RANK SCORE COMPARISON OF SEVERAL REGRESSION PARAMETERS

By J. N. ADICHIE

University of Nigeria, Nsukka

For testing $\beta_i = \beta$, $i = 1, \dots, k$, in the model $Y_{ij} = \alpha + \beta_i x_{ij} + Z_{ij}$ $j = 1, \dots, n_i$ a class of rank score tests is presented. The test statistic is based on the simultaneous ranking of all the observations in the different k samples. Its asymptotic distribution is proved to be chi square under the hypothesis and noncentral chi square under an appropriate sequence of alternatives. The asymptotic efficiency of the given procedure relative to the least squares procedure is shown to be the same as the efficiency of rank score tests relative to the t-test in the two sample problem. The proposed criterion would be an asymptotically most powerful rank score test for the hypothesis if the distribution function of the observations is known.

- **0.** Introduction and summary. For the regression model $Y_{ij} = \alpha_i + \beta_i x_{ij} + Z_{ij}$, $j = 1, \dots, n_i$; $i = 1, \dots, k$ where Z_{ij} are independent random variables, a rank score method of testing that $\beta_i = \beta$ for all i, while α_i are nuisance parameters, was recently studied by Sen (1969). His test criteria are based on the individual ranks of the k different samples. In this paper, we show that for the special case where $\alpha_i = \alpha$ (unknown), suitable rank score tests for $\beta_i = \beta$ may be based on the simultaneous ranking of all the observations.
- 1. Notations and preliminaries. For each $i=1,\dots,k$, let $Y_{ij},\,j=1,\dots,n_i$ be independent random variables with continuous distribution functions F_{ij} given by

(1.1)
$$P(Y_{ij} \leq y) = F_{ij}(y) = F(y - \alpha - \beta_i x_{ij}).$$

The precise functional form of $F(\cdot)$ is not assumed to be known. Here x_{ij} 's are known regression constants, α is a nuisance parameter and the β_i 's are the quantities of interest. Our problem is to test the hypothesis

$$(1.2) H_0: \beta_i = \beta \quad (unknown)$$

against the set of alternatives that β_1, \dots, β_k are not all equal.

To simplify the notation, let us write

(1.3)
$$x_{i\bullet} = n_i^{-1} \sum_j x_{ij};$$
 $C_{ni}^2 = \sum_j (x_{ij} - x_{i\bullet})^2;$ $\gamma_{ni} = \{C_{ni}/C_n\}^2$ $i = 1, \dots, k$

$$(1.4) C_n^2 = \sum_i C_{ni}^2; n = \sum_i n_i$$

where for the summations above, and in fact throughout the rest of the paper,

Received October 1972; revised March 1973.

AMS 1970 subject classifications. Primary 62G10, 62G20; Secondary 62E20, 62J05.

Key words and phrases. Simultaneous ranking, score generating function, bounded in probability, orthogonal transformation, asymptotic normality, asymptotic efficiency.

j goes from 1 to n_i ; i (or s) goes from 1 to k. We shall also write

(1.5)
$$c_{sj}^{(i)} = \gamma_{ni}(x_{sj} - x_{s*}) \qquad s = 1, \dots, i-1, i+1, \dots, k$$
$$= (\gamma_{ni} - 1)(x_{sj} - x_{s*}) \qquad s = i$$

so that

(1.6)
$$\bar{c}^{(i)} = n^{-1} \sum_{s} \sum_{i} c_{si}^{(i)} = 0 \qquad i = 1, \dots, k$$

and

It is assumed that each of the C_{ni}^2 (and hence C_{ni}^2) tends to infinity with n such that

(1.8)
$$\lim \gamma_{ri} = \gamma_i; \qquad 0 < \gamma_0 \le \gamma_1, \dots, \gamma_k \le (1 - \gamma_0) < 1$$

where $\gamma_0 < 1/k$.

