

AD-A077 566

NORTH CAROLINA UNIV AT CHAPEL HILL INST OF STATISTICS
ON ROBUST TESTS FOR HETEROSCEDASTICITY.(U)

F/6 12/1

JUL 79 R J CARROLL , D RUPPERT

AFOSR-75-2796

UNCLASSIFIED Mimeo SER-1241

AFOSR-TR-79-1100

NL

OF
AD
A077566

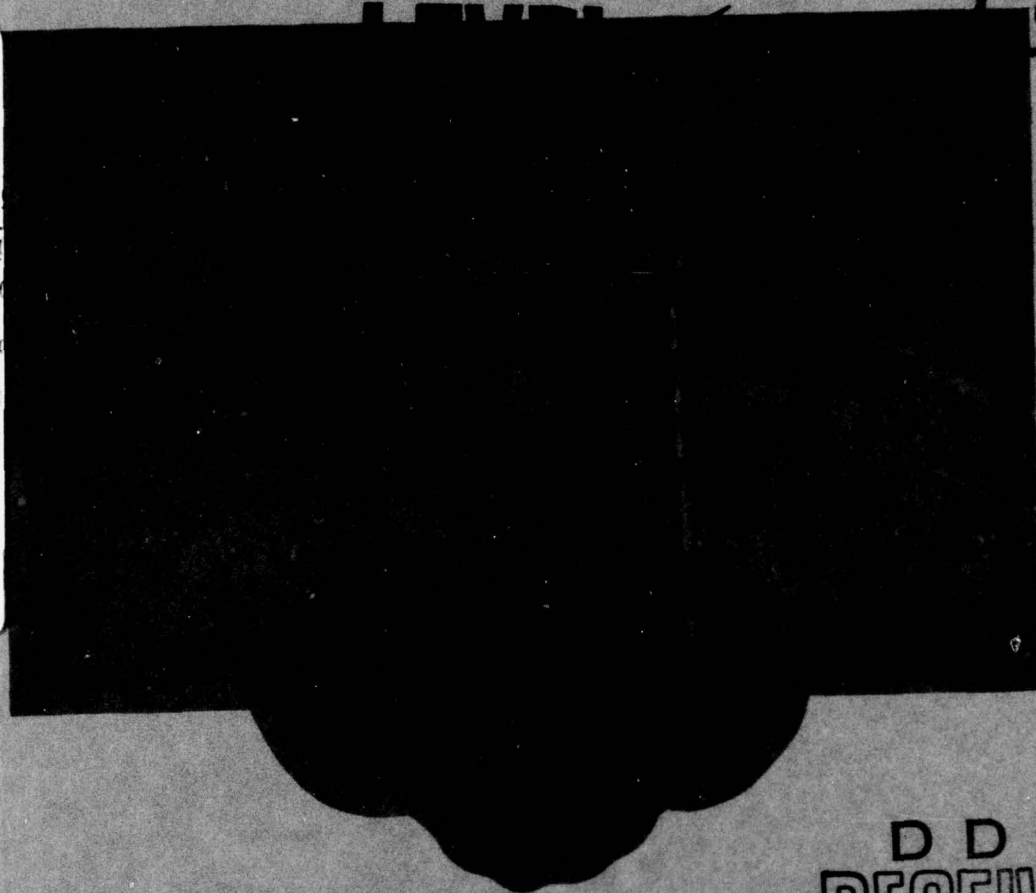


END
DATE
FILMED

1 -80

DDC

AD A 077566



2

DDC
RECEIVED
DEC 3 1979
E

On Robust Tests for Heteroscedasticity

by

Raymond J. Carroll
and
David Ruppert

University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series #1241

July 1979

DDC FILE COPY

79 11 27 062
DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

Approved for public release;
distribution unlimited.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
18 AFOSR-TR-79-1100		
4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED	
6 On Robust Tests for Heteroscedasticity.	9 Interim rept.	
7. AUTHOR(s)	6. PERFORMING ORG. REPORT NUMBER	8. CONTRACT OR GRANT NUMBER(s)
10 Raymond J. Carroll and David Ruppert	Mimeo Series No. 1241	15 AFOSR-75-2796
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
Department of Statistics University of North Carolina Chapel Hill, North Carolina 27514	16 61102F 17 2304/A5	
11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE	
Air Force Office of Scientific Research Bolling AFB, DC 20332	12 Jul 79	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	13. NUMBER OF PAGES	
12 13	13 11	
	15. SECURITY CLASS. (of this report)	
	UNCLASSIFIED	
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report)		
Approved for Public Release: Distribution Unlimited		
14 MIMEO SER-1241		
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)		
Heteroscedasticity, robustness, linear model, M-estimates, regression, asymptotic tests		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)		
We extend Bickel's (1978) tests for heteroscedasticity to include wider classes of test statistics and fitting methods. The test statistics include those based on Huber's function, while the fitting techniques include Huber's Proposal 2 (1977) for robust regression.		
410064		

On Robust Tests for Heteroscedasticity

by

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

We extend Bickel's (1978) tests for heteroscedasticity to include wider classes of test statistics and fitting methods. The test statistics include those based on Huber's function, while the fitting techniques include Huber's Proposal 2 (1977) for robust regression.

AMS 1970 Subject Classifications: Primary 62G35; Secondary 62J05, 62J10.

