The Korean Communications
in Statistics Vol. 9, No. 1, 2002
pp. 101-113

New Dispersion Function in the Rank Regression [1)]

Young Hun Choi [2)]

Abstract

In this paper we introduce a new score generating function for the rank regression in the linear regression model. The score function compares the r'th and s'th power of the tail probabilities of the underlying probability distribution. We show that the rank estimate asymptotically converges to a multivariate normal. Further we derive the asymptotic Pitman relative efficiencies and the most efficient values of r and s under the symmetric distribution such as uniform, normal, cauchy and double exponential distributions and the asymmetric distribution such as exponential and lognormal distributions respectively.

Keywords: Score; Rank estimate, Regression, Asymptotic normality, Efficiency.

1. Introduction

In the last three decades considerable work on the rank based estimates as a robust alternatives to least squares has been pursued for the linear regression model [see, for example, Jureckova(1969, 1971); Jaeckel(1972); McKean and Hettmansperger(1978); Hettmansperger and McKean(1983)]. Recently Naranjo and Hettmansperger(1994) discussed bounded influence, high breakdown rank regression estimate. Witt, Naranjo and McKean(1995) expanded the concept of the influence function for the rank based procedures in the linear model.

Ozturk and Hettmansperger(1996) derived the robust estimates of location and scale parameters from minimizing a minimum distance criterion function. Ozturk(1999) also generated two-sample inference for the ranked set samples.

Further Ahmad(1996) developed a new class of Mann-Whitney-Wilcoxon type test statistics, which only considered the distribution functions of the r'th and s'th power in emphasizing the right tail probabilities. On the other hand, Ozturk and Hettmansperger(1997) state that if

---

1) This research was supported by Hanshin University Research Grant, 2001.
2) Professor, Department of Statistics, Hanshin University, 447-791.
E-mail : choicyh@hanshin.ac.kr

there is no knowledge about the outlier pattern, or if the sample has both small and large outliers, the distribution functions reflecting both right and left tail probabilities would be appropriate. Therefore Ozturk(2001) and Choi and Ozturk(2002) considered another class of Mann-Whitney-Wilcoxon test statistics with having both right and left tail behavior of the underlying distributions, which improved the efficiency for many distributions.

Thus the main purpose of this paper is to extend the Ozturk and Hettmansperger(1997), Ozturk(2001) and Choi and Ozturk(2002)'s concept, where the distribution function reflects both right and left tail probabilities and produces robust estimators with high efficiency, into the rank estimate of regression parameters in the linear regression model.

In Section 2, we propose our score function based on the r'th and s'th power in considering both right and left tail probabilities. We derive that the dispersion function $D(\beta)$ based on our score function is a nonnegative and convex function of β and that the distribution of a rank estimator $\hat{\beta}$ asymptotically converges to a multivariate normal. In Section 3, we compare the efficiency of Wilcoxon rank estimate with the efficiency of our rank estimate. In Section 4, reasonable r and s of our proposed score function are selected, which show one of the most desirable efficiencies for the underlying distributions.

2. Rank-Based Estimate

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. In this model we consider rank regression estimate of the regression parameter β.

In its general form, Jaeckel's(1972) rank dispersion function can be stated as

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

where $R(e_i)$ denotes the rank of $e_i = y_i - x_i^{'} \beta$ and $a(1) \le a(2) \le \cdots \le a(n)$ is a set of scores 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. Under fairly general conditions, minimizer of $D(\beta)$ produces robust estimator in y-space and it has relatively high efficiency at the true model. The property of such estimator is studied in detail for a general score function $\phi(\cdot)$ in Hettmansperger and McKean(1998). We introduce a general score generating function further to improve the efficiency of the rank regression estimator.

We require the following assumptions:

(A1) : f is absolutely continuous and $f > 0$.

(A2) : 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 $\int_0^1 \phi^2(u)\,du = 1$.

(A3) : $\lim_{n\to\infty} n^{-1}X'X = \Sigma > 0$, where X is a $n\times p$ matrix with i th row x_i'.