Let $\psi(u)$, 0 < u < 1, be a smooth non-decreasing function, and let the scores generated by ψ be defined by

(1.9)
$$a_n(p) = \psi\{p/(n+1)\} \qquad p = 1, \dots, n.$$

Also let R_{ij} be the rank of Y_{ij} in the combined ranking of all the *n* observations. We shall need an estimate of β in (1.2); for that purpose, let

$$(1.10) S_n(Y) = \sum_{i} \sum_{j} (x_{ij} - x_{i\bullet}) a_n(R_{ij}).$$

As in Adichie (1967), define the estimate of β based on (1.10) as follows;

(1.11)
$$\beta_n^* = \sup\{b : S_n(Y - bx) > 0\}; \qquad \beta_n^{**} = \inf\{b : S_n(Y - bx) < 0\}$$

$$\hat{\beta}_n = \frac{1}{2}(\beta_n^* + \beta_n^{**})$$

where $S_n(Y = bx)$ denotes the statistic (1.10) when the observations Y_{ij} are replaced by $(Y_{ij} = bx_{ij})$. The estimate defined in (1.12) is consistent in the sense to be made precise in Lemma 2.1 below. Now write $\hat{Y}_{ij} = (Y_{ij} - \hat{\beta}x_{ij})$, and let \hat{R}_{ij} be the ranks of \hat{Y}_{ij} . Define

(1.13)
$$\hat{T}_{ni} = T_{ni}(\hat{Y}) = \sum_{s} \sum_{j} c_{sj}^{(i)} a_{n}(\hat{R}_{sj});$$

the proposed test statistic is

$$\hat{L}_n = \sum_i (\hat{T}_{ni}/AC_{ni})^2$$

where

(1.15)
$$A^{2} = \int \psi^{2}(u) du - \{ \int \psi(u) du \}^{2} .$$

2. Asymptotic distribution of \hat{L}_n . In order to study the asymptotic power properties of the proposed \hat{L}_n test, we shall establish the limiting distribution not only under the hypothesis H_0 but also under the sequence of alternatives,

$$(2.1) H_n: \beta_{ni} = \beta + (\theta_i/C_n), |\theta_i| \leq M_2, i = 1, \dots, k.$$

Now set $Y_{ij}^0 = (Y_{ij} - \beta x_{ij})$, and let R_{ij}^0 be the ranks of Y_{ij}^0 . Also let T_{ni}^0 and L_{ni}^0

be the T_{ni} - and L_n -statistics associated with Y_{ij}^0 . The proof of the limiting distribution of \hat{L}_n depends on the following lemmas.

LEMMA 2.1. Let $\hat{\beta}_n$ be as defined in (1.12). Assume that the score generating function $\psi(u)$, and the regression constants x_{ij} satisfy conditions (i), (iii) and (iv) of Lemma 2.2 below. Then as $n \to \infty$,

(2.2) $|C_n(\hat{\beta}_n - \beta)|$ is bounded in both P_0 and P_n probabilities,

where P_0 and P_n denote probabilities under (1.2) and (2.1) respectively.

PROOF. The proof of boundedness of (2.2) under P_n is similar to that of Lemma 3.1 of Sen (1969).

Lemma 2.2. Let the score generating function $\psi(u)$, 0 < u < 1 have

- (i) a bounded second derivative,
- (ii) $\sup_{y} [(d/dy)\psi \{F(y)\}]$ also bounded; and let the regression constants be such that
- (iii) $\max_{ij} |x_{ij}| \leq M \max_{ij} |x_{ij} x_{i\bullet}|$, M does not depend on n,
- (iv) $\{\max_{i,j} (x_{i,j} x_{i,\bullet})^2 / C_n^2\} \to 0.$

Then for each i, $\{\hat{T}_{ni} - T^0_{ni}\}/C_n\} \rightarrow 0$ in both P_0 - and P_n -probabilities.

PROOF. Without loss of generality, we may take $\alpha = \beta = 0$. Throughout the proof, M will denote a generic constant independent of n. Now let us write

$$(2.3) Y_{ij}^* = Y_{ij} - (b/C_n)x_{ij}, |b| \le M$$

and let $T_{ni}^* = T_{ni}(Y^*)$ be the T_{ni} -statistic defined through Y_{ij}^* 's. To prove the lemma, under H_n , it is sufficient because of Lemma 2.1, to show that as $n \to \infty$,

(2.4)
$$E_n\{(T_{ni}^* - T_{ni}^0)/C_n\}^2 \to 0$$
 uniformly in $|b| \le M$,

where E_n denotes expectation taken under H_n of (2.1). Repeated use will be made of the fact that under H_n , with $\alpha = \beta = 0$, the distribution functions of Y_{ij}^0 and Y_{ij}^* are respectively,

