THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY
OF NORTH CAROLINA

On Bickel's Tests for Heteroscedasticity

by

David Ruppert and Raymond J. Carroll

Institute of Statistics Mimeo Series #1224

April 1979

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

On Bickel's Tests for Heteroscedasticity

by

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

Abstract

The asymptotic distribution theory of Bickel's (1978) tests for heteroscedasticity is extended to a wider class of test statistics and distributions, when the number of regression parameters is fixed.

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

Key Words and Phrases: Heteroscedasticity, robustness, linear model, regression, asymptotic distributions.

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

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

On Bickel's Tests for Heteroscedasticity

1. Introduction. Bickel (1978), generalizing work of Anscombe (1961), considers the general linear model

$$Y_i = \tau_{in} + \sigma(\tau_{in},\theta)e_i, \ \tau_i = \beta' c_{in} \quad \text{for } i = 1,\ldots,n \tag{1.1}$$

where β is an unknown (p×1) vector, c_{in} is an unknown (p×1) vector, and the e_i are i.i.d. with symmetric distribution function F. The function $\sigma(\tau_{in},\theta)$ expresses the (possible) heteroscedasticity in the model. It is assumed that for some function $a(\cdot)$,

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

Also $\sigma(\cdot,\cdot)$ is known but $\theta(= \theta_n)$ is unknown. Bickel developed tests of the hypothesis H: $\theta = 0$ (homoscedasticity) which are robust against gross errors and heavy-tailed F. Let $\{t_i\}$ be the fitted values from either least squares or a robust regression method and let b be an even, bounded (for robustness) function. Bickel's test statistic is

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

with $r_i = Y_i - t_i$ (= ith residual) and

$$\hat{\sigma}_b^2 = \sum_{i=1}^{n} (a(t_i) - a.(t))^2 \, (n-p)^{-1} \sum_{i=1}^{n} (b(r_i) - b.(r))^2 , \tag{1.4}$$

and where for any function g and numbers $x_1,\ldots,x_n$,

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

Let

$$\Delta_b = \theta_n (\sum_{i=1}^{n} (a(\tau_i) - a.(\tau))^2)^{\frac{1}{2}} E\, e_i b'(e_i)(\text{Var } b(e_1))^{-\frac{1}{2}}.$$

Bickel's theorem 3.1 shows that

$$(A_b - \Delta_b) \overset{\mathcal{D}}{\to} N(0,1)$$

under assumptions which include:

$$b \text{ is bounded and has two bounded, continuous derivatives,} \tag{1.5}$$

$$F \text{ has a continuously differentiable density } f, \tag{1.6}$$

$$a \text{ has two bounded, continuous derivatives.} \tag{1.7}$$

$$p\, n^{-\frac{1}{2}} \to 0 \tag{1.8}$$

Bickel states that (1.5) is "unsatisfactory", and is not satisfied for interesting choices of b such as Huber's function squared, that is,

$$b(x) = \min(x^2, k^2)\ . \tag{1.9}$$

Bickel's theorem 3.2 shows that (1.4) holds under considerable weakening of (1.5), if, in addition to the other assumptions of theorem 3.1, the fitting is by least squares, $E\, e_i^2 < \infty$, and

$$p \text{ is fixed.} \tag{1.10}$$

Both the assumption that $E\, e_i^2 < \infty$ and the requirement that least squares be used are restrictive and undesirable in situations where robust methods are deemed necessary, so theorem 3.2 is also unsatisfactory. In another paper (Carroll and Ruppert (1979)), we show that in Bickel's theorem 3.1, one can weaken (1.5) (to include, for example, b given by (1.9)), if (1.8) is strengthened to

$$p\, n^{-\frac{1}{4}} \to 0\ .$$

That paper uses methods of proof very similar to Bickel's. In this note we develop alternate methods, and show that assumptions (1.5)-(1.7) of theorem 3.2 can be weakened, if (1.8) is replaced by (1.10). We do not require that any other assumption besides (1.8) be strengthened. In addition, our results are scale invariant, since in the definition of A_b we replace $b(\cdot)$ by $b(\cdot/\hat{\sigma})$, where $\hat{\sigma}$ is an estimate of scale. Scale invariance is also obtained in Carroll and Ruppert (1979). For example, $\hat{\sigma}$ could be the scale estimate from Huber's (1964, 1973, 1977) proposal 2 or the median of $\{|r_1|, \ldots, |r_n|\}$, possibly normalized through division by a positive constant.

