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 $Y_i = \sum \beta_j x_{ij} + \varepsilon_i$, and studies the problem of testing hypotheses about some of the $\beta$'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 $\beta$ (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 $\beta_i = 0$ 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

(1.1) $Y = X\beta + \varepsilon$,

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

(1.2) $E(\varepsilon) = 0$; $E(\varepsilon\varepsilon') = \sigma^2 I_n$, $\sigma > 0$

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

(1.3) $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$,

Received November 1975; revised August 1977.
Key words and phrases. Asymptotic distribution, asymptotic optimality, contiguous, linear rank statistics.
where $X_i$ and $X_j$ are of order $n \times k$ and $n \times (p - k)$ respectively, while $\beta_1$ and $\beta_2$ are $k \times 1$ and $(p - k) \times 1$ subvectors of $\beta$ respectively. We want to test
\begin{equation}
H_0: \beta_1 = 0, \quad \beta_2 \text{ unspecified},
\end{equation}
against the alternative that $\beta_1 \neq 0$.

The precise functional form of the distribution function $F(y|\sigma)$ of the components of $\varepsilon$ 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 $\{X_n\}$, $n = 1, 2, \ldots$, 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 $X = (x_{ij}) = (x_1, \ldots, x_p)$, where $x_j$ denotes the $j$th column of $X$. Here assume that $X$ satisfies the Kraft and van Eeden (1972) conditions, namely:

**Assumption B.**

(i) $\{\max_i x_{ij}/\sum_i x_{ij}\} \rightarrow 0$, for each $j = 1, \ldots, p$,

(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, $j$, $k$ ($j \neq k$, $j, k = 1, \ldots, p$) there exists a number $\gamma_{jk} \neq 0$, such that for $n > n_0$,

(a) $x_{ij}(x_{ij} + \gamma_{jk}x_{ik}) \geq 0$ for all $i$

(b) $|x_{ij}|$ and $|x_{ij} + \gamma_{jk}x_{ik}|$ are similarly ordered.

Two vectors $u$ and $v$ are similarly ordered if
\begin{equation}
(u_i - u_s)(v_i - v_s) \geq 0 \quad \text{for all} \quad i, \quad s.
\end{equation}

**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
\begin{equation}
\psi(i)(n + 1) = \phi_n(i), \quad i = 1, \ldots, n
\end{equation}
be the scores generated by a function $\phi(u)$ on $(0, 1)$, satisfying the following condition:

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

For later use, define
\begin{equation}
\phi(u, f) = -(f'/f)(F^{-1}((u + 1)/2)), \quad 0 < u < 1,
\end{equation}
and note that \( I(F) \) defined in Assumption A may also be written as

\[
I(F) = \int \phi'(u, f) \, du.
\]

We also need an estimate \( \hat{\beta}_2 \) of the unspecified parameter \( \beta_2 \). For that introduce the notation \( \| \beta \| = (\beta' \beta)^{1/2} \), and assume that the estimate satisfies the following two conditions:

**Assumption D.**

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

(ii) For all \( \beta_2 \), \( \hat{\beta}_2(Y - X_2 \beta_2) = \hat{\beta}_2(Y) - \beta_2 \), 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 \( i = 1, \ldots, n \), set

\[
Y_i(\hat{\beta}_2) = \hat{Y}_i = (Y - X_2 \hat{\beta}_2)_i,
\]

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

Define an \( n \)-component vector by

\[
\Psi(\hat{\beta}_2) = \{ \phi_i(\hat{R}_i) \operatorname{sgn} \hat{Y}_i, i = 1, \ldots, n \}^t
\]

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

\[
s_j(\hat{\beta}_2) = \hat{s}_j = x_j' \Psi(\hat{\beta}_2)/A
\]

where

\[
A^2 = \int \phi^2(u) \, du.
\]

Also write the vector of statistics in (1.11) as

\[
S(\hat{\beta}_2) = \hat{S} = (\hat{s}_1, \ldots, \hat{s}_p)^t = X' \Psi(\hat{\beta}_2)/A
\]