(2.5)
$$F_{ij}^{0}(y) = F\{y - (bx_{ij}/C_n)\}; \qquad F_{ij}^{*}(y) = F\{y - (\theta_i - b)x_{ij}/C_n\}.$$

Following Hájek (1968) define, for each i, a sum of n independent random variables with zero expectation; thus

$$(2.6) Z_{ni} = Z_{ni}(Y) = \sum_{s} \sum_{j} n^{-1} \sum_{t} \sum_{v} (c_{tv}^{(i)} - c_{sj}^{(i)}) B_{tv}(Y_{sj})$$

where

$$(2.7) -B_{tv}(Y_{sj}) = \psi\{F_{sj}(Y_{sj})\} + Q_{sjtv}(Y_{sj}) + \text{const.},$$

and

$$(2.8) |Q_{sjtv}(y)| \leq M \max_{sjt,v,y} |F_{sj}(y) - F_{tv}(y)|.$$

Also for each i, set

(2.9)
$$\mu_{ni} = \sum_{s} \sum_{j} c_{sj}^{(i)} \int \psi\{\bar{F}(y)\} dF_{sj}(y)$$

where

$$(2.10) n\bar{F}(y) = \sum_{s} \sum_{i} F_{si}(y).$$

Correspondingly, let Z_{ni}^* and μ_{ni}^* be the statistic (2.6) and the quantity (2.9) defined through Y_{ij}^* 's. On writing

$$(2.11) E_n(T_{ni}^* - T_{ni}^0)^2 = E_n\{(T_{ni}^* - Z_{ni}^* - \mu_{ni}^*) - (T_{ni}^0 - Z_{ni}^0 - \mu_{ni}^0) - (Z_{ni}^0 - Z_{ni}^* - \mu_{ni}^* + \mu_{ni}^0)\}^2$$

and making repeated use of the elementary inequality $(b-d)^2 \le 2(b^2+d^2)$ we obtain

$$(2.12) E_n(T_{ni}^* - T_{ni}^0)^2 \leq 4E_n(T_{ni}^* - Z_{ni}^* - \mu_{ni}^*)^2 + 4E_n(T_{ni}^0 - Z_{ni}^0 - \mu_{ni}^0)^2 + 2E_n(Z_{ni}^0 - Z_{ni}^*)^2 + 2(\mu_{ni}^0 - \mu_{ni}^*)^2.$$

For the first two terms of the inequality, Hájek (1968) has shown that under our condition (i) alone, and in view of (1.6)

$$E_{n}(T_{ni} - Z_{ni} - \mu_{ni})^{2}$$

$$\leq M_{1}n^{-1} \sum_{s} \sum_{j} \{c_{sj}^{(i)}\}^{2}$$

$$\leq M_{1} \max_{sj} (c_{sj}^{(i)})^{2}$$

$$= M_{1} \max \{\gamma_{ni}^{2} \max_{t \neq i, j} (x_{tj} - x_{t\bullet})^{2}, (1 - \gamma_{ni})^{2} \max_{j} (x_{ij} - x_{i\bullet})^{2}\}$$

$$\leq M_{1} \max_{sj} (x_{sj} - x_{s\bullet})^{2}.$$

As for the term $E_n(Z_{ni}^0 - Z_{ni}^*)^2$, (2.6), (2.7) and (2.8) imply that

$$(2.14) |Z_{ni}^{0}(y) - Z_{ni}^{*}(y)| \leq \sum_{s} \sum_{j} n^{-1} \sum_{t} \sum_{v} |c_{tv}^{(i)} - c_{sj}^{(i)}| \times |\phi\{F_{si}^{*}(y)\} - \phi\{F_{si}^{0}(y)\}| + |Q_{sitv}(y) - Q_{sitv}^{*}(y)|.$$