Key Words and Phrases: Heteroscedasticity, robustness, linear model, M-estimates, regression, asymptotic tests.

*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.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

1. Introduction

We consider the general linear model

$$(1.1) \quad Y_i = \tau_i + \sigma(\tau_i, \theta) \epsilon_i, \quad \tau_i = \underline{c}_i' \underline{\beta}_0 \quad (i = 1, \dots, n),$$

where $\underline{\beta}_0$ is an unknown $(p \times 1)$ vector, the $(p \times 1)$ vectors $\underline{c}_i'$ are known, the error terms ϵ_i are independent and identically distributed (i.i.d.) with common distribution function F , and $\sigma(\tau_i, \theta)$ expresses the possible heteroscedasticity in the model, with

$$(1.2) \quad \sigma(\tau, \theta) = 1 + \theta a(\tau) + o(\theta) \quad \text{as } \theta \rightarrow 0.$$

Bickel (1978), generalizing work of Anscombe (1961), defines robust tests for heteroscedasticity, which in the present context is a test of $H_0: \theta = 0$; the idea is to replace aspects of the usual informal examination of residuals by formal statistical inference about the probability structure of the data. If $\{t_i\}$ are the fitted values (from least squares or possibly a robust regression method (Huber (1973)(1977))) and b is an even function, Bickel's robust test statistic is

$$(1.3) \quad A_b = \frac{n}{\sum_{i=1}^n} (a(t_i) - a_1(t)) b(r_i) / \hat{\sigma}_b^2,$$

where

$$(1.4) \quad r_i = Y_i - t_i = \text{residual},$$

$$(1.5) \quad \hat{\sigma}_b^2 = \frac{n}{\sum_{i=1}^n} (a(t_i) - a_1(t))^2 (n-p)^{-1} \frac{n}{\sum_{i=1}^n} (b(r_i) - b_1(r))^2,$$

and for any function g ,

$$g_1(x) = n^{-1} \sum_{i=1}^n g(x_i).$$

Accession For	
NTIS	Govt
DDC TAB	Unannounced
Justification	
By	
Distribution	
Availability Code	
Dist	Avail and/or Special

Bickel makes the following assumption:

(1.6) b is bounded and has two continuous, bounded derivatives.

Under (1.6) and other assumptions (see Theorem 1 below), Bickel obtains the asymptotic distribution of A_p under $H_0: \theta = 0$ and contiguous alternatives; results are obtained for the case $p^2/n \rightarrow 0$.

One of the most attractive choices of b (well motivated in Bickel's Section 3) is Huber's function squared:

$$(1.7) \quad \begin{aligned} b(x) &= x^2 & |x| \leq k \\ &= k^2 & |x| > k. \end{aligned}$$

This choice of b does not satisfy (1.6) so that Bickel's Theorem 3.1 does not apply. He states that the strong smoothness condition (1.6) is "unsatisfactory" and obtains results for (1.7) only when p is bounded and fitting is by least squares.