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

\[
\hat{S}'_2 = (\hat{s}_{k+1}, \ldots, \hat{s}_p)^t = X_2' \Psi(\hat{\beta}_2)/A,
\]

the signed-rank statistic to be considered, is

\[
M(\hat{\beta}_2) = \hat{M} = \hat{S}'(X'X)^{-1} \hat{S} - \hat{S}'_2(X_2'X_2)^{-1} \hat{S}_2
\]

\[
= \Psi(\hat{\beta})W\Psi(\hat{\beta})/A^2,
\]

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

\[
W = X(X'X)^{-1}X' - X_2(X_2'X_2)^{-1}X_2'.
\]

Observe that \( W \) is orthogonal with \( X_2 \) in the sense that

\[
WX_2 = 0.
\]
Furthermore

\[(1.18) \quad WX_1 = \{ I_n - X_2(X_2'X_2)^{-1}X_2' \}X_1. \]

This property of orthogonality of \( W \) and \( X_1 \) 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:

\[(1.19) \quad H_n: \hat{\beta}_1 = n^{-1}b_1, \quad \|b_1\| < C. \]

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

**Theorem 1.1.** Under Assumptions A—D

\[(1.20) \quad \lim P_n(\hat{M} \leq y) = P(\chi^2_k \leq y) \]

\[(1.21) \quad \lim P_n(\hat{M} \leq y) = p(\chi^2_k(\Delta_n) \leq y) \]

where \( \chi^2_k(\Delta_n) \) denotes the chi-square random variable with \( k \) degrees of freedom and noncentrality parameter.

\[(1.22) \quad \Delta_n = \lim n^{-1}b_1'WX_1b_1/3^2(\hat{\phi}) \]

and

\[(1.23) \quad K_{0}(\hat{\phi}) = \int \psi(u)\psi(u, f) \, du/4 \]

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, below. Such rank tests are briefly discussed in this section.

Consider now the model

\[(2.1) \quad Y = X\theta + \varepsilon \]

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

\[(2.2) \quad Y = X_1\beta_1 + X_2\beta_2 + \varepsilon \]

where \( X_1 \) and \( X_2 \) are of order \( n \times k \) and \( n \times (p - k) \) respectively, while \( \beta_1 \) and \( \beta_2 \) are subvectors of \( \beta \). The problem is to test

\[(2.3) \quad H_0: \beta_1 = 0, \quad \beta_2 \text{ unspecified} \]

against the alternative that \( \beta_1 \neq 0 \).
The common distribution function $F(y|\sigma)$ of the components of $\varepsilon$ shall be in this section assumed to satisfy:

**Assumption A.** 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

$$Z = X - \bar{X} = ((x_{ij} - \bar{x}_{ij})) = ((\bar{z}_{ij})),$$

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

**Assumption B.**

(i) $\max \{z_{ij}/\sum_i z_{ij}^2\} \rightarrow 0$, for each $j = 1, \ldots, p$;

(ii) rank of $Z$, $r(Z) = p$;

(iii) $n^{-1}(Z'Z)$ tends to a positive definite matrix $\Sigma^* = (\sigma_{ij}^*)$;

(iv) for each pair $j, k$ ($j \neq k, j, k = 1, \ldots, p$), 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 $r(X) \leq r(Z) + r(\bar{X})$ and $r(X) \leq p$, B(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(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(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

$$\phi(i/(n + 1)) = \phi_n(i), \quad i = 1, \ldots, n$$

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

**Assumption C.** $\phi(u)$ is nonconstant and is expressible as a difference between two monotone square integrable functions on $(0, 1)$. Put

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

As in (1.7) define for later use the function

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

and set

$$K_p(\phi) = \int \phi(u)\phi(u, f) \, du/A(\phi).$$

The estimate $\hat{\theta}_2$ of the unspecified $\theta_2$ is required to satisfy two conditions:
ASSUMPTION D₁.

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

(ii) For all \( \theta_s \), \( \tilde{\theta}_s(Y - Z_s \theta_s) = \tilde{\theta}_s(Y) - \theta_s \), where \( \tilde{\theta}_s(Y) \) denotes the estimate computed from \( Y \).

