ON THE ASYMPTOTIC NORMALITY OF ROBUST REGRESSION ESTIMATES

by

Raymond J. Carroll* and David Ruppert**
University of North Carolina at Chapel Hill

Abstract

Huber's (1973) proof of asymptotic normality of robust regression estimates is modified to include the estimates used in practice, which have unknown scale and only piecewise smooth defining functions ψ.

Key Words and Phrases: Robustness, M-estimates, regression, linear model, asymptotic theory.

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

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

1. Introduction

We consider the general linear model

$$(1.1) \qquad y_i = \tau_i + z_i/\sigma_0 \quad (i = 1,\ldots,n),$$

where

$$(1.2) \qquad \tau_i = \sum_{j=1}^{p} c_{ij}\ \beta_j^{(0)} \ .$$

Here the $\beta_j^{(0)}$ are unknown parameters, the c_{ij} are known constants, the z_i are independent and identically distributed with common distribution function F (symmetric about zero), and σ_0 is a constant to be chosen later. Huber (1973) proposed estimating β by solving the system of equations

$$(1.3) \qquad \sum_i \psi(\sigma(y_i - t_i(\beta)))c_{ij} = 0 \quad (j = 1,\ldots,p)$$

$$(1.4) \qquad \sum_i \{\psi^2(\sigma(y_i - t_i(\beta))) - \xi\} = 0 \ ,$$

where

$$t_i(\beta) = \sum_{j=1}^{p} c_{ij}\ \beta_j$$

$$\xi = E_\Phi\ \psi^2(z) \ ,$$

(the expectation being taken under the standard normal distribution) and ψ is a monotone nondecreasing function. Let $C = \{c_{ij}\}$, $\Gamma = C(C^T C)^{-1} C^T$ be the projection matrix, and ε be the maximum diagonal element of Γ.

Huber (1973) considers σ fixed and known ($\sigma \equiv 1$) so that one only need solve (1.3). He proves that if $\varepsilon p^2 \to 0$ as $n \to \infty$ (which implies $p^3/n \to 0$) and if ψ is bounded with two continuous bounded derivatives, then all estimates of the form $\hat{\alpha} = \Sigma_{j=1}^p a_j\ \hat{\beta}_j$ $(\Sigma_j\ a_j^2 = 1)$ are asymptotoically normal. It is routine to extend his results to the system (1.3), (1.4) (see below). The smoothness conditions on ψ are not satisfied for three of the most commonly used functions, namely

Huber's function

$$\psi(x) = x \qquad |x| < c$$
$$= c\ \text{sign}(x) \qquad |x| \geq c$$

Hampel's function

$$\psi(x) = -\psi(-x) = x \qquad 0 < x < a$$
$$= a \qquad a \leq x < b$$
$$= a\left(\frac{c-x}{c-b}\right) \qquad b \leq x < c$$
$$= 0 \qquad x \geq c$$

Andrew's function

$$\psi(x) = -\psi(-x) = \text{sine}(x/c) \qquad 0 < x < \pi c$$
$$= 0 \qquad x \geq \pi c\ .$$

In this note we weaken Huber's conditions to include the common ψ-functions, at the cost of slightly strengthened conditions on the rate of growth of p, the dimension of the problem. The results have been applied by Carroll and Ruppert (1979) to the probelm of testing for heteroscedasticity (Bickel (1978)).

As in Huber's proof, we assume $C'C = I$. Because of the invariance of the problem, we can take $\beta_j^{(0)} = 0$ $(j = 1,\ldots,p)$ and $\sigma_0 = 1$, primarily to simplify notation.

