The Annals of Statistics
1978, Vol. 6, No. 2, 266–291

USING RESIDUALS ROBUSTLY I: TESTS FOR HETEROSCEDASTICITY, NONLINEARITY[1]

BY P. J. BICKEL

University of California, Berkeley

We study the asymptotic power functions of tests for heteroscedasticity and nonlinearity in the linear model which were proposed by Anscombe and introduce and study some competitors robust against gross errors.

Introduction. In the past few years a variety of methods has been proposed for estimating the parameters of a linear model which are less sensitive to departures from normality of the error distribution than the classical least squares estimates. See Huber [13], Jaeckel [14] and Bickel [5] for various approaches.

As Anscombe, Tukey [2], [3], [4] and others have stressed the fitting of a linear model is often only a first tentative step in the analysis of structured data. After the parameter values of the model have been fitted by least squares these authors and most practicing statisticians advocate an analysis of how well the *model* fits with an eye to common specific departures such as nonlinearity, heteroscedasticity and dependence. Most such analyses are graphical and rather informal. However, Anscombe [2] has proposed some simple test statistics for nonlinearity and heteroscedasticity which attempt to formalize this process.

This paper is an investigation of the power of Anscombe's procedures when the error distributions are not normal and a comparison of these procedures with some natural alternative tests which are robust against gross errors. The paper is organized as follows. In Section 1 we introduce Anscombe's model and tests for heteroscedasticity and state and discuss rather general asymptotic properties of these procedures. In Section 2 we do the same for his nonlinearity models and procedures. In Section 3 we introduce robust tests for heteroscedasticity and study their asymptotic theory under less general conditions than those of Section 1. In Section 4 we do the same for the nonlinearity problem. Section 5 is an appendix containing the technical results needed for the proofs of the various theorems of the paper.

1. Testing for heteroscedasticity: Anscombe procedures. Our point of departure is the general linear model, $\mathscr{L}$, in the form

$$Y_i = \tau_i + \varepsilon_i\,, \qquad i = 1, \cdots, n\,, \tag{1.1}$$

Received May 1975; revised October 1976.

[1] This work was completed while the author was visiting Imperial College, London. This paper was prepared with support of U.S. Office of Naval Research Contract No. N00014-69-A-0200-1038/NR042-036 and National Science Foundation Grant No. GP-38485.

AMS 1970 *subject classifications.* Primary 62G20, 62G35; Secondary 62J05, 62J10.

Key words and phrases. Heteroscedasticity, nonlinearity, robust, asymptotic tests, linear model, one degree of freedom for nonadditivity.

where

$$\tau_i = \textstyle\sum_{j=1}^{p} c_{ij}\beta_j\,; \tag{1.2}$$

the β_j are free unknown parameters, the c_{ij} are known constants and the ε_i independent identically distributed random errors with common df F and density f. If F is the $\mathscr{N}(0, \sigma^2)$ df we have the classical normal linear model which we shall refer to as $\mathscr{NLM}$. It is convenient at this point to rewrite (1.1) and (1.2) in matrix notation and introduce the usual estimates. Let

$$C = ||c_{ij}||_{n\times p}$$

$\mathbf{Y} = (Y_1, \cdots, Y_n)'$, $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_n)'$, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)'$. Then (1.1) and (1.2) become

$$\mathbf{Y} = \boldsymbol{\tau} + \boldsymbol{\varepsilon} \tag{1.3}$$

$$\boldsymbol{\tau} = C\boldsymbol{\beta}\,. \tag{1.4}$$

Without loss of generality suppose C is of rank p. The least squares estimates of $\boldsymbol{\beta}$ which we denote $\hat{\boldsymbol{\beta}}^L = (\hat{\beta}_1{}^L, \cdots, \hat{\beta}_p{}^L)$ are given by

$$\hat{\boldsymbol{\beta}}^L = [C'C]^{-1}C'\mathbf{Y} \tag{1.5}$$

and the least squares fitted values $\mathbf{t}^L = (t_1{}^L, \cdots, t_n{}^L)$ by

$$\mathbf{t}^L = \Gamma\mathbf{Y} \tag{1.6}$$

where by definition,

$$\Gamma = ||\gamma_{ij}||_{n\times n} = C[C'C]^{-1}C'\,. \tag{1.7}$$

The least squares residuals denoted by $\mathbf{r}^L = (r_1{}^L, \cdots, r_n{}^L)$ are, of course, given by

$$\mathbf{r}^L = \mathbf{Y} - \mathbf{t}^L = \bar{\Gamma}\mathbf{Y}\,, \tag{1.8}$$

where by definition,

$$\bar{\Gamma} = ||\bar{\gamma}_{ij}||_{n\times n} = I_{n\times n} - \Gamma\,, \tag{1.9}$$

and finally the residual mean square is defined by

$$s^2 = (n-p)^{-1} \textstyle\sum_{i=1}^{n} [r_i{}^L]^2\,. \tag{1.10}$$

For convenience, since we only consider least squares fits in this section and the next, we shall drop the L superscript where there is no ambiguity. The following standard notation will prove useful. If $a_1, \cdots, a_n$ is any sequence we write

$$a_{\bullet} = n^{-1}\textstyle\sum_{i=1}^{n} a_i\,.$$

Thus

$$s^2 = \frac{n}{n-p}[r^2]_{\bullet}\,.$$

Inhomogeneity of the variances of the errors is a common phenomenon in practice. When such inhomogeneity occurs the variances often seem to depend on the means τ of the observations. Following Anscombe [2] we write such a model as

$$Y_i = \tau_i + \sigma(\tau_i)\varepsilon_i\,, \qquad i = 1, \cdots, n\,, \tag{1.11}$$

where the τ_i, ε_i are as before and σ is a positive function of τ. We suppose that σ can be approximated by a member of a parametric family of curves $\sigma(\tau, \theta)$, θ real such that

$$\sigma(\tau, \theta) = 1 + \theta a(\tau) + o(\theta) \tag{1.12}$$

as $\theta \downarrow 0$ uniformly for bounded τ. In this section we consider the simplest case: $a(\tau) = \tau$. Thus $\theta = 0$ corresponds to homoscedasticity. The simple model $\sigma = e^{\theta\tau}$, of course, satisfies (1.12) as do many others. More general models are discussed in Section 3.

We consider the one and two sided testing problems $H\colon \theta = 0$. For F normal Anscombe [5] proposed an estimate of θ which we call h,

$$h = \textstyle\sum_i r_i^2(t_i - \bar{t})/s^2 \sum_{i,j} \bar{\gamma}_{ij}^2(t_i - \bar{t})(t_j - \bar{t})$$

where

$$\bar{t} = (n - p)^{-1} \textstyle\sum_i \bar{\gamma}_{ii} t_i\,. \tag{1.13}$$

If $\mathscr{N}\mathscr{L}$ holds, Anscombe shows that

$$\operatorname{Var}(h \mid t_1^L, \cdots, t_n^L) = 2(n - p)(n - p + 2)^{-1}[\textstyle\sum_{i,j} \bar{\gamma}_{ij}^2(t_i - \bar{t})(t_j - \bar{t})]^{-1}\,.$$

He suggests that (under conditions somewhat difficult to make precise), $h/[\operatorname{Var}(h \mid t_1, \cdots, t_n)]^{\frac{1}{2}}$ can be referred approximately to a normal table and used for a test of significance of H. The resulting test statistic which we shall call A is given by

$$A = \textstyle\sum_i r_i^2(t_i - \bar{t})/\tilde{\sigma} \tag{1.14}$$

where

$$\tilde{\sigma}^2 = 2(n - p)(n - p + 2)^{-1}s^4 \textstyle\sum_{i,j} \bar{\gamma}_{ij}^2(t_i - \bar{t})(t_j - \bar{t})\,. \tag{1.15}$$

Here is one way of seeing that A is reasonable. Calculate the locally most powerful test statistic for $H\colon \theta = 0$ vs. $K\colon \theta > 0$ for the model (1.11), (1.12) when it is assumed that F is $\mathscr{N}(0, \sigma^2)$ and the τ_i and σ^2 are known. This statistic is proportional to

$$\textstyle\sum_i \tau_i(\varepsilon_i^2 - \sigma^2)\,.$$

If we estimate τ_i by t_i, ε_i by r_i, σ^2 by s^2 and, as is often the case, $\bar{\gamma}_{ii} \equiv 1 - (p/n)$ we arrive at the numerator of A. The denominator $\tilde{\sigma}$ is just an estimate of the conditional standard deviation of the numerator given $\mathbf{t}$ under $\mathscr{N}\mathscr{L}$. The use of $\bar{t}$ rather than $t_.$ is dictated by the asymptotics. See the proof of Lemma 1.1, particularly the centering on the left of (A4), and (1.27).

We shall study the behavior of the numerator and denominator of A and the related statistic A_{χ^2} which we introduce below as the number of observations gets large. Since we are interested in situations such as the additive two way layout with one observation per cell we shall follow Huber's [13] lead and consider sequences of designs in which the number of parameters may grow as well as the number of observations. We will really be dealing with sequences of design matrices C_n of dimension p_n as well as parameters $(\beta_{1n}, \cdots, \beta_{p_n n})$ and let $n \to \infty$. To simplify matters we drop the subscript n unless confusion could ensue.

We need to put various conditions on the quantities defining the various models appearing in this paper. We shall sometimes express these conditions in terms of "bounds" on the design matrices of the models and on the true values of the parameters entering into the model ((L) conditions) and on the error distribution ((F) conditions) and sometimes in terms of rates of convergence to 0 for these quantities. All our convergence in law and convergence in probability theorems will be uniform in the regions specified by these bounds when these are given even though we may not make this uniformity explicit. We use P_θ throughout this section to denote probabilities calculated under the model given by (1.11) and (1.12). P is used for P_0. We shall use M with and without subscripts throughout as generic finite positive constants with the understanding that they may vary from condition to condition.

Here are the conditions for this section. We use the prefix H to indicate those which are special to the heteroscedasticity problem.

L1: $\max_i |\tau_i| \leqq M$,
L2: $p/n \to 0$,
HL3: $n^{-1} \sum_{i=1}^n (\tau_i - \bar{\tau})^2 \geqq M^{-1} > 0$,
L4: $|\theta n^{\frac{1}{2}}| \leqq M$,
F1: F is a distribution symmetric about 0,
F2: $0 < M^{-1} \leqq E\varepsilon_1^6 \leqq M$,
HF3: $M^{-1} \leqq J_2(f) \leqq M$ where

$$J_2(f) = \int_{-\infty}^{\infty} \left(x \frac{f'}{f}(x) + 1\right)^2 f(x)\, dx$$

if F has an absolutely continuous density f with derivative f', and $J_2(F) = \infty$ otherwise.

The asymptotic theory of this section rests on the following proposition whose proof is given in the appendix.