2. Assumptions and Notations. Throughout this paper we assume (1.1), (1.2), (1.3), and (1.10). Also we assume:

b is bounded, symmetric about 0, Lipschitz continuous, and its Radon-Nikodym derivative b' is bounded, (2.1)

b'' exists and is bounded except at a finite number of points (For simplicity, we take the exceptional set to be $\{\xi, -\xi\}$. This involves no essential loss of generality), (2.2)

$$a \text{ is Lipschitz continuous,} \tag{2.3}$$

$$\liminf n^{-1} \sum_{i=1}^{n} (a(\tau_i) - a_{\cdot}(\tau))^2 > 0, \tag{2.4}$$

F is symmetric and Lipschitz continuous in a neighborhood of ξ, (2.5)

$$E|e_1| < \infty \quad \text{and} \quad \operatorname{Var} b(e_1) > 0 \tag{2.6}$$

$$\sup_n \sup_{i \le n} |\tau_{in}| < \infty \tag{2.7}$$

$$\lim_{n\to\infty} \sup_{i \le n} n^{-1/2} ||\underline{c}_{in}|| = 0 \tag{2.8}$$

where for $x = (x_1, \ldots, x_p)$, $||x|| = \max_{\ell \le p} |x_\ell|$.

$$\sup_n \left(n^{-1} \sum_{i=1}^{n} ||\underline{c}_{in}||^2\right) < \infty, \quad \sup_n |n^{\frac{1}{2}}\theta| < \infty, \tag{2.9}$$

and for some $\sigma > 0$,

$$n^{\frac{1}{2}}(\hat{\sigma}_n - \sigma) = O_p(1) \text{ and } n^{\frac{1}{2}}||\hat{\beta}_n - \beta|| = O_p(1). \tag{2.10}$$

(Without loss of generality we take $\sigma = 1$.) The first relation in (2.10) will hold for most robust scale estimates if $\theta \equiv 0$, and assumption (2.9) will imply (2.10) for most estimates of σ even if $\theta \not\equiv 0$.

Then $t_i = c_{in}\hat{\beta}_n$. Define A_b and $\hat{\sigma}_b^2$ as in (1.3) and (1.4), but with r_i replaced by $\hat{\sigma}_n^{-1} r_i$.

3. Main Results.

Theorem 1: (Compare Bickel (1978), Theorem 3.1). As $n \to \infty$,

$$(A_b - \Delta_b) \xrightarrow{\mathcal{D}} N(0,1) .$$

To prove theorem 1, we need a number of preliminary results.

Lemma 1: For each $i \leq n$ suppose that g_{in} is a function from $R^p \times R^1$ to R^1, bounded uniformly in i and n. Let ζ_{in}, $i \leq n$, $n = 1,2,\ldots$ be non-negative constants such that

$$\lim_{n\to\infty} \sup_{i\leq n} \zeta_{in} = 0 \tag{3.1}$$

and

$$\sup_n \left(\sum_{i=1}^{n} \zeta_{in}^2\right) < \infty . \tag{3.2}$$

Assume that for each $R > 0$ there exists $K(R)$ such that for all $\varepsilon > 0$ and all $e \in R^1$ (in the application of lemma 1, e will be one of the errors),

$$\sup_n \sup_{i\le n} \sup\{|g_{in}(\Delta,e) - g_{in}(\Delta',e)|: \Delta,\ \Delta' \in R^p,\ ||\Delta||, ||\Delta'|| \le R,\ ||\Delta-\Delta'|| \le \varepsilon\}$$
$$\le K(R)(|e| + 1)\zeta_{in}\ \varepsilon. \qquad (3.3)$$