In this note we show by a simple modification of Bickel's proofs (using techniques of Carroll (1978)), that results for A_p can be obtained for b given by (1.7) even when $p^4/n \rightarrow 0$ and fitting is by robust estimates on least squares. This result is given in Section 2. In Section 3 we note extensions which obtain scale invariance by robust estimation with scale estimated by Huber's Proposal 2.

2. Main Results

Where possible we adopt Bickel's notation. Without loss of generality we assume $n^{-1} \sum_{i=1}^n c_i = 1$. To provide a frame of reference we state:

Theorem 1 (Bickel (1978)). Suppose the following hold:

$$(2.1) \quad \max |\tau_i| \leq M,$$

$$(2.2) \quad n^{-1} \sum (a(\tau_i) - a_*(\tau))^2 \geq M^{-1} > 0,$$

$$(2.3) \quad |\theta n^2| \leq M,$$

$$(2.4) \quad F \text{ is symmetric about zero,}$$

$$(2.5) \quad M^{-1} \leq J_2(F) \leq M \text{ where for } F' = f(\text{absolutely continuous}),$$

$$J_2(F) = \int (x f'(x)/f(x) + 1)^2 f(x) dx,$$

$$(2.6) \quad \text{Var}(b(\epsilon_1)) \geq M^{-1} > 0,$$

$$(2.7) \quad \text{The function } a \text{ is twice boundedly and continuously differentiable.}$$

$$(2.8) \quad \text{If } d_i = t_i - \tau_i, \quad \sum d_i^2 = O_p(p),$$

$$(2.9) \quad b(x) = b(-x),$$

$$(2.10) \quad b \text{ is bounded and satisfies (1.6),}$$

$$(2.11) \quad p^2/n \rightarrow 0.$$

Then

$$(2.12) \quad P_\theta \{ \Lambda_b \geq z \} = 1 - \Phi(z - \Lambda_b) + o(1),$$

where

$$(2.13) \quad \Delta_b(\theta, n) = \theta \left[\sum_{i=1}^n (a(\tau_i) - a_*(\tau))^2 \right]^{1/2} E \epsilon_1 b'(\epsilon_1) [\text{Var}(b(\epsilon_1))]^{-1/2}.$$

Our generalization of Theorem 1 to incorporate such functions as (1.7) is

Theorem 2. Suppose (2.1)-(2.9) and the following hold:

$$(2.14) \quad b \text{ is bounded, Lipschitz of order one, and has two bounded continuous derivatives except possibly at a finite number of points, which we take as } \pm c.$$

$$(2.15) \quad p^4/n \rightarrow 0.$$

Then (2.12) holds. (Assumption (2.8) is discussed in the next section.)

Proof of Theorem 2. The key results in Bickel's proof are (A34)-(A37) with

$$\begin{aligned} w_{ij} &= 1 - 1/n & (i = j) \\ &= -1/n & (i \neq j). \end{aligned}$$

Because b is bounded and Lipschitz of order one, (A34)-(A36) follow exactly as given by Bickel. He uses (A37) to prove

$$\begin{aligned} (2.16) \quad n^{-1/2} \sum_{i=1}^n (a(t_i) - a(\tau_i))b(r_i) \\ = n^{-1/2} \sum_{i=1}^n (a(\tau_i) - a(\tau_i))b(\epsilon_i) \\ + n^{-1/2} E b'(\epsilon_1) \sum_{i=1}^n (a(\tau_i) - a(\tau_i))d_i + o_p(1), \end{aligned}$$

where $d_i = t_i - \tau_i$. Instead of proving (A37) we will prove (2.16) directly.