Now let

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

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

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

Define the dispersion function $D_{r,s}(\beta) = \sum_{i=1}^n e_i\, a[R(e_i)]$, where $e_i = y_i - x_i'\beta$. Then β can be estimated by the rank estimator $\widehat{\beta_{r,s}}$ which minimizes $D_{r,s}(\beta)$. Meanwhile from the following Theorem 1, we can show that $D_{r,s}(\beta)$ is a nonnegative and convex function of β immediately. Further Theorem 2 indicates that the distribution of a rank estimator $\widehat{\beta_{r,s}}$ asymptotically converges to a multivariate normal.

The next Lemma 1 is used in Theorem 1.

Lemma 1. The function $D^1_{r,s}(\beta) = (1/\sqrt{\omega_{r,s}})(n+1)^{-r}\sum_{i=1}^n e_i\left[R^r(e_i) - \tau(r)\right]$, where $\tau(r) = \sum_{i=1}^n i^r/n$, $e_i = y_i - x_i'\beta$ and $\omega_{r,s}$ in (2), is nonnegative and satisfies the property of triangle inequality.

Proof. Let $a_1^*(R_i) = (1/\sqrt{\omega_{r,s}})(n+1)^{-r}[R^r(e_i) - \tau(r)]$ and further let t be such that $a_1^*(1) \le \cdots \le a_1^*(t-1) \le 0 \le a_1^*(t) \le \cdots \le a_1^*(n)$ and $e_{(1)} \le \cdots \le e_{(t)} \le \cdots \le e_{(n)}$ are the ordered residual values. Then

$$D^1_{r,s}(\beta) = \sum_{i=1}^{n} e_i \, a_1^*(R_i)$$

$$= \sum_{i=1}^{n} [e_i - e_{(t)}] \, a_1^*(R_i) \tag{3}$$

since $\sum_{i=1}^{n} e_{(t)} \, a_1^*(R_i) = 0$. However, each term of (3) is greater than or equal to zero. Namely $[e_i - e_{(t)}] \, a_1^*(R_i) \geq 0$. Therefore we can say that $D^1_{r,s}(\beta)$ is nonnegative.

We now verify the triangle inequality. Let $||e|| = (1/\sqrt{\omega_{r,s}})(n+1)^{-r} \sum_{i=1}^{n} e_i \, R^r(e_i)$, which is a modified type for the function provided in Hardy, Littlewood and Polya(1952). Then

$$||e+h|| = \frac{1}{\sqrt{\omega_{r,s}}\,(n+1)^r}\left[\sum_{i=1}^{n} (e_i + h_i)\, R^r(e_i + h_i)\right]$$

$$= \frac{1}{\sqrt{\omega_{r,s}}\,(n+1)^r}\left[\sum_{i=1}^{n} e_i\, R^r(e_i + h_i) + \sum_{i=1}^{n} h_i\, R^r(e_i + h_i)\right] \tag{4}$$

$$\leq \frac{1}{\sqrt{\omega_{r,s}}\,(n+1)^r}\left[\sum_{i=1}^{n} e_i\, R^r(e_i) + \sum_{i=1}^{n} h_i\, R^r(h_i)\right] \tag{5}$$

$$= ||e|| + ||h|| \tag{6}$$

Now in order to identify the relationship between (4) and (5), consider the first term on the right hand side of (4). Primarily we know that

$$\sum_{i=1}^{n} e_i\, R^r(e_i + h_i) = \sum_{i=1}^{n} e_{(i)}\, p_i^r \tag{7}$$

where $p_1, \cdots, p_n$ is a permutation on the integers $1, \cdots, n$. Suppose p_i is not in order, then there exists s and t such that $e_{(s)} \leq e_{(t)}$, but $p_t^r \leq p_s^r$. Therefore

$$\begin{aligned}[e_{(s)}\, p_t^r + e_{(t)}\, p_s^r] - [e_{(s)}\, p_s^r + e_{(t)}\, p_t^r] &= e_{(t)}\,(p_s^r - p_t^r) - e_{(s)}\,(p_s^r - p_t^r) \\ &= [e_{(t)} - e_{(s)}]\,(p_s^r - p_t^r) \\ &\geq 0 \end{aligned} \tag{8}$$