For each \( i \) set

\[
Y_i(\tilde{\theta}_s) = \hat{Y}_i = (Y - Z_s \tilde{\theta}_s)_i
\]

and define an \( n \)-component vector by

\[
\Phi(\tilde{\theta}_s) = (\phi_{s_i}(\hat{R}_i), i = 1, \ldots, n)'
\]

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

\[
S(\tilde{\theta}_s) = \bar{S} = (s_1, \ldots, s_p)' = Z' \Phi(\tilde{\theta}_s)/A(\phi)
\]

where

\[
s_j(\tilde{\theta}_s) = s_j = z_j' \Phi(\tilde{\theta}_s)/A(\phi), \quad j = 1, \ldots, p.
\]

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

\[
M(\tilde{\theta}_s) = \bar{M} = \bar{S}'(Z'Z)^{-1}\bar{S} - \bar{S}'(Z'Z)^{-1}\bar{S} = \Phi'(\tilde{\theta}_s)\tilde{W} \Phi(\tilde{\theta}_s)/A'(\phi),
\]

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

If we consider a sequence of alternatives

\[
H_n: \theta = n^{-1} \theta_1, \quad ||\theta|| < c,
\]

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 A₁—D₁,

\[
\lim P_n(\bar{M} \leq y) = P(\chi_k^2 \leq y)
\]

\[
\lim P_n(\bar{M} \leq y) = P(\chi_k^2(\Delta_n^*) \leq y)
\]

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

\[
\Delta_n^* = \lim n^{-1}(\theta_1'Z_s', Z_s'\tilde{W}Z_s' \theta_1)K_s(\phi)
\]

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

3. **Proofs of theorems.** It is to be noticed that the \( s_j \)'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 \( M(\hat{\beta}_\lambda) \), see (1.15), be the signed-rank statistic formed from ranks of absolute values of the unobservable random variables \( Y_i(\hat{\beta}_\lambda) = (Y - x_i\hat{\beta}_\lambda), i = 1, \ldots, n. \) We now prove

**Lemma 3.1.** If the assumptions of Theorem 1.1 hold then \( M(\hat{\beta}_\lambda) \) and \( M(\beta_\lambda) \) have the same limiting distribution under \( H_0 \) and \( H_n \) of (1.4) and (1.19) respectively.

**Proof.** Under \( H_0 \), \( Y(\hat{\beta}_\lambda) \) is a vector of independent identically distributed random variables while \( Y(\beta_\lambda) = Y(\beta_\lambda) - X_i(\beta_\lambda - \beta_\lambda). \) By Assumption D (ii), we may take \( \beta_\lambda = 0. \) By D (i), there exists a number \( K \) such that \( P(||\hat{\beta}_\lambda|| \leq n^{-1}K) \) is arbitrarily close to one for all \( n > n_0. \) It follows that for each \( j = 1, \ldots, p, \) the quantity

\[
n^{-1}||s_j(\hat{\beta}_\lambda) - s_j(0) + x_j'X_j\hat{\beta}_\lambda K_p(\psi)||^2
\]

will be with arbitrarily high probability bounded by

\[
\sup_{||b|| \leq n^{-1}K} ||n^{-1}||s_j(b) - s_j(0) + x_j'X_jb||K_p(\psi)||^2.
\]

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

\[
(3.0) \quad \sup_{||b|| \leq n^{-1}K} ||n^{-1}||s_j(b) - s_j(0) + x_j'X_jb||K_p(\psi)||^2 = o_p(1),
\]

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

\[
(3.1) \quad ||n^{-1}||s_j(\hat{\beta}_\lambda) - s_j(0) + X_j'X_j\hat{\beta}_\lambda K_p(\psi)||^2 = o_p(1).
\]

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

\[
S'(0)(X'X)^{-1}S(0) - \hat{\beta}'(X'X)(X'X)^{-1}S(0)K_p(\psi)
\]

(3.2)

\[
= S'(0)(X'X)^{-1}(X'X\hat{\beta}_\lambda)K_p(\psi)
\]

\[
+ \hat{\beta}'(X'X)(X'X)^{-1}X'X\hat{\beta}_\lambda K_p(\psi) - S'(0)X'X\hat{\beta}_\lambda K_p(\psi)
\]

\[
+ \hat{\beta}'S(0)K_p(\psi) + S'(0)\hat{\beta}_\lambda K_p(\psi) - \hat{\beta}'X'X\hat{\beta}_\lambda K_p(\psi),
\]

converges to zero in probability.

