The Annals of Statistics
1981, Vol. 9, No. 2, 368–380

ASYMPTOTICALLY OPTIMAL TESTS FOR HETEROSCEDASTICITY IN THE GENERAL LINEAR MODEL

BY THOMAS HAMMERSTROM

University of Michigan

Some plausible models for violations of homoscedasticity and linearity in the general linear model have been proposed by Tukey and Anscombe. Bickel has provided robust tests for such violations. In this paper Bickel's tests are shown to be asymptotically optimal as well.

1. Introduction. The first step in analyzing data from a regression or ANOVA experiment is usually to fit a linear parametric model for the means of the observations, assuming the errors are independent $N(0, \sigma^2)$ variables. One then begins to check the validity of the underlying assumptions of the model. Anscombe, Tukey and others have discussed plausible forms of departure from the assumptions and have proposed ways of using the residuals to investigate the validity of the assumptions. Bickel [5], in particular, has proposed tests of the assumptions of homoscedasticity and linearity and examined the robustness of these tests for various distributions of the errors. We will show that Bickel's tests have the additional property of being asymptotically uniformly most powerful.

2. Testing for heteroscedasticity. The first form of departure from the assumptions we will investigate is heteroscedasticity of the errors. In one of the most common forms of this, σ_i increases approximately linearly with τ_i. Here σ_i is the standard deviation and τ_i the expectation of the ith observation Y_i.

Anscombe [2] has proposed an explicit parametric model for this situation. We have independent observations $Y_1, \ldots, Y_n$ with

$$Y_i \sim N(\tau_i, \sigma^2 a_i(\theta)) \tag{2.1}$$

where, for each n, the parameters $(\theta, \log \sigma^2, \tau_1, \ldots, \tau_n)$ lie in a compact subset of R^{n+2}.

We specifically assume $\tau = C\beta$, C being a known $n \times p$ design matrix and that σ^2, θ vary independently of β. Without loss of generality we can take $C^T C = I_{pxp}$. Let

$$\Gamma = C\,C^T$$

$$\bar{\Gamma} = I_{n\times n} - \Gamma.$$

We further assume

$$\text{(A)}\quad \begin{aligned} &\sup\nolimits_{1\le i<\infty} |\tau_i| \text{ and } |\log \sigma^2| \text{ are bounded by } M < \infty;\\ &\lim\nolimits_{n\to\infty} \tau. = \lim\nolimits_{n\to\infty} \frac{1}{n} \sum{}' \tau_i \text{ exists;}\\ &\lim\nolimits_{n\to\infty} \frac{1}{n} \sum (\tau_i - \tau.)^2 \text{ exists and is positive.}\end{aligned}$$

Received February 1979; revised December 1979.
AMS 1970 subject classification. Primary 62F20, 62J10.
Key words and phrases. General linear model, heteroscedasticity, locally asymptotically normal, asymptotically uniformly most powerful.

Finally we assume that $a_i(\theta)$ is twice continuously differentiable in θ, that

$$a_i(0) = 1, \qquad \frac{\partial a_i(0)}{\partial \theta} = \tau_i$$

and that

$$\frac{\partial^2 a_i(\theta)}{\partial \theta^2} = O(1) \qquad \text{as} \quad \theta \to 0.$$

Here we write $A_i(\theta) = O(\theta^k)$ as $\theta \to 0$ to mean that for some $\epsilon > 0$, there exists $M < \infty$ with

$$\sup \left\{ \frac{|A_i(\theta)|}{|\theta|^k} : 1 \le i < \infty, |\theta| < \epsilon \right\} < M.$$

We want to test H: $\theta = 0$ (errors are homoscedastic) against one of the alternatives K: $\theta > 0$ or K: $\theta \neq 0$.

For this model, Bickel has proposed the following test statistic, modifying an estimate suggested by Anscombe.

Let $\mathbf{t} = \Gamma \mathbf{Y}$ = L.S.E. for τ and let $\mathbf{r} = \mathbf{Y} - \mathbf{t} = \bar{\Gamma}\gamma$ = vector of residuals. Let $t. = (1/n)\sum t_i$ and $\bar{t} = (1/(n-p)) \sum \bar{\gamma}_{ii} t_i$. (We will always use dot and bar to denote these two different averages for any n-vector.)

Let
$$s^2 = \frac{1}{n-p} \sum r_i^2.$$

This s^2 is just the standard estimate of σ^2 used in least squares regression. The test statistic is

$$h = \frac{\frac{1}{n} \sum r_i^2 (t_i - \bar{t})}{s^2 \left[\frac{1}{n} \sum\sum \bar{\gamma}_{ij}^2 (t_i - \bar{t})(t_j - \bar{t}) \right]^{1/2}}.$$

The tests will have critical functions ψ_n,

$$\psi_n = 1 \qquad \text{if} \quad \sqrt{n}h > z(1-\alpha)$$
$$= 0 \qquad \text{otherwise}$$

where $z(1-\alpha) = (1-\alpha)$ quantile of the standard normal variable.

We will show

THEOREM 1. *Given the model of* (2.1), *suppose Assumption* A *holds. Then* $(p/n) \to 0$ *is a sufficient condition for the tests with critical functions* $\{\psi_n\}$ *to be locally asymptotically* UMP *level* α *for* H: $\theta = 0$ *against* K: $\theta > 0$.

The proof may be outlined as follows. We consider the special case in which all the expectations, τ_i, are known and find an upper bound for the power function of any test of H against K in this special case. We then compare this upper bound with the actual asymptotic power function of the tests $\{\psi_n\}$, as computed by Bickel and show that they coincide. If the tests, $\{\psi_n\}$, give the best possible asymptotic power even when the τ_i are known, a fortiori, they must give the best power when the τ_i are unknown.

3. Proof of Theorem 1. Let $\gamma = (\theta, \log \sigma^2)$. Suppose $\gamma_0 = (0, \log \sigma_0^2) \in \{0\} \times \mathbb{R}$ and that $\gamma_n = \{(\theta_n, \log \sigma_n^2)\}$ is a sequence with $\|\gamma_0 - \gamma_n\| \le M/\sqrt{n}$ for some finite M. Let $(Y_1, \ldots, Y_n)$ be distributed as in model (2.1) with $\tau_1, \ldots, \tau_n$ as specified constants and let $\{Q_\gamma^{(n)}\}$ be the family of their joint distributions. In order to find an upper bound on the asymptotic power function of any test of H: $\theta = 0$ we will establish that the family $\{Q_\gamma^{(n)}\}$ is uniformly locally asymptotically normal at every $\gamma_0 \in \{0\} \times \mathbb{R}$, as defined in Definition A.1 of the appendix, and then appeal to a theorem of LeCam and Çibişov, Theorem A in the appendix.

