The Korean Communications
in Statistics Vol. 10 No. 3, 2003
pp. 1007-1015

Hypothesis Testing for New Scores in a Linear Model [1)]

Young Hun Choi [2)]

Abstract

In this paper we introduced a new score generating function for the rank dispersion function in a general linear model. Based on the new score function, we derived the null asymptotic theory of the rank-based hypothesis testing in a linear model.

In essence we showed that several rank test statistics, which are primarily focused on our new score generating function and new dispersion function, are mainly distribution free and asymptotically converges to a chi-square distribution.

Keywords: Scores; Linear Model; Dispersion function; Rank Test; Asymptotic Distribution.

1. Introduction

Since Jureckova(1971) and Jaeckel(1972) defined the dispersion function, considerable work on the rank based estimates as a robust alternatives to least squares has been done for the linear regression model. Recently Ozturk and Hettmansperger(1996) and Ozturk(1999) derived the robust estimates of location and scale parameters from minimizing a minimum distance criterion function.

Meanwhile Ahmad(1996) developed a new class of Mann-Whitney-Wilcoxon type test statistics with the right tail probabilities. Ozturk and Hettmansperger(1997) and Choi(1998) considered the distribution functions reflecting both right and left tail probabilities. Ozturk(2001) considered another class of Mann-Whitney-Wilcoxon test statistics with having both right and left tail distributions. Further Choi and Ozturk(2002) introduced a new score generating function for the rank dispersion function in a multiple linear regression model. The score function compares the r'th and s'th power of the tail probabilities of the underlying probability distribution, which improved the efficiency for many distributions.

Now the main purpose of this paper is to extend the Hettmansperger and McKean(1998) and Choi and Ozturk(2002)'s concept to the problem of hypothesis testing in a linear model. In Section 2, we propose our score function based on the r'th and s'th power in

1) This research was supported by Hanshin University Special Research Fund, 2003.
2) Professor, Department of Information and Statistics, Hanshin University, 447-791.
E-mail : choicyh@hanshin.ac.kr

considering both right and left tail probabilities. We show the dispersion function $D_{r,s}(\beta)$ based on our score function. In Section 3, we concentrate on hypothesis testing of β based on our score function. We derive the rank-based test statistics which are mostly distribution free and asymptotically converge to a chi-square distribution.

2. Notation and Assumptions

Consider the linear regression model , $y_i = \alpha + x_i^{'}\beta + e_i$, $i=1,\cdots,n$, where x_i and β are $p\times 1$ vectors of explanatory variables and unknown regression parameters respectively and e_i is a random variable with density f and distribution function F.

Jaeckel's(1972) general rank dispersion function can be defined as

$$D(\beta) = \sum_{i=1}^{n} (y_i - x_i^{'}\beta)\, a[R(y_i - x_i^{'}\beta)]$$

where a set of scores is generated by $a(i) = \phi(i/(n+1))$, the score generating function $\phi(u)$ is defined on $(0,1)$ and is nondecreasing, bounded and square-integrable.

We require the following assumptions:

(H1) : f is absolutely continuous and $f > 0$.

(H2) : scores are generated as $a(i) = \phi(i/(n+1))$ where ϕ is defined on $(0,1)$, nondecreasing, bounded and satisfies the conditions $\int_0^1 \phi(u)\,du = 0$ and $\int_0^1 \phi^2(u)\,du = 1$.

(H3) : $\lim_{n\to\infty} n^{-1}X^{'}X = \Sigma > 0$, where X is a $n\times p$ matrix with i th row $x_i^{'}$.