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

\[
X_2 - X(X'X)^{-1}X'X_2 = 0.
\]

(3.3)

On writing

\[
S(0) = X'\Psi(0)/A, \quad S(0) = X_2'\Psi(0)/A
\]

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

\[
S'(0)(X'X)^{-1}S(0) - \hat{\beta}'X_2'\Psi(0)K_p(\psi)/A - \Psi'(0)X'X\hat{\beta}_\lambda K_p(\psi)/A
\]

\[
+ \hat{\beta}'X'X\hat{\beta}_\lambda K_p(\psi) - S'(0)X_2'X_2\hat{\beta}_\lambda K_p(\psi)
\]

\[
+ \hat{\beta}'X_2'\Psi(0)K_p(\psi)/A + \Psi'(0)X_2'\hat{\beta}_\lambda K_p(\psi)/A - \hat{\beta}'X_2'X_2\hat{\beta}_\lambda K_p(\psi),
\]

which is easily seen to be equal to

\[
S'(0)(X'X)^{-1}S(0) - S'(0)X_2X_2^{-1}S(0) = M(0).
\]
That $M(\hat{\beta}_2)$ and $M(\hat{\beta}_3)$ 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 $M(\hat{\beta}_2)$ it is convenient using well-known transformations to rewrite the matrix $W$, and hence $M(\hat{\beta}_2)$; so put

\begin{equation}
X = LB; \quad X_s = LB_s
\end{equation}

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

\begin{equation}
(X'X) = (B'B); \quad X_s'X_s = B_s'B_s; \quad L'L = I_p.
\end{equation}

On applying this transformation, $W$ reduces to

\begin{equation}
W = L[I_p - DD']L' = LVL',
\end{equation}

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

\begin{equation}
L = ((l_{ij})) = (l_1, \ldots, l_p),
\end{equation}

and define

\begin{equation}
t_j(\hat{\beta}_2) = l_j' \Psi(\hat{\beta}_2)/A \quad j = 1, \ldots, p,
\end{equation}

the statistic $M(\hat{\beta}_2)$ can be written as

\begin{equation}
M(\hat{\beta}_2) = T'(\hat{\beta}_2)VT(\hat{\beta}_2),
\end{equation}

where

\begin{equation}
T'(\hat{\beta}_2) = (t_1(\hat{\beta}_2), \ldots, t_p(\hat{\beta}_2))'.
\end{equation}

We prove

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

\begin{equation}
\lim \{\max_i l_{ij}/\sum_i l_{ij}\} = 0, \quad j = 1, \ldots, p.
\end{equation}

**Proof.** First note that

\begin{equation}
\sum_i l_{ij} = 1, \quad j = 1, \ldots, p.
\end{equation}

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

\begin{equation}
\lim \{\max_i \sum_j x_{ij}^2/n\} = 0,
\end{equation}

and

\begin{equation}
\lim \sum_i (x_{ij}^2/n) = \sigma_j^2, \quad 0 < \sigma_j^2 < \infty, \quad j = 1, \ldots, p.
\end{equation}

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

\begin{equation}
\{\lambda_{\text{max}}(X'X)/\lambda_{\text{min}}(X'X)\} < K,
\end{equation}

where $\lambda_{\text{max}}(\lambda_{\text{min}})$ denotes maximum (minimum) characteristic root. Now from
(3.5) and (3.6), we have, using Schwarz's inequality,

\[ l_{ij} = (\sum_k x_{ik} b_{kj})^2 \leq \sum_k x_{ik}^2 \sum_k b_{kj}^2 , \]

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

\[
\sum_k b_{kj}^2 \leq \sum_k \sum_j b_{kj}^2 = \text{tr} (X'X)^{-1} = \sum_k \{1/\lambda_k(X'X)\}
\leq p/\lambda_{\text{min}}(X'X) \leq Kp^2/\sum_k \lambda_k(X'X) , \quad \text{using} \ (3.15)
\]

\[ = Kp^2/\sum_j x_{ij}^2 , \quad \text{so that} \quad R_{ij} \leq Kp^2 \sum_k x_{ik}^2/\sum_j x_{ij}^2 . \]