Fixing such a γ_0 we need to find a sequence of statistics $\{\mathbf{U}^{(n)}\} = \{(\theta, \log\hat{\sigma}^2)^{(n)}\}$ and a positive definite matrix $\sum(\gamma_0)$ such that the densities, $\{q_n\}$, of $\{Q^{(n)}\}$ satisfy the equation

$$q_n(\mathbf{y}^{(n)}, \gamma) = C_n(\mathbf{y}^{(n)})\exp\left\{-\frac{n}{2}(\mathbf{U}^{(n)} - \gamma)'\textstyle\sum(\gamma_0)(\mathbf{U}^{(n)} - \gamma) + Z_n(\gamma, \gamma_0)\right\} \tag{3.1}$$

with

(i) $\sqrt{n}(\mathbf{U}^{(n)} - \gamma_0) \to N(\mathbf{O}, \sum^{-1}(\gamma_0))$ in law under $Q_{\gamma_0}^{(n)}$;

(ii) $\sup\{|Z_n(\gamma, \gamma_0)|: \|\gamma - \gamma_0\| \le M/\sqrt{n}\} \to 0$ *in probability under* $Q_{\gamma_0}^{(n)}$, for all $M < \infty$.

We find these $\{\mathbf{U}^{(n)}\}$ by taking one step estimates for γ_n, using the method of scoring. Initial estimates for the method of scoring are obtained by solving the likelihood equations under the assumption $\theta = 0$.

The calculations of both the likelihood equations and the Fisher information matrix require the log likelihood function,

$$l_n(\gamma) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum\left\{\frac{(y_i - \tau_i)^2}{\sigma^2 a_i(\theta)} + \log a_i(\theta)\right\} \tag{3.2}$$

where $a_i(\theta)^{-1} = 1 - \theta\tau_i + O(\theta^2)$.

The gradient is given by the equations

$$\begin{aligned} \frac{\partial l_n}{\partial\theta} &= -\frac{1}{2}\sum\left\{\frac{\partial\log a_i(\theta)}{\partial\theta} + \frac{(y_i - \tau_i)^2}{\sigma^2}\frac{\partial a_i(\theta)^{-1}}{\partial\theta}\right\} \\ \frac{\partial l_n}{\partial\log\sigma^2} &= -\frac{n}{2} + \frac{1}{2}\sum\frac{(y_i - \tau_i)^2}{\sigma^2 a_i(\theta)}. \end{aligned} \tag{3.3}$$

The Hessian is given by the equations

$$\begin{aligned} \frac{\partial^2 l_n}{\partial\theta^2} &= -\frac{1}{2}\sum\left\{\frac{\partial^2\log a_i(\theta)}{\partial\theta^2} + \frac{(y_i - \tau_i)^2}{\sigma^2}\frac{\partial^2 a_i(\theta)^{-1}}{\partial\theta^2}\right\} \\ \frac{\partial^2 l_n}{\partial\theta\,\partial\log\sigma^2} &= \frac{1}{2}\sum\frac{(y_i - \tau_i)^2}{\sigma^2}\frac{\partial a_i(\theta)^{-1}}{\partial\theta}, \\ \frac{\partial^2 l_n}{\partial(\log\sigma^2)^2} &= -\frac{1}{2}\sum\frac{(y_i - \tau_i)^2}{\sigma^2 a_i(\theta)}. \end{aligned} \tag{3.4}$$

Notice that both $(\partial a_i(\theta)^{-1}/\partial\theta)$ and $-(\partial\log a_i(\theta)/\partial\theta)$ are of the form $-\tau_i + O(\theta)$.

Taking expectations under $Q_\gamma^{(n)}$ we have that the Fisher information matrix is

$$I_n(\gamma) = \frac{1}{2}\begin{bmatrix} \sum(\tau_i^2 + O(\theta)) & \sum(\tau_i + O(\theta)) \\ \sum(\tau_i + O(\theta)) & n \end{bmatrix}. \tag{3.5}$$

Observe that the family $(1/n)I_n(\gamma)$; $n = 1, \ldots$ is equicontinuous in γ for γ in a $\sqrt{n}$ neighborhood of $\{O\} \times \mathbb{R}$.

We now take as our initial estimates

$$\hat{\theta}_0 = 0$$

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum(Y_i - \tau_i)^2.$$

The one step estimates of the method of scoring are given by the equation

$$\hat{\gamma}_n = \begin{bmatrix} \hat{\theta}_n \\ \log\hat{\sigma}_n^2 \end{bmatrix} = \begin{bmatrix} \hat{\theta}_0 \\ \log\hat{\sigma}_0^2 \end{bmatrix} + I_n^{-1}(\hat{\theta}_0, \log\hat{\sigma}_0^2)\,\mathrm{grad}\, l_n(\hat{\theta}_0, \log\hat{\sigma}_0^2). \tag{3.6}$$

We will be able to establish the ULAN property for $\{Q_\gamma^{(n)}\}$ by taking a Taylor series expansion for $l_n(\gamma)$ about $\hat\gamma$ The two-term Taylor expansion of $l_n(\gamma)$ is given by

$$\begin{aligned} l_n(\gamma) &= l_n(\hat\gamma) + (\gamma-\hat\gamma)' \operatorname{grad} l_n(\hat\gamma) + \frac{1}{2}(\gamma-\hat\gamma)' H_n(\gamma^*)(\gamma-\hat\gamma) \\ &= l_n(\hat\gamma) - \frac{n}{2}(\hat\gamma-\gamma)'\textstyle\sum(\gamma_0)(\hat\gamma-\gamma) + \sqrt{n}(\gamma-\hat\gamma)'\dfrac{1}{\sqrt{n}}\operatorname{grad} l_n(\hat\gamma) \\ &\quad + \frac{n}{2}(\gamma-\hat\gamma)'\left[\textstyle\sum(\gamma_0) + \dfrac{1}{n}H_n(\gamma^*)\right](\gamma-\hat\gamma) \end{aligned} \tag{3.7}$$

where H_n = Hessian of l_n and γ^* is on the line segment from γ to $\hat\gamma$.

The matix $\sum(\gamma_0)$ is defined as $\lim(1/n)I_n(\gamma_0)$. Using equation (3.5) we have

$$\textstyle\sum(\gamma_0) = \dfrac{1}{2}\lim\begin{bmatrix} \dfrac{1}{n}\sum\tau_i^2 & \dfrac{1}{n}\sum\tau_i \\ \dfrac{1}{n}\sum\tau_i & 1 \end{bmatrix} \tag{3.8}$$

To establish part (i) of the ULAN condition we need to show $\sqrt{n}(\hat\gamma_n - \gamma_0) \to N(0, \sum^{-1}(\gamma_0))$ in law under $Q_{\gamma_0}^{(n)}$.

To establish part (ii) we need only show that the supremum over the set of $\{\gamma_n\}$ with $\|\gamma_n - \gamma_0\| \le M/\sqrt{n}$ of the absolute value of each of the last two terms on the right-hand side of (3.7) tends to zero in probability under $Q_{\gamma_0}^{(n)}$. We will accomplish this by showing each of the terms consists of a product of factors bounded in probability with a factor converging to zero in probability.

We begin by observing that for $\{\gamma_n\}$ with $\|\gamma_n - \gamma_0\| \le M/\sqrt{n}$, $\sqrt{n}\,\|\hat\gamma_0 - \gamma_n\|$ is bounded in probability under $Q_{\gamma_0}^{(n)}$.

Observe that the family $(1/n)I_n(\cdot)$ is equicontinuous and nonsingular in a neighborhood of γ_0.

