_j| + \varepsilon d) + (1+\varepsilon)\varepsilon\ ||\underline{c}_i||$$

$$\le \varepsilon(\max|a_j| + \varepsilon d) + (1+\varepsilon)\varepsilon\ \ d/2$$

$$< \varepsilon d \quad \text{(from B.2)} . \qquad \square$$

We say that $X_n \le c$ almost surely as $n \to \infty$ when

$$\Pr\{\text{there exists } N \text{ such that } n \ge N \text{ implies } X_n \le c\} = 1.$$

Throughout this section, all inequalities are taken in the above sense.

Define $g(\varepsilon) = \varepsilon^{\delta_*/4(1+\delta_*)}$.

Proposition B3. There exist positive numbers C, ε_0, δ_{**} such that for any $0 < \delta < \delta_{**}$ and any sequence $\varepsilon_1, \varepsilon_2, \ldots$ in $[0, \varepsilon_0]$ the following holds: defining $a_n = n^{-\frac{1}{2}+\delta}$, for almost all w, there exists N(w) such that $n \ge N(w)$, $||\underline{\beta}|| \le \varepsilon_n$, $|\sigma^{-1} - 1| \le \varepsilon_n$, $nd_n\,\varepsilon_n \ge \log n$ imply

$$(B.4) \quad ||\Sigma_n^{-1} n^{-1} \sum_{i=1}^{n} \underline{c}_i' \{\psi((Y_i - \underline{c}_i \underline{\beta})/\sigma) - \psi(Y_i) - (\sigma^{-1} - 1) E\, Y_1 \psi'(Y_1)\} + \underline{\beta}\, E\, \psi'(Y_1))|| \leq C(g(\varepsilon_n) + n^{-\delta_{**}})(|\sigma^{-1} - 1| + ||\underline{\beta}||)$$

and

$$(B.5) \quad ||\Sigma_n^{-1} n^{-1} \sum_{i=1}^{n} \underline{c}_i' \{\psi(\sigma^{-1}(Y_i - \underline{c}_i \beta)) - (\sigma^{-1} - 1) E\, Y_1 \psi'(Y_1)\} + \underline{\beta}\, E\psi'(Y_1)|| \leq C\{a_n + (g(\varepsilon_n) + n^{-\delta_{**}})(|\sigma^{-1} - 1| + ||\underline{\beta}||) .$$

Similar bounds hold for

$$(B.6) \quad |n^{-1} \sum_{i=1}^{n} \{\psi^2((Y_i - \underline{c}_i \underline{\beta})/\sigma) - \psi^2(Y_i) + 2\, E\, \psi(Y_1) \psi'(Y_1)\underline{c}_i \underline{\beta}\} - 2(\sigma^{-1} - 1) E\, Y_1 \psi(Y_1) \psi'(Y_1)|$$

and

$$(B.7) \quad |n^{-1} \sum_{i=1}^{n} \{\psi^2((Y_i - \underline{c}_i \underline{\beta})/\sigma) - \xi + 2\, E\psi(Y_1) \psi'(Y_1)\underline{c}_i \underline{\beta}\} - 2(\sigma^{-1} - 1) E\, Y_1 \psi(Y_1) \psi'(Y_1)|$$

respectively.

Proof of Proposition B3. Consider the expression

$$(B.8) \quad ||\underline{c}_i'\{\psi((Y_i - \underline{c}_i \underline{\beta})/\sigma) - \psi(Y_i) - (\sigma^{-1} - 1)Y_i \psi'(Y_i) + \underline{c}_i \underline{\beta} \psi'(Y_i)\}|| .$$

By Proposition B2, if $Y_i \notin B(\varepsilon_n)$ and $\underline{c}_i \notin A(\underline{c}_i, \varepsilon_n)$, then (B.8) can be expanded in a Taylor series. Writing $d_n = d(\varepsilon_n)$ and noting that $||\underline{c}_i|| \leq d_n/2$, we bound (B.8) (in this case) by

$$(B.9) \quad C_1\{|\sigma - 1| + ||\underline{c}_i|| \ ||\underline{\beta}||\}^2 ||\underline{c}_i|| I\{||\underline{c}_i|| < d_n/2\} .$$

If $Y_i \in B(\varepsilon_n)$ (in which case $|Y_i| \leq \max|a_j| + \varepsilon_n d_n$) or $\underline{c}_i \varepsilon A(c_i,\varepsilon_n)$, then since ψ is Lipschitz we can bound (B.8) by

$$\text{(B.10)} \quad C_2\{|\sigma-1| + ||\underline{c}_i|| \; ||\underline{\beta}||\} \, ||\underline{c}_i|| \, (I(Y_i \in B(\varepsilon_n)) + I(A(c_i,\varepsilon_n)))$$

$$+ \, |\psi(\sigma(Y_i - \underline{c}_i \, \underline{\beta})) - \psi(Y_i)| \; ||\underline{c}_i|| \; I(A(\underline{c}_i,\varepsilon_n)) \; .$$