Define, for $\Delta \in R^p$,

$$M_n(\Delta) = n^{-\frac{1}{2}} \sum_{i=1}^{n} g_{in}(\Delta, e_i)$$

and

$$N_n(\Delta) = M_n(\Delta) - M_n(0) \ .$$

Then for each $R > 0$,

$$\sup_{||\Delta||\le R} |N_n(\Delta) - E\ N_n(\Delta)| = o_p(1) \ .$$

<u>Proof</u>: Without loss of generality we assume $R = 1$. Fix Δ such that $||\Delta|| \le 1$. By (2.6), (3.1), and (3.3),

$$\text{Var } N_n(\Delta) \to 0 \quad \text{as} \quad n \to \infty,$$

so for each fixed Δ with $|\Delta| \le 1$,

$$N_n(\Delta) - E\ N_n(\Delta) = o_p(1). \qquad (3.4)$$

Define

$$H_n = n^{-\frac{1}{2}} \sum_{i=1}^{n} (|e_i| + 1)\zeta_{in} \ .$$

Then by (3.2)

$$H_n + E\ H_n = O_p(1) \ .$$

Fix $\varepsilon > 0$. By (3.3), if $||\Delta||$, $||\Delta'|| \le 1$ and $||\Delta-\Delta'|| \le \varepsilon$, then

$$|N_n(\Delta) - N_n(\Delta')| \leq K(1)\, H_n\, \varepsilon. \qquad (3.5)$$

For $\eta > 0$ and $x \in R^p$ define $B_x(\eta) = \{y \in R^p;\ ||y-x|| \leq \eta\}$. Then choose $L > 0$ and $a(1),\ldots,a(L) \in B_0(1)$ such that $\bigcup_{\ell=1}^{L} B_{a(\ell)}(\varepsilon) \subset B_0(1)$. Then by (3.4) and (3.5)

$$\begin{aligned}\sup_{||\Delta||\leq 1} |N_n(\Delta) - E\, N_n(\Delta)| &\leq \max_{\ell\leq L} |N_n(a(\ell)) - E\, N_n(a(\ell))| \\ &+ \max_{\ell\leq L} \sup_{\Delta\in B_{a(\ell)}(\varepsilon)} |N_n(\Delta) - N_n(a_i)| + E|N_n(\Delta) - N_n(a_i)| \\ &\leq o_p(1) + (H_n + E\, H_n)\varepsilon\ .\end{aligned}$$

Since H_n is independent of ε, we are done. □

We now introduce additional notation. For $\delta \in R^1$ and $\Delta \in R^p$ define

$$A_{in}(\Delta) = a(\tau_i + n^{-\frac{1}{2}}\, \Delta'\, c_{in}),\quad A^*_{in}(\Delta) = A_{in}(\Delta) - A_{\cdot n}(\Delta)\ ,$$

$$B_{in}(\delta,\Delta,\theta) = b((1 + n^{-\frac{1}{2}}\, \delta)(\sigma(\tau_i,\theta)e_i - n^{-\frac{1}{2}}\, \Delta'\, c_{in}))\ ,$$

and

$$B^*_{in}(\delta,\Delta,\theta) = B_{in}(\delta,\Delta,\theta) - B_{\cdot n}(\delta,\Delta,\theta)\ .$$

Lemma 2. Define

$$U(\delta,\Delta,\theta) = n^{-\frac{1}{2}} \sum_{i=1}^{n} A^*_{in}(\Delta)\, B_{in}(\delta,\Delta,\theta) \quad \text{and}$$

$$V(\delta,\Delta,\theta) = U(\delta,\Delta,\theta) - U(0,0,\theta)\ .$$

Then for each $R > 1$, as $n \to \infty$ and $\sup_n |n^{\frac{1}{2}}\, \theta| < \infty$,

$$\sup_{|\delta|+||\Delta||\leq R} |V(\delta,\Delta,\theta) - E\, V(\delta,\Delta,\theta)| = o_p(1).$$

