The Annals of Statistics
1981, Vol. 9, No. 1, 206–210

ON ROBUST TESTS FOR HETEROSCEDASTICITY

BY RAYMOND J. CARROLL[1] AND DAVID RUPPERT[2]

University of North Carolina at Chapel Hill

We extend Bickel's 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 for robust regression.

1. Introduction. We consider the general linear model

$$Y_i = \tau_i + \sigma(\tau_i, \theta)\epsilon_i, \ \tau_i = \mathbf{c}_i\boldsymbol{\beta}_0 \qquad (i = 1, \cdots, n), \tag{1.1}$$

where $\boldsymbol{\beta}_0$ is an unknown $(p \times 1)$ vector, the $(p \times 1)$ vectors $\mathbf{c}_i'$ 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

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

Bickel (1978), generalizing work of Anscombe (1961), defines robust tests for heteroscedasticity, which in the present context are tests 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

$$A_b = \textstyle\sum_{i=1}^n (a(t_i) - a.(t))b(r_i)/\hat{\sigma}_b, \tag{1.3}$$

where

$$r_i = Y_i - t_i = \text{residual}, \tag{1.4}$$

$$\hat{\sigma}_b^2 = \textstyle\sum_{i=1}^n (at_i) - a.(t))^2(n-p)^{-1} \sum_{i=1}^n (b(r_i) - b.(r))^2, \tag{1.5}$$

and for any function g,

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

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_b under $H_0: \theta = 0$ and contiguous alternatives; results are obtained for the case $p^2/n \to 0$.

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

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

This choice of b does not satisfy (1.6) so that Bickel's Theorem 3.1 does not apply. He states

Received March 1979; revised October 1979.

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

[2] This work was supported by the National Science Foundation Grant NSF MCS78-01240.

AMS 1970 *subject classification.* Primary 62G35; secondary 62J05, 62J10.

Key words and phrases. Heteroscedasticity, robustness, linear model, M-estimates, regression, asymptotic tests.

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_b can be obtained for b given by (1.7) even when $p^4/n \to 0$ and fitting is by robust estimates or 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. We will assume that $Q_n = \Sigma\, \mathbf{c}_i'\mathbf{c}_i$ is invertible and reparameterize to $Q_n = I$; the reparameterization can be ignored here because it does not change the τ_i's. To provide a frame of reference we state:

THEOREM 1. (Bickel (1978)). *Suppose the following hold:*

$$\max |\tau_i| \le M, \tag{2.1}$$

$$n^{-1}\,\Sigma(a(\tau_i) - a\cdot(\tau))^2 \ge M^{-1} > 0, \tag{2.2}$$

$$|\theta n^{1/2}| \le M, \tag{2.3}$$

$$F \text{ is symmetric about zero}, \tag{2.4}$$

$$M^{-1} \le J_2(F) \le M \qquad \text{where for} \quad F' = f(\text{absolutely continuous}),$$
$$J_2(F) = \int (xf'(x)/f(x) + 1)^2 f(x)\,dx, \tag{2.5}$$

$$\operatorname{Var}(b(\epsilon_1)) \ge M^{-1} > 0, \tag{2.6}$$

$$\text{the function } a \text{ is twice boundedly and continuously differentiable}, \tag{2.7}$$

$$\text{if } d_i = t_i - \tau_i,\ \Sigma\, d_i^2 = O_p(p), \tag{2.8}$$

$$b(x) = b(-x), \tag{2.9}$$

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

$$p^2/n \to 0. \tag{2.11}$$

Then

$$P_\theta\{A_b \ge z\} = 1 - \Phi(z - \Delta_b) + o(1), \tag{2.12}$$

where

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

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

THEOREM 2. *Suppose* (2.1)–(2.9) *and the following hold:*

$$\begin{array}{c}\textit{b is bounded, Lipschitz of order one, and has two bounded continuous derivatives}\\ \textit{except possibly at a finite number of points, which we take as } \pm c.\end{array} \tag{2.14}$$

$$p^4/n \to 0. \tag{2.15}$$

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

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

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

$$\begin{aligned} &n^{-1/2}\textstyle\sum_{i=1}^n (a(t_i) - a\cdot(t))b(r_i) \\ &= n^{-1/2}\textstyle\sum_{i=1}^n (a(\tau_i) - a\cdot(\tau))b(\epsilon_i) \\ &\quad + n^{-1/2}Eb'(\epsilon_1)\textstyle\sum_{i=1}^n (a(\tau_i) - a\cdot(\tau))\, d_i + o_p(1), \end{aligned} \tag{2.16}$$

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

$$n^{-1/2}A_n = n^{-1/2}\textstyle\sum_{i=1}^n\sum_{j=1}^n w_{ij}a(\tau_i)(b(r_j) - b(\epsilon_j) + d_jb'(\epsilon_j)) \to_p 0. \tag{2.17}$$

