The Annals of Statistics
1978, Vol. 6, No. 5, 1012–1026

RANK TESTS OF SUB-HYPOTHESES IN THE GENERAL LINEAR REGRESSION

BY J. N. ADICHIE

University of Nigeria, Nsukka and University of Sheffield

This paper considers the general linear regression model $Y_i = \sum_j \beta_j x_{ij} + \varepsilon_i$, and studies the problem of testing hypotheses about some of the β's while regarding others as nuisance parameters. The test criteria discussed, which are based on ranks of residuals, are shown to be asymptotically distribution-free.

0. Introduction and summary. In the general linear model $Y = X\beta + \varepsilon$, rank methods for testing hypotheses about the entire β (e.g., $\beta = 0$) have been discussed under various regularity conditions by many authors, e.g., Adichie (1967a), Koul (1969). But the methods suggested by these authors do not easily carry over to the case where there are nuisance parameters. However, Koul (1970) proposed a rank order test for $\beta_1 = 0$ in the case where $\beta' = (\beta_1, \beta_2)$ has only two components; see also Puri and Sen (1973).

In this paper we construct and study rank order statistics suitable for testing the general subhypotheses in linear regression models of full rank. Sections 1 and 2 contain the construction of a class of signed-rank and rank test statistics respectively, while in Section 3 the asymptotic distribution of the proposed classes of statistics is established. In Section 4, the asymptotic performance of the proposed test is compared with that of the classical procedure, and in Section 5, the asymptotic optimality of the test is discussed. Finally in Section 6, the general result is applied not only to the problem considered by Koul (1970) but also to the important problem of testing linearity in polynomial regression.

1. Signed-rank test statistics. Consider the general linear model

$$Y = X\beta + \varepsilon\,, \tag{1.1}$$

where Y is an $n \times 1$ vector of independent observations, X is an $n \times p$ matrix of known constants, β is a $p \times 1$ vector of unkown regression parameters such that

$$E(\varepsilon) = 0\,; \qquad E(\varepsilon\varepsilon') = \sigma^2 I_n\,, \quad \sigma > 0 \tag{1.2}$$

where I_n is the identity matrix of order n. It is convenient to write $X = (X_1, X_2)$ so that (1.1) may be put in the form

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon\,, \tag{1.3}$$

Received November 1975; revised August 1977.
AMS 1970 *subject classifications*. Primary 62G10; Secondary 62E20.
Key words and phrases. Asymptotic distribution, asymptotic optimality, contiguous, linear rank statistics.

where X_1 and X_2 are of order $n \times k$ and $n \times (p-k)$ respectively, while β_1 and β_2 are $k \times 1$ and $(p-k) \times 1$ subvectors of β respectively. We want to test

$$H_0\colon \beta_1 = 0\,, \qquad \beta_2 \quad \text{unspecified,} \tag{1.4}$$

against the alternative that $\beta_1 \neq 0$.

The precise functional form of the distribution function $F(y/\sigma)$ of the components of ε need not be known, but in this section we shall assume that it satisfies the following:

ASSUMPTION A. The distribution F has a symmetric density f which is absolutely continuous such that the Fisher information $I(F) = \int (f'/f)^2\, dF$ is finite, where f' denotes derivative.

In what follows, we shall be concerned with sequences of vectors of random variables $\{Y_n\}$ and nonrandom matrices $\{X_n\}$, $n = 1, 2, \cdots$, but for simplicity of notation we shall not emphasize the dependence on n. While all limits are taken as n tends to infinity, the number p of parameters remains fixed. We shall write the design matrix variously as $X = ((x_{ij})) = (x_1, \cdots, x_p)$, where x_j denotes the jth column of X. Here assume that X satisfies the Kraft and van Eeden (1972) conditions, namely:

ASSUMPTION B.

(i) $\{\max_i x_{ij}^2/\sum_i x_{ij}^2\} \to 0$, for each $j = 1, \cdots, p$,
(ii) rank of X, $r(X) = p$,
(iii) $n^{-1}(X'X)$ tends to a positive definite matrix $\Sigma = ((\sigma'_{jj}))$,
(iv) for each pair, j, k ($j \neq k, j, k = 1, \cdots, p$) there exists a number $\gamma_{jk} \neq 0$, such that for $n > n_0$,

(a) $x_{ij}(x_{ij} + \gamma_{jk} x_{ik}) \geqq 0$ for all i
(b) $|x_j|$ and $|x_j + \gamma_{jk} x_k|$ are similarly ordered.

Two vectors u and v are similarly ordered if

$$(u_i - u_s)(v_i - v_s) \geqq 0 \qquad \text{for all} \quad i\,, \quad s\,. \tag{1.5}$$

REMARK. Because of B(iv), the regression model described in this section does not cover the whole class of regression models of full rank that are usually treated by the least squares method.

Now let

$$\psi(i/(n+1)) = \psi_n(i)\,, \qquad i = 1, \cdots, n \tag{1.6}$$

be the scores generated by a function $\psi(u)$ on $(0, 1)$, satisfying the following condition:

ASSUMPTION C. $\psi(u)$ is expressible as a difference between two monotone nonnegative square integrable functions, such that $\int \psi^2(u)\, du > 0$.

For later use, define

$$\psi(u, f) = -(f'/f)(F^{-1}((u+1)/2))\,, \qquad 0 < u < 1\,, \tag{1.7}$$

and note that $I(F)$ defined in Assumption A may also be written as

$$I(F) = \int \psi^2(u, f)\, du . \tag{1.8}$$

We also need an estimate $\tilde{\beta}_2$ of the unspecified parameter β_2. For that introduce the notation $||\beta|| = (\beta'\beta)^{\frac{1}{2}}$, and assume that the estimate satisfies the following two conditions:

ASSUMPTION D.

(i) The term $n^{\frac{1}{2}}||\hat{\beta}_2 - \beta_2||$ is $O_p(1)$ as $n \to \infty$, where p refers to probability under (1.4),

(ii) For all β_2, $\hat{\beta}_2(Y - X_2\beta_2) = \hat{\beta}_2(Y) - \beta_2$, where $\hat{\beta}_2(Y)$ denotes the estimate computed from Y.