Furthermore, by (2.5) and conditions (ii) and (iii) of the lemma, we have

$$|\psi\{F_{sj}^{*}(y)\} - \psi\{F_{sj}^{0}(y)\}| \leq \max_{sj} |x_{sj}| (\theta_{s}/C_{n}) \psi'\{F(y)\} f(y)$$

$$\leq M \max_{sj} |x_{sj}|/C_{n}$$

$$\leq M_{1} \max_{sj} |x_{sj} - x_{s \bullet}|/C_{n}.$$

Also (2.5), (2.8) and condition (iii) imply

$$(2.16) |Q_{sjtv}^{0}(y) - Q_{sjtv}^{*}(y)|$$

$$\leq \max_{sjtvy} [f(y)\{|x_{tv} - x_{sj}|(b/C_n) + |\theta_t x_{tv} - \theta_s x_{sj}|/C_n\}]$$

$$\leq M_2 \max_{sj} |x_{sj} - x_{s\bullet}|/C_n.$$

On applying (2.15) and (2.16) in (2.14) and making use of the inequality

$$(2.17) \qquad \sum_{s} \sum_{i} (n^{-1} \sum_{t} \sum_{v} |c_{tv}^{(i)} - c_{si}^{(i)}|)^{2} \leq 2 \sum_{s} \sum_{i} \{c_{si}^{(i)}\}^{2} \leq 2 C_{n}^{2}$$

it follows, since the Z_{ni} 's are sums of independent random variables with zero expectations, that

$$(2.18) E_n(Z_{ni}^0 - Z_{ni}^*)^2 \leq M \max_{sj} (x_{sj} - x_{s\bullet})^2.$$

Finally, we write

$$(\mu_{ni}^0 - \mu_{ni}^*) = \sum_{s} \sum_{j} c_{sj}^{(i)} [\int \phi \{\bar{F}^0(y)\} dF_{sj}^0(y) - \int \phi \{\bar{F}^*(y)\} dF_{sj}^*(y)].$$

On expanding and integrating by parts, making use of (2.5), the terms in the square brackets give

$$\begin{split} C_{n}^{-1}[&-\theta_{s}x_{sj} \oint \psi_{y}'\{F(y)\} \, dF(y) - bx_{\bullet \bullet} \oint \psi_{y}'\{F(y)\} F_{sj}^{0}(y) \\ &+ n^{-1} \sum_{s} \sum_{j} (\theta_{s} - b)x_{sj} \oint \psi_{y}'\{F(y)\} \, dF_{sj}^{*}(y)] \;, \end{split}$$

where $x_{\bullet \bullet} = n^{-1} \sum_{i} \sum_{j} x_{ij}$, and ϕ_{y}' denotes the derivative of ϕ with respect to y. In view of (2.1), and conditions (ii) and (iii) of the lemma,

$$\begin{aligned} |\mu_{ni}^{0} - \mu_{ni}^{*}| &\leq C_{n}^{-1} |M_{2} M_{3} x_{sj} + \{b M_{4} + M_{5} (M_{2} - b)\} x_{\bullet \bullet}| \sum_{s} \sum_{j} |c_{sj}^{(i)}| \\ &\leq (M_{6} / C_{n}) \max_{sj} |x_{sj} - x_{s \bullet}| \sum_{s} \sum_{j} |c_{sj}^{(i)}|, \end{aligned}$$

where M with subscripts are also generic constants. It follows then from (1.7), that

$$(2.19) (\mu_{ni}^0 - \mu_{ni}^*)^2 \le M \max_{si} (x_{si} - x_{s\bullet})^2.$$

The inequalities (2.13), (2.18) and (2.19) together imply (2.4) and the lemma is proved. The convergence of $E_0\{(T_{ni}^*-T_{ni}^*)^2/C_n^2\}$ to zero follows as an obvious corollary.

LEMMA 2.3. If $\hat{U}_{ni} = (\hat{T}_{ni}/AC_{ni})$, then under the conditions of Lemma 2.2, the random vector $\hat{\mathbf{U}}_{n'} = (\hat{U}_{n1}, \dots, \hat{U}_{nk})$ is asymptotically normal $N(\mathbf{0}, \Sigma)$ under P_0 , and $N(\mathbf{v}_n, \Sigma)$ under P_n , where $\mathbf{v}_{n'} = (\mathbf{v}_{n1}, \dots, \mathbf{v}_{nk})$, with

(2.20)
$$\nu_{ni} = \gamma_{ni}^{i} \{\theta_{i} - \sum_{s} (\gamma_{ns}\theta_{s})\} \int \psi_{y}' \{F(y)\} dF(y) / A$$

where δ denotes the Kronecker delta.

