The Korean Communications in Statistics
Vol. 13 No. 2, 2006, pp. 359-368

Rank Scores for Linear Models under Asymmetric Distributions [1)]

Young Hun Choi [2)]

Abstract

In this paper we derived the asymptotic relative efficiency, ARE(ms, rs), of our new score function with respect to the McKean and Sievers scores for the asymmetric error distributions which often occur in practice. We thoroughly explored the asymptotic relative efficiency, ARE(ms, rs), of our score function that provides much improvement over the McKean and Sievers scores for all values of r and s under asymmetric distributions.

Keywords : Rank Scores; Asymmetric Distribution; Dispersion function; Asymptotic Relative Efficiency; Generalized F distribution.

1. Introduction

Since McKean and Sievers (1989) discussed suitable score selection procedures for analyses of linear models under error distributions, Ozturk and Hettmansperger (1996) derived the robust estimates of location and scale parameters by minimizing distance criterion function. Ahmad (1996) developed a new class of Mann-Whitney-Wilcoxon type test statistics which only considered the one side tail probabilities of the underlying distribution. Ozturk and Hettmansperger (1997) considered the distribution functions reflecting both right and left tail probabilities. Ozturk (2001) considered another class of Mann-Whitney-Wilcoxon test statistics by incorporating both right and left tail behavior of the underlying distributions. Further Choi and Ozturk (2002) introduced a new score generating function for the rank dispersion function in a multiple linear regression model which improved the efficiency for many distributions by comparing the score function with the r th and s th power of the tail probabilities of the underlying probability distributions. Choi (2004a) explored efficiency comparison over the Wilcoxon scores under the asymmetric distributions in essence.

Now the main purpose of this paper is to extend the Hettmansperger and

1) This research was supported by Hanshin University Research Grant in 2006.
2) Professor, Department of Information and Statistics, Hanshin University, 447-791. Korea.
E-mail : choicyh@hanshin.ac.kr

McKean (1998) and Choi and Ozturk (2002)'s concept, where the distribution function reflects on both right and left tail probabilities and produces robust estimators with high efficiency, into the rank estimate of regression parameters in a linear model. In addition, this paper is to extend Choi (2004b)'s preferable rank scores characteristics for asymmetric errors by comparing with McKean and Sievers (1989) scores, which is currently known as one of optimal scores in terms of asymptotic relative efficiencies, since the asymmetric distributions are more realistic and applicable with respect to the diversified scheme of distributions in practice.

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 define the dispersion function $D_{rs}(\beta)$ based on our r th and s th power score function. In Section 3, we define the asymptotic relative efficiency, ARE(ms, rs), of our score function with respect to the McKean and Sievers scores. In Section 4, we compare the efficiency of rank estimator based on our proposed score generating function with the efficiency of rank estimator based on the McKean and Sievers scores for the asymmetric distributions. We thoroughly explore the impact of r and s that provides improvement over the McKean and Sievers scores.

2. Score Function

Consider the linear regression model,

$$y_i = \alpha + x_i^{'}\beta + e_i$$

for $i = 1, \cdots, n$, 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 $a(i) = \phi(i/(n+1))$, and the score generating function $\phi(u)$ is a nondecreasing, square integrable, bounded function on $(0,1)$ and satisfies the conditions $\int_0^1 \phi(u)\,du = 0$ and $\int_0^1 \phi^2(u)\,du = 1$.

Now let

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

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

where

$$\omega_{rs} = \frac{r^2}{(2r+1)(r+1)^2} + \frac{s^2}{(2s+1)(s+1)^2} + \frac{2}{(r+1)(s+1)} - 2\frac{\Gamma(r+1)\Gamma(s+1)}{\Gamma(r+s+2)}$$

Define the dispersion function

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

where $R(e_i)$ denotes the rank of $e_i = y_i - x_i^{'}\beta$. Then β can be estimated by the rank estimator $\hat{\beta}_{rs}$ which minimizes the dispersion function.

3. Pitman Efficiency

In this section, we compare the efficiency of our proposed score function with respect to the McKean and Sievers scores. The asymptotic variance of the rank estimate of β based on the McKean and Sievers scores is denoted as $v(\hat{\beta}_{ms})$. Then from Theorem 2 of Choi and Ozturk (2002) and Lemma 3 given below, the asymptotic relative efficiency of our estimator $\hat{\beta}_{rs}$ with respect to McKean and Sievers estimator $\hat{\beta}_{ms}$ is expressed as

$$\mathrm{ARE}(ms,rs) = \left(\frac{|v(\hat{\beta}_{rs})|}{|v(\hat{\beta}_{ms})|}\right)^{1/p} = \frac{\omega_{rs}}{\tau_{rs}} \frac{m_1+m_2+1}{m_1 m_2}, \tag{3.1}$$