Thus, (B.9)-(B.10) show that

$$\text{(B.11)} \quad ||n^{-1} \sum_{i=1}^{n} \underline{c}_i' \, \{\psi((Y_i - \underline{c}_i \, \underline{\beta})/\sigma) - \psi(Y_i) - (\sigma^{-1} - 1)Y_i \, \psi'(Y_i)$$

$$+ \, \underline{c}_i \, \underline{\beta} \, \psi'(Y_i)\}|| \leq C_3(A_{n1} + A_{n2} + A_{n3} + A_{n4}) \; ,$$

where

$$A_{n1} = n^{-1} \sum_{i=1}^{n} ||\underline{c}_i|| \, (|\sigma-1| + ||\underline{c}_i|| \; ||\underline{\beta}||)^2 \; I(||\underline{c}_i|| < d_n/2)$$

$$A_{n2} = n^{-1} \sum_{i=1}^{n} ||\underline{c}_i|| \, (|\sigma-1| + ||\underline{c}_i|| \; ||\underline{\beta}||) \; I(Y_i \in B(\varepsilon_n))$$

$$A_{n3} = n^{-1} \sum_{i=1}^{n} ||\underline{c}_i|| \, (|\sigma-1| + ||\underline{c}_i|| \; ||\underline{\beta}||) \; I(A(\underline{c}_i,\varepsilon_n)) \; .$$

$$A_{n4} = n^{-1} \sum_{i=1}^{n} |\psi(\sigma(Y_i - \underline{c}_i \, \underline{\beta})) - \psi(Y_i)| \; ||\underline{c}_i|| \; I(A(\underline{c}_i,\varepsilon_n)) \; .$$

By (3.7), since $|\sigma-1| \leq \varepsilon_n$, $||\underline{\beta}|| \leq \varepsilon_n$;

$$A_{n1} \leq C_4 \;{}_n(|\sigma-1| + ||\underline{\beta}||)d_n \leq C_4 \; g(\varepsilon_n) \; (|\sigma-1| + ||\underline{\beta}||) \; .$$

By Lemma 1 of Carroll (1978), and (3.4),

$$\text{(B.12)} \qquad \limsup_{n} \; n^{-1} \sum_{i=1}^{n} I(Y_i \in B(\varepsilon_n)) \leq M_*(d_n \; \varepsilon_n)$$

for some constant M_*. Hence

$$A_{n2} \le C_5(|\sigma-1| + ||\underline{\beta}||)(M_* \, d_n \, \varepsilon_n) + n^{-1} \sum_{i=1}^{n} ||\underline{c}_i||^2 \, I(Y_i \in B(\varepsilon_n)) \, .$$

But, by Hölder's inequality,

$$n^{-1} \sum_{i=1}^{n} ||\underline{c}_i||^2 \, I(Y_i \in B)$$

$$\le (n^{-1} \sum_{i=1}^{n} ||\underline{c}_i||^{2(1+\delta_*)})^{1/(1+\delta_*)} (M_* \, d_n \, \varepsilon_n)^{\delta_*/(1+\delta_*)}$$

$$\le C_6 \, g(\varepsilon_n) \, ,$$

so that almost surely as $n \to \infty$,

$$A_{n2} \le C_7(|\sigma-1| + ||\underline{\beta}||) \, g(\varepsilon_n) \, .$$

Also,

$$A_{n3} \le (|\sigma-1| + ||\underline{\beta}||) \, n^{-1} \sum_{i=1}^{n} (||\underline{c}_i|| + ||\underline{c}_i||^2) \, I(A(\underline{c}_i, \varepsilon_n)) \, .$$

But, by Schwarz's and Hölder's inequalities,

$$n^{-1} \sum_{i=1}^{n} ||\underline{c}_i|| \, I(A(\underline{c}_i, \varepsilon_n)) \le C_8(\varepsilon_n)^{\frac{1}{2}} \le C_8 \, g(\varepsilon_n)$$

$$n^{-1} \sum_{i=1}^{n} ||\underline{c}_i||^2 \, I(A(\underline{c}_i, \varepsilon_n)) \le C_9 \, g(\varepsilon_n) \, .$$

Finally,

$$A_{n4} \le n^{-1} \sum_{i=1}^{n} |\psi(\sigma^{-1} (Y_i - \underline{c}_i \, \underline{\beta})) - \psi(\sigma^{-1} Y_i)| \, ||\underline{c}_i|| \, I(A(\underline{c}_i, \varepsilon_n))$$

$$+ n^{-1} \sum_{i=1}^{n} |\psi(\sigma^{-1} Y_i) - \psi(Y_i)| \, ||\underline{c}_i|| \, I(A(\underline{c}_i, \varepsilon_n)) = A_{n4}^{(1)} + A_{n4}^{(2)} \, .$$