We next prove two lemmas that show that the families $(1/n)H_n(\gamma_n)$ and $(1/n)H_n(\hat\gamma_n)$ are equicontinuous with high probability under $Q_{\gamma_0}^{(n)}$ for nonrandom sequences $\{\gamma_n\}$, $\gamma_n \to \gamma_0$, and for random sequences $\{\hat\gamma_n\}$, $\hat\gamma_n \to \gamma_0$ in probability under $Q_{\gamma_0}^{(n)}$. Specifically we prove

LEMMA 3.1. *Each entry of the Hessian matrix, $H_n(\gamma)$, of the log likelihood function can be written as $\sum F(Y_i, \tau_i, \gamma)$ where F is continuous and the family*

$$\{|F(y_i, \tau_i, \hat\gamma_n)|\, q_n(\mathbf{y}^{(n)}, \gamma_0) \colon 1 \le i \le n,\ 1 \le n\}$$

is uniformly integrable with respect to Lebesque measures on $\mathbb{R}^n$.

Also, for any compact neighborhood, L, of γ_0, the family

$$\{|F(y_i, \tau_i, \hat\gamma_n)|\, I(\hat\gamma_n \in L) q_n(\mathbf{y}^{(n)}, \gamma_0) \colon 1 \le i \le n,\ 1 \le n\}$$

is uniformly integrable with respect to Lebesgue measure. Here I is the indicator function.

PROOF. By equations (3.4), F is one of the three functions

$$\left[\frac{\partial^2 \log a_i(\theta)}{\partial\theta^2} + \frac{(Y_i-\tau_i)^2}{\sigma^2}\frac{\partial^2 a_i(\theta)^{-1}}{\partial\theta^2}\right], \quad \frac{(Y_i-\tau_i)^2}{\sigma^2}\frac{\partial a_i(\theta)^{-1}}{\partial\theta^2}, \quad \frac{(Y_i-\tau_i)^2}{\sigma^2}a_i(\theta)^{-1}.$$

These are clearly continuous, by the assumptions on a_i, and hence are uniformly continuous and bounded on compact sets.

If $\gamma = (\theta, \log\sigma^2)$ is within a compact neighborhood, L, of $\gamma_0 = (0, \log\sigma_0^2)$, then $\sup_{\gamma\in L} F(y_i, \tau_i, \gamma)$ is dominated by a bounded factor times $(y_i - \tau_i)^2$ plus a bounded term.

The boundedness of the coefficients follows from the form of $a_i(\theta)$ and the fact that $\sup_i |\tau_i| < \infty$.

Since q_n is the product of $N(\tau_i, \sigma^2 a_i(\theta))$ densities it follows, again using $\sup_i |\tau_i| < \infty$, that

$$\lim_{a\to\infty} \sup_n \sup_{1\le i\le n} \int \cdots \int_{|y_i|>a} \sup_{\gamma\in L} F(y_i, \tau_i, \gamma) q_n(\mathbf{y}^{(n)}, \gamma_0) dy = 0. \tag{3.9}$$

Since $|F(y_i, \tau_i, \hat\gamma_n)| I(\hat\gamma_n \in L) \le \sup_{\gamma\in L} |F(y_i, \tau_i, \gamma)|$ the second uniform integrability result asserted in the statement of the lemma holds.

To prove the first uniform integrability result, one can repeat the same argument with $\{\hat\gamma_n\}$ replaced by $\{\gamma_n\}$. Since $\{\gamma_n\}$ is a deterministic sequence the indicator factor, $I(\gamma_n \in L)$, $\equiv 1$ for n large enough. □

For the next lemma we again let $\{\gamma_n\}$ be any nonrandom sequence with $\gamma_n \to \gamma_0$ and $\{\hat\gamma_n\}$ be any random sequence with $\hat\gamma_n \to \gamma_0$ in $Q_{\gamma_0}^{(n)}$ probability.

LEMMA 3.2. *The quantities* $|(1/n)H_n(\gamma_n) - (1/n)H_n(\gamma_0)|$, $|(1/n)H_n(\hat\gamma_n) - (1/n)H_n(\gamma_0)|$ *both converge to zero in probability under* $Q_{\gamma_0}^{(n)}$.

PROOF. Each entry in H_n can be written as $\sum F(Y_i, \tau_i, \gamma)$ with F uniformly continuous on compacts.

Let $\epsilon > 0$ be given. Let L = unit ball in $\mathbb{R}^2$ with center γ_0. By Lemma 3.1 we can choose a compact $K \subset \mathbb{R}$ so that

(i) $\sup_i E_{Q_{\gamma_0}} |F(Y_i, \tau_i, \gamma_n)| I(Y_i \notin K) < \epsilon$;

(ii) $\sup_i E_{Q_{\gamma_0}} |F(Y_i, \tau_i, \gamma_n)| I(Y_i \notin K, \gamma_n \in L) < \epsilon$.

Choose $\delta > 0$ so that $Y_i \in K$ and $\|\gamma_n - \gamma_0\| < \delta$, imply

$$\gamma_n \in L \quad \text{and} \quad |F(Y_i, \tau_i, \gamma_n) - F(Y_i, \tau_i, \gamma_0)| < \epsilon.$$

Finally, choose N so that $n > N$ implies $\|\gamma_n - \gamma_0\| < \delta$, $Q_{\gamma_0}^{(n)}(\|\hat\gamma_n - \gamma_0\| < \delta) > 1 - \epsilon$. Then the entry of interest in $|(1/n)H_n(\hat\gamma_n) - (1/n)H_n(\gamma_0)|$ can be written as

$$\begin{aligned}
&\left|\frac{1}{n}\sum F(Y_i, \tau_i, \hat\gamma_n) - F(Y_i, \tau_i, \hat\gamma_0)\right| \\
&\quad \le \frac{1}{n}\sum_{Y_i\in K} |F(Y_i, \tau_i, \hat\gamma_n) - F(Y_i, \tau_i, \gamma_0)| \\
&\qquad + \frac{1}{n}\sum |F(Y_i, \tau_i, \hat\gamma_n)| I(Y_i \notin K) I(\hat\gamma_n \notin L) \\
&\qquad + \frac{1}{n}\sum |F(Y_i, \tau_i, \gamma_0)| I(Y_i \notin K) \\
&\qquad + \frac{1}{n}\sum |F(Y_i, \tau_i, \hat\gamma_n')| I(Y_i \notin K) I(\hat\gamma_n \in L).
\end{aligned} \tag{3.10}$$

The first term on the right-hand side of (3.10) is less than ϵ with probability greater than $Q_{\gamma_0}^{(n)}(\|\hat\gamma_n - \gamma_0\| < \delta) > 1 - \epsilon$.

The second term on the right is greater than ϵ with probability less than $Q_{\gamma_0}^{(n)}(\hat\gamma_n \notin L) < \epsilon$.

By conditions (i) and (ii) the third and fourth terms on the right have expectations less than ϵ. By Markov's inequality they are greater than a with probability less than ϵ/a, in particular they are greater than $\sqrt{\epsilon}$ with probability less than $\sqrt{\epsilon}$.

Thus all the terms on the right-hand side of (3.10) can be made arbitrarily small with arbitrarily high probability. Since there are only three distinct entries in the matrix, all of the entries can be made small with high probability.

The proof with γ_n in place of $\hat{\gamma}_n$ is even easier because $\{\gamma_n\}$ is a nonrandom sequence. □

We now point out that $H_n(\gamma_0)$ is a sum of independent terms whose variances are bounded so the law of large numbers for nonidentically distributed random variables, Theorems 5.1.1 and 5.1.2 of Chung [7], implies $|(1/n)H_n(\gamma_0) + (1/n)I_n(\gamma_0)| \to 0$ in probability under $Q_{\gamma_0}^{(n)}$. By definition $|(1/n)I_n(\gamma_0) - \sum(\gamma_0)| \to 0$.