2. Notation, Assumptions and Main Results

Let a be an arbitrary $(p \times 1)$ vector for which $\Sigma a_j^2 = ||a||^2 = 1$. Define $s_i = \sum_{j=1}^{p} c_{ij} a_j$. Note that $||t(\beta)||^2 = ||\beta||^2$. Define for $j = 1,\ldots,p$,

$$\Phi_j(\beta,\sigma) = -\sum_{i=1}^{n} \psi(\sigma(y_i - t_i(\beta)))\, c_{ij}/E\psi'(y_1)$$

$$\Psi_j(\beta,\sigma) = \beta_j - \sum_{i=1}^{n} \psi(y_i)\, c_{ij}/E\psi'(y_1),$$

while

$$\Phi_{p+1}(\beta,\sigma) = -n^{-\frac{1}{2}} \sum_{i=1}^{n} \{\psi^2(\sigma(y_i - t_i(\beta))) - \xi\} \,/2\, Ey_1\, \psi(y_1)\psi'(y_1)$$

$$\Psi_{p+1}(\beta,\sigma) = -n^{-\frac{1}{2}} \Big((\sigma-1) + \sum_{i=1}^{n} \{\psi^2(y_i) - \xi\}/2\, Ey_1\, \psi(y_1)\psi'(y_1)\Big) .$$

Our estimates are solutions to $\Phi_j(\hat\beta,\hat\sigma) = 0$ $(j = 1,\ldots,p+1)$, and we hope to approximate $(\hat\beta,\hat\sigma)$ by $(\tilde\beta,\tilde\sigma)$ which solves $\Psi_j(\tilde\beta,\tilde\sigma) = 0$ $(j = 1,\ldots,p+1)$.

We make the following assumptions.

(2.1) ψ is odd, bounded, and constant outside a finite interval.

(2.2) ψ is Lipschitz of order one and has two continuous bounded derivatives except at a finite number of points, which we take without loss as $\pm c$.

(2.3) F is symmetric about zero.

(2.4) F is Lipschitz in neighborhoods of $\pm c$.

(2.5) $E\,\psi'(y_1) \neq 0$, $Ey_1\, \psi(y_1)\, \psi'(y_1) \neq 0$.

<u>Theorem 1</u>. If (2.1) - (2.5) hold and, in addition, there is a sequence $a_n \to 0$ such that

(2.6) $$(\varepsilon p/a_n^2) \to 0, \quad (\varepsilon n a_n) \to 0 ,$$

then there is a sequence of solutions to (1.3) - (1.4) such that

(2.7) $$||(\hat{\beta},\ (np)^{\frac{1}{2}}\ (\hat{\sigma} - 1))|| = O_p(p).$$

If in addition

(2.8) $$(\varepsilon p^2/a_n^2) \to 0, \quad (\varepsilon n p a_n) \to 0 ,$$

then

(2.9) $$||(\hat{\beta},\ (np)^{\frac{1}{2}}\ (\hat{\sigma} - 1)) - (\tilde{\beta},\ (np)^{\frac{1}{2}}\ (\hat{\sigma} - 1))|| \overset{p}{\to} 0 .$$

Thus, (2.1) - (2.5) and (2.8) imply that all esimtates of the form $\hat{\alpha} = \sum_{j=1}^{p} \alpha_j\ \hat{\beta}_j\ (||\alpha||^2 = 1)$ are asymptotically normal.

Remark: The result (2.7) is the starting point in constructing robust tests for heteroscedasticity (Bickel (1978), Carroll and Ruppert (1979)); it implies the crucial assumption T of Bickel's Theorem 3.1. For balanced designs ($\varepsilon = p/n$), assumption (2.6) is satisfied by choosing $a_n = p^{-(1+\gamma)}$ for small $\gamma > 0$, which then requires $p^{4+2\gamma}/n \to 0$, as compared to Huber's condition $p^2/n \to 0$ when ψ has two derivatives.

3. Proofs

Proposition 1. Suppose that (2.1) - (2.5) hold, that ψ has two bounded continuous derivatives, and that $\varepsilon p \to 0$. Then on the set

$$||(\beta,\ (pn)^{\frac{1}{2}}\ (\sigma - 1))||^2 \leq Kp,$$

the following hold:

$$(3.1) \qquad ||\Phi(\beta,\sigma) - \Psi(\beta,\sigma)|| = O_p((\varepsilon p^2)^{\frac{1}{2}}) ,$$

$$(3.2) \qquad ||\Phi(\beta,\sigma) - (\beta, (np)^{\frac{1}{2}} (\sigma - 1))|| = O_p(p^{\frac{1}{2}} + (\varepsilon p^2)^{\frac{1}{2}}) .$$