Now, since ψ is Lipschitz, for some $M > 0$,

$$A_{n4}^{(1)} \le M \, n^{-1} \sum_{i=1}^{n} |\underline{c}_i \, \underline{\beta}| \, ||\underline{c}_i|| \, I(A(\underline{c}_i, \varepsilon_n)) \le ||\underline{\beta}|| \, g(\varepsilon_n) \, .$$

Further, since ψ is constant outside an interval and $|\sigma-1| \le \varepsilon_n$, there is a constant K_0 for which $\psi(\sigma Y_i) - \psi(Y_i) = 0$ if $|Y_i| > K_0$. Hence, using this and the fact that ψ is Lipschitz,

$$A_{n4}^{(2)} \le M K_0 \, |\sigma-1| \, n^{-1} \sum_{i=1}^{n} ||\underline{c}_i|| \, I(A(\underline{c}_i, \varepsilon_n)) \le |\sigma-1| \, g(\varepsilon_n) .$$

We have thus obtained for (B.11) the bound we wish to obtain for (B.4). Bounding the difference in the two terms requires two steps, the first considering

$$A_{n5} = n^{-1} \sum_{i=1}^{n} \underline{c}_i' \, \underline{c}_i \, \underline{\beta}(\psi'(Y_i) - E\,\psi'(Y_1)) .$$

Now there is a $\delta_{**} > 0$, depending only upon δ_* (in 3.7) for which if $a_{nK} = c_{Kj}^2/n^{1-\delta_{**}}$, there exists $\alpha_1, \alpha_2 > 0$ for which

$$|a_{nK}| \le n^{-\alpha_1} \qquad \text{(from (3.6))}$$

$$\sum_{K=1}^{n} a_{nK}^2 \le n^{-\alpha_2} \qquad \text{(from (3.6) and (3.7))} .$$

From Proposition B1, this means that almost surely as $n \to \infty$, $||A_{n5}|| \le n^{-\delta_{**}} \, ||\underline{\beta}||$. By similar arguments (noting that ψ' vanishes outside an interval) we obtain that almost surely as $n \to \infty$,

$$||n^{-1} \sum_{i=1}^{n} (\sigma^{-1} - 1)\underline{c}_i \, (Y_i \, \psi'(Y_i) - E\, Y_1 \, \psi'(Y_1))|| \le n^{-\delta_{**}} \, |\sigma-1| \quad \text{(proving (B.4))}$$

$$||n^{-1} \sum_{i=1}^{n} \underline{c}_i \, \psi(Y_i)|| \le n^{-\frac{1}{2}+\delta} = a_n \quad \text{(proving (B.5))}.$$

The results (B.6) and (B.7) follow similarly. □

References

Andrews, D. (1974). A robust method for multiple linear regression. *Technometrics 16*, 523-532.

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

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

Carroll, R.J. (1977). On the asymptotic normality of stopping times based on robust estimators. *Sankhya, Series A* 39, 355-377.

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

Carroll, R.J. (1978b). On estimating variances of robust estimators when the errors are asymmetric. To appear in J. Am. Statist. Assoc.

Carroll, R.J. (1979). Robust methods for factorial experiments with outliers. Unpublished manuscript.

Gleser, L.J. (1965). On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters. *Ann. Math. Statist.* 36, 463-467 (correction 37, 1053-1055).

Gross, A.M. (1977). Confidence intervals for bisquare regression estimates. *J. Am. Statist. Assoc.* 72, 341-354.

Huber, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* 35, 73-101.

Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Prob. 1*, 221-233. Univ. of California Press.

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

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

Lai, T.L., Robbins, H. and Wei, C.Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Natl. Acad. Sci.* 75, 3034-3036.

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

Stout, W.F. (1974). Almost Sure Convergence. Academic Press, New York.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Almost Sure Properties of Robust Regression Estimates

5. TYPE OF REPORT & PERIOD COVERED
Technical

6. PERFORMING ORG. REPORT NUMBER
Mimeo Series #1240

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

8. CONTRACT OR GRANT NUMBER(s)
AFOSR 75 2796

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Dept. of Statistics
University of North Carolina at Chapel Hill

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

11. CONTROLLING OFFICE NAME AND ADDRESS
Directorate of Mathematics & Statistics
AFOSR
Bolling AFB, DC

12. REPORT DATE
July 1979

13. NUMBER OF PAGES
24

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)
Distribution unlimited - approved for public release

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, linear model, almost sure properties, consistency, M-estimates, Fixed-width confidence intervals.

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

We consider Huber's Proposal 2 for robust regression estimates in the general linear model. The estimates are first shown to be strongly consistent. We then develop an almost sure expansion of the estimates, approximating them (to order $o(n^{-1/2})$) by a weighted sum of bounded random variables. The approximation is sufficiently strong to permit construction of sequential fixed-width confidence regions for the regression parameter.