Proof: Take R = 1. We will show that lemma 1 applies with $g_{in}((\delta,\Delta), e_i) = A^*_{in}(\Delta)\ B_{in}(\delta,\Delta,\theta)$ and $\zeta_{in} = 2n^{-\frac{1}{2}}\ (1 + ||c_{in}||)$. By (2.3), (2.7), and (2.8)

$$\sup_{|\delta|+\|\Delta\|\le 1}\ \sup_n\ \sup_{i\le n}\ |A^*_{in}(\Delta)| < \infty\ , \tag{3.6}$$

and since b is bounded

$$\sup_{|\delta|+\|\Delta\|\le 1}\ \sup_n\ \sup_{i\le n}\ |B_{in}(\delta,\Delta,\theta)| < \infty\ . \tag{3.7}$$

Moreover, by (2.1) and (2.3) there exists K such that $|\delta| + ||\Delta|| \le 1$, $|\delta'| + ||\Delta'|| \le 1$, and $|\delta-\delta'| + ||\Delta-\Delta'|| \le \varepsilon$ implies

$$\begin{aligned} &|B_{in}(\delta,\Delta,\theta) - B_{in}(\delta',\Delta',\theta)| \le K\ \varepsilon\ \zeta_{in}\ (|e_i| + 1) \text{ and} \\ &|A^*_{in}(\Delta) - A^*_{in}(\Delta')| \le n^{-\frac{1}{2}}\ K\ \varepsilon\ \zeta_{in}\ . \end{aligned} \tag{3.8}$$

Since

$$\begin{aligned} &|A^*_{in}(\Delta)\ B_{in}(\delta,\Delta,\theta) - A^*_{in}(0)\ B_{in}(0,0,\theta)| \\ &\le |A^*_{in}(\Delta)(B_{in}(\delta,\Delta,\theta) - B_{in}(0,0,\theta))| + |B_{in}(0,0,0)(A^*_{in}(\Delta) - A^*_{in}(0))| \end{aligned}$$

the assumptions of lemma 1 are implied by (3.6)-(3.8). □

Lemma 3. For all R > 0,

$$\sup_{|\delta|+\|\Delta\|\le R} |E\ V(\delta,\Delta,\theta)| = o(1)\ .$$

Proof: Take R = 1. Then

$$\begin{aligned} E\ V(\delta,\Delta,\theta) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} (A^*_{in}(\Delta) - A^*_{in}(0))E\ B_{in}(0,0,\theta) \\ &+ n^{-\frac{1}{2}} \sum_{i=1}^{n} A^*_{in}(\Delta)(E\ B_{in}(\delta,\Delta,\theta) - E\ B_{in}(0,0,\theta))\ . \end{aligned} \tag{3.9}$$

Let $D(i,n) = \{|\xi - e_1| < n^{-\frac{1}{4}},\ |\xi + e_1| < n^{-\frac{1}{4}}$, or $|e_1| > n^{-\frac{1}{4}}\}$. Define

$$T_{1,i} = n^{-\frac{1}{2}} E\, b'(\sigma(\tau_i,\theta)e_i)(\delta\, \sigma(\tau_i,\theta)e_i - (1 + n^{-\frac{1}{2}}\delta)\Delta' c_{in})$$

$$T_{2,i} = E(B_{in}(\delta,\Delta,\theta) - B_{in}(0,0,\theta) - T_{1,i})(1 - I_{D(i,n)}) \text{ and}$$

$$T_{3,i} = E\{B_{in}(\delta,\Delta,\theta) - B_{in}(0,0,\theta)\}I_{D(i,n)}. \text{ Then,}$$

$$E\, B_{in}(\delta,\Delta,\theta) - E\, B_{in}(0,0,\theta) = T_{1,i}(1 - I_{D(i,n)}) + T_{2,i} + T_{3,i}. \tag{3.10}$$

By (2.2), for some K,

$$|E\, T_{2,i}| \le n^{-1}\ (\delta^2\ \sigma(\tau_i,\theta)^2\ E|e_1|n^{\frac{1}{4}} + (\Delta'\ c_{in})^2)K$$

and therefore by (3.6)

$$\sup_{|\delta|+||\Delta||\le 1} n^{-\frac{1}{2}} \sum_{i=1}^{n} |A^*_{in}\ E\ T_{2,i}| = o(1)\ . \tag{3.11}$$