PROOF. The proof will be given only for P_n . By virtue of Lemma 2.2 $\{\hat{T}_{ni}/C_n\}$ and $\{T^0_{ni}/C_n\}$ have the same limiting distribution. Now, since under (2.1) with $\beta=0$

$$\max_{sjtvy} |F_{sj}(y) - F_{tv}(y)| \le M \max_{sj} |x_{sj} - x_{s*}|/C_n$$

and

$$\begin{array}{l} \max_{sj} \left[\{ c_{sj}^{(i)} \}^2 / \sum_{s} \sum_{j} \{ c_{sj}^{(i)} \}^2 \right] \leq M_1 \max_{sj} (x_{sj} - x_{s \bullet})^2 / (\gamma_{ni} - \gamma_{ni}^2) C_{n^2} \\ \leq M \max_{sj} (x_{sj} - x_{s \bullet})^2 / C_{n^2} \end{array}$$

it follows from Theorem 2.2 of Hájek (1968) that $(T_{ni}^0 - \mu_{ni})/A(1 - \gamma_{ni})^{\frac{1}{2}}C_{ni}$, and hence $(\hat{T}_{ni} - \mu_{ni})/A(1 - \gamma_{ni})^{\frac{1}{2}}C_{ni}$ is asymptotically normal N(0, 1). But under (2.1),

$$(\mu_{ni}/AC_{ni}) = [\{\mu_{ni}(\beta_{ni}) - \mu_{ni}(0)\}/AC_{ni}] \sim \nu_{ni}$$

where $\mu_{ni}(\beta_{ni})$ and $\mu_{ni}(0)$ are the values of (2.9) computed under (2.1) and (1.2) respectively. Hence under (2.1),

(2.22)
$$\hat{U}_{ni}$$
 is asymptotically $N(\nu_{ni}, 1 - \gamma_{ni})$ $i = 1, \dots, k$.

Furthermore, any linear combination of the \hat{T}_{ni} 's is of the form

$$\hat{T}_n = \sum_i \lambda_i T_{ni} = \sum_i \sum_s \sum_j \lambda_i c_{sj}^{(i)} a_n(\hat{R}_{sj})$$
.

It is easy to see that the new constants $(\lambda_i c_{sj}^{(i)})$ satisfy the Noether condition with $c_{sj}^{(i)}$, so that under the assumptions of Lemma 2.2 \hat{T}_n has a limiting normal distribution under (2.1). Hence \hat{U}_n is asymptotically normal. Now if we write

$$W_{ni} = \sum_{s} \sum_{j} c_{i} \psi \{F_{sj}(Y_{sj})\}/AC_{ni}, \qquad i = 1, \dots, k$$

it can be shown by arguments similar to those used in the proof of Theorem 2.2 of Hájek (1968), that under (2.1),

(2.23)
$$\operatorname{Cov}(U_{ni}^{0}, U_{ns}^{0}) \sim \operatorname{Cov}(W_{ni}, W_{ns}) = -(\gamma_{ni}\gamma_{ns})^{\frac{1}{2}}.$$

The symbol ~ denotes asymptotic equivalence. In view of Lemma 2.2,

(2.24)
$$\operatorname{Cov}(\hat{U}_{ni}, \hat{U}_{ns}) \sim -(\gamma_{ni}\gamma_{ns})^{\frac{1}{2}}.$$

The proof of the lemma is complete.

The main result of the paper is given in the following

THEOREM 2.1. Consider model (1.1) and assume that the conditions of Lemma 2.2 are satisfied. Then

(2.25)
$$\lim P_0(\hat{L}_n \le y) = P(\chi_{k-1}^2 \le y)$$

(2.26)
$$\lim P_n(\hat{L}_n \leq y) = P(\chi^2_{k-1}(\Delta_L) \leq y)$$

where $\chi^2_{k-1}(\Delta_L)$ denotes the noncentral chi-square random variable with (k-1) degrees of freedom and noncentrality parameter

(2.27)
$$\Delta_{L} = \sum_{i} \nu_{ni}^{2} = \{ \sum_{i} \gamma_{ni} \theta_{i}^{2} - (\sum_{i} \gamma_{ni} \theta_{i})^{2} \} [\int \psi_{y}' \{F(y)\} dF(y)]^{2} / A^{2} .$$