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

The asymptotic relative efficiencies ARE(ms,rs), where ARE$(ms,rs) < 1$ implies that the efficiency of our score function is superior to that of the McKean and Sievers scores, are discussed below for several asymmetric distributions by using the generalized F distribution. Let F be a random variable having an $F_{2m_1,2m_2}$ distribution with $2m_1$ and $2m_2$ degrees of freedoms. Then $T = log(F)$ is said to have the generalized F distribution, $GF(2m_1, 2m_2)$, with $2m_1$ and $2m_2$ degrees of freedoms with a variety of shape and tail behaviors.

Lemma 1. The generalized F distribution, $GF(2m_1, 2m_2)$, with degrees of freedoms $2m_1$ and $2m_2$ has the following probability density function.

$$f_{2m_1,2m_2}(y) = \frac{\Gamma(m_1+m_2)(m_1/m_2)^{m_1}}{\Gamma(m_1)\,\Gamma(m_2)} \cdot \frac{e^{y\,m_1}}{[1+(m_1/m_2)\,e^y]^{m_1+m_2}} \,. \tag{3.2}$$

Proof. First consider a random variable F having the ordinary $F_{h,k}$ distribution with h and k degrees of freedoms. Then the probability density function can be written as

$$f_{h,k}(F) = \frac{\Gamma((h+k)/2)(h/k)^{h/2}}{\Gamma(h/2)\,\Gamma(k/2)} \cdot \frac{F^{h/2-1}}{(1+(h/k)F)^{(h+k)/2}} . \tag{3.3}$$

Let $y=\log(F)$, $h=2m_1$ and $k=2m_2$. By using integration by parts and substituting $F=e^y$ and $dF=e^y dy$ into (3.3), we can show that

$$f_{2m_1, 2m_2}(y) = \frac{\Gamma(m_1+m_2)(m_1/m_2)^{m_1}}{\Gamma(m_1)\,\Gamma(m_2)} \cdot \frac{e^{y\,m_1}}{[1+(m_1/m_2)\,e^y]^{m_1+m_2}} .$$

By using Lemma 1, we state the following Lemma 2.

Lemma 2. The McKean and Sievers score generating function $\phi_{ms}(y) = [m_1m_2(e^y-1)]/[m_2+m_1e^y]$ has the following properties for random errors with $GF(2m_1, 2m_2)$ distribution.

$$\text{(i)} \quad \phi'_{ms}(y) = \frac{m_1\,(m_1+m_2)\,e^y}{m_2\left(1+\frac{m_1}{m_2}e^y\right)^2} , \tag{3.4}$$

$$\text{(ii)} \quad \int \phi'_{ms}(y)f(y)\,dy = \frac{m_1\,m_2}{m_1+m_2+1} . \tag{3.5}$$

Proof. Let $F_{2m_1, 2m_2}(y) = F(y)$. Then for the McKean and Sievers scores,

$$\phi_{ms}(y) = \frac{m_1\,m_2\,(e^y-1)}{m_2+m_1\,e^y}$$

$$\text{(i)} \quad \phi'_{ms}(y) = m_1m_2\,e^y(m_2+m_1e^y)^{-1} - (m_2+m_1e^y)^{-2}\,m_1e^y \cdot m_1m_2(e^y-1)$$

$$= \frac{m_1\,m_2\,e^y\,(m_2+m_1)}{(m_2+m_1e^y)^2}$$

$$= \frac{m_1}{m_2} \cdot \frac{e^y\,(m_2+m_1)}{\left(1+\frac{m_1}{m_2}e^y\right)^2} .$$

(ii) Substituting the result (3.2) of Lemma 1 and (3.4) of Lemma 2 yields that

$$\int \phi'_{ms}(y)f(y)\,dy = \frac{m_1(m_2+m_1)}{m_2}\frac{\Gamma(m_1+m_2)(m_1/m_2)^{m_1}}{\Gamma(m_1)\,\Gamma(m_2)} \cdot \int \frac{e^y\,e^{ym_1}}{[1+(m_1/m_2)\,e^y]^{2+m_1+m_2}}dy$$

$$= \frac{m_1(m_2+m_1)}{m_2}\frac{\Gamma(m_1+m_2)(m_1/m_2)^{m_1}}{\Gamma(m_1)\,\Gamma(m_2)} \cdot \int \frac{e^{y(m_1+1)}}{[1+(m_1/m_2)\,e^y]^{m_1+m_2+2}}dy$$