By (2.6), $E\ |e_i|\ I_{D(i,n)} \to 0$ as $n \to \infty$. By (2.1), for some K $|T_{3,i}| \le Kn^{-\frac{1}{2}}(E|e_i|\ I_{D(i,n)} + |\Delta'\ c_{in}|\ P(D(i,n))$. Therefore

$$\sup_{|\delta|+||\Delta||\le 1} n^{-\frac{1}{2}}\ |\sum_{i=1}^{n} A^*_{in}(\Delta)\ T_{3,i}(\Delta,\delta)| \to 0. \tag{3.12}$$

Since $E\ b'\ (\sigma(\tau_i,\theta)e_i) = 0$ by (2.1) and (2.5),

$$T_{1,i}(\delta,\Delta,\theta) = E\ b'(\sigma(\tau_i,\theta)e_i)e_i\ n^{-\frac{1}{2}}\ \delta\ \sigma(\tau_i,\theta)\ . \tag{3.13}$$

Now $T_{1,i}(\delta,\Delta,0)$ is independent of i, so for all δ and Δ

$$\sum_{i=1}^{n} A^*_{in}(\Delta)\ T_{1,i}(\delta,\Delta,0) = 0\ . \tag{3.14}$$

By (1.2) and (3.13),

$$\sup_{|\delta|+||\Delta||\le 1} |T_{1,i}(\delta,\Delta,\theta) - T_{1,i}(\delta,\Delta,0)| = o(n^{-\frac{1}{2}})\ . \tag{3.15}$$

By (3.14) and (3.15),

$$\sup_{|\delta|+||\Delta||\le 1} |n^{-\frac{1}{2}} \sum_{i=1}^{n} A^*_{in}(\Delta)\ T_{1,i}(\delta,\Delta,\theta)| = o(1)\ . \tag{3.16}$$

By (3.10), (3.11), (3.12), and (3.16)

$$\sup_{|\delta|+||\Delta||\le 1} n^{-\frac{1}{2}} \left| \sum_{i=1}^{n} A^*_{in}(\Delta)(E\, B_i(\delta,\Delta,\theta) - E\, B_i(0,0,\theta)) \right| = o(1). \quad (3.17)$$

For all Δ, $n^{-\frac{1}{2}} \sum_{i=1}^{n} (A^*_{in}(\Delta) - A^*_{in}(0))\, E\, B_i(0,0,0) = 0$, and since $\sup_{i\le n} |E(B_{in}(0,0,\theta) - B_{in}(0,0,0))| = o(1)$, $\sup_{|\Delta|\le 1} |n^{-\frac{1}{2}} \sum_{i=1}^{n} (A^*_{in}(\Delta) - A^*_{in}(0))(E\, B_{in}(0,0,\theta) - E\, B_{in}(0,0,0))| = o(1)$, so that $\sup_{||\Delta||+|\delta|\le 1} n^{-\frac{1}{2}} \sum_{i=1}^{n} (A^*_{in}(\Delta) - A^*_{in}(0))\, E\, B_i(0,0,\theta)| = o(1)$, and then by (3.9) and (3.17) the lemma is proven. □

Combining Lemmas 2 and 3 we have:

<u>Lemma 4</u>: For all $R > 0$,

$$\sup_{||\Delta||+|\delta|\le R} |V(\delta,\Delta,\theta)| = o_p(1). \quad (3.18)$$

<u>Lemma 5</u>. For all $R > 0$,

$$\sup_{|\delta|+||\Delta||\le R} \left| n^{-1} \sum_{i=1}^{n} (B^*_{in}(\Delta,\delta,\theta))^2 - \mathrm{Var}\, b(e_1) \right| = o_p(1) \quad (3.19)$$

and

$$\left| n^{-1} \sum_{i=1}^{n} (A^*_i(\Delta))^2 - n^{-1} \sum_{i=1}^{n} (a(\tau_i) - a_{.}(\tau))^2 \right| = o(1). \quad (3.20)$$

<u>Proof</u>: By the law of large numbers and the boundedness of b

$$n^{-1} \sum_{i=1}^{n} (B^*_i(0,0,0))^2 \xrightarrow{P} \mathrm{Var}\, b(e_i). \quad (3.21)$$