Such an interchange never decreases the sum. Therefore when combining the above results (7) and (8) with $\sum_{i=1}^{n} e_{(i)}\, i^r = \sum_{i=1}^{n} e_i\, R^r(e_i)$, we can yield the following result.

$$\sum_{i=1}^{n} e_i\ R^r(e_i + h_i) \le \sum_{i=1}^{n} e_i\ R^r(e_i) \tag{9}$$

Finally substituting the result (9) into (4) and from the fact that a similar result can be obtained for the second term on the right hand side of (4), we can generate the result (5) as mentioned above. This completes the proof of the triangle inequality of (6). That is $||e+h|| \le ||e|| + ||h||$.

Theorem 1. Under the score generating function in (1), the dispersion function

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

is a nonnegative and convex function of β, where $\tau(r) = \sum_{i=1}^{n} i^r/n$, $\tau(s) = \sum_{i=1}^{n} i^s/n$, $e_i = y_i - x_i'\beta$ and $\omega_{r,s}$ in (2).

Proof. First of all redefine the function $D_{r,s}(\beta) = D_{r,s}^1(\beta) + D_{r,s}^2(\beta)$, where

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

and

$$D_{r,s}^2(\beta) = \frac{1}{\sqrt{\omega_{r,s}}\,(n+1)^s}\sum_{i=1}^{n} e_i\,[-\{(n+1) - R(e_i)\}^s + \tau(s)].$$

Then since $D_{r,s}^1(\beta)$ is symmetric about $D_{r,s}^2(\beta)$, we are sufficient to show that $D_{r,s}^1(\beta)$ is a norm. Further $D_{r,s}^1(\beta)$ is equivalent to the function $||e|| = (1/\sqrt{\omega_{r,s}})(n+1)^{-r} \sum_{i=1}^{n} e_i R^r(e_i)$ since $\sum_{i=1}^{n} e_i\,\tau(r) = 0$. Consequently, it is adequate to prove that the function $||e||$ is a norm. To show that $||e||$ has a norm, Hettmansperger and McKean(1998)'s Definition 2.2.1 should be met. Obviously $||\alpha \cdot e|| = |\alpha| \cdot ||e||$ for $\alpha > 0$. Also $||e|| = 0$ if and only if $e_1 = \cdots = e_n = 0$. So when combining the results of Lemma 1 we can say that $D_{r,s}^1(\beta)$, in turn $D_{r,s}(\beta)$, is immediately a nonnegative and convex function of β.

Theorem 2. Let $\widehat{\beta_{r,s}}$ be a rank estimator which minimizes the dispersion function

$D_{r,s}(\beta)$ defined in Theorem 1 and let β_0 be the true regression parameter value. Then under the assumptions in (1),

$$\sqrt{n}\,(\widehat{\beta_{r,s}} - \beta_0) \xrightarrow{d} Z \sim MVN\left(0,\ \frac{\omega_{r,s}}{\tau_{r,s}}\ \Sigma^{-1}\right),$$

where
$$\omega_{r,s} = \frac{r^2}{(2r+1)(r+1)^2} + \frac{s^2}{(2s+1)(s+1)^2} + \frac{2}{(r+1)(s+1)} - 2\frac{\Gamma(r+1)\,\Gamma(s+1)}{\Gamma(r+s+2)},$$
$$\tau_{r,s} = \left(\int [\,rF^{r-1}(t) + s(1-F(t))^{s-1}]\,f^2(t)\,dt\right)^2 \text{ and } \Sigma = \lim_{n\to\infty} n^{-1}X'X.$$

Proof. The result of Theorem 1 indicates that $D_{r,s}(\beta)$ is a nonnegative and convex function of β. Therefore we can apply Hettmansperger's(1991) Theorem 5.2.3 into our rank estimate of regression parameters.

In essence, we obtain the following linearity result by using $S_{r,s}(\beta) = -\partial D_{r,s}(\beta)/\partial\beta$.