PROPOSITION 1.1. *Suppose that* L1, 2 *and* F1, 2 *hold. Then,*

$$n^{-\frac{1}{2}} \sum_i (t_i - \bar{t}) r_i^2 = n^{-\frac{1}{2}} \sum (\tau_i - \bar{\tau})(\varepsilon_i^2 - E\varepsilon_1^2) + o_p(1) \tag{1.16}$$

$$s^2 = E\varepsilon_1^2 + o_p(1) \tag{1.17}$$

$$\begin{aligned} n^{-1} \sum_{i,j} \gamma_{ij}^2 (t_i - \bar{t})(t_j - \bar{t}) &= n^{-1} \sum_i (t_i - \bar{t})^2 + o_p(1) \\ &= n^{-1} \sum_i (\tau_i - \bar{\tau})^2 + o_p(1) \end{aligned} \tag{1.18}$$

$$(n-p)^{-1} \sum_i (r_i^2 - [r^2]_\bullet)^2 = \operatorname{Var} \varepsilon_1^2 + o_p(1)\,. \tag{1.19}$$

As a consequence we find that if the assumptions of the proposition and HL3 hold, then an application of the Lindeberg-Feller theorem yields

$$P_0[A \geqq z] = 1 - \Phi(z(2/\mathrm{Var}\,\varepsilon_1^2)^{\frac{1}{2}}) + o(1)\,. \tag{1.20}$$

Thus,

(i) if F is normal or more generally the kurtosis of F is 0, using normal critical values for the one and two sided tests based on A is appropriate,

(ii) if the kurtosis of F does not vanish, these tests do not have robustness of validity and, in particular, for long tailed (leptokurtic) distributions have a greater probability of type I error than under $\mathscr{N}\mathscr{L}$.

This is in agreement with Box's [9] findings for the F test for equality of scale of two populations.

Remark (ii) and (1.16) lead us (as it did Box) to construct the following "studentized" version of A which we call A_{x^2}:

$$A_{x^2} = \sum_i (t_i - \bar{t}) r_i^2 / \tilde{\sigma}_{x^2} \tag{1.21}$$

where

$$\tilde{\sigma}_{x^2}^2 = \sum_i (t_i - \bar{t})^2 (n-p)^{-1} \sum_i (r_i^2 - [r^2]_\cdot)^2\,. \tag{1.22}$$

Clearly by (1.16), (1.18), and (1.19) if L1, 2, HL3 and F1, 2 hold,

$$A_{x^2} = (\tfrac{1}{2}\,\mathrm{Var}\,\varepsilon_1^2)^{-\frac{1}{2}} A + o_p(1) \tag{1.23}$$

has a limiting $\mathscr{N}(0, 1)$ distribution under $\mathscr{L}$, and can be used to construct valid tests.

We can investigate the power of tests based on A_{x^2} using Proposition 1.1 and the following proposition whose proof is given in the appendix.

PROPOSITION 1.2. *Suppose* L1 *and* L4 *and* HF3 *hold. Then the measures* $P_\theta \mathbf{Y}^{-1}$ *and* $P_0 \mathbf{Y}^{-1}$ *are contiguous and*

$$\log \frac{dP_\theta \mathbf{Y}^{-1}}{dP_0 \mathbf{Y}^{-1}}(\mathbf{Y}) = -\theta \sum_{i=1}^n \tau_i \left(1 + \varepsilon_i \frac{f'}{f}(\varepsilon_i)\right) - \frac{\theta^2}{2} J_2(f) \sum_{i=1}^n \tau_i^2 + o_p(1)\,. \tag{1.24}$$

Then:

THEOREM 1.1. *Suppose* L1, 2, HL3, L4, F1, 2, HF3 *hold. Then,*

$$P_\theta[A_{x^2} \geqq z] = 1 - \Phi(z - \Delta_{x^2}) + o(1) \tag{1.25}$$

where

$$\Delta_{x^2}(\theta, n) = 2\theta\left[\sum_{i=1}^n (\tau_i - \tau_\cdot)^2\right]^{\frac{1}{2}} E\varepsilon_1^2 / [\mathrm{Var}\,\varepsilon_1^2]^{\frac{1}{2}}\,.$$

PROOF. By Proposition 1.1, under L1, 2, HL3, F1, 2,

$$A_{x^2} = \sum_i (\tau_i - \bar{\tau})(\varepsilon_i^2 - E\varepsilon_1^2) / \left[\sum_i (\tau_i - \bar{\tau})^2\, \mathrm{Var}\,\varepsilon_1^2\right]^{\frac{1}{2}} + o_p(1)\,. \tag{1.26}$$

Since

$$|\hat{\tau} - \tau_{\bullet}| \leqq M(n-p)^{-1}\{p + \textstyle\sum_i \gamma_{ii}\} = O(pn^{-1}) \tag{1.27}$$

we may (using L1, 2, HL3, and F2) replace $\hat{\tau}$ by $\tau_{\bullet}$ on the right-hand side of (1.26).

Let

$$W_{i1} = (\tau_i - \tau_{\bullet})(\varepsilon_i^2 - E(\varepsilon_i^2))/[\textstyle\sum_i (\tau_i - \tau_{\bullet})^2 \operatorname{Var} \varepsilon_1^2]^{\frac{1}{2}}$$
$$W_{i2} = -\theta\tau_i\left(1 + \varepsilon_i \frac{f'}{f}(\varepsilon_i)\right).$$

In view of L1, L3, and L4 we may suppose without loss of generality that $\theta^2 \sum_i \tau_i^2$ and $\sigma^2 \sum_i (\tau_i - \tau_{\bullet})^2$ converge to nonzero limits. Then we can apply the vector Lindeberg–Feller theorem to conclude that, under $\mathscr{L}$, $\sum_{i=1}^n (W_{i1}, W_{i2})$ has a limiting normal distribution with covariance

$$\lim_n - \theta[\textstyle\sum (\tau_i - \tau_{\bullet})^2/\operatorname{Var} \varepsilon_1^2]^{\frac{1}{2}} \int_{-\infty}^{\infty} \chi^2\left(1 + x\frac{f'}{f}(x)\right) f_0(x)\,dx = \lim_n \Delta_{x2}(\theta, n),$$

after an integration by parts. In view of (1.26) and (1.24) A_{x2} and $\log dP_\theta \mathbf{Y}^{-1}/dP_0 \mathbf{Y}^{-1}$ have the same limiting joint distribution. The theorem now follows by Le Cam's third lemma ([11], page 208). □

Statistical implications of the theorem.

(1) The one and two sided tests based on A_{x2} are asymptotically unbiased under the assumptions of the theorem.

(2) If F is normal, A and A_{x2} have equivalent power behavior by (1.23) and Proposition 1.2. Work in progress at Berkeley suggests that if F is normal these statistics are in a suitable sense asymptotically best among asymptotically unbiased tests (e.g., for p fixed in the sense of Neyman [15]).

(3) The power of the tests based on A_{x2} is low if F is leptokurtic since Δ_{x2} is inversely proportional to (kurtosis)$^{\frac{1}{2}}$.

Comments on regularity conditions.

(1) The theorem as stated has minimum conditions on the designs, $p/n \to 0$, but excessive symmetry and moment conditions on F. An examination of the proof of Lemma 1.1 given in the appendix reveals that we can replace F1 by the minimal condition

F1$'$: $E\varepsilon_i = 0$

and have the conclusion of Theorem 1.1 still hold provided that we replace L2 by

$$\frac{\gamma p}{n^{\frac{1}{2}}} \to 0. \tag{1.28}$$

This last condition is not onerous. In the balanced case $\gamma_{ii} \equiv p/n$ it corresponds to $p^4/n^3 \to 0$. If we replace F2 by the natural minimal condition,

F2$'$: $M^{-1} \leqq E\varepsilon_1^4 \leqq M$

and (1.27) holds we can show that all assertions of Proposition 1.1 hold save (1.19). I do not know whether reduction to F2$'$ is possible or (1.28) can be dispensed with. A sketch of the argument needed for the assertions of this comment is in the appendix.

2. Testing for nonlinearity: Anscombe–Tukey tests. Nonlinearity and nonadditivity are two different kinds of departures from $\mathscr{L}$ which are difficult to distinguish from each other. By nonlinearity we mean that the $E(Y_i)$ are nonlinear functions of the parameters of interest. By nonadditivity, in accordance with Tukey and Anscombe, we mean that the Y's themselves do not follow a linear model but that there is a nonlinear transformation T such that the $T(Y_i)$ do follow $\mathscr{L}$. In this section we consider some special nonlinear models. For a special case see Andrews [1]. Nonadditivity will be treated in a later paper. Let

$$Y_i = s(\tau_i) + \varepsilon_i\,, \qquad i = 1, \cdots, n\,, \tag{2.1}$$

where the τ_i are given in (1.2) and the ε_i are as before. For example, suppose our linear model is an additive two way layout with one observation per cell. Thus,

$$Y_{jk} = \mu + \alpha_j + \beta_k + \varepsilon_{jk}\,, \qquad \alpha_\bullet = \beta_\bullet = 0,\ 1 \leqq j \leqq J,\ 1 \leqq k \leqq K\,. \tag{2.2}$$

If s is quadratic then as μ, α_j, β_k vary (2.1) sweeps out a family of models of a form related to one given by Scheffé [16],

$$E(Y_{jk}) = \mu + \tilde{\alpha}_j + \tilde{\beta}_k + C\alpha_j\beta_k\,, \qquad \tilde{\alpha}_\bullet = \tilde{\beta}_\bullet = 0\,.$$

As in the previous section we introduce a parameter θ and suppose that

$$s(\tau, \theta) = \tau + \theta a(\tau) + o(\theta) \tag{2.3}$$

as $\theta \downarrow 0$ uniformly on compact sets of τ. *For this section we shall consider the simplest case: a is quadratic with leading coefficient* 1.

In this case if there is an additive main effect and the observations are normally distributed it is natural to use a test statistic discussed by Anscombe [2] for a related nonadditive model. This statistic is just

$$B_x = \textstyle\sum_i t_i^2 r_i / [\sum_{i,j} \bar{\gamma}_{ij} t_i^2 t_j^2]^{\frac{1}{2}} s$$

where s^2 and the t_i, r_i are as in Section 1. If $\theta = 0$ and F is $\mathscr{N}(0, \sigma^2)$ the distribution of $B_x s$ is $\mathscr{N}(0, \sigma^2)$ and thus critical values for B can be approximated by normal critical values. In fact, if we define

$$a(\mathbf{t}) = (a(t_1), \cdots, a(t_n))$$

and

$$\tilde{s}^2 = (n - p - 1)^{-1}[(n - p)s^2 - (a(\mathbf{t})\mathbf{r})^2 / a(\mathbf{t})\bar{\Gamma} a(\mathbf{t})'] \tag{2.4}$$

then

$$\tilde{B}_x = B_x \frac{s}{\tilde{s}}$$

is just Tukey's one degree of freedom for nonadditivity and has, under $\mathscr{N}\mathscr{L}$, a t distribution with $n - p - 1$ degrees of freedom (see, e.g., Scheffé [16], Problem 4.19). B_x can be motivated on the same grounds as A, as a studentized estimate of the LMP test statistic for $H: \theta = 0$ vs. $K: \theta > 0$ under the assumption F normal and τ_i known.