PROOF. The asymptotic covariance matrix (2.21) of the vector $\hat{\mathbf{U}}_n$ is singular of rank (k-1). On using an orthogonal transformation, e.g.

$$(2.28) V_0 = \sum_s \gamma_{ns}^{\frac{1}{2}} \hat{U}_{ns}$$

$$V_i = \sum_s e_{is} \hat{U}_{ns} i = 1, \dots, (k-1)$$

where the e's are properly chosen to make the transformation orthogonal, it follows from Lemma 2.3 that under (1.2) the sum of squares $\hat{L}_n = \sum_i \hat{U}_{n_i}^2$ has asymptotically a central chi-square distribution with (k-1) degrees of freedom, and under (2.1) has asymptotically a noncentral chi-square distribution with (k-1) degrees of freedom, and noncentrality parameter $\sum_i \nu_{n_i}^2$ given in (2.27).

In view of (2.25) an asymptotically level ε test rejects the hypothesis (1.2) if \hat{L}_n is greater than the upper $100\varepsilon\%$ point of the chi-square distribution with (k-1) degrees of freedom.

3. Asymptotic efficiency. The usual method of testing the hypothesis (1.2) in the model (1.1) is based on the least squares estimates $\tilde{\beta}_n$ and $\tilde{\beta}_{ni}$ of the parameters β and β_i . In computing the estimates, it is easier to work with $x'_{ij} = (x_{ij} - x_{i\bullet})$ instead of the original x_{ij} . With this reparametrization, the test

statistic becomes

(3.1)
$$M_n = \sum_i C_{ni}^2 (\bar{\beta}_{ni} - \bar{\beta}_n)^2 / (k-1) s_e^2$$

where

and s_e^2 is the mean square due to error. It is well known that for any distribution function F(y) for which $\sigma^2(F) = \{ \int y^2 dF(y) - (\int y dF(y))^2 \} < \infty$,

(3.3)
$$\lim P_0\{(k-1)M_n \le v\} = P(\chi_{k-1}^2 \le v)$$

(3.4)
$$\lim P_n\{(k-1)M_n \le y\} = P\{\chi_{k-1}^2(\Delta_M) \le y\}$$

where

(3.5)
$$\Delta_y = \{\sum_i \gamma_{ni} \theta_i^2 - (\sum_i \gamma_{ni} \theta_i)^2\} / \sigma^2(F).$$

By the conventional method of measuring asymptotic efficiency, the efficiency of \hat{L}_n test relative to the usual least squares test is therefore

(3.6)
$$\Delta_L/\Delta_M = \sigma^2(F)[\int \psi_N'\{F(y)\} dF(y)]^2/A^2.$$

The efficiency in (3.6) is the same as that obtained by Sen, and it is the familiar efficiency of rank tests relative to the classical tests. The connection between efficiency and asymptotic power of tests is now well established (see e.g. Theorem 6.1 of Hájek (1962)). If the functional form of the distribution function F is known, one can improve on both the classical and the rank score tests. Provided F has a finite Fisher information, an asymptotically optimum parametric test is not M_n , but the likelihood ratio test (see Section 2 of Sen (1969) for details). For the rank score tests, if we choose the score generating function

(3.7)
$$\psi(u) = -[f'\{F^{-1}(u)\}/f\{F^{-1}(u)\}],$$

it follows that under the assumptions of Lemma 2.2 our \hat{L}_n test based on (3.7), with $\int \phi^2(u) du < \infty$, provides an asymptotically most powerful rank order test for the hypothesis (1.2) in model (1.1).

REFERENCES

- [1] ADICHIE, J. N. (1967). Estimates of ergression parameters based on rank tests. *Ann. Math. Statist.* 38 894-904.
- [2] HAJEK, J. (1962). Asymptotically most powerful rank order tests. Ann. Math. Statist. 33 1124-1147.
- [3] На́JEK, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives.

 Ann. Math. Statist. 39 325-346.
- [4] SEN, P. K. (1969). On a class of rank order tests for the parallelism of several regression lines. Ann. Math. Statist. 40 1668-1683.

DEPARTMENT OF MATHEMATICS AND STATISTICS UNIVERSITY OF NIGERIA

EAST CENTRAL STATE, NIGERIA