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

Then we have a linear approximation to the partial derivatives of $D_{r,s}(\beta)$.

$$\frac{1}{\sqrt{n}} S_{r,s}(\beta) \simeq \frac{1}{\sqrt{n}} S_{r,s}(\beta_0) - \sqrt{\frac{\tau_{r,s}}{\omega_{r,s}}}\ \Sigma\,\sqrt{n}\,(\beta - \beta_0)\ . \tag{10}$$

From (10), we can construct a quadratic approximation to $D_{r,s}(\beta)$ as follows.

$$Q_{r,s}(\beta) = D_{r,s}(\beta_0) - (\beta - \beta_0)'\,S_{r,s}(\beta_0) + \frac{1}{2}\sqrt{\frac{\tau_{r,s}}{\omega_{r,s}}}\ n\,(\beta - \beta_0)'\,\Sigma(\beta - \beta_0). \tag{11}$$

(11) shows the property that $Q_{r,s}(\beta_0) = D_{r,s}(\beta_0)$ and the gradient of $Q_{r,s}(\beta)$ is the linear approximation on the right hand side of (10). Jaeckel(1972) shows that $Q_{r,s}(\beta)$ provides a useful approximation to $D_{r,s}(\beta)$.

Therefore the value $\tilde{\beta_{r,s}}$ which minimizes the quadratic approximation $Q_{r,s}(\beta)$ in (11) solves the right hand side of (10) as follows.

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

If assumptions in (1) hold, we have a limiting multivariate normal distribution that $(1/\sqrt{n})\, S_{r,s}(\beta_0) \xrightarrow{d} Z \sim MVN(0, \Sigma)$, where $\Sigma = \lim_{n\to\infty} n^{-1}X'X$. Hence from (12) it can be shown that $\sqrt{n}\,(\tilde{\beta}_{r,s} - \beta_0) \xrightarrow{d} Z \sim MVN(0, (\omega_{r,s}/\tau_{r,s})\,\Sigma^{-1})$. Finally we can say that $\widehat{\beta_{r,s}}$ behaves asymptotically like $\tilde{\beta}_{r,s}$ from Jaeckel(1972) and thus $\sqrt{n}\,(\widehat{\beta_{r,s}} - \beta_0)$ has the same limiting distribution as $\sqrt{n}\,(\tilde{\beta}_{r,s} - \beta_0)$. This completes the proof.

3. Asymptotic Relative Efficiencies

In this section we compare the efficiency of the rank estimator based on Wilcoxon scores with the efficiency of our rank estimator $\widehat{\beta_{r,s}}$. The asymptotic relative efficiencies of the Wilcoxon rank estimator with respect to our rank estimator, which is denoted as ARE(11, rs), are given below in terms of ω_{rs} in (2) and τ_{rs} defined in Theorem 2 for the regression parameters of underlying distributions such as uniform, normal, cauchy, double exponential, exponential and lognormal distributions.

3.1 Uniform Distribution

Let $f(t) = 1$, for $0 < t < 1$. Then $F(t) = t$. So we have

$$\tau_{rs} = 4 \quad \text{and} \quad \text{ARE(11, rs)} = 3\,\omega_{rs}.$$

We evaluated ARE(11, rs) for several values of $r, s = (0.1)(3)(0.1)$. Our estimator performs better than Wilcoxon estimator for $r, s < 1$ or $r, s > 1$. Some of the selected values of ARE(11, rs) are given in Table 1. Thus we should choose r, s as low as possible for $r, s < 1$ or r, s as high as possible for $r, s > 1$.

3.2 Exponential Distribution

Let $f(t) = \exp(-t)$, for $t > 0$. Then $F(t) = 1 - \exp(-t)$. So we have

$$\tau_{rs} = \left(\frac{1}{r+1} + \frac{s}{s+1} \right)^2$$

and

$$\text{ARE}(11, \text{rs}) = 3 \frac{\omega_{rs}}{\tau_{rs}} .$$

Table 1 shows that our procedure has better efficiency than Wilcoxon score if r is decreased or s is increased. The efficiency of our procedure appears to be better high for small r and large s. Thus we should choose r as low as possible or s as high as possible.