We study B_x and $\tilde{B}_x$ using the same large sample framework and conditions as in the previous section. Of course P_θ now denotes probabilities under the model given by (2.1) and (2.3). We shall need the following new conditions. The prefix N indicates they are special to the nonlinearity problem.

NL3: $\sum_{i,j} \tilde{\gamma}_{ij} \tau_i^2 \tau_j^2 \geqq M^{-1} > 0$,

NL5: $\sum_j \tilde{\gamma}_{ij} = 0$ for all i (the regression has a constant term),

NF3: $0 < M^{-1} \leqq J_1(f) \leqq M$ where

$$J_1(f) = \int_{-\infty}^{\infty} \left(\frac{f'}{f}(x)\right)^2 f(x)\, dx \tag{2.5}$$

if F has an absolutely continuous density f with derivative f' and $J_1(f) = \infty$ otherwise.

The proofs of the following two propositions are given in the appendix.

PROPOSITION 2.1. *Suppose* L1, 2, NL5 *and* F1, 2, *hold. Then,*

$$n^{-\frac{1}{2}} \sum_i t_i^2 r_i = n^{-\frac{1}{2}} \sum_{i,k} \tilde{\gamma}_{ik} \tau_i^2 \varepsilon_k + o_p(1) \tag{2.6}$$

$$n^{-1} \sum_{i,k} \tilde{\gamma}_{ik} t_i^2 t_k^2 = n^{-1} \sum_{i,k} \tilde{\gamma}_{ik} \tau_i^2 \tau_k^2 + o_p(1) \tag{2.7}$$

$$\tilde{s}^2 = s^2 + O_p(n^{-1}) = E\varepsilon_1^2 + o_p(1)\,. \tag{2.8}$$

PROPOSITION 2.2. *Suppose* L1, L4 *and* NF3 *hold. Then,* $P_\theta \mathbf{Y}^{-1}$ *and* $P_0 \mathbf{Y}^{-1}$ *are contiguous and*

$$\log \frac{dP_\theta \mathbf{Y}^{-1}}{dP_0 \mathbf{Y}^{-1}}(\mathbf{Y}) = -\theta \sum_i a(\tau_i) \frac{f'}{f}(\varepsilon_i) - \frac{\theta^2}{2} J_1(f) \sum_i a^2(\tau_i) + o_p(1)\,. \tag{2.9}$$

As a consequence of the propositions we have

THEOREM 2.1. *Suppose* L1, 2, 4, NL3, 5 *and* F1, 2, NF3 *hold. Then,*

$$P_\theta[B_x \geqq z] = 1 - \Phi(z - \eta_x) + o(1) \tag{2.10}$$

where

$$\eta_x(\theta, n) = \theta[\sum_{i,j} \tau_i^2 \tau_j^2 \tilde{\gamma}_{ij}]^{\frac{1}{2}} [E\varepsilon_1^2]^{-\frac{1}{2}}\,. \tag{2.11}$$

The same assertion holds for $\tilde{B}_x$.

The proof follows the same lines as that of Theorem 1.1 and will not be repeated. We only note the provenance of η_x from

$$\begin{aligned} [E\varepsilon_1^2]^{-\frac{1}{2}} \operatorname{Cov} &\left\{ \frac{\sum_i \sum_k \tilde{\gamma}_{ik} \tau_i^2 \varepsilon_k}{[\sum_{i,k} \tau_i^2 \tau_k^2 \tilde{\gamma}_{ik}]^{\frac{1}{2}}},\ -\theta \sum_k a(\tau_k) \frac{f'}{f}(\varepsilon_k) \right\} \\ &= \theta [E\varepsilon_1^2]^{-\frac{1}{2}} [\sum_{i,k} \tau_i^2 \tau_k^2 \tilde{\gamma}_{ik}]^{-\frac{1}{2}} \sum_{i,k} \tilde{\gamma}_{ik} \tau_i^2 a(\tau_k) \\ &= \eta_x(\theta, n) \end{aligned} \tag{2.12}$$

by NL5 and the structure of a. □

Statistical implications of the theorem.

(1) The one and two sided tests based on B_x or $\tilde{B}_x$ have robustness of validity and are asymptotically unbiased.

(2) Work in progress at Berkeley suggests that if F is normal these tests are asymptotically UMP unbiased. However, better procedures can and will be found if F is leptokurtic.

(3) The condition NL5 guarantees only that the test statistic B_x is sensible. If it does not hold t_i^2 should be replaced by $a(t_i)$ in B_x. The appropriate statistics for general a and F and their theory are developed in Section 4.

Comment on regularity conditions. As in Section 1 we can show that the conclusion of Theorem 3.1 holds if F1 is replaced by F1′ and L2 is replaced by (1.28). F2 can also be weakened to F2′ but in this case F2′ still involves moments of higher order than those appearing in (2.11). I do not know whether second moments are enough.

3. Testing for heteroscedasticity: general models and robust procedures. In this section we consider models satisfying (1.11) and (1.12) for general $a(\tau)$. More significantly we introduce procedures which in the special case $a(\tau) = \tau$ and more generally are robust against gross errors, i.e., have better power behavior than A_{x2} for leptokurtic errors and perform almost as well when F is normal. The following notation is useful. Let g be a function of a real variable. If $x_1, \cdots, x_n$ is an arbitrary sequence let

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

With $a(\cdot)$ given by (1.12) and assumed known we consider statistics A_b defined by

$$A_b = \sum_i (a(t_i) - a_{\bullet}(t))b(r_i)/\tilde{\sigma}_b \tag{3.1}$$

where

$$\tilde{\sigma}_b^2 = \sum_i (a(t_i) - a_{\bullet}(t))(n-p)^{-1} \sum_i (b^2(r_i) - [b^2(r)]_{\bullet})^2 .$$

We do not assume in this section and the next that $\mathbf{t} = \mathbf{t}^L$, $\mathbf{r} = \mathbf{r}^L$. *Rather* $(t_1, \cdots, t_n)$ *is a vector of fitted values obtained in some way*, i.e., *an n-dimensional statistic taking values in the column space of C and* $\mathbf{r} = \mathbf{Y} - \mathbf{t}$. Natural conditions to be put on $\mathbf{t}$, b will become evident as we proceed.

Such statistics can be motivated readily. If we assume the τ_i, F known the LMP test statistic of $H: \theta = 0$ vs. $K: \theta > 0$ is proportional to

$$-\sum_i a(\tau_i)\left(\varepsilon_i \frac{f'}{f}(\varepsilon_i) - E\varepsilon_i \frac{f'}{f}(\varepsilon_i)\right) .$$

If we let $b(x) = -x(f'/f)(x)$ and estimate the τ_i, ε_i appropriately we are led to the numerator of (3.1).

If f is symmetric about 0 and strongly unimodal $((f'/f)\downarrow)$ b is an increasing function of $|x|$. This class of b's seems particularly natural and as we shall see below only members of this class lead to statistics having "reliable" power behavior.

Another approach proceeds from (M) estimation in the linear model (scale known). Here we obtain the fitted values $\mathbf{t}^R$ as minimizing $\sum_{i=1}^n \rho(Y_i - t_i)$ where $(t_1, \cdots, t_n)$ ranges over the column space of C. For robustness ρ is chosen to emphasize large residuals less than x^2. If $\rho' = \psi$ exists and the minimum $\mathbf{t}^R$ is assumed then $\mathbf{t}^R$ satisfies

$$C'\boldsymbol{\psi}(\mathbf{Y} - \mathbf{t}^R) = 0$$

where $\boldsymbol{\psi}((u_1, \cdots, u_n)') = (\psi(u_1), \cdots, \psi(u_n))'$. Equivalently we can think of $\mathbf{t}^R$ as being least squares fitted values for the "pseudo observations,"

$$\tilde{\mathbf{Y}} = \mathbf{t}^R + \hat{\lambda}^{-1}\boldsymbol{\psi}(\mathbf{Y} - \mathbf{t}^R) \tag{3.2}$$

where

$$\hat{\lambda} = n^{-1} \textstyle\sum_i \psi'(Y_i - t_i^R) .$$

(The choice of $\hat{\lambda}$ is unimportant for this discussion. A rationale for its use and further discussion will be found in the reply to discussants in [7].)

The residual vector of the pseudo observations is proportional to $\boldsymbol{\psi}(\mathbf{Y} - \mathbf{t}^R)$. If we substitute these residuals in A_{x^2} given in Section 1 we obtain essentially A_b with $b = \psi^2$ (and $a(\tau) = \tau$). The only difference is in the use of $t_\bullet$ rather than $\bar{t}$ but that is asymptotically negligible as we have seen in Section 1. If we require that ψ be odd and increasing $b = \psi^2$ is an increasing function of $|x|$.

Some interesting choices of b.

The power family:

$$b(x) = |x|^\alpha , \qquad 1 \leqq \alpha \leqq 2 . \tag{3.3}$$

These b's correspond to the LMP tests for $f(x) \propto e^{-|x|^\alpha}$ and to the (M) estimates for $\psi(x) = |x|^{\alpha/2} \operatorname{sgn} x$.

Huber's function squared:

$$\begin{aligned} b(x) &= x^2 , \quad |x| \leqq k \\ &= k^2 , \quad |x| > k . \end{aligned} \tag{3.4}$$

Here b corresponds to Huber's classical

$$\begin{aligned} \psi_k(x) &= x , \qquad\quad |x| \leqq k \\ &= k \operatorname{sgn} x , \quad |x| > k . \end{aligned} \tag{3.5}$$

It was proposed in a different context (for estimation of scale for a single sample) in Huber [12].

A smooth bounded increasing function of $|x|$:

$$b(x) = \tanh^2 x = \left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right)^2 . \tag{3.6}$$

This b corresponds to $\psi = -f'/f$ where f is the density of the logistic distribution.

Unfortunately an asymptotic theory of the generality presented in Section 1 can easily be obtained only for functions such as (3.5). We shall develop the theory for smooth bounded b and general p here and discuss its shortcomings, extensions, etc. at the end.

We need some further conditions on the model. We use the prefix G to indicate that these conditions correspond to weaker or special conditions in Section 1.

GHL2: $pn^{-\frac{1}{2}} \to 0$,
GHL3: $n^{-1} \sum_{i=1}^{n} (a(\tau_i) - a_\bullet)^2 \geqq M^{-1} > 0$.

Smoothness conditions on a:

S: a is continuously twice differentiable.
 (i) $|a'(x)| \leqq M$
 (ii) $|a''(x)| \leqq M$.

Regularity conditions on b:

HR1: (i) $b(x) = b(-x)$ for all x,
 (ii) b is increasing for $x \geqq 0$;
R2: $\text{Var}\, b(\varepsilon_1) \geqq M^{-1} > 0$;
R3: b is twice continuously differentiable.
 (i) $|b(x)| \leqq M$ for all x,
 (ii) $|b'(x)| \leqq M$ for all x,
 (iii) $|b''(x)| \leqq M$ for all x.

A basic condition on $\mathbf{t}$:
Let $d_i = t_i - \tau_i$.