Note that the usual least squares estimate computed under H_0 satisfies Assumption D.

Now for each $i = 1, \cdots, n$, set

$$Y_i(\hat{\beta}_2) = \hat{Y}_i = (Y - X_2\hat{\beta}_2)_i \tag{1.9}$$

where the extreme right-hand side of (1.9) denotes the ith component of the residual vector $Y - X_2\hat{\beta}_2$.

Define an n-component vector by

$$\Psi(\hat{\beta}_2) = \{\psi_n(\hat{R}_i) \operatorname{sgn} \hat{Y}_i, i = 1, \cdots, n\}' \tag{1.10}$$

where $\operatorname{sgn} y = 1$ or -1 according as $y >$ or < 0, and $\hat{R}_i$ is the rank of the absolute value $|\hat{Y}_i|$ among $|\hat{Y}_1|, \cdots, |\hat{Y}_n|$. For each $j = 1, \cdots, p$, set

$$s_j(\hat{\beta}_2) = \hat{s}_j = x_j'\Psi(\hat{\beta}_2)/A \tag{1.11}$$

where

$$A^2 = \int \psi^2(u)\, du . \tag{1.12}$$

Also write the vector of statistics in (1.11) as

$$S(\hat{\beta}_2) = \hat{S} = (\hat{s}_1, \cdots, \hat{s}_p)' = X'\Psi(\hat{\beta}_2)/A \tag{1.13}$$

and let $\hat{S}' = (\hat{S}_1', \hat{S}_2')$ be its partition such that

$$\hat{S}_2 = (\hat{s}_{k+1}, \cdots, \hat{s}_p)' = X_2'\Psi(\hat{\beta}_2)/A , \tag{1.14}$$

the signed-rank statistic to be considered, is

$$\begin{aligned} M(\hat{\beta}_2) = \hat{M} &= \hat{S}'(X'X)^{-1}\hat{S} - \hat{S}_2'(X_2'X_2)^{-1}\hat{S} \\ &= \Psi'(\hat{\beta})W\Psi(\hat{\beta})/A^2 , \end{aligned} \tag{1.15}$$

where W is a symmetric idempotent matrix of order $n \times n$ defined by

$$W = X(X'X)^{-1}X' - X_2(X_2'X_2)^{-1}X_2' . \tag{1.16}$$

Observe that W is orthogonal with X_2 in the sense that

$$WX_2 = 0 . \tag{1.17}$$

Furthermore

$$WX_1 = \{I_n - X_2(X_2'X_2)^{-1}X_2'\}X_1 . \tag{1.18}$$

This property of orthogonality of W and X_2 which is crucial in the distribution theory of the least squares criterion, will also be very useful (see proof of Lemma 3.1 below) in the distribution theory of our rank statistics. It is primarily to achieve this orthogonality (and avoid imposing the unnecessary condition $X_1'X_2 = 0$) that motivates our use of W as the weighting function. It will be shown in Section 3 that $\hat{M}$ provides an asymptotically distribution-free statistic for testing the hypothesis (1.4). The test rejects the hypothesis if $\hat{M}$ is large. In order to consider the asymptotic power performance of $\hat{M}$, it will be necessary to find its limiting distribution not only under (1.4) but also under a sequence of Pitman alternatives:

$$H_n : \beta_1 = n^{-\frac{1}{2}}b_1 , \quad \|b_1\| < C . \tag{1.19}$$

We now state the main theorem, the proof of which is given in Section 3.

THEOREM 1.1. *Under Assumptions* A—D

$$\lim P_0(\hat{M} \leqq y) = P(\chi_k^2 \leqq y) \tag{1.20}$$

$$\lim P_n(\hat{M} \leqq y) = p(\chi_k^2(\Delta_{\hat{M}}) \leqq y) \tag{1.21}$$

where $\chi_k^2(\Delta_{\hat{M}})$ denotes the chi-square random variable with k degrees of freedom and noncentrality parameter.

$$\Delta_{\hat{M}} = \lim n^{-1}\{b_1'X_1'WX_1b_1\}K_F^2(\psi) \tag{1.22}$$

and

$$K_F(\psi) = \textstyle\int \psi(u)\psi(u, f)\, du/A \tag{1.23}$$

while P_0 and P_n denote probabilities under (1.4) *and* (1.19) *respectively.*

2. Rank test statistics. Rank statistics, as different from signed-rank statistics of Section 1, may also be used to construct the test statistic in the case where the design matrix X satisfies a set of assumptions specified in B_1 below. Such rank tests are briefly discussed in this section.

Consider now the model

$$Y = X\theta + \varepsilon \tag{2.1}$$

where Y is an $n \times 1$ vector of independent observations, X is an $n \times p_1$ design matrix, θ is a $p_1 \times 1$ vector of unknown regression parameters and ε satisfies (1.2). Rewrite (2.1) as

$$Y = X_1\theta_1 + X_2\theta_2 + \varepsilon \tag{2.2}$$

where X_1 and X_2 are of order $n \times k$ and $n \times (p_1 - k)$ respectively, while θ_1 and θ_2 are subvectors of θ. The problem is to test

$$H_0 : \theta_1 = 0 , \qquad \theta_2 \quad \text{unspecified} \tag{2.3}$$

against the alternative that $\theta_1 \neq 0$.

The common distribution function $F(y/\sigma)$ of the components of ε shall be in this section assumed to satisfy:

ASSUMPTION A_1. The distribution F has a density f which is absolutely continuous such that the Fisher information $I(F)$ is finite. As for the design matrix X, let

$$Z = X - \bar{X} = ((x_{ij} - \bar{x}_j)) = ((z_{ij})) ,$$

where $\bar{x}_j = n^{-1} \sum_i x_{ij}$, and let $Z = (Z_1, Z_2)$ correspond to $X = (X_1, X_2)$. We shall also write $Z = (z_1, \cdots, z_{p_1})$, and assume that X is such that Z satisfies the Kraft and van Eeden (1972) conditions, namely:

ASSUMPTION B_1.