Next we remark that from Theorem 4.2 in Chapter 2 of Roussas [15] we know that

$$\frac{1}{\sqrt{n}} \operatorname{grad} l_n(\gamma_0) \to N(0, \textstyle\sum(\gamma_0))$$

in law under $Q_{\gamma_0}^{(n)}$. Then we may observe that

$$\frac{1}{\sqrt{n}} \operatorname{grad} l_n(\hat{\gamma}_0) = \frac{1}{\sqrt{n}} \operatorname{grad} l_n(\gamma_0) + \sqrt{n}(\hat{\gamma}_0 - \gamma_0) \int_0^1 \frac{1}{n} H_n(t\hat{\gamma}_0 + (1-t)\gamma_0)dt.$$

As just remarked $\frac{1}{\sqrt{n}} \operatorname{grad} l_n(\gamma_0)$ converges in law under $Q_{\gamma_0}^{(n)}$. Clearly $\sqrt{n}(\hat{\gamma}_0 - \gamma_0)$ is bounded in probability under $Q_{\gamma_0}^{(n)}$. By Lemma 3.2, the integral converges in probability to $-\sum(\gamma_0)$ under $Q_{\gamma_0}^{(n)}$. We can therefore conclude that $\left\{\frac{1}{\sqrt{n}} \operatorname{grad} l_n(\hat{\gamma}_0)\right\}$ is bounded in probability under $Q_{\gamma_0}^{(n)}$.

From equation (3.6) we get

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) = \sqrt{n}(\hat{\gamma}_0 - \gamma_0) + \frac{1}{\sqrt{n}} \operatorname{grad} l_n(\hat{\gamma}_0) n I_n^{-1}(\hat{\gamma}_0).$$

We just saw that $\left\{\frac{1}{\sqrt{n}} \operatorname{grad} l_n(\hat{\gamma}_0)\right\}$ was bounded in $Q_{\gamma_0}^{(n)}$ probability, as was $\sqrt{n}(\hat{\gamma}_0 - \gamma_0)$. By the equicontinuity of $\{nI_n^{-1}(\cdot)\}$, $\{nI_n^{-1}(\hat{\gamma}_0)\}$ is also bounded in probability under $Q_{\gamma_0}^{(n)}$ and hence $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ is likewise.

The last preliminary calculation we need for establishment of the ULAN condition is that

$$\frac{1}{\sqrt{n}} \operatorname{grad} l_n(\hat{\gamma}_n) \to 0 \text{ in probability}$$

under $Q_{\gamma_0}^{(n)}$.

To see this we observe

$$\frac{1}{\sqrt{n}} \operatorname{grad} l_n(\hat{\gamma}_n) = \frac{1}{\sqrt{n}} \operatorname{grad} l_n(\hat{\gamma}_0) + \frac{1}{\sqrt{n}} (\hat{\gamma}_n - \hat{\gamma}_0) \int_0^1 H_n(\hat{\gamma}_n t + (1+t)\hat{\gamma}_0)dt \tag{3.11}$$

$$= \sqrt{n}(\hat{\gamma}_n - \hat{\gamma}_0) \int_0^1 \frac{1}{n} [H_n(\hat{\gamma}_n t + (1-t)\hat{\gamma}_0) + I_n(\hat{\gamma}_0)]dt$$

$$\sqrt{n}(\hat{\gamma}_n - \hat{\gamma}_0) = \sqrt{n}(\hat{\gamma}_n - \hat{\gamma}_0) + \sqrt{n}(\gamma_n - \hat{\gamma}_0) \tag{3.12}$$

and both terms on the right-hand side of (3.12) are bounded in $Q_{\gamma_0}^{(n)}$ probability. It follows from Lemma 3.2 and the equicontinuity of $(1/n)I_n(\cdot)$ that the integral in (3.11) tends to zero in $Q_{\gamma_0}^{(n)}$ probability. Thus

$$\frac{1}{\sqrt{n}} \operatorname{grad} l_n(\hat{\gamma}_n) \to 0 \text{ in probability under } Q_{\gamma_0}^{(n)}.$$

With this result, we have demonstrated that both of the last two terms on the right-hand side of equation (3.7) tend to zero in $Q_{\gamma_0}^{(n)}$ probability when $\|\gamma_n - \gamma_0\| \le M/\sqrt{n}$. Thus, part (ii) of the ULAN condition holds.

Finally, to show part (i) we observe that

$$\frac{1}{\sqrt{n}} \operatorname{grad} l_n(\hat{\gamma}_n) = \frac{1}{\sqrt{n}} \operatorname{grad} l_n(\gamma_0) + \sqrt{n}(\hat{\gamma}_n - \gamma_0) \int_0^1 \frac{1}{n} H_n(\gamma_t)dt$$

where $\gamma_t = t\hat{\gamma}_n + (1-t)\gamma_0$ so

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) = -\left[\int_0^1 \frac{1}{n} H_n(\gamma_t)dt\right]^{-1} \cdot \left[\frac{1}{\sqrt{n}} \operatorname{grad} l_n(\gamma_0) - \frac{1}{\sqrt{n}} \operatorname{grad} l_n(\hat{\gamma}_n)\right]. \tag{3.13}$$

Using the convergences established above and Theorem 2.4.2 of Roussas, the right-hand side of (3.13) converges in law under $Q_{\gamma_0}^{(n)}$ to $N(0, \sum(\gamma_0)^{-1})$, as required by part (i) of the ULAN condition.

We therefore have that the sequence $\{Q_\gamma^{(n)}\}$ is ULAN at any $\gamma_0 \in \{0\} \times \mathbb{R}$ and, by Theorem A in the appendix

$$\sup_M \limsup_n \sup\left\{\beta_n(\gamma, \psi_n^*) - \left[1 - \Phi\left(z(1-\alpha) - \frac{\sqrt{n}\theta}{\sqrt{\sigma^{11}(\gamma_0)}}\right)\right] : 0 < \theta < Mn^{-1/2}\right\} \le 0$$

for any asymptotically level α tests $\{\psi_n^*\}$.

But Bickel [5], Theorem 1.1, shows that, provided $p/n \to 0$, the asymptotic power function is

$$\beta(\gamma, \psi_n) = 1 - \Phi\left(z(1-\alpha) - \sqrt{}\theta\left[\frac{1}{n}\sum(\tau_i - \bar{\tau})^2\right]^{1/2}\right) + o(1)$$

so the tests $\{\psi_n\}$ are surely asymptotically level α. To complete the proof it suffices to show

$$\left[\frac{1}{n}\sum(\tau_i - \bar{\tau})^2\right]^{1/2} - [\sigma^{11}(\gamma_0)]^{1/2} \to 0.$$

But from equation (3.8) for $\sum(\hat{\gamma}_0)$ we see that $\sigma^{11}(\hat{\gamma}_0) = \lim_n (1/n)\sum(\tau_i - \tau.)^2$. Since $p/n \to 0$, $(1/n)\sum(\tau_i - \tau.)^2 - (1/n)\sum(\tau_i - \bar{\tau})^2 \to 0$.