T: $\sum_{i=1}^{n} d_i^2 = O_p(p)$.

Again P_θ corresponds to computations under (1.11) (1.12) and $P = P_0$ is the distribution under $\mathscr{L}$.

THEOREM 3.1. *Suppose* L1, GHL2, GHL3, L4, F1, HF3, HR1(i), R2, R3, S *and* T *all hold. Then,*

$$P_\theta[A_b \geqq z] = 1 - \Phi(z - \Delta_b) + o(1) \tag{3.7}$$

where

$$\Delta_b(\theta, n) = \theta \sum_i (a(\tau_i) - a_\bullet(\tau))^2 E(\varepsilon_1 b'(\varepsilon_1))[\text{Var}\, b(\varepsilon_1)]^{-\frac{1}{2}} . \tag{3.8}$$

PROOF. In view of R3 and S we can apply (A37) of the appendix with $\omega_{ij} = \delta_{ij} - (1/n)$ (which satisfies (A29)) to get

$$\begin{aligned} \sum_i (a(t_i) - a_\bullet(t))b(r_i) = n^{-\frac{1}{2}} & \sum_i (a(\tau_i) - a_\bullet(\tau))b(\varepsilon_i) \\ & + n^{-\frac{1}{2}} Eb'(\varepsilon_1) \sum_i (a(\tau_i) - a_\bullet(\tau))(t_i - \tau_i) \\ & + o_p(1) . \end{aligned} \tag{3.9}$$

By F1 and HR1(i), $Eb'(\varepsilon_1) = 0$, and the second term in (3.9) vanishes. Now apply (A34) of the appendix to get

$$(3.10) \qquad n^{-1} \sum_i (a(t_i) - a_{\bullet}(t))^2 = n^{-1} \sum_i (a(\tau_i) - a_{\bullet}(\tau))^2 + o_p(1) .$$

Finally, since b^2 satisfies R3 if b does, we can apply (A36) to get

$$(3.11) \qquad n^{-1} \sum_i (b^2(r_i) - [b^2]_{\bullet}(r))^2 = \operatorname{Var} b^2(\varepsilon_1) + o_p(1) .$$

The rest of the proof of the theorem uses Proposition 1.2 and parallels the proof of Theorem 1.1 exactly. □

Comments on the regularity conditions.

(1) Condition T is satisfied by least squares estimates if $E\varepsilon_1 = 0$, $E\varepsilon_1^2 < \infty$, since then

$$E(\textstyle\sum_{i=1}^n d_i^2) = \sum_{i=1}^n \operatorname{Var} t_i = pE\varepsilon_1^2 .$$

Huber [13] has extended this property to (M) estimates with known scale under some regularity conditions on ψ and the condition $\gamma p = o(1)$.

Results similar to Huber's theorem are not at present available for the other methods of estimation except for the case p bounded (see [6], [14]). It has, however, been stated by Huber in [13] that this property as well as the other important linear approximation properties of his theorem carry over to (M) estimates with unknown scale.

(2) Conditions L1, GHL3, L4, HF3, and R2 seem dictated by the nature of the problem.

(3) Conditions S and GL2 are probably not necessary but do not rule out many interesting situations.

(4) Conditions F1 and HR1(i) or at least $Eb'(\varepsilon_1) = 0$ are *necessary* for the conclusion of the theorem. See (A37).

(5) Condition R3 is unsatisfactory. The power family does not belong since R3(i) is always violated and R3(ii), (iii) fail if $\alpha < 2$. Huber's function squared does not satisfy R3(ii), (iii). Unfortunately I am unable to obtain results without R3 unless *p is bounded* and fitting is by least squares or at least we can estimate quantities such as $\sum_i d_i^4$.

Even so the results are inelegant. Here is a typical theorem.

Introduce

GL2′: (i) p is fixed,
(ii) $n^{-1}C'C \to \Sigma_0$,
(iii) $\max_{i,j} |c_{ij}| = o(n^{\frac{1}{2}})$.

GHF3: F3 holds and f' is continuous.

R2′: (i) $M^{-1} \leqq \operatorname{Var} b(\varepsilon_1) \leqq M$
(ii) $E(b(\varepsilon_1) - Eb(\varepsilon_1))^2\varepsilon_1^2 \leqq M$.

R3′: b is absolutely continuous with R–N derivative b' such that

$$|b'(x)| \leqq M \qquad \text{for all} \quad x .$$

T′: $\mathbf{t} = \mathbf{t}^L$ and $E\varepsilon_1 = 0$, $E(\varepsilon_1^2) \leqq M$.

THEOREM 3.2. *Suppose* L1, GL2′, GHL3, L4, F1, GHF3, HR1 (i), R2′, R3′, S *and* T′ *hold. Then* (3.7) *holds.*

Thus (3.7) holds for the Huber squares and $b(x) = |x|$ under mild conditions since R3′ is satisfied for such b. The proofs of these assertions are sketched in the appendix.

Statistical implications of Theorems 3.1 *and* 3.2.

(1) If we refer A_b to normal critical values and b satisfies HR1, then, under the conditions of the theorem, the resulting tests are asymptotically unbiased since F1 and HR1 imply $E\varepsilon_1 b'(\varepsilon_1) > 0$.

(2) The power is independent of the method of estimation of the parameters of $\mathscr{S}$ and depends only on the choice of b. Thus using robust estimates in A_{x^2} would not improve its performance in leptokurtic situations. On the other hand least squares estimates could be used with appropriate b's to give better performance in such situations (see below).

(3) The power depends on the design sequence only through the τ_i.

(4) The theorems enable us to compare statistics based on different b's. We can perform the usual Pitman efficiency computation. That is, we calculate the limiting reciprocal ratio of sample sizes needed by tests based on A_{b_1}, A_{b_2} respectively to reach the same asymptotic power at the same level for a sequence of alternatives $\theta_n = O(n^{-\frac{1}{2}})$. If $e(b_1, b_2)$ denotes the Pitman efficiency we find, provided that $E(b_i'(\varepsilon_1)\varepsilon_1) > 0$, $i = 1, 2$,

$$e(b_1, b_2) = \frac{\operatorname{Var} b_2(\varepsilon_1)}{\operatorname{Var} b_1(\varepsilon_1)} \frac{E^2(b_1'(\varepsilon_1)\varepsilon_1)}{E^2(b_2'(\varepsilon_1)\varepsilon_1)}, \tag{3.12}$$

which depends only on b_1, b_2. If we specialize to $b_1(x) = |x|$, $b_2(x) = x^2$ we find

$$e(b_1, b_2) = \frac{\frac{1}{4}\left(\dfrac{E\varepsilon_1^4}{(E\varepsilon_1^2)^2} - 1\right)}{\left(\dfrac{E\varepsilon_1^2}{(E|\varepsilon_1|)^2} - 1\right)}. \tag{3.13}$$

This expression occurs in Bickel–Lehmann [8] (among other places) as the efficiency of the mean deviation to the standard deviation as measures of scale. The bounds and numerical results of that paper are thus available.

If F is normal, $e = .88$. In general e can be arbitrarily large but $e \geqq .25$ whatever be f. For scale mixtures of normal distributions with mean 0, $e \geqq .48$.

Tables 5.1, 5.2 of [8] show just how favorably the mean deviation statistic A_{b_1} compares with Anscombe's statistic for reasonable slightly heavy tailed distributions. As was noted by Tukey in [17] the classical normal procedures for estimation (and testing) for scale are much more sensitive to nonnormality than the corresponding procedures in location problems. Application of these procedures to selected data sets suggests that the asymptotic predictions are borne out. This and other numerical work will appear elsewhere.

(5) From (3.8) we see that it is proportional to the square root of the reciprocal asymptotic variance of the (M) estimate of log scale based on $\chi = b$. Thus the minimax theory of Huber [12] can be carried over and b given by (3.4) is suggested for moderate contamination.

(6) Suppose f is known. From (3.8) it is clear that the unique most powerful test within the class of all based on A_b satisfying the regularity conditions of the theorems would have to have

$$b(x) = -\left(x \frac{f'}{f}(x) + 1\right).$$

Of course, the regularity conditions of Theorem 3.1 typically rule out b but the optimum power can always be approached arbitrarily closely. Work in progress at Berkeley suggests that this is the best that can be done among all asymptotically unbiased or invariant tests.

The asymmetric case. As Sukhatme and other authors (see Crouse [10]) found, the asymmetric case poses peculiar difficulties. The asymptotic behavior of the numerator of A_b now depends through the second term in (A37) on the method of estimation of the τ_i. It is possible to modify the denominator of A_b to obtain an asymptotically standard normal test statistic under the hypothesis. However, its power behavior is complicated, efficiency comparisons depend on the design, etc. We do not pursue this although the question is clearly an important one.

Extensions and related procedures.

(1) It seems reasonable that, if instead of observing $\tau_i + (1 + \theta\tau_i)\varepsilon_i$, $i = 1, \cdots, n$, we observe $\tau_i + (1 + \theta\tau_i)\sigma\varepsilon_i$, $i = 1, \cdots, n$ our inferences about θ and hence our power curve should be the same at least asymptotically. It may, however, be shown that of the tests based on A_b essentially only those in which $b(x) = |x|^\alpha$ for some α have this invariance property. This is, of course, related to the fact that the statistics A_b were motivated from locally most powerful tests in which the distribution was assumed known and from (M) estimates with scale known. Invariant tests may be constructed in the obvious way by replacing $b(\cdot)$ in A_b with $b(\cdot/\hat{\sigma})$ where $\hat{\sigma}$ is a data dependent estimate of the scale of the observations satisfying at least

$$\hat{\sigma}(\sigma\mathbf{Y}) = \sigma\hat{\sigma}(\mathbf{Y}) \tag{3.14}$$

and

$$\hat{\sigma}(\mathbf{Y}) = \sigma(F) + o_p(n^{-\frac{1}{2}}) \tag{3.15}$$

where $\sigma(F)$ is some positive measure of scale. A commonly used choice in estimation of location is

$$\hat{\sigma}(\mathbf{r}) = \text{median}\,\{|r_1|, \cdots, |r_n|\}/\Phi^{-1}(.75)\,. \tag{3.16}$$

Another less robust possibility is, of course, $\hat{\sigma}^2 = s_L^{\,2}$.

It may be shown that if the conditions of Theorems 3.1 or 3.2 hold as well as (3.15) then

$$P_\theta[A_{b_{\hat\sigma}} \geqq z] = 1 - \Phi(z - \Delta_{b_\sigma}) + o(1)\,, \tag{3.17}$$

where

$$b_{\hat\sigma}(x) = b\left(\frac{x}{\hat\sigma}\right) \tag{3.18}$$

$$b_\sigma(x) = b\left(\frac{x}{\sigma(F)}\right). \tag{3.19}$$

A proof is sketched in the appendix.