As seen in Bickel's (A41)-(A47), (2.16) is verified by proving either (A48) (as Bickel has done) or

$$(2.17) \quad n^{-1/2} A_n = n^{-1/2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} a(\tau_i) (b(r_j) - b(\epsilon_j) + d_j b'(\epsilon_j)) \xrightarrow{p} 0.$$

We will prove (2.17). Note that in Bickel's proofs of (A41)-(A47) the assumption (1.6) is not needed; the weaker assumption (2.14) suffices. Since $r_j = \epsilon_j - d_j$, with I being the indicator function, rewrite

$$\begin{aligned} (2.18) \quad A_n &= \sum_i \sum_j w_{ij} a(\tau_i) (b(\epsilon_j - d_j) - b(\epsilon_j) + d_j b'(\epsilon_j)) \\ &\times \{I(-c + a_n \leq \epsilon_j \leq c - a_n) + I(c - a_n \leq \epsilon_j \leq c + a_n) \\ &+ I(-c - a_n \leq \epsilon_j \leq -c + a_n) + I(\epsilon_j \geq c + a_n) + I(\epsilon_j \leq -c - a_n)\} \\ &= A_{n1} + A_{n2} + A_{n3} + A_{n4} + A_{n5}, \end{aligned}$$

where $a_n \rightarrow 0$ will be specified later. We also write $\Lambda_n = \sum_i \sum_j K_{ij}(n)$. We can further write

$$(2.19) \quad \Lambda_{n1} = \sum_i \sum_j K_{ij}(n) \left\{ \begin{array}{l} I(-c < \epsilon_j - d_j < c) \\ + I(|\epsilon_j - d_j| > c) \end{array} \right\} I(-c + a_n < \epsilon_j < c - a_n) = \Lambda_{n1}^{(1)} + \Lambda_{n2}^{(2)}$$

As in Bickel's (A48), since b is differentiable on $(-c, c)$,

$$|\Lambda_{n1}^{(1)}| = O_p(\sum d_i^2) = O_p(p).$$

Note that if $-c + a_n < \epsilon_j < c - a_n$ and $|\epsilon_j - d_j| > c$ then $|d_j| > a_n$. Then, by (2.1), since b is Lipschitz and $w_{ij} = \delta_{ij} - 1/n$,

$$\begin{aligned} |\Lambda_{n1}^{(2)}| &\leq M_1 \sum_{j=1}^n |b(\epsilon_j - d_j) - b(\epsilon_j) + d_j b'(\epsilon_j)| I\{-c + a_n < \epsilon_j < c - a_n, |\epsilon_j - d_j| \geq c\} \\ &\leq M_2 \sum_{j=1}^n |d_j| I\{|d_j| \geq a_n\} \leq M_2 \left(\sum_{j=1}^n d_j^2 \right)^{1/2} \left(\sum_{j=1}^n I\{|d_j| \geq a_n\} \right)^{1/2} = O_p(p/a_n). \end{aligned}$$

Thus $|\Lambda_{n1}| = O_p(p/a_n)$. Similarly, $|\Lambda_{n4}| = O_p(p/a_n)$, $|\Lambda_{n5}| = O_p(p/a_n)$.

Further, by (2.1) and since b is Lipschitz,

$$(2.20) \quad |\Lambda_{n2}| \leq M_1 \left(\sum d_j^2 \right)^{1/2} \left(\sum_j I\{c - a_n \leq \epsilon_j \leq c + a_n\} \right)^{1/2}.$$

A similar bound holds for $|\Lambda_{n3}|$. Since by (2.5) F is Lipschitz in neighborhoods of $\pm c$, Lemma 1 of Carroll (1978) shows

$$(2.21) \quad \sum_{j=1}^n I\{c - a_n \leq \epsilon_j \leq c + a_n\} = O_p(\max(\log n, n(F(c + a_n) - F(c - a_n)))) .$$

Provided $na_n \geq \log n$, (2.20) and (2.21) give

$$|\Lambda_{n2}| = O_p((n pa_n)^{1/2}) \quad |\Lambda_{n1}| = O_p((n pa_n)^{1/2}).$$

This yields

$$(2.22) \quad n^{-1/2} \Lambda_n = O_p((pa_n)^{1/2} + n^{-1/2} (p/a_n)).$$

If we take $a_n = n^{-1/4}$, (2.15) and (2.22) yield

$$n^{-1/2} A_n = o_p(1),$$

completing the proof. $\square$

3. Extensions

A. Scale invariance

The test statistic A_b is not scale invariant. To obtain such invariance, one would rewrite the model (1.1)-(1.2) so that $a(\tau, \theta) = (1 + a(\tau)\theta + \rho(\theta))/\sigma_0$, where σ_0 is a scale parameter consistently estimated (when $\theta = 0$) by a scale estimate $\hat{\sigma}$ (provided by least squares or Huber's Proposal 2 for robust regression). To obtain scale invariance, Bickel suggests replacing $b(r_i)$ by $b(r_i/\hat{\sigma})$. The statements and proofs of Theorems 1 and 2 must be modified for this new test statistic which we denote $A_b(\hat{\sigma})$. An analogue of Theorem 2 is