(i) $\max_i \{z_{ij}^2 / \sum_i z_{ij}^2\} \to 0$, for each $j = 1, \cdots, p_1$;
(ii) rank of Z, $r(Z) = p_1$;
(iii) $n^{-1}(Z'Z)$ tends to a positive definite matrix $\Sigma^* = ((\sigma_{jj'}^*))$;
(iv) for each pair j, k ($j \neq k, j, k = 1, \cdots, p_1$), there exists a number $\gamma_{jk} \neq 0$ such that for $n > n_0$, z_j and $z_j + \gamma_{jk} z_k$ are similarly ordered.

REMARKS.

1. Because $r(X) \leqq r(Z) + r(\bar{X})$ and $r(X) \leqq p_1$, B_1(ii) will be satisfied only for some X in (2.1) for which $r(X) < p_1 + 1$, i.e., for some X with full rank p_1 or less. A particular class of X for which B_1(ii) holds is any orthogonal design matrix with $\bar{x}_1 = \cdots = \bar{x}_{p_1}$. Observe on the other hand that B(ii) of Section 1 holds for all X of full rank.
2. Because of B_1(ii), (iii), and (iv), the rank score method described in this section cannot be used in all linear models of full rank where the least squares method usually succeeds.
3. The testing procedure considered in this section is also valid under Jurečková (1971) conditions on X.

We shall require that the scores

$$\phi(i/(n+1)) = \phi_n(i) , \qquad i = 1, \cdots, n \tag{2.4}$$

are generated by a function $\phi(u)$ on (0, 1) that satisfies

ASSUMPTION C_1. $\phi(u)$ is nonconstant and is expressible as a difference between two montone square integrable functions on (0, 1). Put

$$A^2(\phi) = \int (\phi(u) - \bar{\phi})^2 \, du ; \quad \bar{\phi} = \int \phi(u) \, du . \tag{2.5}$$

As in (1.7) define for later use the function

$$\phi(u, f) = -(f'/f)(F^{-1}(u)) , \qquad 0 < u < 1 , \tag{2.6}$$

and set

$$K_F(\phi) = \int \phi(u)\phi(u, f) \, du / A(\phi) . \tag{2.7}$$

The estimate $\tilde{\theta}_2$ of the unspecified θ_2 is required to satisfy two conditions:

ASSUMPTION D_1.

(i) The term $n^{\frac{1}{2}}\|\tilde{\theta}_2 - \theta_2\|$ is $O_p(1)$ as $n \to \infty$; where p refers to probability under (2.3);

(ii) For all θ_2, $\tilde{\theta}_2(Y - Z_2\theta_2) = \tilde{\theta}_2(Y) - \theta_2$, where $\tilde{\theta}_2(Y)$ denotes the estimate computed from Y.

For each i set

$$Y_i(\tilde{\theta}_2) = \tilde{Y}_i = (Y - Z_2\tilde{\theta}_2)_i$$

and define an n-component vector by

$$\Phi(\tilde{\theta}_2) = \{\phi_n(\tilde{R}_i), i = 1, \cdots, n\}' \tag{2.8}$$

where $\tilde{R}_i$ is the rank of $\tilde{Y}_i$ in the ranking of n variables $\tilde{Y}_1, \cdots, \tilde{Y}_n$. Observe that the vector $\{\tilde{R}_i, i = 1, \cdots, n\}$ remains unchanged if instead of $\tilde{Y}_i$ we rank $(Y - X_2\tilde{\theta}_2)_i$, $i = 1, \cdots, n$. Now write

$$S(\tilde{\theta}_2) = \tilde{S} = (\tilde{s}_1, \cdots, \tilde{s}_{p_1})' = Z'\Phi(\tilde{\theta}_2)/A(\phi) \tag{2.9}$$

where

$$s_j(\tilde{\theta}_2) = \tilde{s}_j = z_j'\Phi(\tilde{\theta}_2)/A(\phi)\,, \qquad j = 1, \cdots, p_1\,. \tag{2.10}$$

Form the partition $\tilde{S} = (\tilde{S}_2'\tilde{S}_2')$, and the proposed rank statistics can then be written as

$$\begin{aligned} M(\tilde{\theta}_2) = \tilde{M} &= \tilde{S}'(Z'Z)^{-1}\tilde{S} - \tilde{S}_2'(Z_2'Z_2)^{-1}\tilde{S}_2 \\ &= \Phi'(\tilde{\theta}_2)\tilde{W}\Phi(\tilde{\theta}_2)/A^2(\phi)\,, \end{aligned} \tag{2.11}$$

where $\tilde{W}$ is the symmetric idempotent matrix of order $n \times n$ obtained from (1.16) by writing Z instead of X. In using $M(\tilde{\theta})$ for the testing problem, the hypothesis (2.3) is rejected for large values of $\tilde{M}$.

If we consider a sequence of alternatives

$$H_n: \theta_1 = n^{-\frac{1}{2}}\theta_1\,, \quad \|\theta_1\| < c\,, \tag{2.12}$$

we can state the main result of this section, which is analogous to the result of Section 1, as follows:

THEOREM 2.1. *Under the Assumptions* A_1—D_1,

$$\lim P_0(\tilde{M} \leqq y) = P(\chi_k^2 \leqq y) \tag{2.13}$$

$$\lim P_n(\tilde{M} \leqq y) = P(\chi_k^2(\Delta_M^*) \leqq y) \tag{2.14}$$

where $\chi_k^2(\Delta_M^*)$ *is the chi-square random variable with k degrees of freedom and noncentrality parameter*

$$\Delta_M^* = \lim n^{-1}\{\theta_1'Z_1'\tilde{W}Z_1\theta_1\}K_F^2(\phi) \tag{2.15}$$

while P_0 *and* P_n *denote probabilities under* (2.3) *and* (2.12) *respectively.*