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 \( M(\hat{\beta}_2) \) 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 \( H_0, t_j(\hat{\beta}_2) \) defined in (3.9) is asymptotically \( N(0, 1) \), for each \( j \). From the way \( t_j(\hat{\beta}_2) \) is defined, any linear combination \( \sum_j \lambda_j t_j(\hat{\beta}_2) \) is again a linear rank statistic whose weights \( \sum_j \lambda_j \) satisfy (3.11). Hence under \( H_0, T(\hat{\beta}_2) \) is asymptotically normally distributed with mean zero and covariance matrix \( I_p \). As for the statistic \( M(\hat{\beta}_2) \) of (3.10), we may fact with (3.7) write

\[ T'(\hat{\beta}_2)T(\hat{\beta}_2) = T'(\hat{\beta}_2)DD'T(\hat{\beta}_2) + T'(\hat{\beta}_2)VT(\hat{\beta}_2) , \]

where both \( DD' \) and \( V \) are idempotent matrices with rank \( p - k \) and \( k \) respectively. Furthermore, it is clear from Assumption B(iii) that \( DD' = (n^{-1}B_2nB_2)^{-1}(n^{-1}B_2') \) tends to a \( p \times p \) matrix, while \( V \) by definition also tends to a limiting \( p \times p \) matrix. Because \( T'(\hat{\beta}_2) \) 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 \( T'(\hat{\beta}_2)VT(\hat{\beta}_2) \) 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 \( H_\circ \) of (1.19). It follows in the same way as in Theorem VI, 2.5 page 220 of [7] that under (1.19), \( T(\hat{\beta}_2) \) still has a limiting normal distribution with the same covariance matrix \( I_p \), but with different mean vector \( \mu \) given by

\[ \mu = \lim n^{-1}(L'X, b_\circ)K_\phi(\psi) . \]

From Theorem 4.16 of [6], it follows that under (1.19) \( M(\hat{\beta}_2) \) 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 \( \Delta_\circ \) 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 \( \varepsilon \) is normal, the usual statistic for testing (1.4) is based on the maximum likelihood ratio

\[
Q = (n - p)D_1/kD_0 ,
\]

where

\[
D_1 + D_0 = (Y - X_2 \hat{\beta}_2)'(Y - X_2 \hat{\beta}_2),
\]

and

\[
D_0 = (Y - X \hat{\beta})'(Y - X \hat{\beta}) = Y'[I_n - X(X'X)^{-1}X']Y
\]

with \( \hat{\beta}_2 \) and \( \hat{\beta} \) being the least squares estimates of \( \beta \) 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 \( \varepsilon \) is dropped, \( Q \) loses its optimality and its exact distribution is not even known. However, for any marginal distribution \( F \) of the components of \( \varepsilon \), 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 \( Y(\hat{\beta}_2) = Y - X_2 \hat{\beta}_2 \), \( D_1 \) may be written as

\[
D_1 = Y'WY = Y'(\hat{\beta}_2)WY(\hat{\beta}_2) + \hat{\beta}_2'X_2'WX_2 \hat{\beta}_2
\]

\[
= Y'(\hat{\beta}_2)LVY(\hat{\beta}_2), \quad \text{due to (1.17),}
\]

\[
= Y'(\hat{\beta}_2)LVL'Y(\hat{\beta}_2) \quad \text{by (3.7),}
\]

where \( Y'(\hat{\beta}_2)L = (\sum_l I_{ij} Y_i(\hat{\beta}_2), j = 1, \ldots, p) \). It follows from (3.11) (see, for example, Theorem 3 of Gnedenko and Kolmogorov (1954), page 103) that under \( H_0 \), \( L'Y(\hat{\beta}_2)/\sigma \) is asymptotically normal with mean zero and covariance matrix \( I_p \), and under \( H_a \), of (1.19), has asymptotic mean \( n^{-1}L'X_1b_1/\sigma \). We have therefore proved the following

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

\[
\lim P_0(kQ \leq y) = P(X^2_{\nu} \leq y),
\]

\[
\lim P_n(kQ \leq y) = P(X^2_{\nu}(\Delta) \leq y)
\]

where \( P_0 \), \( P_n \), and \( X^2_{\nu}(\Delta) \) are as defined in Theorem 2.1 and

\[
\Delta_Q = \lim n^{-1}[b_1'X_1'WX_1b_1]/\sigma^2.
\]