let $(m_1/m_2)e^y = x$ and $(m_1/m_2)e^y dy = dx$, then

$$= \frac{m_1(m_2+m_1)}{m_2}\frac{\Gamma(m_1+m_2)(m_1/m_2)^{m_1}}{\Gamma(m_1)\,\Gamma(m_2)} \cdot \int \frac{(m_2/m_1)^{m_1}\, x^{m_1}}{(1+x)^{m_1+m_2+2}}\left(\frac{m_2}{m_1}\right)dx$$

$$= \frac{m_1(m_2+m_1)}{m_2}\frac{\Gamma(m_1+m_2)(m_2/m_1)}{\Gamma(m_1)\,\Gamma(m_2)} \cdot \int \frac{x^{m_1}}{(1+x)^{m_1+m_2+2}}\,dx$$

by using the property of probability density function of $B(m_1+1, m_2+1)$ distribution

$$= \frac{m_1(m_2+m_1)}{m_2}\frac{\Gamma(m_1+m_2)(m_2/m_1)}{\Gamma(m_1)\,\Gamma(m_2)} \cdot \frac{\Gamma(m_1+1)\,\Gamma(m_2+1)}{\Gamma(m_1+m_2+1)}$$

$$= \frac{m_1\; m_2}{m_1+m_2+1}\ .$$

By using Lemma 2, we derive the following Lemma 3.

Lemma 3. The asymptotic relative efficiency of our estimator with respect to the McKean and Sievers scores for the generalized F distribution with $2m_1$ and $2m_2$ degrees of freedoms is expressed as

$$\mathrm{ARE}(ms, rs) = \frac{\omega_{rs}}{\tau_{rs}}\frac{m_1+m_2+1}{m_1 m_2}\ ,$$

where $\omega_{rs} = \dfrac{r^2}{(2r+1)(r+1)^2} + \dfrac{s^2}{(2s+1)(s+1)^2} + \dfrac{2}{(r+1)(s+1)} - 2\dfrac{\Gamma(r+1)\Gamma(s+1)}{\Gamma(r+s+2)}$

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

Proof. For easy reference, from Hettmansperger and McKean (1998) we define the McKean and Sievers scale parameter as $\int \phi_{ms}(u)\phi_f(u)du$ where $\phi_f(u) = -f'(F^{-1}(u))/f(F^{-1}(u))$. Further let $u = F_{2m_1, 2m_2}(y) = F(y)$. Then

$$\int \phi_{ms}(u)\,\phi_f(u)du = -\int \phi_{ms}(u)\,\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}\,du$$

$$= -\int \phi_{ms}(y)\,f'(y)\,dy$$

with partial integration

$$= -\,\phi_{ms}(y)f(y) + \int \phi'_{ms}(y)f(y)\,dy$$

$$= \int \phi'_{ms}(y)f(y)\,dy$$

from the result (3.5) of Lemma 2

$$= \frac{m_1 \ m_2}{m_1 + m_2 + 1}$$

Accordingly we can sat that

$$\frac{1}{\int \phi_{ms}(u) \ \phi_f(u) du} = \frac{m_1 + m_2 + 1}{m_1 \ m_2}$$

Therefore when substituting the above result into (3.1), ARE(ms, rs) for the generalized F distribution can be obtained straightforwardly.

$$\text{ARE}(ms, rs) = \frac{\omega_{rs}}{\tau_{rs}} \frac{m_1 + m_2 + 1}{m_1 m_2}$$

4. Efficiency Comparisons

In this section, we evaluate the efficiency of our score function with respect to the McKean and Sievers scores for asymmetric distributions by using the generalized F distribution. We basically explore the impact of r and s that provides improvement over the McKean and Sievers scores. For comparison purposes, ARE(ms, rs) provided in Lemma 3 are calculated for asymmetric distributions such as right-skewed and left-skewed distributions respectively. We evaluated ARE(ms, rs) for several values of $r, s = 0\,(3)\,0.1$. Tables and figures depict ARE(ms, rs) and corresponding perspective plots as a function of r and s.

The generalized F distribution is a very flexible distribution that covers a variety of shape and tail behaviors. It produces asymmetric distributions which are positively skewed distributions for $m_1 > m_2$ and negatively skewed distributions for $m_1 < m_2$.

4.1 Right-Skewed Distributions

<Table 1>, <Figure 1> and <Table 2>, <Figure 2> show ARE(ms, rs) and corresponding pdf and perspective plots for the right-skewed distributions such as $GF(3, 0.3)$ and $GF(4, 0.2)$. The computations were made for all $r, s = 0\,(3)\,0.1$.