3.3 Double Exponential Distribution

Let $f(t) = \exp(-|t|)/2$, for $-\infty < t < \infty$. Then $F(t) = \exp(t)/2$, for $t < 0$, $= 1 - \exp(-t)/2$, for $t > 0$. So we have

$$\tau_{rs} = \left[\frac{2^r - 1}{(r+1)\, 2^r} + \frac{2^s - 1}{(s+1)\, 2^s} \right]^2$$

and

$$\text{ARE}(11, \text{rs}) = \frac{3}{4} \frac{\omega_{rs}}{\tau_{rs}} .$$

Again Table 1 shows that our procedure has better efficiency than Wilcoxon score if r and s are both close to 1.5. Thus we should choose r, $s = 1.5$.

3.4 Normal Distribution

Let $f(t) = (1/\sqrt{2\pi})\exp(-t^2/2)$, for $-\infty < t < \infty$. Then

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

and

$$\text{ARE}(11, \text{rs}) = \frac{3}{\pi} \frac{\omega_{rs}}{\tau_{rs}} .$$

For normal distribution, it appears to be there is not much difference between our estimator and Wilcoxon estimator. On the other hand, for some values of r and s our estimator is slightly better than Wilcoxon estimator, see Table 1.

3.5 Cauchy Distribution

Let $f(t) = 1/[\pi(1+t^2)]$, for $-\infty < t < \infty$. Then

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

and

$$\text{ARE}(11, \text{rs}) = \frac{3}{\pi^2} \frac{\omega_{rs}}{\tau_{rs}} .$$

Again as in the double exponential distribution, Table 1 shows that our estimator outperforms Wilcoxon score estimator when r and s are both close to 1.5. Thus we should choose $r, s = 1.5$.

3.6 Lognormal Distribution

Let $f(t) = \exp[-\{\log(t)\}^2/2] / (\sqrt{2\pi}\, t)$, for $t > 0$. Then $F(t) = \Phi\{\log(t)\}$, where $\Phi(t)$ is cdf of the standard normal distribution. So we have

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

and

$$\text{ARE}(11, \text{rs}) = 12 \left(\int f^2(t)\, dt \right)^2 \frac{\omega_{rs}}{\tau_{rs}} .$$

As in the exponential distribution, our estimator outperforms Wilcoxon score estimator for small r and large s, see, for example, Table 1.

4. Pitman Efficiency Comparision

We now conduct to explore the properties of the newly proposed rank estimate of regression parameters. The main purpose of this section is to investigate the asymptotic relative efficiencies of rank estimator based on Wilcoxon score denoted as Rank(1,1) relative to our rank estimator denoted as Rank(r,s) for the underlying distributions, uniform, normal, cauchy, double exponential, exponential and lognormal distributions. In addition, we will explore the selection of r and s which improves the efficiency of rank estimator based on Wilcoxon score.

When looking over the results of Table 1, we can say that uniform distribution and for the light tailed distribution such as normal distribution, generally low $r, s < 1$ yield higher asymptotic relative efficiencies for Rank(r,s). Meanwhile for the heavy tailed distribution such as cauchy and double exponential distributions, $1 < r, s < 2$ yield higher asymptotic relative efficiencies for Rank(r,s). In addition as r and s is changed toward 1.5, Rank(r,s) tends to give much improved efficiencies.

Table 1. Asymptotic relative efficiencies, ARE = Var(Rank(r,s)) / Var(Rank(1,1)), of Rank(1,1) with respect to Rank(r,s) for the regression parameters of underlying distributions. [r, s = (0.1) (3) (0.1)]