Proof of Proposition 1. By a Taylor series expansion,

$$(3.3) \quad \left| \sum_{j=1}^{p} a_j(\Phi_j(\beta,\sigma) - \Psi_j(\beta,\sigma)) + (\sigma - 1) \sum_{i=1}^{n} s_i y_i \, \psi'(y_i)/E\, \psi'(y_1) \right.$$
$$\left. -\sigma \sum_{i=1}^{n} s_i \, t_i(\beta)(\psi'(y_i) - E\, \psi'(y_1))/E\, \psi'(y_1) \right|$$
$$= |A_n|$$
$$\leq |\sigma - 1| \, \left|\sum_{i=1}^{n} s_i \, t_i(\beta) y_i \, \psi''((1 + \eta_{2i}) y_i)/E\, \psi'(y_1)\right|$$
$$+ \sigma^2 \left|\sum_{i=1}^{n} s_i \, t_i^2(\beta) \, \psi''(\sigma\, y_i + \eta_{1i})/E\, \psi'(y_1)\right| .$$

On the set $||\beta||^2 \leq Kp$, $|\sigma - 1| \leq \frac{1}{2}$, $||a|| = 1$, Huber shows that the second term on the r.h.s. of (3.3) is $O(\varepsilon^{\frac{1}{2}} \, ||\beta||^2)$. Since $\psi'' = 0$ outside a finite set and $|\eta_{2i}| > \frac{1}{4}$, the first term on the r.h.s. of (3.3) is bounded by

$$M|\sigma - 1| \, (\Sigma \, s_i^2)^{\frac{1}{2}} \, (\Sigma \, t_i^2(\beta))^{\frac{1}{2}} = O(||\beta|| \, |\sigma - 1|) .$$

Thus,

$$|A_n| = O(\varepsilon^{\frac{1}{2}} \, ||\beta||^2 + |\sigma - 1| \, ||\beta||) .$$

We next consider A_n, in particular its last two terms. Since $\Sigma \, s_i^2 = 1$ and $E\, y_1 \, \psi'(y_1) = 0$,

$$(3.4) \qquad (\sigma - 1) \sum_{i=1}^{n} s_i \, y_i \, \psi'(y_i)/E\, \psi'(y_1) = O_p(|\sigma - 1|).$$

Huber shows that

$$(3.5) \qquad \sum_{i=1}^{n} s_i t_i(\beta)(\psi'(y_i) - E\,\psi'(y_1))/E\,\psi'(y_1) = O_p(\varepsilon^{\frac{1}{2}}\,||\beta||^2).$$

Thus, uniformly on the set $||\beta||^2 \leq Kp$, $|\sigma - 1| \leq Kn^{-\frac{1}{2}}$, $||a|| = 1$ we have from (3.3) - (3.5)

$$(3.6) \qquad |\sum_{j=1}^{p} a_j(\Phi_j(\beta,\sigma) - \Psi_j(\beta,\sigma)| = O_p(\varepsilon^{\frac{1}{2}}p + pn^{-\frac{1}{2}}) .$$

Another Taylor series expansion shows that

$$(3.7) \qquad \begin{aligned} |B_n| &= |\Phi_{p+1}(\beta,\sigma) - \Psi_{p+1}(\beta,\sigma) + C_{n1} + C_{n2}| \\ &\leq M\{|\sigma - 1|\ ||\beta|| + pn^{-\frac{1}{2}} + n^{-\frac{1}{2}}\ |\sigma - 1|^2\} , \end{aligned}$$

where

$$-\{2\,E\,y_1\,\psi(y_1)\,\psi'(y_1)\}C_{n1} = 2n^{-\frac{1}{2}}\,(\sigma - 1)\sum_{i=1}^{n}(y_i\,\psi(y_i)\,\psi'(y_i) - E\,y_1\,\psi(y_1)\,\psi'(y_1))$$

and

$$\{2\,E\,y_1\,\psi(y_1)\,\psi'(y_1)\}C_{n2} = 2\,\sigma\,n^{-\frac{1}{2}}\sum_{i=1}^{n} t_i(\beta)\,\psi(y_i)\,\psi'(y_i) .$$