Since b is Lipschitz, for some $K > 0$

$$\sup_{|\delta|+||\Delta||\le 1} |B_{in}(\Delta,\delta,\theta) - B_{in}(0,0,0)| \le n^{-\frac{1}{2}} K(|e_i| + |c_{in}|).$$

Thus, $E(\sup_{|\delta|+||\Delta||\le 1} |B_{in}(\Delta,\delta,\theta) - B_{in}(0,0,0)|) = o(1)$, which implies that

$$\sup_{|\delta|+||\Delta||\le 1} |n^{-1} \sum_{i=1}^{n} (B^*_{in}(\Delta,\delta,\theta))^2 - (B^*_{in}(0,0,0))^2 = o_p(1). \quad (3.22)$$

Now (3.21) and (3.22) imply (3.19). It is easy to proof (3.20). □

<u>Lemma 6</u>: $U(0,0,\theta) - U(0,0,0) = n^{-\frac{1}{2}} E(b'(e_1)e_1)\theta \sum_{i=1}^{n} (a(\tau_i) - a_{.}(\tau))^2 + o(1)$.

<u>Proof</u>: Since $A^*_{in}(0)$ is bounded uniformly in i and n and $\theta \to 0$, we have $E\, A^*_{in}(0)(B_{in}(0,0,\theta) - B_{in}(0,0,0)) = o(1)$, uniformly in i and n. Therefore $\mathrm{Var}(U(0,0,\theta) - U(0,0,0)) = o(1)$. By a Taylor expansion and (1.2), (2.9), and (2.10), for some ν_i

$$|E(B_{in}(0,0,\theta) - B_{in}(0,0,\theta)) - E(b'(e_1)e_1)(\sigma(\tau_i,\theta) - 1)|$$

$$\le |E\, b''(\nu_i)(\sigma(\tau_i,\theta) - 1)^2 + O(P(B_{in})\,|\sigma(\tau_i,\theta) - 1|\,)$$

$$= O(n^{-3/4}) ,$$

where $B_{in} = \{|e_i - \xi| \le n^{-\frac{1}{4}}$ or $|e_i + \xi| \le n^{-\frac{1}{4}}\}$.

Since $\sup_n \sup_{i\le n} |A^*_{in}(0)| < \infty$, the proof is easily finished. □

<u>Proof of theorem 1</u>: First note that since $r_i = e_i\, \sigma(\tau_i,\theta) - (\hat{\beta}_n-\beta)'\, c_{in}$,

$$A_b = U(0,0,\theta)(n^{-1} \sum_{i=1}^{n} (a(\tau_i) - a_{.}(\tau))^2 \mathrm{Var}\, b(e_1))^{-\frac{1}{2}} + o_p(1)$$

for (2.10) allows us to substitute $\delta = n^{\frac{1}{2}} (\hat{\sigma}_n^{-1} - 1)$ and $\Delta = n^{\frac{1}{2}} (\hat{\beta}_n-\beta)$ into (3.18)-(3.20). By (2.1), (2.4), (2.7), and the Lindeberg Central Limit Theorem,

$$U(0,0,0)(n^{-1} \sum_{i=1}^{n} (a(\tau_i) - a_{.}(\tau))^2 \mathrm{Var}\, b(e,1)^{-\frac{1}{2}} \overset{\mathcal{D}}{\to} N(0,1) .$$

Theorem 1 follows from lemma 6. □

References

Anscombe, F.J. (1961). Examination of Residuals. In *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* (J. Neyman, ed.), pp. 1-36. University of California Press. Berkeley and Los Angeles, California.

Bickel, Peter J. (1978). Using Residuals Robustly I: Tests for Heteroscedasticity, Nonlinearity. *Ann. Statist. 6*, 266-291.

Huber, Peter J. (1964). Robust Estimation of a Location Parameter. *Ann. Math. Statist. 35*, 73-101.

Huber, Peter J. (1973). Robust Regression: Asymptotics, Conjectures and Monte Carlo. *Ann. Statist. 5*, 799-821.

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