Now let $\phi(u) = \frac{1}{\sqrt{\omega_{r,s}}}\left[u^r - \frac{1}{r+1} - (1-u)^s + \frac{1}{s+1}\right]$,

$$a(i) = \frac{1}{\sqrt{\omega_{r,s}}}\left[\left(\frac{i}{n+1}\right)^r - \frac{1}{r+1} - \left(1-\frac{i}{n+1}\right)^s + \frac{1}{s+1}\right], \tag{1}$$

where $$\omega_{r,s} = \frac{r^2}{(2r+1)(r+1)^2} + \frac{s^2}{(2s+1)(s+1)^2} + \frac{2}{(r+1)(s+1)} - 2\frac{\Gamma(r+1)\Gamma(s+1)}{\Gamma(r+s+2)}. \tag{2}$$

Define the dispersion function

$$D_{r,s}(\beta) = \Sigma_{i=1}^{n} e_i\, a[R(e_i)], \tag{3}$$

where $R(e_i)$ denotes the rank of $e_i = y_i - x_i' \beta$. Then β can be estimated by the rank estimator $\widehat{\beta_{r,s}}$ which minimizes (3). Then from the following Theorem 1, we can show that test statistic based on (3) asymptotically converges to a chi-square distribution. Further Theorem 2 and Theorem 3 indicate that test statistic based on the partial derivatives of (3) and quadratic form of rank statistic are distribution free and asymptotically converge to a chi-square distribution respectively.

3. Rank-Based Tests

This section deals with a hypothesis testing in the linear model. We decompose regression parameters β into β_1 with $(p-q)\times 1$ and β_2 with $q \times 1$ vectors. The hypotheses of interest are

$$H_0 : \beta_2 = 0 , \qquad \beta_1 \text{ unspecified} \tag{4}$$

and

$$H_1 : \beta_2 \neq 0 , \qquad \beta_1 \text{ unspecified.}$$

Let β_0 denote the $p \times 1$ vector specified by (4) so that we can rewrite the null hypothesis (4) as $H_0 : \beta = \beta_0$.

3.1 Test Based on $D_{r,s}(\beta)$

Theorem 1 shows that test statistic D based on $D_{r,s}(\beta)$ is not distribution free because of $\tau_{r,s}$ and asymptotically converges to a chi-square distribution.

Theorem 1. Let $\widehat{\beta_0}$ and $\widehat{\beta_{r,s}}$ be rank estimators of regression parameters β that minimize the dispersion function $D_{r,s}(\beta)$ for the reduced and full linear model respectively. Suppose the null hypothesis (4) holds. Then under the assumptions (H1)-(H3) and scores in (1),

$$D = \frac{D_{r,s}(\widehat{\beta_0}) - D_{r,s}(\widehat{\beta_{r,s}})}{\sqrt{\frac{\omega_{r,s}}{\tau_{r,s}}} / 2} \tag{5}$$

where $D_{r,s}(\beta)=\frac{1}{\sqrt{\omega_{r,s}}}\left[\frac{1}{(n+1)^r}\sum_{i=1}^{n}e_i\{R^r(e_i)-\tau(r)\}-\frac{1}{(n+1)^s}\sum_{i=1}^{n}e_i\{(n+1-R(e_i))^s-\tau(s)\}\right]$,

and $\qquad \tau_{r,s}=\left(\int[rF^{r-1}(t)+s(1-F(t))^{s-1}]f^2(t)\,dt\right)^2$,

has an asymptotic χ^2 distribution with q degrees of freedom with $\tau_{r,s}$ and $\omega_{r,s}$ provided in (2).

Proof. We can rewrite the numerator of (5) as

$$\begin{aligned} D_{r,s}(\widehat{\beta_0})-D_{r,s}(\widehat{\beta_{r,s}}) = & \ [D_{r,s}(\widehat{\beta_0})-Q_{r,s}(\widehat{\beta_0})]+[Q_{r,s}(\widehat{\beta_0})-Q_{r,s}(\tilde{\beta_0})]+[Q_{r,s}(\tilde{\beta_0})-Q_{r,s}(\tilde{\beta}_{r,s})] \\ & + \ [Q_{r,s}(\tilde{\beta}_{r,s})-Q_{r,s}(\widehat{\beta_{r,s}})]+[Q_{r,s}(\widehat{\beta_{r,s}})-D_{r,s}(\widehat{\beta_{r,s}})]\,. \end{aligned} \tag{6}$$