3. Proofs of theorems. It is to be noticed that the $\hat{s}_j$'s defined in (1.11) are not the ordinary linear rank statistics because the residuals $\hat{Y}_i$ in (1.9) are not independent random variables. Now let $M(\hat{\beta}_2)$, see (1.15), be the signed-rank statistic formed from ranks of absolute values of the unobservable random variables $Y_i(\beta_2) = (Y - \chi_2\beta_2)_i$, $i = 1, \cdots, n$. We now prove

LEMMA 3.1. *If the assumptions of Theorem* 1.1 *hold then* $M(\hat{\beta}_2)$ *and* $M(\beta_2)$ *have the same limiting distribution under* H_0 *and* H_n *of* (1.4) *and* (1.19) *respectively.*

PROOF. Under H_0, $Y(\beta_2)$ is a vector of independent identically distributed random variables while $Y(\hat{\beta}_2) = Y(\beta_2) - X_2(\hat{\beta}_2 - \beta_2)$. By Assumption D(ii), we may take $\beta_2 = 0$. By D(i), there exists a number K such that $P\{\|\hat{\beta}_2\| \leqq n^{-\frac{1}{2}}K\}$ is arbitrarily close to one for all $n > n_0$. It follows that for each $j = 1, \cdots, p$, the quantity

$$n^{-\frac{1}{2}}|s_j(\hat{\beta}_2) - s_j(0) + x_j'X_2\hat{\beta}_2K_F(\psi)|$$

will be with arbitrarily high probability bounded by

$$\sup_{\|b_2\| \leqq n^{-\frac{1}{2}}K} |n^{-\frac{1}{2}}\{s_j(b_2) - s_j(0) + x_j'\chi_2 b_2 K_F(\psi)\}| .$$

But by Theorem 7.2 of Kraft and van Eeden (1972), as $n \to \infty$,

$$\sup_{\|b_2\| \leqq n^{-\frac{1}{2}}K} \|n^{-\frac{1}{2}}\{S(b_2) - S(0) + X'X_2K_F(\psi)\}\| = o_p(1) , \tag{3.0}$$

so that under D(i) as $n \to \infty$,

$$\|n^{-\frac{1}{2}}\{S(\hat{\beta}_2) - S(0) + X'X_2\hat{\beta}_2K_F(\psi)\}\| = o_p(1) . \tag{3.1}$$

Observing that $\hat{S}(X'X)^{-1}\hat{S}$ may be written as $(n^{-\frac{1}{2}}\hat{S})'\{n(X'X)^{-1}\}(n^{-\frac{1}{2}}\hat{S})$, and that $n(X'X)^{-1} \to \Sigma$, it follows from (1.15) and (3.1) that the difference between $M(\hat{\beta}_2)$ and

$$\begin{aligned} S'(0)(X'X)^{-1}S(0) &- \hat{\beta}_2'(X_2'X)(X'X)^{-1}S(0)K_F(\psi) \\ &- S'(0)(X'X)^{-1}(X'X_2)\hat{\beta}_2K_F(\psi) \\ &+ \hat{\beta}_2'(X_2'X)(X'X)^{-1}X'X_2\hat{\beta}_2K_F^2(\psi) - S_2'(X_2'X_2)^{-1}S_2(0) \\ &+ \hat{\beta}_2'S_2(0)K_F(\psi) + S_2'(0)\hat{\beta}_2K_F(\psi) - \hat{\beta}_2'X_2'X_2\hat{\beta}_2K_F^2(\psi) , \end{aligned} \tag{3.2}$$

converges to zero in probability.

Now from the identity $\{I - X(X'X)^{-1}X'\}X = 0$ we have

$$X_2 - X(X'X)^{-1}X'X_2 = 0 . \tag{3.3}$$

On writing

$$S(0) = X'\Psi(0)/A , \qquad S_2(0) = X_2'\Psi(0)/A \tag{3.4}$$

as in (1.13) and (1.14), and making repeated use of (3.3) and (3.4), the quantity in (3.2) reduces to

$$\begin{aligned} S'(0)(X'X)^{-1}S(0) &- \hat{\beta}_2'X_2'\Psi(0)K_F(\psi)/A - \Psi'(0)X_2\hat{\beta}_2K_F(\psi)/A \\ &+ \hat{\beta}_2'X_2'X_2\hat{\beta}_2K_F^2(\psi) - S_2'(0)(X_2'X_2)^{-1}S_2(0) \\ &+ \hat{\beta}_2'X_2'\Psi(0)K_F(\psi)/A + \Psi'(0)X_2\hat{\beta}_2K_F(\psi)/A - \hat{\beta}_2'X_2'X_2\hat{\beta}_2K_F^2(\psi) , \end{aligned}$$

which is easily seen to be equal to

$$S'(0)(X'X)^{-1}S(0) - S_2'(0)(X_2'X_2)^{-1}S_2(0) = M(0) .$$

That $M(\hat{\beta}_2)$ and $M(\beta_2)$ have the same limiting distribution under H_n follows from the fact that the sequence of distributions under H_n is contiguous to that under H_0. The proof of the lemma is thus complete.

For the asymptotic distribution of $M(\beta_2)$ it is convenient using well-known transformations to rewrite the matrix W, and hence $M(\beta_2)$; so put

(3.5) $$X = LB\,; \qquad X_2 = LB_2$$

where B is a $p \times p$ upper triangular matrix with positive diagonal elements, L is an $n \times p$ semi-orthogonal matrix and B_2 is a $p \times (p-k)$ matrix with

(3.6) $$(X'X) = (B'B)\,; \qquad X_2'X_2 = B_2'B_2 \colon L'L = I_p\,.$$

On applying this transformation, W reduces to

(3.7) $$W = L\{I_p - DD'\}L' = LVL'\,,$$

where we have written D for $B_2(B_2'B_2)^{-\frac{1}{2}}$. Because V is symmetric and idempotent, if we write the matrix