Symmetric Distribution											
Uniform			Normal			Cauchy			Double Exponential		
r	s	ARE	r	s	ARE	r	s	ARE	r	s	ARE
0.1	0.1	0.1113	0.1	0.1	0.9553	1.1	1.1	0.9898	1.1	1.1	0.9954
0.3	0.3	0.3755	0.3	0.3	0.9606	1.3	1.3	0.9784	1.3	1.3	0.9902
0.5	0.5	0.6442	0.5	0.5	0.9738	1.5	1.5	0.9763	1.5	1.5	0.9891
0.7	0.7	0.8383	0.7	0.7	0.9866	1.7	1.7	0.9814	1.7	1.7	0.9914
0.9	0.9	0.9605	0.9	0.9	0.9964	1.9	1.9	0.9925	1.9	1.9	0.9965
0.5	1.0	0.8169	0.5	1.0	0.9937	1.5	1.0	0.9930	1.5	1.0	0.9996
0.5	3.0	0.7984	1.1	0.8	0.9997	1.5	2.0	0.9907	1.5	2.0	0.9973
2.0	2.5	0.9546	1.8	2.5	0.9995	1.0	1.5	0.9930	1.0	1.5	0.9996
2.0	3.0	0.9077	2.0	2.5	0.9959	2.0	1.5	0.9907	2.0	1.5	0.9973
2.5	3.0	0.8608	2.0	3.0	0.9929						
3.0	3.0	0.8142	3.0	3.0	0.9784						

Asymmetric Distribution					
Exponential			Lognormal		
r	s	ARE	r	s	ARE
0.1	0.1	0.1113	0.1	0.1	0.5545
0.3	0.3	0.3755	0.3	0.3	0.7029
0.5	0.5	0.6442	0.5	0.5	0.8252
0.7	0.7	0.8383	0.7	0.7	0.9166
0.9	0.9	0.9605	0.9	0.9	0.9786
0.5	1.0	0.6003	0.5	1.0	0.7355
0.5	1.5	0.5353	0.5	1.5	0.6469
0.5	2.0	0.4796	0.5	2.0	0.5801
0.5	3.0	0.3979	0.5	3.0	0.4915
1.0	1.5	0.8547	1.0	1.5	0.8574
1.0	2.0	0.7469	1.0	2.0	0.7575
1.0	3.0	0.6022	1.0	3.0	0.6316
1.5	2.0	0.9069	1.5	2.0	0.8991
1.5	3.0	0.7144	1.5	3.0	0.7354
2.0	2.0	1.0000	2.0	2.0	1.0000
2.0	2.5	0.8698	2.0	2.5	0.8860
2.0	3.0	0.7734	2.0	3.0	0.8045
2.5	2.5	0.9079	2.5	2.5	0.9387
3.0	3.0	0.8142	3.0	3.0	0.8728

On the other hand, the results for the asymmetric distribution can be summarized as follows. Especially, right-skewed distribution such as exponential and lognormal which we encounter in practice commonly have the following patterns. Exponential and lognormal distributions indicate that if $r, s < 1$, Rank(r,s) is relatively much more efficient than Rank(1,1). Secondly in general if $1 < r < s$, Rank(r,s) is relatively more efficient. Futhermore as long as s becomes greater than r, Rank(r,s) shows much improved efficiencies. In other words, we can say that Rank(r,s) is usually more efficient than Rank(1,1) if s is increased for a lower r.

Table 2. Reasonable selection of r and s which improves the efficiency of our rank estimator $\widehat{\beta_{r,s}}$ with respect to Wilcoxon estimator $\widehat{\beta_{1,1}}$. [$r, s = (0.1)(3)(0.1)$]

Distribution		Symmetric				Asymmetric	
		Uniform	Normal	Cauchy	Double Exp.	Exponential	Lognormal
$r, s < 1$ ARE		$r, s = 0.1$ 0.1113	$r, s = 0.1$ 0.9553	None	None	$r, s = 0.1$ 0.1113	$r, s = 0.1$ 0.5545
$1 \leq r, s \leq 3$ ARE		$r, s = 3.0$ 0.8142	$r, s = 3.0$ 0.9784	$r, s = 1.5$ 0.9763	$r, s = 1.5$ 0.9891	$r=1.0$ $s=3.0$ 0.6022	$r=1.0$ $s=3.0$ 0.6316
OVERALL SUGGEST	primary	$r, s = 0.1$ 0.1113	$r, s = 0.1$ 0.9553	$r, s = 1.5$ 0.9763	$r, s = 1.5$ 0.9891	$r, s = 0.1$ 0.1113	$r, s = 0.1$ 0.5545
	alternative	$r, s = 0.3$ 0.3755	$r, s = 0.3$ 0.9606	$r, s = 1.3$ 0.9784	$r, s = 1.3$ 0.9902	$r=0.5$ $s=3.0$ 0.3979	$r=0.5$ $s=3.0$ 0.4915
		$r, s = 3.0$ 0.8142	$r, s = 3.0$ 0.9784	$r, s = 1.7$ 0.9814	$r, s = 1.7$ 0.9914	$r=1.0$ $s=3.0$ 0.6022	$r=1.0$ $s=3.0$ 0.6316