where $\quad Q_{r,s}(\beta) = D_{r,s}(\beta_0)-(\beta-\beta_0)'S_{r,s}(\beta_0)+\frac{1}{2}\sqrt{\frac{\tau_{r,s}}{\omega_{r,s}}}\,n(\beta-\beta_0)'\Sigma(\beta-\beta_0) \qquad (7)$

and $\quad S_{r,s}(\beta)=\frac{1}{\sqrt{\omega_{r,s}}}\left[\frac{1}{(n+1)^r}\sum_{i=1}^{n}\{R^r(e_i)-\tau(r)\}-\frac{1}{(n+1)^s}\sum_{i=1}^{n}\{(n+1-R(e_i))^s-\tau(s)\}\right]. \quad (8)$

Then Jaeckel(1972)'s result can be applied to insist that the first and fifth difference on the right side of (6) tend to zero in probability.

Next substituting $\tilde{\beta}_{r,s}$ which minimizes the quadratic approximation $Q_{r,s}(\beta)$ in (7) and $\widehat{\beta_{r,s}}$ into (7), we have

$$\begin{aligned} Q_{r,s}(\tilde{\beta}_{r,s})-Q_{r,s}(\widehat{\beta_{r,s}}) = & \ (\widehat{\beta_{r,s}}-\tilde{\beta}_{r,s})'S_{r,s}(\beta_0) \\ & +\frac{1}{2}\sqrt{\frac{\tau_{r,s}}{\omega_{r,s}}}\,n\,[(\tilde{\beta}_{r,s}-\beta_0)'\Sigma(\tilde{\beta}_{r,s}-\beta_0)-(\widehat{\beta_{r,s}}-\beta_0)'\Sigma(\widehat{\beta_{r,s}}-\beta_0)]\,. \end{aligned} \tag{9}$$

The term in square bracket of (9) can be expressed as $(\tilde{\beta}_{r,s}-\widehat{\beta_{r,s}})\Sigma(\tilde{\beta}_{r,s}-\beta_0+\widehat{\beta_{r,s}}-\beta_0)$. Thus (9) can be rewritten as

$$Q_{r,s}(\tilde{\beta}_{r,s})-Q_{r,s}(\widehat{\beta_{r,s}})=\sqrt{n}(\widehat{\beta_{r,s}}-\tilde{\beta}_{r,s})'\left[\frac{1}{\sqrt{n}}S_{r,s}(\beta_0)-\frac{1}{2}\sqrt{\frac{\tau_{r,s}}{\omega_{r,s}}}\,\Sigma\sqrt{n}(\tilde{\beta}_{r,s}-\beta_0+\widehat{\beta_{r,s}}-\beta_0)\right]. \tag{10}$$

From Theorem 5.2.1 of Hettmansperger(1991), we know that $\sqrt{n}(\widehat{\beta_{r,s}}-\tilde{\beta}_{r,s})$ tends to zero in probability and $\sqrt{n}(\tilde{\beta}_{r,s}-\beta_0)$ and $\sqrt{n}(\widehat{\beta_{r,s}}-\beta_0)$ are asymptotically normally distributed.

Also we know that $(1/\sqrt{n})S_{r,s}(\beta_0)$ are asymptotically normally distributed. Hence we say that (10), in turn the fourth difference on the right side of (6), tends to zero in probability. Same result can be obtained for the second difference on the right side of (6).

The equation (6) can be simplified as

$$D_{r,s}(\widehat{\beta_0}) - D_{r,s}(\widehat{\beta_{r,s}}) \simeq Q_{r,s}(\tilde{\beta_0}) - Q_{r,s}(\tilde{\beta_{r,s}}) . \tag{11}$$

Further substituting $\tilde{\beta_{r,s}} = \beta_0 + \sqrt{\omega_{r,s}/\tau_{r,s}}\ \Sigma^{-1}(1/n)S_{r,s}(\beta_0)$ into (7) yields

$$Q_{r,s}(\tilde{\beta_{r,s}}) = D_{r,s}(\beta_0) - \frac{1}{2n}\sqrt{\frac{\omega_{r,s}}{\tau_{r,s}}}\, S_{r,s}(\beta_0)'\ \Sigma^{-1} S_{r,s}(\beta_0) . \tag{12}$$