(3.8) $$L = ((l_{ij})) = (l_1, \cdots, l_p)\,,$$

and define

(3.9) $$t_j(\beta_2) = l_j'\Psi(\beta_2)/A \qquad j = 1, \cdots, p\,,$$

the statistic $M(\beta_2)$ can be written as

(3.10) $$M(\beta_2) = T'(\beta_2)VT(\beta_2)\,,$$

where

$$T'(\beta_2) = (t_1(\beta_2), \cdots, t_p(\beta_2))'\,.$$

We prove

LEMMA 3.2. *Let L be as defined in* (3.5). *If X satisfies Assumption* B(i) *and* B(iii), *then*

(3.11) $$\lim \{\max_i l_{ij}^2/\textstyle\sum_i l_{ij}^2\} = 0\,, \qquad j = 1, \cdots, p\,.$$

PROOF. First note that

(3.12) $$\textstyle\sum_i l_{ij}^2 = 1\,, \qquad j = 1, \cdots, p\,.$$

Furthermore, Assumptions B(i) and B(ii) together imply

(3.13) $$\lim \{\max_i \textstyle\sum_j x_{ij}^2/n\} = 0\,,$$

and

(3.14) $$\lim \textstyle\sum_i (x_{ij}^2/n) = \sigma_j^2\,, \qquad 0 < \sigma_j^2 < \infty, j = 1, \cdots, p\,.$$

It is also known (see, e.g., Albert (1966), page 1606), that B(iii) implies

(3.15) $$\{\lambda_{\max}(X'X)/\lambda_{\min}(X'X)\} < K\,,$$

where $\lambda_{\max}(\lambda_{\min})$ denotes maximum (minimum) characteristic root. Now from

(3.5) and (3.6), we have, using Schwarz's inequality,

$$l_{ij}^2 = (\sum_k x_{ik} b_{kj})^2 \leqq \sum_k x_{ik}^2 \sum_k b_{kj}^2 ,$$

where $B^{-1} = ((b_{ij}))$ and the summation over k is from 1 to p. Now

$$\begin{aligned} \sum_k b_{kj}^2 &\leqq \sum_k \sum_j b_{kj}^2 = \operatorname{tr}(X'X)^{-1} = \sum_k \{1/\lambda_k(X'X)\} \\ &\leqq p/\lambda_{\min}(X'X) \leqq Kp^2/\sum_k \lambda_k(X'X) , \quad \text{using (3.15)} \\ &= Kp^2/\sum_i \sum_j x_{ij}^2 , \quad \text{so that} \quad l_{ij}^2 \leqq Kp^2 \sum_k x_{ik}^2/\sum_i \sum_j x_{ij}^2 . \end{aligned}$$

The maximum over i of the right-hand side tends to zero because of (3.13) and (3.14). This fact together with (3.12) proves (3.11).

PROOF OF THEOREM 1.1. In view of Lemma 3.1, we restrict attention to $M(\beta_2)$ as defined in (3.10). Due to Assumption C and (3.11), it follows in the same way as in Hájek and Šidák (1967), page 166, that under H_0, $t_j(\beta_2)$ defined in (3.9) is asymptotically $N(0, 1)$, for each j. From the way $t_j(\beta_2)$ is defined, any linear combination $\sum_j \lambda_j t_j(\beta_2)$ is again a linear rank statistic whose weights $\sum_j \lambda_j l_{ij}$ satisfy (3.11). Hence under H_0, $T(\beta_2)$ is asymptotically normally distributed with mean zero and covariance matrix I_p. As for the statistic $M(\beta_2)$ of (3.10), we may from (3.7) write

$$T'(\beta_2)T(\beta_2) = T'(\beta_2)DD'T(\beta_2) + T'(\beta_2)VT(\beta_2) , \tag{3.16}$$

where both DD' and V are idempotent matrices with rank $p - k$ and k respectively. Furthermore, it is clear from Assumption B(iii) that $DD' = (n^{-\frac{1}{2}}B_2)n(B_2'B_2)^{-1}(n^{-\frac{1}{2}}B_2')$ tends to a $p \times p$ matrix, while V by definition also tends to a limiting $p \times p$ matrix. Because $T'(\beta_2)$ is asymptotically normal, and the matrices DD' and V are idempotent, it follows from a well-known theorem on distribution of quadratic forms (see, e.g., Theorem 4.16 of Graybill (1961)), that under H_0 the quadratic form $T'(\beta_2)VT(\beta_2)$ has asymptotically a chi-square distribution with k degrees of freedom. This together with (3.10) and Lemma 3.1 proves (1.20).

To prove (1.21), note that Lemma 3.1 is valid under H_n of (1.19). It follows in the same way as in Theorem VI, 2.5 page 220 of [7] that under (1.19), $T(\beta_2)$ still has a limiting normal distribution with the same covariance matrix I_p, but with different mean vector μ given by

$$\mu = \lim n^{-\frac{1}{2}}(L'X_1 b_1)K_F(\psi) .$$

From Theorem 4.16 of [6], it follows that under (1.19) $M(\beta_2)$ has asymptotically a noncentral chi-square distribution with k degrees of freedom, and noncentrality parameter $\mu' V\mu$, which in view of (3.7) reduces to $\Delta_{\hat{M}}$ given in (1.22). The proof is thus complete.

PROOF OF THEOREM 2.1. The proof, which depends on Theorem 7.1 of [11], is omitted because it is similar to the proof of Theorem 1.1.