Hence we have completed the proof of Theorem 1. □

4. The case of nonnormal errors. Bickel has also showed how the test statistic in Section 2 may be modified so as to achieve desirable robustness properties when the errors are merely symmetrically distributed about zero. Our model now is

$$Y_i = \tau_i + \epsilon_i, \qquad 1 \le i \le n \tag{4.1}$$

with ϵ_i being independent with density

$$\frac{1}{\sigma a_i(\theta)} f\left(\frac{x}{\sigma a_i(\theta)}\right).$$

We retain the same assumptions about the vector of means, τ, and the function $a_i(\theta)$ as in Section 2. The density $f(x)$ is assumed to be known and to satisfy

(A.1) $f(-x) = f(x)$, i.e., f is symmetric about 0

(A.2) $\log f(x)$ exists, i.e., f is positive everywhere

(A.3) $f'(x), f''(x), f'''(x)$ exist and $\lim_{|x|\to\infty} f(x) = 0$

(A.4) $\left|\frac{f'}{f}(x)\right|, \left|x\left(\frac{f'}{f}\right)'(x)\right|$ and $\left|x^2\left(\frac{f'}{f}\right)''(x)\right|$ are all $\leq M < \infty$ for all x,

$$\int x\frac{f'}{f}\left(\frac{x}{\sigma}\right)f(x)dx < \infty \text{ and } \int x^2\left(\frac{f'}{f}\right)'\left(\frac{x}{\sigma}\right)f(x)dx < \infty \text{ for any fixed } \sigma > 0$$

(A.5) $F\left(\frac{1}{2}\right) = \frac{3}{4}, F\left(-\frac{1}{2}\right) = \frac{1}{4}$ where $F(x)$

$= \int_{-\infty}^{x} f(u)du$, i.e. the scale of f is chosen so that the interquartile range is 1.

Assumption (A.5) can always be satisfied. Notice that for normal errors, $(f'/f)(x) = -x$ so that (A.4) fails. Thus, the model in Section 2 is not a special case of the model in this section. Also, the double exponential density violates assumption (A.3). I do not know whether an analogue of Theorem 2 below holds for double exponential errors. However, both the logistic and the Cauchy densities are easily seen to satisfy these assumptions.

We are thus considering distributions $\{P_n\}$ for $\{\mathbf{Y}^{(n)} = (Y_1, \ldots, Y_n)\}$ that satisfy the model of line (4.1) and assumptions (A.1)–(A.5) and also

(A.6) $|\sqrt{n}\theta| \leq M$; $\sup_n \sup_{1\leq i\leq n} |\tau_i| \leq M$; $|\log\sigma| \leq M$; $\lim \frac{1}{n}\sum \tau_i$ exists;

$$\lim \frac{1}{n}\sum (\tau_i - \tau.)^2 \text{ exists and is greater than zero.}$$

Assumption (A.6) is the same as Assumption (A) of the previous section.

Let $b(x) = x(f'(x)/f(x) + 1$. If the means, $\{\tau_i\}$, and the scale factor, σ, were all known it would follow from results in, e.g., Hájek and Sidák [10] that the optimal test would be based on the statistic

$$\frac{\frac{1}{n}\sum b\left(\frac{Y_i - \tau_i}{\sigma}\right)(\tau_i - \tau.)}{\left[\frac{1}{n}\sum(\tau_i - \tau.)^2\right]^{1/2}\left[\frac{1}{n-p}\sum\left(b\left(\frac{Y_i - \tau_i}{\sigma}\right) - b.\right)^2\right]^{1/2}}$$

where $b. = (1/n)\sum b((Y_i - \tau_i)/\sigma)$.

Bickel has suggested using the test statistic obtained from this expression by replacing the $\{\tau_i\}$ and the σ by suitable robust estimates. We would like to prove his conjecture that this will lead to an asymptotically UMP test.

Since we are no longer assuming that the ϵ_i are normal, the least squares estimates are no longer the appropriate estimates for τ_i. Instead we assume that we have robust estimates, $\hat{\beta}_k$, for β_k, $1 \leq k \leq p$, with

(A.7) $$\sum_{k=1}^{p} (\hat{\beta}_k - \beta_k)^2 = O_{P_n}(p).$$

For example, Huber's M estimates can be used for the $\{\hat{\beta}_k\}$, as Yohai and Maronna (19) show in their Theorem 2.2.

Let $\mathbf{t} = C\hat{\boldsymbol{\beta}}$ be the estimate of $\boldsymbol{\tau}$ corresponding to the estimate, $\hat{\boldsymbol{\beta}}$, of $\boldsymbol{\beta}$. Let $\mathbf{r} = \mathbf{Y} - \mathbf{t}$ be the vector of residuals. Assume we have

(A.8) s is a $\sqrt{n}$ consistent estimate of σ.

Finally we assume

(A.9) $p/\sqrt{n} \to 0$ as p and $n \to \infty$.

Our proposed test statistics are now

$$h = \frac{\frac{1}{n}\sum b\left(\frac{r_i}{s}\right)(t_i - t.)}{\left[\frac{1}{n}\sum (t_i - t.)^2\right]^{1/2}\left[\frac{1}{n-p}\sum\left(b\left(\frac{r_i}{s}\right) - b.\right)^2\right]^{1/2}} \tag{4.2}$$

where $b. = (1/n)\sum b(r_i/s)$ and our proposed tests have critical functions ψ_n,

$$\begin{aligned}\psi_n &= 1 \quad \text{if} \quad \sqrt{n}h > z(1-\alpha)\\ &= 0 \quad \text{otherwise.}\end{aligned}$$

We will prove

THEOREM 2. *Given the model of line* (4.1), *assume that* (A.1) – (A.9) *hold. Then the tests with critical functions* $\{\psi_n\}$ *are locally asymptotically* UMP *level* α *for* H: $\theta = 0$ *against* K: $\theta > 0$.

5. Proof of Theorem 2. We prove Theorem 2 in the same way we proved Theorem 1. We establish that when the expectations, τ_i, are given constants the family of distributions of $(Y_1, \cdots, Y_n)$ are uniformly locally asymptotically normal at any $(\theta_0, \log\sigma_0) = (0, \log\sigma)$. Consequently, any test of H: $\theta = 0$ has asymptotic power function with an upper bound given by the theorem of Le Cam and Çibişov. Since Bickel's proposed test achieves this upper bound even without knowledge of the τ_i, his test is asymptotically uniformly most powerful.

As in Section 3, let $\boldsymbol{\gamma} = (\theta, \log\sigma)$ and let $\{Q_\gamma^{(n)}\}$ be the family of distributions of $(Y_1, \cdots, Y_n)$ with $\tau_1, \cdots, \tau_n$ as specified constants.

To prove Theorem 2 we first establish that $\{Q_\gamma^{(n)}\}$ is ULAN at every point of the line $\theta = 0$ with the quadratic form $\sum(\boldsymbol{\gamma})$ = the limit of $n^{-1}\times$ Fisher information matrix at $\boldsymbol{\gamma}$, that is,

$$\textstyle\sum(\boldsymbol{\gamma}) = \lim_n A\begin{bmatrix}(\tau^2). & \tau./\sigma\\ \tau./\sigma & 1/\sigma\end{bmatrix} \tag{5.1}$$

where $A = -\int xb'(x)f(x)\,dx$.