(2) An alternative approach to the problem discussed in (1) above is to use other random transforms of the residuals than $b(r_i/\hat\sigma)$. One possibility is to replace $b(r_i)$ by the square of the rank of r_i among $r_1, \cdots, r_n$. This procedure was considered by Sukhatme and others (see Crouse [10]) when $\mathscr{L}$ is the two sample model and $a(\tau) = \tau$. Unfortunately as with the estimates based on ranks studied by Jaeckel [14] and others only bounded p results seem attainable for such statistics. However, where these have been obtained, the qualitative picture agrees with ours, e.g., lack of dependence on the method of fitting, difficulties with asymmetry.

4. Testing for nonlinearity: general models and robust procedures. In this section we consider models satisfying (2.1) and (2.3) for general $a(\tau)$ and introduce generalizations of the Tukey–Anscombe tests which are robust against gross errors. Naive local power considerations lead to studentizing $-\sum_i a(t_i)(f'/f)(r_i)$. On the other hand, applying the Anscombe–Tukey statistic in the form given in Scheffé [16], problem 4.19, to pseudo observations similarly leads to studentizing $\sum_i a(t_i)\psi(r_i)$. Thus we again arrive at statistics of the form A_b where, however, the function b is naturally taken to be antisymmetric and increasing. Actually because we do not necessarily want to match the method of fitting and b we are led to the following form:

$$B_b = \hat\sigma_b^{-1} \textstyle\sum_{i,k} \tilde\gamma_{ik} a(t_i) b(r_k)\,, \tag{4.1}$$

where

$$\hat\sigma_b^2 = (n-p)^{-1} \textstyle\sum_i (b(r_i) - b_\cdot(r))^2 \sum_{i,k} \tilde\gamma_{ik} a(t_i) a(t_k)\,.$$

This form agrees with A_b if the vector $(b(r_1), \cdots, b(r_n))'$ is orthogonal to the column space of Γ.

Some interesting choices of b.

The power family:

$$b(x) = |x|^\alpha \operatorname{sgn} x\,, \qquad 0 \leqq \alpha \leqq 1\,. \tag{4.2}$$

Huber's ψ_k given by (3.5).

The logistic score function:

$$b(x) = \tanh x\,. \tag{4.3}$$

We shall need the conditions

GNL2: $\gamma p^2 \to 0$,
GNL3: $\sum_{i,j} \tilde{\gamma}_{ij} a(\tau_i)a(\tau_j) \geqq M^{-1} > 0$,
NR1: (i) $b(x) = -b(-x)$,
(ii) b is increasing.

THEOREM 4.1. *Suppose* L1, GNL2, GNL3, L4, NF3, R2, R3, S *and* T *all hold. Suppose also that* either NL5 *holds* or NR1 (i) *and* F1 *hold. Then,*

$$P_\theta[B_b \geqq z] = 1 - \Phi(z - \eta_b) + o(1) \tag{4.4}$$

where

$$\eta_b(\theta, n) = \theta[\sum_{i,j} a(\tau_i)a(\tau_j)\tilde{\gamma}_{ij}]^{\frac{1}{2}} Eb'(\varepsilon_1)[\mathrm{Var}\, b(\varepsilon_1)]^{-\frac{1}{2}} . \tag{4.5}$$

PROOF. Using R3 and S expand the numerator of B_b using (A37) with $\omega_{ij} = \tilde{\gamma}_{ij}$. Note that NL5 is just (A29) while NR1 (i) and F1 imply (A30).

We obtain only the leading term

$$\sum_{i,j} \tilde{\gamma}_{ij} a(\tau_i) b(\varepsilon_j) ,$$

since $\sum_j \tilde{\gamma}_{ij}(t_j - \tau_j) = 0$ for all i. The remainder is $O_p(p(1 + [n \max_i \gamma_{ii}]^{\frac{1}{2}}))$ from T and the following estimation

$$\begin{aligned} \max_i \sum_j |\tilde{\gamma}_{ij}| &\leqq 1 + \max_i \sum_j |\gamma_{ij}| \\ &\leqq 1 + n^{\frac{1}{2}} \max_i [\sum_j \gamma_{ij}^2]^{\frac{1}{2}} \\ &= 1 + n^{\frac{1}{2}} \max \gamma_{ii}^{\frac{1}{2}} . \end{aligned} \tag{4.6}$$

If we similarly apply (A34) and (A36) to the terms of the denominator we arrive at

$$B_b = \{\sum_{i,j} \tilde{\gamma}_{ij} a(\tau_i) b(\varepsilon_j) / [\mathrm{Var}^{\frac{1}{2}}\, b(\varepsilon_i)[\sum_{i,k} a(\tau_i)a(\tau_k)\tilde{\gamma}_{ik}]^{\frac{1}{2}}]\} + o_p(1) . \tag{4.7}$$

Moreover

$$\begin{aligned} \max_j n^{-\frac{1}{2}} |\sum_i \tilde{\gamma}_{ij} a(\tau_i)| &\leqq n^{-\frac{1}{2}} \{\max_j |a(\tau_j)| + [\sum_i a^2(\tau_i)]^{\frac{1}{2}} \max_i \gamma_{ii}^{\frac{1}{2}}\} \\ &= o(1) \end{aligned}$$

since $\gamma \to 0$.

Therefore the leading term on the right of (4.6) has a limiting $N(0, 1)$ distribution by the Lindeberg–Feller theorem. The rest of the argument proceeds as that of Theorem 2.1.

Comments on the regularity conditions.

(1) Again L1, GNL3, L4, NF3, and R2 are dictated by the problem while S is reasonable.

(2) NL2 is surprisingly strong. It is forced by the estimate (4.6) and can undoubtedly be improved at least in special cases. Note, however, that even when we can write $\sum_{i,j} \tilde{\gamma}_{ij} a(t_i) b(r_j)$ as $\sum_i a(t_i) b(r_i)$ we must use the former form in the expansion in order to conclude that the "$Eb'(\varepsilon_1)$" term vanishes. Thus some norm of Γ must appear in the remainder estimates.

(3) The pair F1, NR1 (i) can be replaced by $Eb(\varepsilon_1) = 0$.
(4) Condition R3 is again unsatisfactory. It is easy to establish

THEOREM 4.2. *Suppose* L1, GL2′, GNL3, L4, NF3, R2′, R3′ *and* either NL5 or NR1 (i) *and* F1 *hold. Then* (4.4) *is valid.*

Thus (4.4) holds for the Huber functions under mild conditions. Unfortunately $b(x) = \operatorname{sgn} x$ is not covered.

Statistical implications of Theorems 4.1 *and* 4.2.

(1) If we refer B_b to normal critical values and b satisfies NR1 (ii) the resulting tests are asymptotically unbiased.

(2) The asymptotic power is independent of the method by which the τ_i are estimated. It does, however, depend on the designs rather than just on the τ_i.

(3) The Pitman efficiency of tests based on b_1, b_2 respectively is, if $Eb_i'(\varepsilon_1) > 0$, $i = 1, 2$,

$$e(b_1, b_2) = \frac{\operatorname{Var} b_2(\varepsilon_1)}{\operatorname{Var} b_1(\varepsilon_1)} \frac{E^2 b_1'(\varepsilon_1)}{E^2 b_2'(\varepsilon_1)} \tag{4.8}$$

and is independent of the design. This is the same as the relative efficiency of the (M) estimates of location based on b_1 and b_2.

(4) In view of Theorem 4.2, we can apply the results of [12]. For instance, if there is an additive main effect we conclude that $b = \psi_k$ given by (3.5) maximizes the minimum asymptotic power in neighborhood of the normal distribution of the form $\{F: F = (1 - \varepsilon)\Phi + \varepsilon H\}$ where ε is known. (The restriction to F symmetric about 0 is unnecessary since the tests based on b, $b + c$ coincide if there is an additive main effect.)

(5) If f is known it is clear from (4.5) that the most powerful test within the class of all tests based on B_b satisfying the regularity conditions of the theorems would have

$$b(x) = -\frac{f'}{f}(x) .$$

Again such tests are typically ruled out by the regularity conditions but can be approached arbitrarily closely and seem to be "optimal" within a larger class.

Extensions and related procedures.

(1) Again it seems appropriate to replace $b(\cdot)$ by $b(\cdot/\hat{\sigma})$. If $\hat{\sigma}$ satisfies (3.15) and the conditions of Theorems 4.1 or 4.2 hold then

$$P_\theta[B_{b_{\hat{\sigma}}} \geqq z] = 1 - \Phi(z - \eta_{b_\sigma}) + o(1) . \tag{4.9}$$

On the basis of asymptotic efficiency these procedures with $b = \psi_k$, $k = 1, 1.5$, $\hat{\sigma}$ estimated using Huber's proposal 2 or given by (3.16) (say) perform much more satisfactorily than Tukey's one degree of freedom (B_x). Again, application to specific data sets seems to support the asymptotics.

(2) Rank procedures are possibilities here as well. Their properties have

not been studied even for the case p bounded although it seems clear how they should behave.

5. Appendix.

5A. *Contiguity results.* Lemmas 1.2 and 2.2 follow readily from Theorem A1 below, which is an immediate consequence of the discussion on pages 210–214 of Hájek–Sidák [11]. Consider the following perturbations of model $\mathscr{L}$.

$$\mathscr{L}_1\colon\ Y_i = \tau_i + \mu_i + \varepsilon_i\,, \qquad i = 1, \cdots, n$$

where the ε_i and τ_i are as before and $\mu_1, \cdots, \mu_n$ are arbitrary constants.

$$\mathscr{L}_2\colon\ Y_i = \tau_i + \sigma_i\varepsilon_i\,, \qquad i = 1, \cdots, n$$

where the σ_i are arbitrary positive constants. As with the τ_i we suppose the μ_i and σ_i are really a double array depending on n.

Let g_{ij} denote the density of Y_i under $\mathscr{L}_j$ and

$$f_i(y) = f(y - \tau_i)$$

denote the density under $\mathscr{L}$.

THEOREM A1. (a) *The model $\mathscr{L}_1$ is contiguous to $\mathscr{L}$ if*

$$\max\nolimits_{1\leqq i\leqq n} |\mu_i| = o(1)$$
$$M_1^{-1} \leqq \sum \mu_i^2 \leqq M_1$$

for all n, f is absolutely continuous, and

$$M_2^{-1} \leqq J_1(f) \leqq M_2\,.$$

Moreover, in that case, under $\mathscr{L}$,

$$\text{(A1)} \qquad \textstyle\sum_{i=1}^n \log \dfrac{g_i}{f_i}(Y_i) = -\sum_{i=1}^n \mu_i \dfrac{f'}{f}(Y_i) - \dfrac{J_1(f)}{2} \sum_{i=1}^n \mu_i^2 + o_p(1)\,.$$

(b) *The model $\mathscr{L}_2$ is contiguous to $\mathscr{L}$ if*

$$\max\nolimits_{1\leqq i\leqq n} |\sigma_i - 1| = o(1)$$
$$M_3^{-1} \leqq \textstyle\sum_{i=1}^n (\sigma_i - 1)^2 \leqq M_3$$

for all n, f is absolutely continuous and

$$M_4^{-1} \leqq J_2(f) \leqq M_4\,.$$

Moreover, in that case, under $\mathscr{L}$,

$$\text{(A2)} \qquad \textstyle\sum_{i=1}^n \log \dfrac{g_{i2}}{f_i}(Y_i) = -\sum_{i=1}^n (\sigma_i - 1)\left(1 + \varepsilon_i \dfrac{f'}{f}(\varepsilon_i)\right)$$
$$- \frac{J_2(f)}{2} \textstyle\sum_{i=1}^n (\sigma_i - 1)^2 + o_p(1)\,.$$

PROOF.

$$\textstyle\sum_{i=1}^n \log \dfrac{g_{i1}}{f_i}(Y_i) = \sum_{i=1}^n \log \dfrac{f(\varepsilon_i - \mu_i)}{f(\varepsilon_i)}$$

and we can apply the theorem of Section 2.2 of [11] directly. Similarly,

$$\sum_{i=1}^{n} \log \frac{g_{i2}}{f_i}(Y_i) = \sum_{i=1}^{n} \log \sigma_i^{-1} \frac{f(\varepsilon_i \sigma_i^{-1})}{f(\varepsilon_i)} .$$

Translating the conditions of the theorem of Section 2.3 of [11] into ours is then straightforward.