4. Asymptotic relative efficiency. If the model given in (1.1) and (1.2) is of full rank and if the distribution of ε is normal, the usual statistic for testing (1.4) is based on the maximum likelihood ratio

$$Q = (n-p)D_1/kD_0\,, \tag{4.1}$$

where

$$D_1 + D_0 = (Y - X_2\bar{\bar{\beta}}_2)'(Y - X_2\bar{\bar{\beta}}_2)\,, \tag{4.2}$$

and

$$D_0 = (Y - X\bar{\beta})'(Y - X\bar{\beta}) = Y'\{I_n - X(X'X)^{-1}X'\}Y \tag{4.3}$$

with $\bar{\bar{\beta}}_2$ and $\bar{\beta}$ being the least squares estimates of β under (1.4) and (1.1) respectively. In this setup, Q has the variance ratio distribution with $(k, n-p)$ degrees of freedom, and the test that rejects H_0 for large values of Q is the most powerful invariant test. When the basic assumption of normality of ε is dropped, Q loses its optimality and its exact distribution is not even known. However, for any marginal distribution F of the components of ε, for which the variance $\sigma^2 = \sigma^2(F)$ is finite, it can be shown that $|n^{-1}D_0 - \sigma^2| = o_p(1)$, as $n \to \infty$ (see, for example, Theorem 3.4 of Gleser (1966), where a stronger result is proved). Furthermore, on setting $Y(\beta_2) = Y - X_2\beta_2$, D_1 may be written as

$$\begin{aligned} D_1 &= Y'WY = Y'(\beta_2)WY(\beta_2) + \beta_2'X_2'WX_2\beta_2 \\ &= Y'(\beta_2)WY(\beta_2)\,, \quad \text{due to (1.17)}\,, \\ &= Y'(\beta)LVL'Y(\beta_2) \quad \text{by (3.7)}\,, \end{aligned}$$

where $Y'(\beta_2)L = (\sum_i l_{ij}Y_i(\beta_2), j = 1, \cdots, p)$. It follows from (3.11) (see, for example, Theorem 3 of Gnedenko and Kolmogorov (1954), page 103) that under H_0, $L'Y(\beta_2)/\sigma$ is asymptotically normal with mean zero and covariance matrix I_p, and under H_n, of (1.19), has asymptotic mean $n^{-\frac{1}{2}}L'X_1b_1/\sigma$. We have therefore proved the following

THEOREM 4.1. *If the components of* ε *in model* (1.1) *and* (1.2) *have common distribution function* $F(y/\sigma)$ *with* $0 < \sigma < \infty$, *then*

$$\lim P_0(kQ \leqq y) = P(\chi_k^2 \leqq y.)$$
$$\lim P_n(kQ \leqq y) = P(\chi_k^2(\Delta_Q) \leqq y)$$

where P_0, P_n, *and* $\chi_k^2(\Delta)$ *are as defined in Theorem* 2.1 *and*

$$\Delta_Q = \lim n^{-1}\{b_1'X_1'WX_1b_1\}/\sigma^2\,. \tag{4.4}$$

Thus Q provides an asymptotically distribution-free test for the class of F for which $\sigma^2(F) < \infty$.

By the conventional method of measuring the relative asymptotic efficiency of two test statistics that have chi-square distributions with the same degree of freedom, it follows from (1.22) and (4.4) that the asymptotic efficiency of $\hat{M}$ relative to the least squares criterion is

$$e_{\hat{M},Q} = \Delta_{\hat{M}}/\Delta_Q = \sigma^2K_F^2(\psi)\,, \tag{4.5}$$

which is the standard asymptotic efficiency of rank score tests relative to the t-test in the two-sample problem.

The results of this section hold for the rank statistic $\tilde{M}$ of (2.11), if Assumption A_1 through D_1 hold. More precisely, the asymptotic efficiency of $\tilde{M}$ relative to the least squares criterion Q computed with Z instead of X, and ϕ instead of ψ is

$$e_{\tilde{M},Q} = \sigma^2 K_F{}^2(\phi) . \tag{4.6}$$

5. Asymptotic optimality. If the functional form of F is known, the asymptotic performance of the $\hat{M}$-tests can be improved upon. To be specific, suppose that in addition to Assumption A of Section 1, F satisfies Assumptions I—V of Wald (1943) viz:

ASSUMPTION A*.

(i) The maximum likelihood estimates $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2)'$ exist and are uniformly consistent.

(ii) $f(y, \beta)$ is twice differentiable with respect to β and $f''(y, \beta)$ is continuous in β, where $f(y, b)$ denotes $f((y - \sum_j b_j x_j)/\sigma)$.

(iii) Let $h(y, \beta)$ denote $((f''/f) - (f'/f)^2)(y, \beta)$.

(a) For any sequences $\{\beta_{n1}\}$, $\{\beta_{n2}\}$, and δ_n such that $\lim \beta_{n1} = \lim \beta_{n2} = \beta$, and $\delta_n \to 0$, we have $\lim E_{\beta_{n1}}\{\sup h(Y, \beta)\} = \lim E_{\beta_{n1}}\{\inf h(Y, \beta)\} = I(F) < \infty$ where the sup (inf) is over β in $|\beta - \beta_{n2}| \leqq \delta_n$.

(b) There exists $\varepsilon > 0$, such that $E_{\beta_1}\{\sup h(Y, \beta)\}$ and $E_{\beta_1}\{\inf h(Y, \beta)\}$ are bounded for $\|\beta_1 - \beta_2\| < \varepsilon$ and $|\delta| < \varepsilon$ where the sup (inf) is over β in $\|\beta - \beta_2\| < \delta$.

(iv) $f(y, \beta)$ is twice differentiable with respect to β under the integral sign.

(v) There exists $n > 0$, such that $E_\beta|(f'/f)(Y, \beta)|^{2+n}$ is bounded.

For testing (1.4) on the basis of n observations Y of model (1.1), Wald's test statistic ((115) of Wald (1943), page 457) becomes

$$\begin{aligned} W_n{}^* &= \tilde{\beta}_1'[X_1'X_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1]\tilde{\beta}_1 I(F) \\ &= \tilde{\beta}_1'X_1'WX_1\tilde{\beta}_1 I(F) , \quad \text{in view of (1.18)} . \end{aligned} \tag{5.1}$$

The test rejects (1.4) for large values of $W_n{}^*$. To study the optimality of $W_n{}^*$, define a surface $S_c(b)$ by

$$\begin{aligned} S_c(b) = \{b: \; & b_1'[X_1'X_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1]b_1 I(F) = c, \\ & b_2 = \beta_2 - \gamma_{22}^{-1}\gamma_{21} b_1\} \end{aligned} \tag{5.2}$$

where γ_{21}, γ_{22} are parts of a partitioned $p \times p$ nonsingular matrix

$$\gamma = \begin{pmatrix} \gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \tag{5.3}$$

satisfying $\gamma(X'X)\gamma' = I_p$.