Assumption (A.6) guarantees that $\sum(\boldsymbol{\gamma})$ is nonsingular.

Once we have verified this, we may conclude that the asymptotic power of any sequence $\{\psi_n^*\}$ of asymptotically level α tests for H against K in the model with the distributions $\{Q_\gamma^{(n)}\}$ obey the bound given by Theorem A of the appendix. That is,

$$\beta(\boldsymbol{\gamma}, \psi_n^*) \le 1 - \Phi\left(z(1-\alpha) - \frac{\sqrt{n}\theta}{\sqrt{\sigma^{11}(\boldsymbol{\gamma})}}\right) + o(1) \tag{5.2}$$

where $[\sigma^{ij}(\boldsymbol{\gamma})] = \sum^{-1}(\boldsymbol{\gamma})$.

The ULAN property is verified by modifying the proof in Section 3 where necessary to account for the fact that the tails of the density $f(x)$ fall off less rapidly than those of the normal density. This will manifest itself in our choosing different initial estimates for the method of scoring, in different forms for the gradient of the log likelihood and the Fisher information matrix and in a different proof of the analogue of Lemma 3.1.

For the first part of the proof we fix a $\boldsymbol{\gamma}_0 = (0, \log\sigma_0) \in \{0\} \times \mathbb{R}$ and show that $\{Q_\gamma^{(n)}\}$ are ULAN at $\boldsymbol{\gamma}_0$. To do so we need to find a sequence of statistics $\{\mathbf{U}^{(n)}\} = \{(\hat\theta, \log\hat\sigma)^{(n)}\}$ and a positive definite matrix $\sum(\boldsymbol{\gamma}_0)$ such that the densities, q_n, of $\{Q_\gamma^{(n)}\}$ satisfy the equation (3.1).

We find these $\{\mathbf{U}^{(n)}\}$ by taking one step estimates for $\boldsymbol{\gamma}$, using the method of scoring. Again we obtain our initial estimates for the method of scoring by solving the likelihood equation for σ under the assumption that $\theta = 0$.

For the calculation of the maximum likelihood estimate for σ and of the Fisher information matrix we will need the log likelihood function, $l_n(\boldsymbol{\gamma}) = \log q_n(y^{(n)}, \boldsymbol{\gamma})$ and its gradient and Hessian.

The log likelihood is given by

$$l_n(\boldsymbol{\gamma}) = -n \log \sigma - \sum \log a_i(\theta) + \sum \log f\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right). \tag{5.3}$$

Its gradient is given by

$$\begin{aligned} \frac{\partial l_n}{\partial \theta} &= -\sum b\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right)\frac{\partial \log a_i(\theta)}{\partial \theta} \\ \frac{\partial l_n}{\partial \log \sigma} &= -\sum b\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right). \end{aligned} \tag{5.4}$$

The Hessian matrix is given by

$$\begin{aligned} \frac{\partial^2 l_n}{\partial \theta^2} &= \sum b'\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right)\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right)\left(\frac{\partial \log a_i(\theta)}{\partial \theta}\right)^2 - \sum b\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right)\left(\frac{\partial^2 \log a_i(\theta)}{\partial \theta^2}\right). \\ \frac{\partial^2 l_n}{\partial \theta \partial \log \sigma} &= \sum b'\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right)\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right)\frac{\partial \log a_i(\theta)}{\partial \theta}, \\ \frac{\partial^2 l_n}{\partial (\log \sigma)^2} &= \sum b'\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right)\left(\frac{y_i - t_i}{\sigma a_i(\theta)}\right). \end{aligned} \tag{5.5}$$

Notice that $(\partial \log a_i(\theta))/\partial \theta = \tau_i + O(\theta)$ as $\theta \to 0$.

Taking expectations under $Q_\gamma^{(n)}$ we find that the Fisher information matrix is

$$I_n(\boldsymbol{\gamma}) = A \begin{bmatrix} \sum\left(\frac{\partial \log a_i(\theta)}{\partial \theta}\right)^2 & \sum \frac{\partial \log a_i(\theta)}{\partial \theta} \\ \sum \frac{\partial \log a_i(\theta)}{\partial \theta} & n \end{bmatrix} \tag{5.6}$$

where $A = -\int xb'(x)f(x)dx$. Here we have used assumption (A.3) to integrate $\int [x(f'/f)(x) + 1]f(x)dx$ by parts and conclude that $Eb(Y_i - \tau_i)/\sigma a_i(\theta)) = 0$.

Since $(\partial \log a_i(\theta))/\partial \theta = \tau_i + O(\theta)$ we find

$$I_n(\boldsymbol{\gamma}) = A \begin{bmatrix} \sum [\tau_i^2 + O(\theta)] & \sum [\tau_i + O(\theta)] \\ \sum [\tau_i + O(\theta)] & n \end{bmatrix}. \tag{5.7}$$

Observe also that $\{(1/n)I_n(\boldsymbol{\gamma})\}$ is an equicontinuous family of functions of $\boldsymbol{\gamma}$ for $\boldsymbol{\gamma}$ in a $n^{1/2}$ neighborhood of a fixed $\boldsymbol{\gamma}_0$ in $\{0\} \times \mathbb{R}$, provided the $\{\tau_i\}$ satisfy assumption (A.6).

We now select initial estimates for the method of scoring. We take $\hat{\theta}_0 = 0$; $\hat{\sigma}_0 =$ interquartile range of $\{Y_i - \tau_i\}$.

Next we define the one-step estimates by the equation

$$\hat{\boldsymbol{\gamma}} = \begin{bmatrix} \hat{\theta} \\ \log \hat{\theta} \end{bmatrix} = \begin{bmatrix} \hat{\theta}_0 \\ \log \hat{\sigma}_0 \end{bmatrix} + I_n^{-1}(\hat{\theta}_0, \log \hat{\sigma}_0)\operatorname{grad} l_n(\hat{\theta}_0, \log \hat{\sigma}_0). \tag{5.8}$$

To establish that $\{Q_\gamma^{(n)}\}$ is ULAN at $\boldsymbol{\gamma}_0$ we need to show that the initial estimates, $\hat{\boldsymbol{\gamma}}_0$, are locally $\sqrt{n}$ consistent and that the conclusions of Lemma 3.1 still hold with error densities satisfying assumptions (A.1) to (A.5).

The first of these claims follows from showing that

$$\sqrt{n}(\hat{\sigma}_0 - \sigma_n(1 + \theta_n \tau.))$$

converges in law under $Q_{\gamma_n}^{(n)}$ to a normal distribution with mean zero. Since $\hat{\sigma}_0$ is the interquartile range of the $\{\epsilon_i\}$, this is a straight-forward consequence of a theorem of Weiss [18] on the joint distribution of the order statistics of nonidentically distributed samples.

LEMMA 5.1. *Each entry of the Hessian matrix, $H_n(\boldsymbol{\gamma})$, of* $\log q_n(\mathbf{Y}^{(n)}, \boldsymbol{\gamma})$ *can be written as* $\sum F(Y_i, \tau_i, \boldsymbol{\gamma})$ *where F is continuous and the family* $\{|F(y_i, \tau_i, \boldsymbol{\gamma}_n)|q_n(\mathbf{y}^{(n)}, \boldsymbol{\gamma}_0); 1 \le i \le n, 1 \le n\}$ *is uniformly integrable with respect to Lebesgue measure on* $\mathbb{R}^n$ *for any determinate sequence* $\{\boldsymbol{\gamma}_n\}$ *with* $\boldsymbol{\gamma}_n \to \boldsymbol{\gamma}_0$.