REMARK. Clearly one can have perturbations both of type $\mathscr{L}_1$ and $\mathscr{L}_2$ and contiguity of this super model to $\mathscr{L}$ is obtained by combining the conditions of (a), (b).

5B. *Proofs of Propositions* 1.1 *and* 2.1.

PROOF OF PROPOSITION 1.1: Let $d_i = t_i - \tau_i$, $i = 1, \cdots, n$.

PROOF OF (1.16). Let $E(\varepsilon_1^2) = 1$. Write

$$\sum_i (t_i - \bar{t}) r_i^2 = \sum (\tau_i - \bar{\tau}) r_i^2 + \sum (d_i - \bar{d}) r_i^2 \tag{A3}$$

where $\bar{d} = (n-p)^{-1} \sum_i \bar{\gamma}_{ii} d_i$. Then,

$$\begin{aligned} \sum_i (\tau_i - \bar{\tau}) r_i^2 &= \sum_i (\tau_i - \bar{\tau}) \varepsilon_i^2 - 2 \sum_i (\tau_i - \bar{\tau}) \varepsilon_i d_i + \sum_i (\tau_i - \bar{\tau}) d_i^2 \\ &= \sum_i (\tau_i - \bar{\tau}) \varepsilon_i^2 - 2 \sum_{i,k} (\tau_i - \bar{\tau}) \gamma_{ik} \varepsilon_i \varepsilon_k \\ &\quad + \sum_{i,k,l} (\tau_i - \bar{\tau}) \gamma_{ik} \gamma_{il} \varepsilon_k \varepsilon_l . \end{aligned} \tag{A4}$$

The expectation of the left-hand side of (A4) is 0 if $E(\varepsilon_1) = 0$. Therefore if we center the last two variables at their expectations and call them R_1, R_2 we obtain

$$\sum_i (\tau_i - \bar{\tau}) r_i^2 = \sum_i (\tau_i - \bar{\tau})(\varepsilon_i^2 - 1) + R_1 + R_2 . \tag{A5}$$

Now

$$E(R_1^2) = 4\{(E(\varepsilon_1^4) - 1) \sum_i \gamma_{ii}^2 (\tau_i - \bar{\tau})^2 + \sum_{i \neq j} \gamma_{ij}^2 [(\tau_i - \bar{\tau})(\tau_j - \bar{\tau}) + (\tau_i - \bar{\tau})^2]\} .$$

Therefore, using $\Gamma^2 = \Gamma$ and $\max_{i,j} |\gamma_{ij}| \leqq 1$ we get

$$E(R_1^2) = O(\sum_i \gamma_{ii}) = O(p) . \tag{A6}$$

Similarly,

$$\begin{aligned} E(R_2^2) &= O(\sum_{i,j} |(\tau_i - \bar{\tau})(\tau_j - \bar{\tau})| [\sum_k \gamma_{ik}^2 \gamma_{jk}^2 + \gamma_{ij}^2]) \\ &= O(\sum_i \gamma_{ii}) = O(p) . \end{aligned} \tag{A7}$$

Therefore,

$$R_1 + R_2 = O_p(p^{\frac{1}{2}}) . \tag{A8}$$

Next consider

$$\begin{aligned} \sum_i (d_i - \bar{d}) r_i^2 &= \sum_i (d_i - \bar{d})(r_i^2 - \bar{\gamma}_{ii}) \\ &= \sum_i (d_i - \bar{d})[(\varepsilon_i^2 - 1 - 2(d_i \varepsilon_i - \gamma_{ii}) + (d_i^2 - \gamma_{ii}))] . \end{aligned} \tag{A9}$$

Since

$$\begin{aligned} \bar{d} &= (n-p)^{-1} \sum_i (\sum_i \gamma_{ik} \bar{\gamma}_{ii}) \varepsilon_k , \\ E(\bar{d}) &= 0 , \\ E(\bar{d}^2) &= (n-p)^{-2} \sum_k (\sum_i \bar{\gamma}_{ii} \gamma_{ik})^2 \leqq (n-p)^{-2} \sum_i \bar{\gamma}_{ii}^2 = O(n^{-1}) . \end{aligned}$$

Therefore

(A10) $$\bar{d} \sum_i (\varepsilon_i^2 - 1) = O_p(1) .$$

The estimates used for R_1 yield

(A11) $$\bar{d} \sum (d_i \varepsilon_i - \gamma_{ii}) = O_p((pn^{-1})^{\frac{1}{2}})$$

while those used for R_2 give

(A12) $$\bar{d} \sum (d_i^2 - \gamma_{ii}) = O_p((pn^{-1})^{\frac{1}{2}}) .$$

Simple computations of the same type yield

(A13) $$\sum d_i(\varepsilon_i^2 - 1) = O_p(p^{\frac{1}{2}}) ,$$

(A14) $$\sum d_i \gamma_{ii} = O_p(p^{\frac{1}{2}}) .$$

Thus

(A15) $$\sum_i (d_i - \bar{d}) r_i^2 = 2 \sum_i d_i^2 \varepsilon_i + \sum_i d_i^3 + O_p(p^{\frac{1}{2}}) .$$

Now consider

$$\sum_i d_i^2 \varepsilon_i = \sum_{i,k,l} \gamma_{ik} \gamma_{il} \varepsilon_i \varepsilon_k \varepsilon_l .$$

Since f is symmetric about 0, $\sum_i d_i^2 \varepsilon_i$ has expectation 0 and variance which is

$$O(\sum_{i,k} \gamma_{ik}^4 + \sum_{i,k} \gamma_{ik}^3 \gamma_{kk} + \sum_{i,k} \gamma_{ik}^2 \gamma_{ii}^2 + \sum_{i,k} \gamma_{ii} \gamma_{ik} \gamma_{kk}^2 + \sum_{i,k,m} \gamma_{ik}^2 \gamma_{im}^2 + \sum_{i,k,m} \gamma_{ii} \gamma_{ik} \gamma_{km}^2 + \sum_{i,k,m} \gamma_{ii} \gamma_{kk} \gamma_{im} \gamma_{km}) .$$

The terms in the 0 expression (which correspond to nonvanishing moments) can be bounded as follows:

$$\begin{aligned}
\sum_{i,k} \gamma_{ik}^4 &\leqq \sum_{i,k} \gamma_{ik}^2 = \sum_i \gamma_{ii} = p \\
|\sum_{i,k} \gamma_{ik}^3 \gamma_{kk}| &\leqq \sum_{i,k} \gamma_{ik}^2 \\
\sum_{i,k} \gamma_{ik}^2 \gamma_{ii}^2 &= \sum_i \gamma_{ii}^3 \leqq \sum_i \gamma_{ii} \\
|\sum_{i,k} \gamma_{ii} \gamma_{ik} \gamma_{kk}^2| &\leqq [\sum_k \gamma_{kk}^4]^{\frac{1}{2}} [\sum_k [\sum_i \gamma_{ik} \gamma_{ii}]^2]^{\frac{1}{2}} \\
&\leqq \sum_k \gamma_{kk}^2 \qquad \text{since} \quad \Gamma \quad \text{is a projection matrix} \\
|\sum_{i,k,m} \gamma_{ii} \gamma_{kk} \gamma_{im} \gamma_{km}| &= |\sum_{i,k} \gamma_{ii} \gamma_{kk} \gamma_{ik}| \leqq \sum_k \gamma_{kk}^2 \\
|\sum_{i,k,m} \gamma_{ii} \gamma_{ik} \gamma_{km}^2| &= |\sum_{i,k} \gamma_{ii} \gamma_{kk} \gamma_{ik}| \\
\sum_{i,k,m} \gamma_{ik}^2 \gamma_{im}^2 &= \sum_i \gamma_{ii}^2 .
\end{aligned}$$

Therefore

(A16) $$\sum_i d_i^2 \varepsilon_i = O_p(p^{\frac{1}{2}}) .$$

Finally we note, using the symmetry of f again, that

$$E(\sum_i d_i^3) = 0$$

and after some simplification,

$$\operatorname{Var} (\sum_i d_i^3) = O(\sum_{i,j,k} \gamma_{ii} \gamma_{ik} \gamma_{jk}^3 + \sum_{i,j,k} \gamma_{ik}^2 \gamma_{jk}^2 \gamma_{ij} + \sum_{i,j} \gamma_{ij} \gamma_{ii} \gamma_{jj} + \sum_{i,j} \gamma_{ij}^3) .$$

The terms in the 0 expression can be bounded as follows:

$$\begin{aligned}
|\sum_{i,j,k} \gamma_{ii}\gamma_{ik}\gamma_{jk}^3| &= |\sum_k (\sum_i \gamma_{ii}\gamma_{ik} \sum_j \gamma_{jk}^3)| \\
&\leqq \{\sum_k [\sum_i \gamma_{ii}\gamma_{ik}]^2 \sum_k (\sum_j \gamma_{jk}^3)^2\}^{\frac{1}{2}} \\
&\leqq \sum_k \gamma_{kk}^2 \\
|\sum_{i,j,k} \gamma_{ik}^2\gamma_{jk}^2\gamma_{ij}| &= \sum_k \sum_i [\sum_j \gamma_{jk}^2\gamma_{ij}]^2 \\
&\leqq \sum_k \sum_i \gamma_{ik}^2 = \sum_k \gamma_{kk} \\
|\sum_{i,j} \gamma_{ii}\gamma_{ij}\gamma_{jj}| &= \sum_i (\sum_j \gamma_{jj}\gamma_{ij})^2 \leqq \sum_i \gamma_{ii}^2 \\
|\sum_{i,j} \gamma_{ij}^3| &\leqq \sum_{i,j} \gamma_{ij}^2 = \sum_i \gamma_{ii} .
\end{aligned}$$

In conclusion,

$$\sum_i d_i^3 = O_p(p^{\frac{1}{2}}) \tag{A17}$$

and combining (A15)—(A17) with (A5) and (A8), (1.16) follows. □

PROOF OF (1.17).

$$\begin{aligned}
s^2 &= (n-p)^{-1}\{\sum_i \varepsilon_i^2 + 2\sum_i \varepsilon_i d_i + \sum_i d_i^2\} \\
&= E(\varepsilon_1^2) + O_p(pn^{-1})
\end{aligned} \tag{A18}$$

by arguing as for (A11) and the law of large numbers.