Theorem 3. Suppose the conditions of Theorem 2 hold and, in addition,

$$(3.1) \quad n^{1/2} (\hat{\sigma} - \sigma_0) = o_p(1),$$

$$(3.2) \quad E\{b'(e_1)e_1\}^2 < \infty,$$

$$(3.3) \quad E\{b''(e_1)e_1\}^2 < \infty.$$

Then (2.12) holds for $A_b(\hat{\sigma})$.

Remark. Assumption (3.1) is the subject of part B of this section. Assumptions (3.2) and (3.3) hold if b is constant outside an interval (as is (1.7)).

Sketch of the proof of Theorem 3. We need to verify substitutes for Bickel's (A35) and (A37) when $b(r_i)$ is replaced by $b(r_i/\hat{\sigma})$, $b(e_i)$ is replaced by $b(e_i/\sigma_0)$ and the remainder terms are (respectively) $o_p((np)^{1/2})$ and $o_p(p)$. To prove the

substitute for (A35) one must show that

$$(3.4) \quad \sum w_{ij} b(r_i/\hat{\sigma}) b(r_j/\hat{\sigma}) - \sum w_{ij} b(\epsilon_i/\hat{\sigma}) b(\epsilon_j/\hat{\sigma}) = O_p((np)^{1/2})$$

$$(3.5) \quad \sum w_{ij} b(\epsilon_i/\hat{\sigma}) b(\epsilon_j/\hat{\sigma}) - \sum w_{ij} b(\epsilon_i/\sigma_0) b(\epsilon_j/\sigma_0) = O_p((np)^{1/2})$$

Using the special form of w_{ij} , (3.4) follows from (2.8) and the fact that b is bounded and Lipschitz; (3.5) is a consequence of (3.1) and (3.2). Statement (A37) is more complex. The analogue of (A42)-(A45) is to show

$$\sum_{i,j} w_{ij} (a(\tau_i) - a(t_i)) b(\epsilon_j/\hat{\sigma}) = O_p(p),$$

for which (using Bickel's proof) it suffices to show

$$(3.6) \quad \sum_{i,j} w_{ij} a'(\tau_i) d_i (b(\epsilon_j/\hat{\sigma}) - b(\epsilon_j/\sigma_0)) = O_p(p).$$

We rewrite (3.6) as

$$\begin{aligned} (3.7) \quad & \sum_{i,j} w_{ij} a'(\tau_i) d_i (\epsilon_j b'(\epsilon_j/\sigma_0) - E \epsilon_1 b'(\epsilon_1/\sigma_0) (1/\hat{\sigma} - 1/\sigma_0)) \\ & + \sum_{i,j} w_{ij} a'(\tau_i) d_i (b(\epsilon_j/\hat{\sigma}) - b(\epsilon_j/\sigma_0) - (1/\hat{\sigma} - 1/\sigma_0) \epsilon_j b'(\epsilon_j/\sigma_0)) \\ & = B_{1n} + B_{2n}. \end{aligned}$$

That $B_{1n} = O_p(p)$ follows from using the Schwarz inequality, the boundedness of a , and then applying (3.1) and (3.2). That $B_{2n} = O_p(p)$ is complicated notationally but is a consequence of a weakened version of Lemma 2 of Carroll (1978). This verifies (3.6).