Also for any compact neighborhood, L, of $\boldsymbol{\gamma}_0$, and any random sequence $\hat{\boldsymbol{\gamma}}_n$ converging to $\boldsymbol{\gamma}_0$ in $Q_{\gamma_0}^{(n)}$ probability, the family

$$\{|F(y_i, \tau_i, \hat{\boldsymbol{\gamma}}_n)|I(\hat{\boldsymbol{\gamma}}_n \in L)q_n(\mathbf{y}^{(n)}, \boldsymbol{\gamma}_0); 1 \le i \le n, 1 \le n\}$$

is uniformly integrable with respect to Lebesgue measure.

PROOF. The entries of H_n are of the form $\sum F(Y_i, \tau_i, \gamma)$ with F being one of the three functions

$$\left[b'\left(\frac{Y_i - \tau_i}{\sigma a_i(\theta)}\right)\left(\frac{Y_i - \tau_i}{\sigma a_i(\theta)}\right)\left(\frac{\partial \log a_i(\theta)}{\partial \theta}\right)^2 - b\left(\frac{Y_i - \tau_i}{\sigma a_i(\theta)}\right)\left(\frac{\partial^2 \log a_i(\theta)}{\partial \theta^2}\right)\right],$$
$$b'\left(\frac{Y_i - \tau_i}{\sigma a_i(\theta)}\right)\left(\frac{Y_i - \tau_i}{\sigma a_i(\theta)}\right)\left(\frac{\partial \log a_i(\theta)}{\partial \theta}\right),$$
$$b'\left(\frac{Y_i - \tau_i}{\sigma a_i(\theta)}\right)\left(\frac{Y_i - \tau_i}{\sigma a_i(\theta)}\right).$$

By our assumptions on $a_i(\theta)$ and f these are all continuous and hence uniformly continuous and bounded on compact sets. By assumption (A.4) $b'(x) = x(f'/f)'(x) + (f'/f)(x)$ is bounded. Hence, if $\boldsymbol{\gamma} = (\theta, \log \sigma)$ is within a compact neighborhood, L, of $\boldsymbol{\gamma}_0 = (0, \log \sigma_0)$ then

$$\sup_{\boldsymbol{\gamma} \in L} |F(y_i, \tau_i, \boldsymbol{\gamma})|$$

is dominated by

$$\sup_{\boldsymbol{\gamma} \in L}\left\{K_1 \left| b'\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right)\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right)\right| + K_2 \left| b\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right)\right|\right\}$$

for some constants $K_1, K_2 < \infty$.
We have

$$\int_{|y_i|>a} \sup_{\boldsymbol{\gamma} \in L}|F(y_i, \tau_i, \boldsymbol{\gamma})|q_n(\mathbf{y}, \boldsymbol{\gamma}_0)\,dy$$
$$\le K \int_{|y_i|>a}\left\{\sup_{\boldsymbol{\gamma} \in L}\left| b'\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right)\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right)\right| + \sup_{\boldsymbol{\gamma} \in L}\left| b\left(\frac{y_i - \tau_i}{\sigma a_i(\theta)}\right)\right|\right\}:$$
$$f\left(\frac{y_i - \tau_i}{\sigma_0}\right) dy_i \cdot \prod_{j \ne i} \int f\left(\frac{y_j - \tau_j}{\sigma_0}\right) dy_j.$$

Since we both know that for any fixed $\sigma > 0$, $\int (x/\sigma)b'(x/\sigma)f(x)\,dx$ and $\int b(x/\sigma)f(x)\,dx$ are both finite and since $\sup|\tau_i| < \infty$, the right-hand side tends to zero as $a \to \infty$.

The proof of the lemma now concludes exactly as that of Lemma 3.1

$$F(y_i, \tau_i, \hat{\boldsymbol{\gamma}})|\mathrm{I}(\hat{\boldsymbol{\gamma}} \in L) \le \sup_{\boldsymbol{\gamma} \in L}|F(y_i, \tau_i, \boldsymbol{\gamma})|$$

and for the deterministic sequence $\{\boldsymbol{\gamma}_n\}$, $\boldsymbol{\gamma}_n \in L$ for all sufficiently large n. □

Lemma 3.2 now holds for the model of this chapter as well since only the truth of Lemma 3.1 was used in the proof of Lemma 3.2.

Likewise, the same proof used in Section 3 now implies $\sqrt{n}\,(\hat{\boldsymbol{\gamma}}_n - \hat{\boldsymbol{\gamma}}_0)$ converges in law to $N(0, \sum(\boldsymbol{\gamma}_0)^{-1})$ under $Q_{\gamma_0}^{(n)}$ and that the suprema over $\|\boldsymbol{\gamma}_n - \boldsymbol{\gamma}_0\| \le M/\sqrt{n}$ of the absolute values of the last two terms of the Taylor expansion (3.7) of the log likelihood $l_n(\boldsymbol{\gamma})$ both tend to zero under $Q_{\gamma_0}^{(n)}$. Consequently $\{Q_{\gamma}^{(n)}\}$ is ULAN at $\boldsymbol{\gamma}_0$.

By the theorem of Le Cam and Çibişov we now know that for any sequence, $\{\psi_n^*\}$, of asymptotically level α tests, the asymptotic power function obeys the relation

$$\beta_n(\gamma, \psi_n^*) \le 1 - \Phi\left(z(1-\alpha) - \frac{\sqrt{n}\theta}{\sqrt{\sigma^{11}(\boldsymbol{\gamma}_0)}}\right) + o(1). \tag{5.9}$$

To complete the proof of Theorem 3 it is now only necessary to show that the proposed tests $\{\psi_n\}$ have asymptotic power function achieving the upper bound of (5.9), which will imply that they are asymptotically UMP level α.

But by Theorem 3.1 of Bickel, the asymptotic power function of the $\{\psi_n\}$ is given by

$$\beta_n(\boldsymbol{\gamma}, \psi_n) = 1 - \Phi\left(z(1-\alpha) - n^{1/2}\theta\left[\frac{1}{n}\sum(\tau_\iota - \bar{\tau})^2\right]^{1/2}\left[\int b^2(x)f(x)dx\right] + o(1)\right).$$

Incidentally, the first three bounds of assumption (A.4) are only used at this point in the proof. Theorem 2 will now follow if we show that

$$\frac{1}{\sigma^{11}(\boldsymbol{\gamma}_0)} = \frac{1}{n}(\tau_\iota - \bar{\tau})^2 \int b^2(x)f(x)dx + o(1) \tag{5.10}$$

If we compute the inverse of $\sum(\boldsymbol{\gamma}_0)$ from equation (3.14) we find

$$\frac{1}{\sigma^{11}(\boldsymbol{\gamma}_0)} = \frac{1}{n}\sum(\tau_\iota - \tau.)^2\left[-\int xb'(x)f(x)dx\right].$$

But $(1/n)\sum(\tau_\iota - \bar{\tau})^2 - (1/n)\sum(\tau_\iota - \tau.)^2 \to 0$ if $(p^2/n) \to 0$. $-\int xb'(x)f(x)dx = \int b^2(x)f(x)dx$ follows by substituting the definition of $b(x)$ and integrating by parts.