Table 2 indicates that the reasonable values of r and s for the symmetric distribution are simultaneously the same and as follows. Uniform distribution and for the light tailed distribution such as normal distribution, suggested values are (i) $r, s = 0.1$ (the lowest value if possible) if $r, s < 1$ or (ii) $r, s = 3.0$ (the highest value under the given range) if $1 \leq r, s \leq 3$. Meanwhile for the heavy tailed distribution such as cauchy and double exponential distributions, suggested values are $r, s = 1.5$.

Moreover for asymmetric distributions, especially for the right-skewed distributions such as exponential and lognormal, suggested values which show much improved efficiencies of Rank(r,s) are (i) $r, s = 0.1$ (the lowest) if $r, s < 1$ or (ii) $r = 1.0$ (the lowest) with $s = 3.0$ (the highest) if $1 \leq r, s \leq 3$.

Beyond the primary suggested values of r and s, Table 2 presents the secondary alternative values of r and s respectively. For instance, for the positively skewed distribution such as exponential and lognormal distributions, the alternative values of r and s are $r=0.5$ (as low as possible) with $s=3.0$ (as high as possible). These results are generated from Table 1.

5. Conclusions

In this paper we introduce the score generating function in rank regression for the linear regression model. We show that the distribution of our rank estimator $\widehat{\beta_{r,s}}$ based on this score function asymptotically converges to a multivariate normal.

Efficiency study shows that for the light and heavy tailed distributions, in general $r,s<1$ and $1<r,s<2$ yield higher asymptotic relative efficiencies respectively. Right-skewed distribution has many efficient possibilities; exponential and lognormal distributions show that in general if $r,s<1$ or $1<r<s$, our rank estimator is relatively more efficient than Wilcoxon score rank estimator.

The suggested values of r and s are (i) the lowest $r,s=0.1$ if $r,s<1$ or the highest $r,s=3.0$ if $1\leq r,s\leq 3$ for uniform distribution and the light tailed distribution(normal), (ii) $r,s=1.5$ for the heavy tailed distribution(cauchy, double exponential) and (iii) the lowest $r,s=0.1$ if $r,s<1$ or the lowest $r=1.0$ with the highest $s=3.0$ if $1\leq r,s\leq 3$ for the right-skewed distribution(exponential, lognormal).

References

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

[2] Choi, Y. H. and Ozturk, O. (2002). A New Class of Score Generating Functions for Regression Models, *Statistics and Probability Letters*, Accepted to Publish.

[3] Hardy, G. H., Littlewood, J. E. and Polya, G. (1952). *Inequalities*, Cambridge University Press, Cambridge, Massachusetts.

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

[5] Hettmansperger, T. P. and McKean J. W. (1983). A Geometric Interpretation of Inferences Based on Ranks in the Linear Model, *The Journal of American Statistical Association*, 78, 885-893.

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

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

[8] Jureckova, J. (1969). Asymptotic Linearity of a Rank Statistic in Regression Parameter, *The Annals of Mathematical Statistics*, 40, 1889-1900.

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

[10] McKean J. W. and Hettmansperger, T. P. (1978). A Robust Analysis of the General Linear Model Based on One Step R-Estimates, *Biometrika*, 65, 571-579.

[11] Naranjo, J. D. and Hettmansperger, T. P. (1994). Bounded Influence Rank Regression, *Journal of Royal Statistical Society*, 56, 209-220.

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

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

[14] 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.

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

[16] Witt, L. D., Naranjo, J. D. and McKean, J. W. (1995). Influence Functions for Rank-Based Procedures in The Linear Model, *Journal of Nonparametric Statistics*, 5, 339-358.

[2001년 11월 접수, 2002년 2월 채택]