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 \( M \) relative to the least squares criterion is

\[
eff_{\hat{\beta}, Q} = \Delta_{\hat{\beta}}/\Delta_Q = \sigma^2 K_F(\psi),
\]
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 $\hat{M}$ of (2.11), if Assumption A$_1$ through D$_3$ hold. More precisely, the asymptotic efficiency of $\hat{M}$ relative to the least squares criterion $Q$ computed with $Z$ instead of $X$, and $\phi$ instead of $\phi$

\[
e_{\hat{M},Q} = \sigma^2 K_p^2(\phi).
\]

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 $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)'$ exist and are uniformly consistent.

(ii) $f(y, \hat{\beta})$ is twice differentiable with respect to $\beta$ and $f''(y, \hat{\beta})$ is continuous in $\hat{\beta}$, where $f(y, b)$ denotes $f((y - \sum_i b_i x_i)/\sigma)$.

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

(a) For any sequences $\{\hat{\beta}_{n1}\}, \{\hat{\beta}_{n2}\}$, and $\delta_n$ such that $\lim \hat{\beta}_{n1} = \lim \hat{\beta}_{n2} = \beta$, and $\delta_n \to 0$, we have $\lim E_{\hat{\beta}_{n1}}[\sup h(Y, \hat{\beta})] = \lim E_{\hat{\beta}_{n1}}[\inf h(Y, \hat{\beta})] = I(F) < \infty$

where the sup (inf) is over $\hat{\beta}$ in $|\hat{\beta} - \beta_{n1}| \leq \delta_n$.

(b) There exists $\epsilon > 0$, such that $E_{\hat{\beta}_1}[\sup h(Y, \hat{\beta})]$ and $E_{\hat{\beta}_1}[\inf h(Y, \hat{\beta})]$ are bounded for $||\beta_1 - \beta_2|| < \epsilon$ and $|\delta| < \delta$ where the sup (inf) is over $\beta$ in $||\hat{\beta} - \beta|| < \delta$.

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

(v) There exists $n > 0$, such that $E_{\hat{\beta}}[(f''/f)(Y, \hat{\beta})]^{2n+1}$ 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

\[
W_n^* = \hat{\beta}_1'[X'_1 X_1 - X'_1 X_2 (X_2' X_2)^{-1} X_2' X_1] \hat{\beta}_1 I(F)
\]

\[
= \hat{\beta}_1' X'_1 W X_1 \hat{\beta}_1 I(F), \text{ in view of (1.18)}.
\]

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

\[
S_a(b) = \{b : b_1'[X'_1 X_1 - X'_1 X_2 (X_2' X_2)^{-1} X_2' X_1] b_1 I(F) = c, b_2 = \beta_2 - \gamma_{21} \gamma_{11} b_1 \}
\]

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

\[
\gamma = \begin{pmatrix} \gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \end{pmatrix}
\]

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

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

\[
b_1^*b_2^* = C, \quad b_2^* = \gamma_1b_2 + \gamma_2b_2.
\]

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 \( S_\alpha(b) \) as \( b_0 \) and for which \( |b - b_0| < \delta \). Let

\[
\gamma(b) = \lim_{\delta \to 0} \left\{ \frac{A(\omega'(b, \delta))}{A(\omega(b, \delta))} \right\},
\]

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 \( \omega \).

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

**Theorem 5.1 (Wald).** Let \( S_\alpha(b) \) be the surface defined in (5.2), and \( \gamma(b) \) 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 \( S_\alpha(b) \) and \( \gamma(b) \),
(b) has asymptotically best constant power on \( S_\alpha(b) \),
(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 \( \lambda_n \) 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]),

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

where \( P_\beta \) denotes probability under the assumption that \( \beta \) 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

\[
\Delta_L = \lim n^{-1}b_1'\left[ X_1'X_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1 \right]b_1I(F)
= \lim n^{-1}b_1'XWb_1I(F).
\]

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

\[
e_{\tilde{M}, L} = \Delta_M/\Delta_L = K_F(\psi)/I(F)
\]
which is unity if \( K_F^*(\psi) = I(F) \), 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 \( L_u \) 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 \( L_u \), is

\[
e_{\hat{M}_{\text{opt}}, L} = 1.
\]