Since $C_{n1} = O_p(|\sigma - 1|)$ and $C_{n2} = O_p((p/n)^{\frac{1}{2}})$, we obtain

$$|\Phi_{p+1}(\beta,\sigma) - \Psi_{p+1}(\beta,\sigma)| = O_p(pn^{-\frac{1}{2}}) .$$

Since $(\varepsilon p^2)^{\frac{1}{2}} \geq (p^3/n)^{\frac{1}{2}} \geq pn^{-\frac{1}{2}}$, we thus obtain (3.1). Equation (3.2) follows from (3.20) of Huber (1973).

Proof of Theorem 1. The proof of Proposition 1 makes it clear that we need to obtain bounds for $|A_n|$ and $|B_n|$ uniformly on the set $||\beta|| \leq Kp$, $|\sigma - 1| \leq Kn^{-\frac{1}{2}}$ and $||a|| = 1$. Rewrite

$$E\,\psi'(y_1)\ |A_n| = |\sum_{i=1}^{n} H(i,\beta,\sigma)|$$

$$\left| \sum_{i=1}^{n} \left\{ \begin{array}{l} s_i(\psi(\sigma(y_i - t_i(\beta))) - \psi(y_i)) \\ + (\sigma - 1)\ s_i\ y_i\ \psi'(y_i) \\ - \sigma\ s_i\ t_i(\beta)(\psi'(y_i) - E\ \psi'(y_1)) \end{array} \right\} \right| .$$

Let I be the indicator function and let $a_n \to 0$, $n^{\frac{1}{2}} a_n \to \infty$. Then

$$|A_n| = \left| \sum_{i=1}^{n} H(i,\beta,\sigma) \left\{ \begin{array}{l} I(-c + a_n \le y_i \le c - a_n) \\ + I(y_i \ge c + a_n) + I(y_i \le -c - a_n) \\ + I(-c - a_n < y_i < -c + a_n) + I(c - a_n < y_i < c + a_n) \end{array} \right\} \right|$$

$$= |A_{n1} + A_{n2} + A_{n3} + A_{n4} + A_{n5}| .$$

For notational purposes, define $d_i = \sigma(y_i - t_i(\beta))$. Then

$$A_{n1} = \sum_{i=1}^{n} H(i,\beta,\sigma)\ I(-c + a_n < y_i < c - a_n)$$

$$\times \{I(-c < d_i < c) + I(|d_i| > c)\} = A_{n1}^{(1)} + A_{n1}^{(2)} .$$

By Proposition 1, $|A_{n1}^{(1)}| = O(\varepsilon^{\frac{1}{2}} p)$. Since ψ is Lipschitz and constant outside a finite interval,

$$|A_{n1}^{(2)}| \le M \sum_{i=1}^{n} |s_i|\{|\sigma - 1| + |t_i(\beta)|\}\ I(-c + a_n < y_i < c - a_n\quad |d_i| > c) .$$

However, since $|\sigma - 1| \le Kn^{-\frac{1}{2}}$ and $n^{\frac{1}{2}} a_n \to 0$,

$$|A_{n1}^{(2)}| \le M\ \varepsilon^{\frac{1}{2}} \sum_{i=1}^{n} \{|\sigma - 1| + |t_i(\beta)|\}\ I\{|t_i(\beta)| > a_n/2\} .$$

Now

$$\sum_{i=1}^{n} I\{|t_i(\beta)| > a_n/2\} \le 4p/a_n^{\,2} \tag{3.8}$$

$$(3.9) \qquad \sum_{i=1}^{n} |t_i(\beta)| \; I\{|t_i(\beta)| > a_n/2\} \leq 4p/a_n \; ,$$

so that (3.8) and (3.9) imply

$$|A_{n1}^{(2)}| \leq 4M \; \varepsilon^{\frac{1}{2}}\{|\sigma - 1| \; p/a_n^{\;2} + p/a_n\} \; .$$