Also consider the transformation $b^* = \gamma b$ where γ is as defined in (5.3). This transformation transforms the surface $S_c(b)$ into a sphere $S_c'(b)$ given by

$$b_1^{*\prime} b_1^* = C , \qquad b_2^* = \gamma_{21} b_2 + \gamma_{22} b_2 .$$

Finally, for any point b_0 and any $\delta > 0$ consider the set $\omega(b_0, \delta)$ consisting of all points b which lie on the same $S_c(b)$ as b_0 and for which $|b - b_0| < \delta$. Let

$$\eta(b) = \lim_{\delta \to 0} \{A(\omega'(b, \delta))/A(\omega(b, \delta))\} , \tag{5.4}$$

where $\omega'(b, \delta)$ is the image of $\omega(b, \delta)$ by the transformation $b^* = \gamma b$, and $A(\omega)$ denotes the area of the set ω.

Collecting together Theorems IV, V, and VI (pages 459, 461, and 462) of Wald (1943), we have

THEOREM 5.1 (Wald). *Let* $S_c(b)$ *be the surface defined in* (5.2), *and* $\eta(b)$ *the weight function in* (5.4). *If Assumptions* A*, B(i), *and* B(ii) *hold, then for testing* (1.4), *the* W_n^**-test given in* (5.1)

(a) *has asymptotically best average power with respect to* $S_c(b)$ *and* $\eta(b)$,
(b) *has asymptotically best constant power on* $S_c(b)$,
(c) *is an asymptotically most stringent test.*

For the definitions of the asymptotic optimality in (a), (b), and (c) of the above theorem, see Definitions VIII, X, and XII at pages 453, 454, and 455 respectively of Wald (1943).

Now let $L_n = -2 \log \lambda_n$, where λ_n is the likelihood ratio statistic for testing (1.4). It is shown in Wald ((1943), page 478, (199)), that under the conditions of Theorem 5.1, and on the assumption that the L_n-test is uniformly consistent (Assumption VII, page 472 of [13]),

$$W_n^* + 2 \log \lambda_n \to 0 \quad \text{in } P_\beta\text{-probability, uniformly in} \quad \beta , \tag{5.5}$$

where P_β denotes probability under the assumption that β is the true parameter point.

It follows from (5.5) and Theorem 5.1 that the L_n-test has the same asymptotic optimality properties as W_n^*. Furthermore it is proved in Theorem IX, page 480 of Wald (1943), that if Assumptions A*, B(i), and B(ii) hold, and the L_n-test is uniformly consistent, then under (1.4), L_n (or W_n^*) has asymptotically a chi-square distribution with k degrees of freedom and under (1.19) has asymptotically a noncentral chi-square distribution with k degrees of freedom and noncentrality parameter

$$\begin{aligned} \Delta_L &= \lim n^{-1} b_1'[X_1'X_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1]b_1 I(F) \\ &= \lim n^{-1} b_1' X_1 W X_1 b_1 I(F) . \end{aligned} \tag{5.6}$$

Now on comparing our signed-rank test statistic $\hat{M}$ with L_n, it follows from (1.22) and (5.6) that the asymptotic efficiency of $\hat{M}$ relative to L_n is

$$e_{\hat{M}, L} = \Delta_{\hat{M}}/\Delta_L = K_F^2(\psi)/I(F) \tag{5.7}$$

which is unity if $K_F{}^2(\psi) = I(F)$, and from (1.8), (1.12), and (1.23), this equation holds if $\psi(u) = \psi(u, f)$. Thus given F that satisfies Assumptions A and A* and for which L_n is uniformly consistent, if we choose $\psi(u) = \psi(u, f)$, the method described in Section 1 will yield an asymptotically optimal test in the sense that the asymptotic efficiency of $\hat{M}_{\text{opt.}}$, the resulting signed-rank test statistic relative to (Theorem 5.1) asymptotically optimal test L_n, is

$$e_{\hat{M}_{\text{opt.}}, L} = 1 . \tag{5.8}$$

A similar result holds for the rank test $\hat{M}$ of (2.11) if Assumptions A*, A_1—D_1, and uniform consistency of L_n hold, and we take $\phi(u)$ of Assumption C_1 to be $\phi(u, f)$ defined in (2.6) and compute the L_n statistic with Z instead of X.

6. Application and example. First let us apply the method of rank statistic of Section 2 to the testing problem considered by Koul (1970), i.e., testing $\theta_1 = 0$ in the model defined in (2.1) and (1.2) with $p_1 = 2$ and $k = 1$. We then have

$$\tilde{s}_j = \textstyle\sum_i z_{ij}\phi_n(\tilde{R}_i)/A(\phi) , \quad z_{ij} = (x_{ij} - \bar{x}_j) \qquad j = 1, 2 , \tag{6.1}$$

where the estimate $\tilde{\theta}_2$ used in obtaining the ranks $\tilde{R}_i$ of $Y_i(\tilde{\theta}_2)$, $i = 1, \cdots, n$, could be either the least squares estimate or the estimate considered by Puri and Sen (1973), since each of them satisfies Assumption D_1. The statistic given in (2.11) may now be written as

$$\tilde{M} = |Z'Z|^{-1}\{\tilde{s}_1{}^2 z_2'z_2 - 2\tilde{s}_1\tilde{s}_2 z_1'z_2 + \tilde{s}_2{}^2 z_1'z_1\} - \tilde{s}_2{}^2/z_2'z_2 . \tag{6.2}$$

Observe that Koul's statistic is $n^{-1}\sum_i x_{i1}\phi_n(\tilde{R}_i)$, which is equivalent to $\tilde{s}_1$ given in (6.1). However, if $z_1'z_2 = 0$, our test can be based on $\tilde{s}_1'(z_1'z_1)^{-1}\tilde{s}_1 = \tilde{s}_1{}^2/z_1'z_1$, and Koul's test is a special case of this. Note also that $z_1'z_2 = 0$ is one of the sufficient conditions for Koul's test to be asymptotically distribution-free (see Lemma 2.4 of Koul (1970)).