A similar result holds for the rank test \( \hat{M} \) of (2.11) if Assumptions \( A^*, A_1-D_1 \), and uniform consistency of \( L_u \) hold, and we take \( \psi(u) \) of Assumption \( C_1 \) to be \( \phi(u, f) \) defined in (2.6) and compute the \( L_u \) 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

\[
s_j = \sum_i z_{ij} \phi_n(\hat{R}_i)/A(\phi), \quad z_{ij} = (x_{ij} - \bar{x})_j = 1, 2,
\]

where the estimate \( \hat{R}_i \) used in obtaining the ranks \( \hat{R}_i \) of \( Y_i(\hat{\theta}_i), i = 1, \ldots, n \), 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

\[
\tilde{M} = |Z'|Z^{-1}[\tilde{s}_1^2\tilde{z}_1'\tilde{z}_2 - 2\tilde{s}_1\tilde{s}_2\tilde{z}_1'\tilde{z}_2 + \tilde{s}_2^2\tilde{z}_1'\tilde{z}_1'] - \tilde{s}_2^2/\tilde{z}_2'\tilde{z}_2.
\]

Observe that Koul's statistic is \( n^{-1} \sum_i x_{ii} \phi_n(\hat{R}_i) \), which is equivalent to \( s_i \) given in (6.1). However, if \( z_1'z_2 = 0 \), our test can be based on \( s_i(z_1'z_1)^{-1}s_i = s_i^2/z_1'z_1 \), 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

\[
b_{11}^2 = z_1'z_1; \quad b_{12}^2 = (z_1'z_2)^2/z_1'z_1; \quad b_{22}^2 = z_2'z_2 - b_{22}^2.
\]

With this, \( \tilde{M} \) reduces to \( T'V^*T \), with \( T = (i_1, i_2) \) where \( V^* = [I_z - D^*D^*'] \), \( i_1 = \tilde{s}_1/b_{11} \), \( i_2 = (\tilde{s}_2/b_{22}) - (b_{12}/b_{11}b_{22})\tilde{s}_1 \) 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 \( \Delta^*_n \) defined in (2.15) reduces in this case to

\[
\lim n^{-1}e_{\hat{M}, L}^2[b_{11}^2 - (z_1'z_2)^2/z_2'z_2]K_F^2(\phi).
\]

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 sub-
hyphoteses in polynomial regression models provided the powers of \( x \)'s satisfy
Assumption B. More precisely, consider the model

\[
Y_i = \alpha + \beta x_i + \gamma x_i^2 + \varepsilon, \quad i = 1, \ldots, n
\]

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

\[
(X'X) = \begin{pmatrix}
\sum_{i=1}^n x_i \\
\sum_{i=1}^n x_i^2 \\
\sum_{i=1}^n x_i^3
\end{pmatrix}
\]

and \((X'_s X_s)\) 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 \( n, x_1, \ldots, x_n \) take a fixed set of values \( x_1, \ldots, x_n \) with frequ-
cencies \( n_1, \ldots, n_n \). Let \( \gamma_{ni} = (n_i/n) \); then it is easy to see that Assumptions
B(i), B(ii), and B(iii) are satisfied if (a) \( \max_{(i \leq j \leq n)} |x_i - x_j| < K \); (b) for each \( n,
\gamma_{ni} < 1, i = 1, \ldots, c \); (c) \( n \to \infty \) and \( n \) tend of infinity such that \( \gamma_{ni} \to \gamma_i < 1 \).

To use \( \tilde{M} \), we need estimates of \( \alpha \) and \( \beta \) 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 \( \alpha \) and \( \beta \) 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 \( \tilde{M} \) are:

\[
s_j(\tilde{\alpha}, \tilde{\beta}) = \delta_j = \sum_i x_i \Phi_n(\tilde{Y}_i/A), \quad j = 1, 2, 3.
\]

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

Math. Statist. 38 894–904.
pendent Random Variables. Addison-Wesley, Cambridge.
Statist. 42 1328–1338.


Department of Statistics
University of Nigeria
Nsukka Campus, Nigeria