PROOF OF (1.18).

$$\begin{aligned}
&n^{-1} \sum_{i,j} \gamma_{ij}^2(t_i - \bar{t})(t_j - \bar{t}) \\
&\quad = n^{-1} \sum_{i=1}^n (t_i - \bar{t})^2 \\
&\quad\quad + O_p(n^{-1}[\sum_{i,j} \gamma_{ij}^2(t_i - \bar{t})(t_j - \bar{t}) + \sum_i \gamma_{ii}(t_i - \bar{t})^2]) .
\end{aligned} \tag{A19}$$

The expression under the O_p sign is nonnegative since $||\gamma_{ij}^2||$ is nonnegative definite by Schur's theorem. Its expectation is easily seen to be $O(pn^{-1})$ under L1 and F2.

Moreover

$$\begin{aligned}
n^{-1} \sum_i (t_i - \bar{t})^2 &= \sum_i (\tau_i - \bar{\tau})^2 + 2\sum_i (\tau_i - \bar{\tau})(d_i - \bar{d}) + \sum_i (d_i - \bar{d})^2 \\
&= n^{-1} \sum_i (\tau_i - \bar{\tau})^2 + O_p((p/n)^{\frac{1}{2}})
\end{aligned} \tag{A20}$$

and (1.18) follows.

PROOF OF (1.19). We need to show (in view of (1.17))

$$(n-p)^{-1} \sum_i r_i^4 = (n-p)^{-1} \sum_i \varepsilon_i^4 + o_p(1) . \tag{A21}$$

But (A21) follows readily upon applying the estimate

$$E(\sum_i d_i^4) \leqq [\sum_{i,j} \gamma_{ij}^4 + \sum_i \gamma_{ii}^2]E\varepsilon_1^4 = O_p(\gamma p) \tag{A22}$$

and Schwartz's inequality to the various terms of the difference between the leading term of the right- and left-hand sides of (A21). □

Sketch of proof of comment at end of Section 1. We sketch the modifications needed for proving (1.16) under the new conditions. All that is needed are revised estimates of $\sum_i d_i^2\varepsilon_i$ and $\sum_i d_i^3$. We illustrate that for $\sum_i d_i^3$. Estimate separately $\sum_i d_i^3 - \sum_{i,k} \gamma_{ik}^3 \varepsilon_k^3$ and $\sum_{i,k} \gamma_{ik}^3 \varepsilon_k^3$. The first of these terms has expectation 0 under F1′ and a variance involving only moments of the ε_i of order 4 or lower. Moreover the coefficients appearing in the variance are just those appearing in Var $(\sum_i d_i^3)$ under F1, F2. Thus,

$$\sum_i d_i^3 - \sum_{i,k} \gamma_{ik}^3 \varepsilon_k^3 = O_p(p^{\frac{1}{2}}) .$$

On the other hand,

$$E|\sum_{i,k} \gamma_{ik}^3 \varepsilon_k^3| \leqq E|\varepsilon_1|^3 \gamma \sum_{i,k} \gamma_{ik}^2 = O(\gamma p) . \qquad \square$$

PROOF OF PROPOSITION 2.1. Again take $E\varepsilon_1^2 = 1$.

PROOF OF (2.6). Write

$$\begin{aligned} \sum_i t_i^2 r_i &= \sum_i \tau_i^2 r_i + 2 \sum_i \tau_i d_i r_i + \sum_i d_i^2 r_i \\ &= \sum_{i,k} \bar{\gamma}_{ik} \tau_i^2 \varepsilon_k + 2[\sum_i \tau_i(d_i \varepsilon_i - \gamma_{ii}) - \sum_i \tau_i(d_i^2 - \gamma_{ii})] \\ &\quad + \sum_i d_i^2 \varepsilon_i - \sum_i d_i^3 . \end{aligned} \tag{A23}$$

The terms following the factor 2 are $O_p(p^{\frac{1}{2}})$ by the arguments used for (A11) and (A12) while the last two terms are $O_p(p^{\frac{1}{2}})$ by (A16) and (A17). (2.6) follows.

PROOF OF (2.7).

$$\sum_{i,k} \bar{\gamma}_{ik} t_i^2 t_k^2 = \sum_{i,k} \bar{\gamma}_{ik} \tau_i^2 \tau_k^2 + O_p((\sum_i t_i^4)^{\frac{1}{2}}(\sum d_i^2)^{\frac{1}{2}} + \sum_i d_i^4) . \tag{A24}$$

The remainder is $O_p(p^{\frac{1}{2}} + p\gamma^{\frac{1}{2}})$.

PROOF OF (2.8). By NL5 and (2.6),

$$\begin{aligned} a(\mathbf{t})\mathbf{r} &= \sum_i t_i^2 r_i = O_p(n^{\frac{1}{2}}) \\ a(\mathbf{t})\bar{\Gamma} a(\mathbf{t})' &= \sum_{i,k} \bar{\gamma}_{ik} t_i^2 t_k^2 \geqq M^{-1}n + o_p(n) . \end{aligned} \tag{A25}$$

The result follows. $\square$

5C. *Taylor expansion results needed for the proofs of Theorems* 3.1 *and* 4.1. Let $(t_1, \cdots, t_n)$ be a set of fitted values as in Sections 3 and 4, i.e., any statistic taking values in the column space of C.

Let a, b be real valued functions of a real variable. The statistics of Sections 1 and 4 are all based on expressions of the form

$$T^{12} = \sum_{i=1}^n \sum_{j=1}^n w_{ij} a(t_i) b(r_j) \tag{A26}$$

$$T^{11} = \sum_{i=1}^n \sum_{j=1}^n w_{ij} a(t_i) a(t_j) \tag{A27}$$

$$T^{22} = \sum_{i=1}^n \sum_{j=1}^n w_{ij} b(r_i) b(r_j) \tag{A28}$$

where $||w_{ij}||$ is a projection matrix (symmetric and idempotent) depending on n. We shall use the conditions R specified in Section 3 and also

$$\sum_{i=1}^n w_{ij} = 0 , \qquad j = 1, \cdots, n . \tag{A29}$$

$$Eb(\varepsilon_1) = 0 . \tag{A30}$$

For this subsection assume without loss of generality that

(A31) $$C'C = I_{p\times p}$$

and define $\hat{\boldsymbol\beta} = (\hat\beta_1, \cdots, \hat\beta_p)'$ by

(A32) $$\hat{\boldsymbol\beta} = C'\mathbf{t}\,.$$

Also let

(A33) $$d_i = t_i - \tau_i\,, \qquad i = 1, \cdots, n\,.$$

THEOREM A2.

(i) *Suppose* L1 *and* S *hold. Then*

(A34) $$T^{11} = \sum_{i=1}^n \sum_{j=1}^n w_{ij}a(\tau_i)a(\tau_j) + O_p(n^{\frac12}[\sum_i d_i^2]^{\frac12} + \sum_i d_i^2)\,.$$

(ii) (a) *Suppose* L1 *and* R3 *hold. Then*

(A35) $$T^{22} = \sum_{i=1}^n \sum_{j=1}^n w_{ij}b(\varepsilon_i)b(\varepsilon_j) + O_p(n^{\frac12}[\sum_i d_i^2]^{\frac12} + \sum_i d_i^2)\,.$$

(b) *If, in addition,* (A29) *or* (A30) *hold then*

(A36) $$T^{22} = \operatorname{Var} b(\varepsilon_1) \sum_{i=1}^n w_{ii} + O_p(n^{\frac12}(1 + [\sum_i d_i^2]^{\frac12}) + \sum_i d_i^2)\,.$$

(iii) *Suppose* R3, S *and either* (A29) *or* (A30) *hold. Then*

(A37) $$T^{12} = \sum_{i=1}^n \sum_{j=1}^n w_{ij}a(\tau_i)b(\varepsilon_j) - E(b'(\varepsilon_1)) \sum_{i=1}^n \sum_{j=1}^n w_{ij}a(\tau_i)d_j + O_p\{[\sum_i d_i^2]^{\frac12}(p^{\frac12} + [\sum_i d_i^2]^{\frac12}(1 + \max_j \sum_i |w_{ij}|))\}\,.$$

PROOF:

PROOF OF (A34). Similar to that of (A35) below and will not be given.

PROOF OF (A35). Write

$$T^{22} = \sum_{i,j} w_{ij}b(\varepsilon_i)b(\varepsilon_j) + 2\sum_{i,j} w_{ij}b(\varepsilon_i)(b(r_j) - b(\varepsilon_j)) + \sum_{i,j} w_{ij}(b(r_i) - b(\varepsilon_i))(b(r_j) - b(\varepsilon_j))\,.$$

By a standard inequality on projection matrices,

(A38) $$|\sum_{i,j} w_{ij}b(\varepsilon_i)(b(r_j) - b(\varepsilon_j))| \leqq [\sum_i b^2(\varepsilon_i)]^{\frac12}[\sum_j (b(r_j) - b(\varepsilon_j))^2]^{\frac12}\,.$$

Similarly

(A39) $$|\sum_{i,j} w_{ij}(b(r_i) - b(\varepsilon_i))(b(r_j) - b(\varepsilon_j))| \leqq \sum_i (b(r_i) - b(\varepsilon_i))^2\,.$$

By R3(i)

$$\sum_i b^2(\varepsilon_i) = O_p(n)$$

while by R3(ii)

$$\sum_j (b(r_j) - b(\varepsilon_j))^2 = O_p(\sum_j d_j^2)\,.$$

The result follows.

PROOF OF (A36). If either (A29) or (A30) holds,

(A40) $$E\sum_{i,j} w_{ij}b(\varepsilon_i)b(\varepsilon_j) = \operatorname{Var} b(\varepsilon_1) \sum_i w_{ii}\,.$$

An elementary calculation using $\sum_{j=1}^n w_{ij}^2 = w_{ii}$ shows that the variance of $\sum_{i,j} w_{ij} b(\varepsilon_i) b(\varepsilon_j)$ is $O_p(\sum_i w_{ii}) = O_p(n)$. (A36) now follows from (A35) and (A40).

PROOF OF (A37). Write

$$
\begin{aligned}
T^{12} &= \sum_{i=1}^n \sum_{j=1}^n w_{ij} a(\tau_i) b(\varepsilon_j) + \sum_{i=1}^n \sum_{j=1}^n w_{ij} a(\tau_i)(b(r_j) - b(\varepsilon_j)) \\
&\quad + \sum_{i=1}^n \sum_{j=1}^n w_{ij} b(\varepsilon_j)(a(t_i) - a(\tau_i)) \\
&\quad + \sum_{i=1}^n \sum_{j=1}^n w_{ij}(a(t_i) - a(\tau_i))(b(r_j) - b(\varepsilon_j)) \,. \qquad \text{(A41)}
\end{aligned}
$$

The last term on the right in (A41) is

$$
O_p([\textstyle\sum_i (a(t_i) - a(\tau_i))^2]^{\frac12} [\sum_i (b(r_i) - b(\varepsilon_i))^2]^{\frac12})
$$

since W is a projection and hence is $O_p(\sum_i d_i^2)$ by S and R3.