The analogue of (A48) is to show that

$$(3.8) \quad \sum_{i,j} w_{ij} a(\tau_i) (b(r_j/\hat{\sigma}) - b(\epsilon_j/\sigma_0)) = O_p(p).$$

First note that (3.1) and the proof of Theorem 2 can be used to show that

$$(3.9) \quad \sum_{i,j} w_{ij} a(\tau_i) (b(r_j/\hat{\sigma}) - b(\epsilon_j/\hat{\sigma}) + (d_i/\hat{\sigma})b'(\epsilon_j/\hat{\sigma})) = O_p(p).$$

To verify (3.8) we must show that the difference between (3.8) and (3.9) is $O_p(p)$; this is a consequence of the following:

$$(3.10) \quad \sum_{i,j} w_{ij} a(\tau_i) (b(\epsilon_j/\hat{\sigma}) - b(\epsilon_j/\sigma_0)) = O_p(p)$$

$$(3.11) \quad \sum_{i,j} w_{ij} a(\tau_i) \frac{d_j}{\hat{\sigma}} (b'(\epsilon_j/\hat{\sigma}) - b'(\epsilon_j/\sigma_0)) = O_p(p)$$

$$(3.12) \quad \sum_{i,j} w_{ij} a(\tau_i) d_j b'(\epsilon_j/\sigma_0) (1/\hat{\sigma} - 1/\sigma_0) = O_p(p).$$

Equations (3.11) and (3.12) follow by applying the Schwarz inequality, (2.8), (2.14), (3.1) and (3.3). We can rewrite (3.10) as

$$(3.13) \quad \sum_{i,j} w_{ij} a(\tau_i) (b(\epsilon_j/\hat{\sigma}) - b(\epsilon_j/\sigma_0) - (1/\hat{\sigma} - 1/\sigma_0)\epsilon_j b'(\epsilon_j/\sigma_0)) \\ + \sum_{i,j} w_{ij} a(\tau_i) [\epsilon_j b'(\epsilon_j/\sigma_0) - E \epsilon_1 b'(\epsilon_1/\sigma_0)] (1/\hat{\sigma} - 1/\sigma_0) = B_{n1}^* + B_{n2}^*,$$

the last step following since $\sum w_{ij} a(\tau_i) = 0$. That $B_{n1}^* = O_p(p)$ follows as in the proof of Theorem 2, while $B_{n2}^* = O_p(p)$ follows from (3.1) and the Chebychev inequality. $\square$

B. On assumption (2.8). Huber's Proposal 2 for robust regression is to solve

$$(3.12) \quad \sum_{i=1}^n \psi((Y_i - \underline{c}_i \underline{\beta})/\sigma) \underline{c}_i = 0$$

$$(3.13) \quad \sum_{i=1}^n \psi^2((Y_i - \underline{c}_i \underline{\beta})/\sigma) = \tau = E \psi^2(Z),$$

the last expectation taken under the standard normal distribution function. Huber (1973) shows (2.8) under the following conditions:

- (3.14) ψ is odd and non-decreasing,
 ψ has two bounded continuous derivatives,
 $\alpha = 1$ and only (3.12) is solved,
 $\gamma p \rightarrow 0$, where γ is the maximum diagonal element of the projection matrix $C(C'C)^{-1}C'$ (since $\gamma \geq p/n$, $p^2/n \rightarrow 0$ is necessary.)

The above conditions are very restrictive in assuming α known, and Huber's function ($\psi(x) = \max(-k, \min(k, x))$) does not satisfy the smoothness condition.

Carroll and Ruppert (1979) have generalized Huber's result by means of the techniques used in the proof of Theorem 2. They utilize the full system (3.12)-(3.13) and assume only that ψ satisfies (2.14), which is true for all functions used in practice. The price paid is a stronger condition on the growth rate of p ; they require that for some sequence $a_n \rightarrow 0$, both $\gamma p/a_n^2 \rightarrow 0$ and $n\gamma a_n \rightarrow 0$. When the design is balanced ($\gamma = p/n$), then $a_n = p^{-(1+\epsilon)}$ for any $\epsilon > 0$ suffices, but that requires $p^{4+2\epsilon}/n \rightarrow 0$.

C. Smoothness of F. Condition (2.5) is rather strong. Ruppert and Carroll (1979) show by entirely different methods that when p is fixed and b satisfies (2.14), (2.5) can be relaxed by requiring only that F is Lipschitz of order one in neighborhoods of $\pm c$.

References

- Anscombe, F.J. (1961). Examination of residuals. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* 1, 1-36
- 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.