Thus equation (5.10) is verified and Theorem 2 is established.

APPENDIX

The following definitions and theorems of Hájek, Le Cam, and Çibişov were needed in this paper. We reproduce the statements of the essential results here.

Let $X^{(n)} = (X_1, \cdots, X_n)$ be a sequence of random variables or random vectors with $P_{\eta}^{(n)}$ being the distribution of $X^{(n)}$. The parameter $\boldsymbol{\eta} \in \Theta \subset \mathbb{R}^k$.

DEFINITION A.1. $\{P_{\eta}^{(n)}\}$ is uniformly locally asymptotically normal (ULAN) at $\boldsymbol{\eta}_0$ if there exist

(1) statistics $\mathbf{U}^{(n)}$ (possibly depending on $\boldsymbol{\eta}_0$);
(2) coefficients $c_n(x^{(n)})$;
(3) a positive definite matrix $\sum(\boldsymbol{\eta}_0)$;
(4) constants $\delta_n \downarrow 0$, such that, with respect to some measure $\mu^{(n)}$, $P_{\eta}^{(n)}$ has the density

$$P_n(x^{(n)}, \boldsymbol{\eta}) = c_n(x^{(n)})\exp\left[-\frac{\delta_n^{-2}}{2}(\mathbf{U}^{(n)} - \boldsymbol{\eta})'\sum(\boldsymbol{\mu}_0)(\mathbf{U}^{(n)} - \boldsymbol{\eta}) + Z_n(\boldsymbol{\eta}, \boldsymbol{\eta}_0)\right]$$

with

(i) $\delta_n^{-1}(\mathbf{U}^{(n)} - \boldsymbol{\eta}_0) \to N(0\sum^{-1}(\boldsymbol{\eta}_0))$ in law under $P_{\eta_0}^{(n)}$;
(ii) $\sup\{|Z_n(\boldsymbol{\eta}, \boldsymbol{\eta}_0)|: |\boldsymbol{\eta} - \boldsymbol{\eta}_0| \le M\delta_n\} \to 0$ in probability under $P_{\eta_0}^{(n)}$ for all $M < \infty$.

Consider the case of $\boldsymbol{\eta} \in \Theta \subset \mathbb{R}^k$, $k \ge 1$ and $\boldsymbol{\eta}_0 = (\eta_{10}, \cdots, \eta_{K0})'$ Interior Θ. We wish to test H: $\eta_1 = \eta_{10}$ against K: $\eta_1 > \eta_{10}$.

For any sequence $\{\psi_n\}$ of critical functions of tests of H against K, the asymptotic power functions are defined as $\beta(\boldsymbol{\eta}, \psi_n) = E_{\eta}\psi_n(X^{(n)})$.

DEFINITION A.2. The sequence of tests $\{\psi_n\}$ is asymptotically level a for H if

$$\sup_M \lim\sup_n \sup\{\beta(\boldsymbol{\eta}, \psi_n): \eta_1 = n_{10} \text{ and } |\boldsymbol{\eta}| \le M_{\delta_n}\} \le \alpha.$$

THEOREM A. *If* $\{P_\eta^{(n)}\}$ *is* ULAN *at* $\boldsymbol{\eta}_0 \in$ *Interior* Θ *then*

$$\sup_M \lim\sup_n \sup\left\{\beta(\boldsymbol{\eta}, \psi_n) - \left[1 - \phi\left(z(1-\alpha) - \delta_n^{-1}\frac{(\eta_1 - \eta_{10})}{\sqrt{\sigma^{11}(\eta_0)}}\right)\right]: \right.$$
$$\left. \eta_{10} < \eta_1 < \eta_{10} + M\delta_n, |\boldsymbol{\eta} - \boldsymbol{\eta}_0| \le M\delta_n\right\} \le 0$$

where $(\sigma^{ij}(\boldsymbol{\eta}_0)) = \sum (\boldsymbol{\eta}_0)^{-1}$ *for all sequences of asymptotically level* α *tests for* H, $\{\psi_n\}$.

Definition A.1 is essentially due to Hájek [9], Assumption A, page 144. Theorem A was stated by Le Cam [12], Proposition 2, page 155, for all sequences of asymptotically level α similar tests but the proof he gave in fact made no use of similarity. The theorem in the i.i.d. case is both stated and proved without mention of similarity by Çibişov [7], Theorem 9.1, page 40.

REFERENCES

[1] ANDREWS, D. A., et al. (1972). *Robust Estimates of Location: Survey and Advances.* Princeton Univ. Press.
[2] ANSCOMBE, F. J. (1961). Examination of residuals. In *Proc. 4th Berkeley Symp. Math. Statist. Prob.* I 1-36. Univ. California Press, Berkeley.
[3] ANSCOMBE, F. J. AND TUKEY, J. W. (1963). Analysis of residuals. *Technometrics* **5** 141-160.
[4] BICKEL, P. J. (1975). One step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428-434.
[5] BICKEL, P. J. (1978). Using residuals robustly I: Tests for heteroscedasticity, nonlinearity and nonadditivity. *Ann. Statist.* **6** 266-291.
[6] CHUNG, K. L. (1974). *A Course in Probability Theory.* Academic, New York.
[7] ÇIBIŞOV, D. M. (1973). Asymptotic expansions for Neyman's $C(\alpha)$ test. In *Proc. Second Japan-USSR Symp. Probability. Springer Lecture Notes in Math.* **330.**
[8] HÁJEK, J. (1970). Local asymptotic minimax and admissibility in estimation. *Proc. 6th Berkeley Symp. Math. Statist. Prob.* **I.** Univ. California, Berkeley.
[9] HÁJEK, J. (1970). Limiting properties of likelihoods and inference. In *Symposium on the Foundations of Statistical Inference.* (V. Godambe, ed.). Univ. Waterloo, Waterloo.
[10] HÁJEK, J. AND SIDÁK, Z. (1967). *Theory of Rank Tests.* Academic, New York.
[11] HUBER, P. J. (1973). Robust regression. *Ann. Statist.* **1** 799-821.
[12] LE CAM, L. (1956). On the asymptotic theory of estimation and testing hypotheses. In *Proc. 3rd Berkeley Symp. Math. Statist. Prob.* **1** 129-158. Univ. California, Berkeley.
[13] LE CAM, L. (1960). *Locally Asymptotically Normal Families of Distributions.* Univ. California, Berkeley.
[14] LE CAM, L. (1968). *Theorie Asymptotique de la Decision Statistique.* Univ. Montreal, Montreal.
[15] ROUSSAS, G. (1972). *Contiguity of Probability Measures: Some Applications in Statistics.* Cambridge Univ., Cambridge.
[16] WALD, A. (1941). Asymptotically most powerful tests of statistical hypotheses. *Ann. Math. Statist.* **12** 1-19.
[17] WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426-482.
[18] WEISS, L. (1969). The asymptotic distribution of quantiles from mixed samples. Sankhyā Ser. *A* **31** 313-318.
[19] YOHAI, V. AND MARONNA, R. (1979). Asymptotic behavior of *M*-estimates for the linear model. *Ann. Statist.* **7** 258-268.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
1447 MASON HALL
419 SOUTH STATE ST.
ANN ARBOR, MI 48109