If we expand the third term around $d_1 = \cdots = d_n = 0$, we obtain

$$
\begin{aligned}
&\textstyle\sum_{i,j} w_{ij} b(\varepsilon_j)(a(t_i) - a(\tau_i)) \\
&\quad = \textstyle\sum_{i,j} w_{ij} a'(\tau_i) d_i b(\varepsilon_j) \qquad \text{(A42)} \\
&\qquad + \int_0^1 [\textstyle\sum_{i,j} w_{ij} a''(\tau_i + \lambda d_i) c_i^2 b(\varepsilon_j)](1 - \lambda)\, d\lambda \,.
\end{aligned}
$$

Now, by (A32),

$$
\sum_{i=1}^n \sum_{j=1}^n w_{ij} a'(\tau_i) d_i b(\varepsilon_j) = \sum_{k=1}^p (\hat\beta_k - \beta_k) W_k \qquad \text{(A43)}
$$

where

$$
W_k = \sum_{j=1}^n b(\varepsilon_j)(\textstyle\sum_{i=1}^n w_{ij} a'(\tau_i) c_{ik}) \,.
$$

But, by (A29) or (A30), $E(W_k) = 0$ for all k. Thus

$$
\begin{aligned}
E(W_k^2) &= \sum_{j=1}^n (\textstyle\sum_{i=1}^n w_{ij} a'(\tau_i) c_{ik})^2 \operatorname{Var} b(\varepsilon_1) \\
&\leqq \textstyle\sum_{i=1}^n [a'(\tau_i)]^2 c_{ik}^2 \qquad \text{(A44)} \\
&\leqq M^2 \operatorname{Var} b(\varepsilon_1)
\end{aligned}
$$

by (A31). By (A31) and (A44)

$$
\begin{aligned}
\sum_{k=1}^p (\hat\beta_k - \beta_k) W_k &= O_p(p^{\frac12}[\textstyle\sum_{k=1}^p (\hat\beta_k - \beta_k)^2]^{\frac12}) \qquad \text{(A45)} \\
&= O_p(p^{\frac12}[\textstyle\sum_i d_i^2]^{\frac12}) \,.
\end{aligned}
$$

For the second term on the right of (A42) we use S and R3 to obtain a bound

$$
M^2(\max_j \textstyle\sum_i |w_{ij}|) \sum_i d_i^2 \,. \qquad \text{(A46)}
$$

Combining (A45) and (A46) we find that

$$
\begin{aligned}
&\sum_{i=1}^n \sum_{j=1}^n w_{ij} b(\varepsilon_j)(a(t_i) - a(\tau_i)) \qquad \text{(A47)} \\
&\quad = O_p\{p^{\frac12}[\textstyle\sum_i d_i^2]^{\frac12} + (\max_i \sum_j |w_{ij}|) \sum_i d_i^2\} \,.
\end{aligned}
$$

Finally expand the second term in (A41) around $d_1 = \cdots = d_n = 0$.

$$
\begin{aligned}
&\sum_{i=1}^n \sum_{j=1}^n w_{ij} a(\tau_i)(b(r_j) - b(\varepsilon_j)) \\
&\quad = -\textstyle\sum_{i=1}^n \sum_{j=1}^n w_{ij} a(\tau_i) d_j Eb'(\varepsilon_1) \\
&\qquad - \textstyle\sum_{i=1}^n \sum_{j=1}^n d_j w_{ij} a(\tau_i)(b'(\varepsilon_j) - Eb'(\varepsilon_j)) \qquad \text{(A48)} \\
&\qquad + \int_0^1 \textstyle\sum_{i,j} w_{ij} a(\tau_i) d_j^2 b''(\varepsilon_j - \lambda d_j)(1 - \lambda)\, d\lambda \\
&\quad = -Eb'(\varepsilon_1) \textstyle\sum_{i=1}^n \sum_{j=1}^n w_{ij} a(\tau_i) d_j \\
&\qquad + O_p\{p^{\frac12}[\textstyle\sum_i d_i^2]^{\frac12} + (\max_i \sum_j |w_{ij}|) \sum_i d_i^2\}
\end{aligned}
$$

by an argument similar to that used for (A42). If we combine our remark following (A41), (A47) and (A48), the result follows. □

5D. *Sketch proofs of Theorem* 3.2 *and* (3.17).

PROOF OF THEOREM 3.2. Note

(i) R3′ $\Rightarrow |b(x) - b(y)| \leqq M|x - y|$, for all x, y.
(ii) Define $\bar{b}(t) = Eb(\varepsilon_1 + t)$. GR3 and GR2 imply $\bar{b}$ is well defined.
(iii) GF3 implies that $\int_{-\infty}^{\infty} |f'(x)|\, dx < \infty$. From GR3 we find that $\bar{b}$ is twice continuously differentiable

$$\bar{b}'(t) = Eb'(\varepsilon_1 + t)$$
$$\bar{b}''(t) = -\int_{-\infty}^{\infty} b'(x)f'(x - t)\, dx$$

and hence

$$|\bar{b}'(t)| \leqq M \qquad \text{for all} \quad t \tag{A49}$$
$$|\bar{b}''(t)| \leqq M \qquad \text{for all} \quad t\,.$$

To establish the theorem we need to prove analogues of (A34)—(A36). Clearly (A34) continues to hold and so does (A35) since its proof only requires a Lipschitz condition on b. The analogue of (A36) is

$$\begin{aligned} &n^{-\frac{1}{2}} \textstyle\sum_{i,j} w_{ij} a(t_i) b(r_j) \\ &\quad = n^{-\frac{1}{2}} \textstyle\sum_{i,j} w_{ij} a(\tau_i) b(\varepsilon_j) - E\bar{b}'(\varepsilon_1) \textstyle\sum_{i,j} w_{ij} a(\tau_i) d_j \\ &\qquad + o_p(n^{\frac{1}{2}} + p(1 + \max_j \textstyle\sum_i |w_{ij}|))\,. \end{aligned} \tag{A50}$$

If we examine the proof of (A36) it is easy to see that in view of (A49), (A50) will follow if we can prove

$$\textstyle\sum_{i,j} w_{ij}(a(t_i) - a(\tau_i)) b(\varepsilon_j) = o_p(p(1 + \max_j \sum_i |w_{ij}|)) \tag{A51}$$

and

$$\textstyle\sum_{i,j} w_{ij} a(\tau_i)(b(r_j) - b(\varepsilon_j) - \bar{b}(r_j) + \bar{b}(\varepsilon_j)) = o_p(n^{\frac{1}{2}})\,. \tag{A52}$$

Assertion (A51) can be established by Taylor expanding as in (A41) but replacing the argument in (A46) by use of the form of the least squares estimates and L_2 estimates as in the proof of Proposition 1.1.

Assertion (A52) is essentially Lemma 4.1 of [6] with b replacing ψ and the c_j replaced by $\sum_i w_{ij} a(\tau_i)$. It is not hard to check that in that lemma the only properties of c_j used are $\max_j |c_j| = o(n^{\frac{1}{2}})$ and $\sum_j c_j^2 = O(n)$ and that the $\sum_i w_{ij} a(\tau_i)$ have these properties. Moreover, condition GL2 is just G of [6] while R3′ implies C1 of [6]. The theorem follows. □

PROOF OF (3.17). Under the assumptions of Theorem 3.1 we can parallel the arguments of Proposition 2.1 without difficulty as long as we operate with $b_{\hat{\sigma}}$. In the end, to prove the theorem we need to establish assertions such as

$$\textstyle\sum_{i,j} w_{ij} a(\tau_i) b\left(\frac{\varepsilon_j}{\hat{\sigma}}\right) = \sum_{i,j} w_{ij} a(\tau_i) b\left(\frac{\varepsilon_j}{\sigma}\right) + o_p(n^{\frac{1}{2}}) \tag{A53}$$

for suitable $||w_{ij}||$. Taylor expansion to two terms around σ will establish (A47) in view of (3.15).

Under the assumptions of Theorem 3.2 we need to prove (A51) and (A52) with b replaced by $b_{\hat{\sigma}}$, $\bar{b}$ replaced by $\bar{b}_{\hat{\sigma}}$. (A52) follows by an appeal to Lemma 4.2 of [6]. As for (A51), we estimate

$$\text{(A54)} \qquad \sum_{i,j} w_{ij}(a(t_i) - a(\tau_i))b\left(\frac{\hat{\varepsilon}_j}{\hat{\sigma}}\right)$$

$$= \sum_{i,j} w_{ij}(a(t_i) - a(\tau_i))b\left(\frac{\hat{\varepsilon}_j}{\sigma}\right) + O_p([\sum d_i^2]^{\frac{1}{2}})$$

using the Lipschitz conditions on b and a and (3.15). The rest of the argument is straightforward. □

REFERENCES

[1] ANDREWS, D. A. (1971). Significance tests based on residuals. *Biometrika* **58** 139-148.
[2] ANSCOMBE, F. J. (1961). Examination of residuals. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 1-36. Univ. of California Press.
[3] ANSCOMBE, F. J. (1967). Topics in the investigation of relations fitted by the method of least squares. *J. Roy. Statist. Soc. Ser. B* **29** 1-52.
[4] ANSCOMBE, F. J. and TUKEY, J. W. (1963). Analysis of residuals. *Technometrics* **5** 141-160.
[5] BICKEL, P. J. (1973). Analogues of linear combinations of order statistics, etc. *Ann. Statist.* **1** 597-616.
[6] BICKEL, P. J. (1975). One step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428-434.
[7] BICKEL, P. J. (1976). Another look at robustness. *Scand. J. Statist.* **3** 145-168.
[8] BICKEL, P. J. and LEHMANN, E. L. (1975-1976). Descriptive statistics for nonparametric models II, III. *Ann. Statist.* **3** 1038-1070 and *Ann. Statist.* **4** 1139-1158.
[9] BOX, G. E. P. (1953). Nonnormality and tests of variances. *Biometrika* **40** 318-335.
[10] CROUSE, C. F. (1964). Note on Mood's test. *Ann. Math. Statist.* **35** 1825-1826.
[11] HÁJEK, J. and SIDÁK, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.
[12] HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73-101.
[13] HUBER, P. J. (1973). Robust regression. *Ann. Statist.* **1** 799-821.
[14] JAECKEL, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43** 1449-1458.
[15] NEYMAN, J. (1958). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics* (U. Grenander, ed.). Wiley, New York.
[16] SCHEFFÉ, H. (1959). *The Analysis of Variance.* Wiley, New York.
[17] TUKEY, J. W. (1950). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics* (I. Olkin, ed.) 448-485. Stanford Univ. Press.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720