Using the fact that partition Σ to correspond to the partition of β is $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, we can easily get the following result from (12).

$$Q_{r,s}(\tilde{\beta_0}) - Q_{r,s}(\tilde{\beta_{r,s}}) = \frac{1}{2n}\sqrt{\frac{\omega_{r,s}}{\tau_{r,s}}}\, S_{r,s}(\beta_0)'\left[\Sigma^{-1} - \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}\right] S_{r,s}(\beta_0) . \tag{13}$$

Hence, after substituting (13) into (11), we can have from (5)

$$D \simeq \frac{1}{\sqrt{n}} S_{r,s}'(\beta_0)\left[\ \Sigma^{-1} - \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}\right] \frac{1}{\sqrt{n}} S_{r,s}(\beta_0). \tag{14}$$

Further we know that (14) becomes from Arnold(1981)

$$(14) = \frac{1}{\sqrt{n}}[(-\Sigma_{21}\Sigma_{11}^{-1}, \mathrm{I})S_{r,s}(\beta_0)]'(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}\frac{1}{\sqrt{n}}[(-\Sigma_{21}\Sigma_{11}^{-1}, \mathrm{I})S_{r,s}(\beta_0)] .$$

By the way we know that $(-\Sigma_{21}\Sigma_{11}^{-1}, \mathrm{I})\frac{1}{\sqrt{n}} S_{r,s}(\beta_0) \xrightarrow{d} Z \sim MVN(0, V)$, where $V = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. Therefore we can say that (14), in turn D, converges in distribution to $Z'V^{-1}Z$, which is a quadratic form in a multivariate normal vector. Consequently D converges in chi-square distribution with q degrees of freedom since V is nonsingular.

3.2 Test Based on $S_{r,s}(\beta)$

Theorem 2 shows that test statistic $S_{r,s}^*$ based on $S_{r,s}(\beta)$ in (8) is distribution free and asymptotically converges to a chi-square distribution.

Theorem 2. Suppose the null hypothesis (4) holds. Then under the assumptions (H1)-(H3) and scores in (1),

$$S^*_{r,s} = S_2'(\widehat{\beta}_0)\,[\,X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2\,]^{-1}\,S_2(\widehat{\beta}_0)$$

has an asymptotic χ^2 distribution with q degrees of freedom, where X_1 and X_2 are partition of X to correspond to partition of β.

Proof. Recall that from a linear approximation to the partial derivatives of $D_{r,s}(\beta)$, we have the partitioned form

$$\frac{1}{\sqrt{n}}\begin{pmatrix} S_1(\beta) \\ S_2(\beta) \end{pmatrix} \simeq \frac{1}{\sqrt{n}}\begin{pmatrix} S_1(\beta_0) \\ S_2(\beta_0) \end{pmatrix} - \sqrt{\frac{\tau_{r,s}}{\omega_{r,s}}}\,\frac{1}{n}\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}\sqrt{n}\begin{pmatrix} \beta_1-\beta_{10} \\ \beta_2-\beta_{20} \end{pmatrix}, \qquad (15)$$

where $\beta_0' = (\beta_{10}', 0')$. Since $\sqrt{n}\,(\widehat{\beta}_0-\beta_0)$ is bounded in probability by Theorem 2 of Choi(2002), we have from (15)

$$\frac{1}{\sqrt{n}} S_2(\widehat{\beta}_0) \simeq \frac{1}{\sqrt{n}} S_2(\beta_0) - \sqrt{\frac{\tau_{r,s}}{\omega_{r,s}}}\,\frac{1}{n}(X_2'X_1)\sqrt{n}\,(\widehat{\beta}_{10}-\beta_{10}) . \qquad (16)$$