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

$$\begin{aligned} A_n &= \Sigma_i\Sigma_j w_{ij}a(\tau_i)(b(\epsilon_j - d_j) - b(\epsilon_j) + d_jb'(\epsilon_j)) \\ &\qquad \times \{I(-c + a_n \le \epsilon_j \le c - a_n) + I(c - a_n \le \epsilon_j \le c + a_n) \\ &\qquad + I(-c - a_n < \epsilon_j < -c + a_n) + I(\epsilon_j \ge c + a_n) + I(\epsilon_j \le -c - a_n)\} \\ &= A_{n1} + A_{n2} + A_{n3} + A_{n4} + A_{n5}, \end{aligned} \tag{2.18}$$

where $a_n \to 0$ will be specified later. We also write $A_n = \Sigma_i\Sigma_j K_{ij}(n)$. We can further write

$$A_{n1} = \Sigma_i\Sigma_j K_{ij}(n)\begin{Bmatrix} I(-c < \epsilon_j - d_j < c \\ + I(|\epsilon_j - d_j| > c)\end{Bmatrix} I(-c + a_n < \epsilon_j < c - a_n) = A_{n1}^{(1)} + A_{n1}^{(2)} \tag{2.19}$$

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

$$|A_{n1}^{(1)}| = O_p(\Sigma\, d_i^2) = O_p(p).$$

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 $w_{ij} = \delta_{ij} - 1/n$,

$$\begin{aligned} |A_{n1}^{(2)}| &\le M_1\textstyle\sum_{j=1}^n |b(\epsilon_j - d_j) - b(\epsilon_j) + d_jb'(\epsilon_j)| \\ &\qquad \cdot I\{-c + a_n < \epsilon_j < c - a_n, |\epsilon_j - d_j| \ge c\} \\ &\le M_2\textstyle\sum_{i=1}^n |d_j|\, I\{|d_j| \ge a_n\} \\ &\le M_2(\textstyle\sum_{j=1}^n d_j^2)^{1/2}(\sum_{j=1}^n I\{|d_j| > a_n\})^{1/2} = O_p(p/a_n). \end{aligned}$$

Thus $|A_{n1}| = O_p(p/a_n)$. Similarly, $|A_{n4}| = O_p(p/a_n)$, $|A_{n5}| = O_p(p/a_n)$. Further, by (2.1) and since b is Lipschitz,

$$|A_{n2}| \le M_1(\Sigma d_j^2)^{1/2}(\Sigma_j I\{c - a_n \le \epsilon_j \le c + a_n\})^{1/2}. \tag{2.20}$$

A similar bound holds for $|A_{n3}|$. By Markov's inequality,

$$\textstyle\sum_{j=1}^n I\{c - a_n \le \epsilon_j \le c + a_n\} = O_p(n(F(c + a_n) - F(c - a_n))). \tag{2.21}$$

Then (2.20) and (2.21) yield

$$|A_{n2}| = O_p((npa_n)^{1/2}),\ |A_{n3}| = O_p((npa_n)^{1/2}).$$

This yields

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

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

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

completing the proof. □

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 $\sigma(\tau, \theta) = (1 + a(\tau)\theta + o(\theta))/\sigma_0$, 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,*

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

$$E\{b'(\epsilon_1)\epsilon_1\}^2 < \infty, \tag{3.2}$$

$$E\{b''(\epsilon_1)\epsilon_1\}^2 < \infty. \tag{3.3}$$

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

REMARK. Assumption (3.1) is discussed in 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 $b(r_i/\hat{\sigma})$, $b(\epsilon_i)$ is replaced by $b(\epsilon_i/\sigma_0)$ 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

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

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

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

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

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

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

We rewrite (3.6) as

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

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

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

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

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

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:

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

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

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

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

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

the last step following since $\Sigma\, w_{ij}a(\tau_i) = 0$. 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. □

B. ON ASSUMPTION (2.8). M-estimates with estimated scale $\hat{\sigma}$ are solutions to (Huber (1973))

$$\sum_{i=1}^{n} \psi((Y_i - \mathbf{c}_i\boldsymbol{\beta})/\hat{\sigma})\mathbf{c}_i = 0,$$

where ψ is an odd function. Huber (1973) verifies (2.8) for M-estimates when scale is not estimated ($\hat{\sigma} = 1$) and $p\gamma \to 0$, where γ is the maximum diagonal element of the projections matrix $C(C'C)^{-1}C'$. Yohai and Maronna (1979) verify (2.8) when $p\gamma \to 0$ and (3.1) holds. Carroll and Ruppert (1979) verify (2.8) and (3.1) for Huber's Proposal 2 (Huber (1973)), but under more restrictive conditions on the size of $p\gamma$.

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. Stastist.* **6** 314–318.

CARROLL, R. J. and RUPPERT, D. (1979). On the Asymptotic Normality of Robust Regression Estimates. Inst. Statist. Mimeo Series #1242, Univ. North Carolina.

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.

RUPPERT, D. AND CARROLL, R. J. (1979). On Bickel's tests for heteroscedasticity. Inst. Statist. Mimeo Series #1224, Univ. North Carolina.

YOHAI, U. J. AND MARONNA, R. A. (1979). Asymptotic behavior of M-estimators for the linear model. *Ann. Statist.* **7** 258–268.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
321 PHILLIPS HALL 039 A
CHAPEL HILL, NORTH CAROLINA 27514