However, $|\sigma - 1|/a_n \leq K(na_n^{\;2})^{-\frac{1}{2}}$ so that

$$(3.10) \qquad |A_{n1}| = O(\varepsilon^{\frac{1}{2}} \; p/a_n \; (1 + (na_n^{\;2})^{-\frac{1}{2}}) \; .$$

The same bound (3.10) holds for $|A_{n2}|$ and $|A_{n3}|$. Noting once again that ψ is Lipschitz and constant outside a finite interval,

$$\begin{aligned} |A_{n5}| &\leq \sum_{i=1}^{n} |s_i| \; \{|\sigma - 1| + |t_i(\beta)|\} \; I\{|y_i - c| < a_n\} \\ &\leq M\varepsilon^{\frac{1}{2}} \; \{|\sigma - 1| \; G_n + p^{\frac{1}{2}} \, G_n^{\;\frac{1}{2}}\}, \end{aligned}$$

where

$$(3.11) \qquad G_n = \sum_{i=1}^{n} I\{|y_i - c| < a_n\} \; .$$

From Lemma 1 of Carroll (1978), provided $na_n \geq \log n$,

$$G_n = O_p(na_n) \; .$$

This gives

$$(3.12) \qquad |A_{n5}| = O_p((\varepsilon \; npa_n)^{\frac{1}{2}}) \; .$$

The bound (3.12) also holds for $|A_{n4}|$. Thus,

$$(3.13) \qquad |A_n| = O_p(\varepsilon^{\frac{1}{2}} \; p/a_n \; (1 + (na_n^{\;2})^{-\frac{1}{2}}) + (\varepsilon \; npa_n)^{\frac{1}{2}}).$$

The same bound holds for $|B_n|$. Thus,

$$(3.14)\quad ||\Phi(\beta,\sigma) - \Psi(\beta,\sigma)|| = O_p(\varepsilon^{\frac{1}{2}}\, p(1 + (na_n^{\;2})^{-\frac{1}{2}})/a_n + (\varepsilon\, npa_n)^{\frac{1}{2}})$$

$$= O_p(r(p,n))\ .$$

Equation (3.14) is the generalization of Huber's (3.18). As he shows, for sufficiently large K, on the set $||\beta||^2 \leq Kp$, $|\sigma - 1| \leq Kn^{-\frac{1}{2}}$,

$$(3.15)\qquad ||\Phi(\beta,\sigma) - (\beta,\ (np)^{\frac{1}{2}}\,(\sigma - 1))|| \leq r(p,n) + \tfrac{1}{2}(Kp)^{\frac{1}{2}}\ .$$

Thus, if

$$(3.16)\qquad r(p,n)/p^{\frac{1}{2}} \to 0\ ,$$

Brouwer's fixed point theorem enables us to conclude that Φ has a zero inside the ball $||(\beta,\ (np)^{\frac{1}{2}}\,(\sigma - 1))|| < Kp$. Equation (3.16) is true if (2.6) is true. The rest of the proof parallels that of Huber. □

References

Bickel, P.J. (1978). Using residuals robustly I: tests for heteroscedasticity, non-linearity. *Ann. Statist.* *6*, 266-291.

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

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.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) On the Asymptotic Normality of Robust Regression Estimates	5. TYPE OF REPORT & PERIOD COVERED Technical
	6. PERFORMING ORG. REPORT NUMBER Mimeo Series #1242
7. AUTHOR(s) Raymond J. Carroll and David Ruppert	8. CONTRACT OR GRANT NUMBER(s) AFOSR-75-2796
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics UNC Chapel Hill, NC 27514	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling AFB, DC	12. REPORT DATE July 1979
	13. NUMBER OF PAGES 10
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)

Robustness, M-estimates, regression, linear model, asymptotic theory

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

Huber's (1973) proof of asymptotic normality of robust regression estimates is modified to include the estimates used in practice, which have unknown scale and only piecewise smooth defining functions ψ.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)