Further since $(1/\sqrt{n}\,)\,S_1(\widehat{\beta}_{10}) \doteq 0$ for the reduced model, we have from (15)

$$\sqrt{n}\,(\widehat{\beta}_{10}-\beta_{10}) = \sqrt{\frac{\omega_{r,s}}{\tau_{r,s}}}\left(\frac{1}{n}X_1'X_1\right)^{-1}\frac{1}{\sqrt{n}} S_1(\beta_{10}) . \qquad (17)$$

Substituting (17) into (16) becomes

$$\frac{1}{\sqrt{n}} S_2(\widehat{\beta}_0) \simeq \frac{1}{\sqrt{n}} S_2(\beta_0) - \frac{1}{n}(X_2'X_1)\left(\frac{1}{n}X_1'X_1\right)^{-1}\frac{1}{\sqrt{n}} S_1(\beta_0) . \qquad (18)$$

Then (18) can be rewritten as

$$\frac{1}{\sqrt{n}} S_2(\widehat{\beta}_0) \simeq (-\Sigma_{21}\Sigma_{11}^{-1}, \mathrm{I})\begin{pmatrix} \frac{1}{\sqrt{n}} S_1(\beta_0) \\ \frac{1}{\sqrt{n}} S_2(\beta_0) \end{pmatrix} = (-\Sigma_{21}\Sigma_{11}^{-1}, \mathrm{I})\,\frac{1}{\sqrt{n}} S_{r,s}(\beta_0) . \qquad (19)$$

This result agrees with that shown in the end of proof of Theorem 1. That is, (19) is a limiting multivariate normal vector with $MVN(0, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$. Now $S^*_{r,s}$ given in Theorem 2 is a quadratic form using the inverse of covariance matrix and has thus asymptotic chi-square distribution with q degrees of freedom, which is the rank of matrix.

3.3 One-Way Layout

We consider the one-way layout. The additive one-way model is specified by the equivalent regression model : $y_{ij} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{k-1} x_{i,k-1} + e_{ij}$ for $i = 1, \cdots, n_j$, $j = 1, \cdots, k$, $N = \sum_{j=1}^{k} n_j$, where $x_{ij} = 1$ if observation i is from treatment j and $x_{ij} = 0$ otherwise. In general if there are k treatments, the regression model will have $k-1$ variables. And let R_{ij} be the rank of e_{ij} in the combined data. We test $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$ where $\beta' = (\beta_1, \cdots, \beta_{k-1})$. Then Theorem 3 shows that quadratic form of rank statistic T is distribution free and asymptotically converges to a chi-square distribution.

Theorem 3. Suppose the null hypothesis (4) holds. Then under the assumptions (H1)-(H3) and scores in (1),

$$T = \frac{1}{\omega_{r,s}} \sum_{j=1}^{k} \frac{1}{n_j} \left[\frac{1}{(N+1)^r} \sum_{i=1}^{n_j} \{ R_{ij}^r - \tau(r) \} - \frac{1}{(N+1)^s} \sum_{i=1}^{n_j} \{ (N+1-R_{ij})^s - \tau(s) \} \right]^2$$

has an asymptotic χ^2 distribution with $k-1$ degrees of freedom, where $\tau(r) = \sum_{i=1}^{N} i^r / N$, $\tau(s) = \sum_{i=1}^{N} i^s / N$ and $\omega_{r,s}$ in (2).

Proof. From (8) we know that the vector $S_{r,s}(\beta_0) = S_{r,s}(0)$ has j th component

$$S_{j(r,s)}(0) = \frac{1}{\sqrt{\omega_{r,s}}} \left[\frac{1}{(N+1)^r} \sum_{i=1}^{n_j} \{ R_{ij}^r - \tau(r) \} - \frac{1}{(N+1)^s} \sum_{i=1}^{n_j} \{ (N+1-R_{ij})^s - \tau(s) \} \right]. \tag{20}$$

The quadratic form of (14) can be written as

$$\frac{1}{N} S'_{r,s}(0) \ \Sigma^{-1} S_{r,s}(0) = \frac{1}{N} \sum_{j=1}^{k} \frac{N}{n_j} S_{j(r,s)}^2(0) . \tag{21}$$