Now, on using the transformation (3.7) as it applies to $\tilde{W}$ of (2.11), we have

$$B^* = \begin{pmatrix} b_{11} & b_{12} \\ 0 & b_{22} \end{pmatrix} ,$$

where

$$b_{11}^2 = z_1'z_1 ; \qquad b_{12}^2 = (z_1'z_2)^2/z_1'z_1 ; \qquad b_{22}^2 = z_2'z_2 - b_{12}^2 . \tag{6.3}$$

With this, $\tilde{M}$ reduces to $\tilde{T}'V^{*\prime}\tilde{T}$, with $\tilde{T}' = (\tilde{t}_1, \tilde{t}_2)$ where $V^* = \{I_2 - D^*D^{*\prime}\}$, $\tilde{t}_1 = \tilde{s}_1/b_{11}$, $\tilde{t}_2 = (\tilde{s}_2/b_{22}) - (b_{12}/b_{11}b_{22})\tilde{s}_1$ and D^* is just the second column of B^*. Under the conditions of Section 2, $\tilde{M}$ has asymptotically a chi-square distribution with one degree of freedom, whether or not $z_1'z_2 = 0$. The noncentrality parameter $\Delta_M{}^*$ defined in (2.15) reduces in this case to

$$\lim n^{-1}\theta_1{}^2\{b_{11}^2 - (z_1'z_2)^2/z_2'z_2\}K_F{}^2(\phi) . \tag{6.4}$$

The test is of course consistent, since Assumption C_1 does not require symmetric $\phi(u)$ (see Theorem 2 of [12]).

Secondly, the method described in Section 1 could be used to test subhypotheses in polynomial regression models provided the powers of x's satisfy Assumption B. More precisely, consider the model

$$Y_i = \alpha + \beta x_i + \gamma x_i^2 + \varepsilon\,, \qquad i = 1, \cdots, n \tag{6.5}$$

which is the same as the one in (1.3) which $p = 3$, $k = 1$. Here interest is on testing $H_0\colon \gamma = 0$. The matrices $(X'X)^{-1}$ and $(X_2'X_2)^{-1}$ in the definition of $\hat{M}$ (1.15) are inverses of

$$(X'X) = \begin{pmatrix} n & \sum x & \sum x^2 \\ \sum x & \sum x^2 & \sum x^3 \\ \sum x^2 & \sum x^3 & \sum x^4 \end{pmatrix}$$

and $(X_2'X_2)$ which is the first principal minor of $(X'X)$.

To see what Assumption B means in this example consider a replicated design in which for each n, $x_1, \cdots, x_n$ take a fixed set of values $x_1^0, \cdots, x_c^0$ with frequencies $n_1^0, \cdots, n_c^0$. Let $\gamma_{ni}^0 = (n_i^0/n)$; then it is easy to see that Assumptions B(i), B(ii), and B(iii) are satisfied if (a) $\max_{1 \leq 1 \leq c} |x_i^0| < K$; (b) for each n, $\gamma_{ni}^0 < 1$, $i = 1, \cdots, c$; (c) n_i^0 and n tend of infinity such that $\gamma_{ni}^0 \to \gamma_i^0 < 1$.

To use $\hat{M}$, we need estimates of α and β that satisfy Assumption D. These could be either the least squares estimates or the rank estimates defined in [2] computed under H_0. It is not difficult to check that the least squares estimates of α and β in the model (6.5) with $\gamma = 0$ satisfy D. That the "rank" estimates also satisfy D(ii) is a consequence of Lemma 4.1 of [2].

The three basic signed rank statistics needed for the definition of $\hat{M}$ are:

$$s_j(\hat{\alpha}, \hat{\beta}) = \hat{s}_j = \textstyle\sum_i x_i^{j-1} \psi_n(\hat{R}_i) \operatorname{sgn} \hat{Y}_i/A\,, \qquad j = 1, 2, 3\,.$$

Acknowledgment. I would like to thank the referee whose liberal comments led to a substantial improvement. I am also grateful to Professor I. Richard Savage for his many valuable suggestions.

REFERENCES

[1] ADICHIE, J. N. (1967a). Asymptotic efficiency of a class of nonparametric tests for regression parameters. *Ann. Math. Statist.* **38** 884–893.

[2] ADICHIE, J. N. (1967b). Estimates of regression parameters based on rank tests. *Ann. Math. Statist.* **38** 894–904.

[3] ALBERT, A. (1966). Fixed size confidence ellipsoids for linear regression parameters. *Ann. Math. Statist.* **37** 1602–1630.

[4] GLESER, L. J. (1966). Correction to "On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters." *Ann. Math. Statist.* **37** 1053–1055.

[5] GNEDENKO, B. V. and KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Cambridge.

[6] GRAYBILL, F. (1961). *An Introduction to Linear Statistical Models*, **1**. McGraw-Hill.

[7] HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

[8] JUREČKOVÁ, J. (1971). Nonparametric estimate of of regression coefficients. *Ann. Math. Statist.* **42** 1328–1338.

[9] KOUL, H. L. (1969). Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. *Ann. Math. Statist.* **40** 1950–1979.

[10] KOUL, H. L. (1970). A class of ADF tests for subhypotheses in the multiple linear regression. *Ann. Math. Statist.* **41** 1273–1281.

[11] KRAFT, C. H. and VAN EEDEN, C. (1972). Linearised rank estimates and signed-rank estimates for the general linear hypothesis. *Ann. Math. Statist* **43** 42–57.

[12] PURI, M. L. and SEN, P. K. (1973). A note on asymptotic distribution-free tests for subhypotheses in multiple linear regression. *Ann. Statist.* **1** 553–556.

[13] WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426–482.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NIGERIA
NSUKKA CAMPUS, NIGERIA