Hence when plugging the result (20) into (21), we can establish the following result.

$$\frac{1}{N} S'_{r,s}(0) \ \hat{\Sigma}^{-1} S_{r,s}(0) = \frac{1}{\omega_{r,s}} \sum_{j=1}^{k} \frac{1}{n_j} \left[\frac{1}{(N+1)^r} \sum_{i=1}^{n_j} \{R_{ij}^r - \tau(r)\} - \frac{1}{(N+1)^s} \sum_{i=1}^{n_j} \{(N+1-R_{ij})^s - \tau(s)\} \right]^2 .$$

Numerical Example. Random samples from each of three different types(A, B, C) of light bulbs are tested to see how long days they last, with the following results. A: 73 64 67 62, B: 84 80 81 77, C: 82 79 71 75.

Assuming the exponential distribution with reasonable value of $r=1$ and $s=3$, we can obtain $T=6.869 > \chi^2_{2,\,0.95}=5.991$. The null hypothesis that all population distribution functions are identical is rejected. The critical level is estimated to be about $\hat{\alpha} \simeq 0.034$.

4. Conclusions

In this paper we developed the null asymptotic theory of the rank-based tests focused on a new score function. We obtained some properties of these tests. First we established that test statistic D based on our dispersion function is not distribution free and asymptotically converges to a chi-square distribution. Further test statistic $S^*_{r,s}$ generated by the partial derivatives of our dispersion function, which is a quadratic form using the inverse of covariance matrix, has thus asymptotic chi-square distribution.

Further we showed that quadratic form of rank statistic T for the one-way layout is distribution free and asymptotically converges to a chi-square distribution.

References

[1] Ahmad, I. A. (1996). A Class of Mann-Whitney-Wilcoxon Type Statistics, *The American Statistician*, 50, 324-327.

[2] Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis*, Wiley, New York, New York.

[3] Choi, Y. H. (1998). A Study of the Power of the Rank Transform Test in a 2^3 Factorial Experiment, *Communications in Statistics-Simulation and Computation,* 27(1), 259-274.

[4] Choi, Y. H. (2002). New Dispersion Function in the Rank Regression, *The Korean Communications in Statistics*, 9(1), 101-113.

[5] Choi, Y. H. and Ozturk, O (2002). A New Class of Score Generating Functions for Regression Models, *Statistics & Probability Letters*, 57, 205-214.

[6] Hettmansperger, T. P. (1991). *Statistical Inference Based on Ranks*, Krieger Publishing Company, Malabar, Florida.

[7] Hettmansperger, T. P. and McKean J. W. (1998). *Robust Nonparametric Statistical Methods*, Wiley & Jones Inc., New York, New York.

[8] Jaeckel, L. A. (1972). Estimating Regression Coefficients by Minimizing the Dispersion of Residuals, *The Annals of Mathematical Statistics*, 43, 1449-1458

[9] Jureckova, J. (1971). Nonparametric Estimate of Regression Coefficients, *The Annals of Mathematical Statistics*, 42, 1328-1338.

[10] Ozturk, O (1999). Two-Sample Inference Based on One-Sample Ranked Set Sample Sign Statistics, *Journal of Nonparametric Statistics*, 10, 197-212.

[11] Ozturk, O (2001). A Generalization of Ahmad's Class of Mann-Whitney-Wilcoxon Statistics, *Australian and New Zealand Journal of Statistics*, 43(1), 67-74.

[12] Ozturk, O and Hettmansperger, T. P. (1996). Almost Fully Efficient and Robust Simultaneous Estimation of Location and Scale Parameters: A Minimum Distance Approach, *Statistics & Probability Letters*, 29, 233-244.

[13] Ozturk, O and Hettmansperger, T. P. (1997). Generalized Weighted Cramer-Von Mises Distance Estimators, *Biometrika*, 84, 283-294.

[Received July 2003, Accepted November 2003]