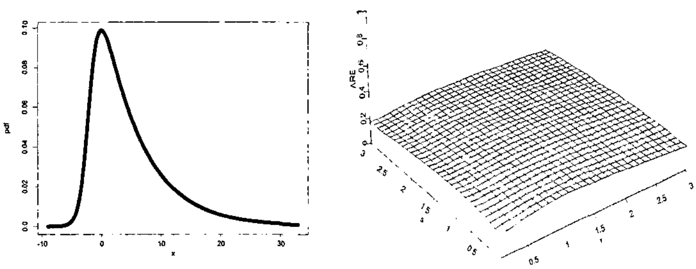
The results of <Table 1>, <Figure 1> and <Table 2>, <Figure 2> can be summarized as follows. For the right-skewed distributions, they indicate that $v(\hat{\beta}_{rs})$ is smaller than $v(\hat{\beta}_{ms})$ for all values of r and s. Namely for the right-skewed distributions with high positive skewness, our proposed score generating function provides much improved efficiency over the McKean and Sievers scores for all values of r and s.

<Table 1> ARE(ms, rs) for $GF(3, 0.3)$ distribution

r \ s	0.1	0.5	1.0	1.5	2.0	2.5	3.0
0.1	0.166	0.204	0.221	0.218	0.205	0.186	0.166
0.5	0.202	0.215	0.225	0.223	0.215	0.203	0.190
1.0	0.231	0.233	0.243	0.244	0.238	0.229	0.218
1.5	0.231	0.234	0.246	0.249	0.246	0.238	0.229
2.0	0.217	0.224	0.240	0.245	0.243	0.236	0.227
2.5	0.198	0.212	0.231	0.237	0.236	0.230	0.221
3.0	0.179	0.200	0.221	0.229	0.228	0.222	0.213

ARE(ms, rs)

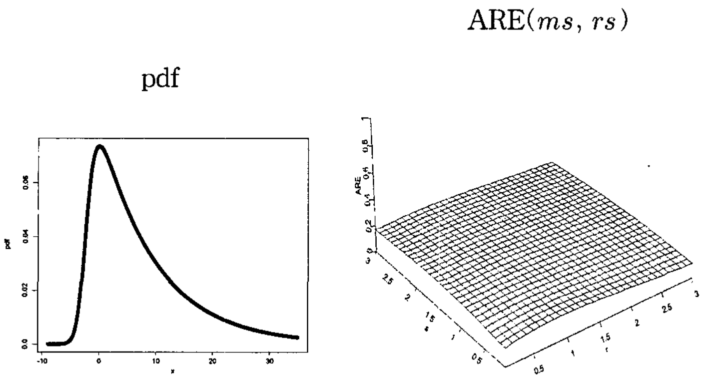
pdf



<Figure 1> pdf and ARE(ms, rs) for $GF(3, 0.3)$ distribution

<Table 2> ARE(ms, rs) for $GF(4, 0.2)$ distribution

r \ s	0.1	0.5	1.0	1.5	2.0	2.5	3.0
0.1	0.132	0.157	0.180	0.189	0.189	0.182	0.173
0.5	0.166	0.171	0.186	0.192	0.193	0.189	0.184
1.0	0.175	0.176	0.191	0.199	0.202	0.200	0.197
1.5	0.163	0.168	0.184	0.194	0.199	0.199	0.196
2.0	0.144	0.155	0.174	0.186	0.191	0.191	0.189
2.5	0.126	0.142	0.164	0.176	0.182	0.183	0.181
3.0	0.111	0.132	0.155	0.169	0.175	0.176	0.173



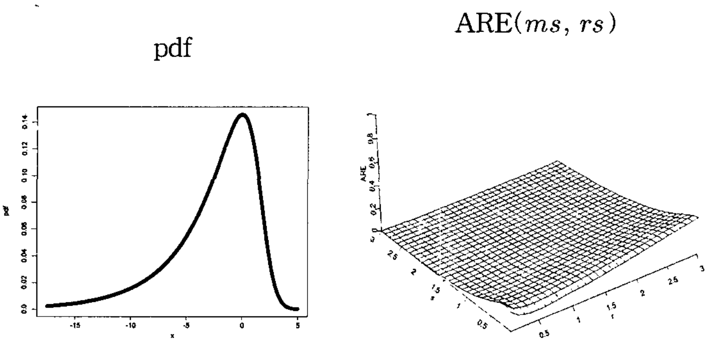
<Figure 2> pdf and ARE(ms, rs) for $GF(4, 0.2)$ distribution

4.2 Left-Skewed Distributions

<Table 3>, <Figure 3> and <Table 4>, <Figure 4> show ARE(ms, rs) and corresponding pdf and perspective plots for the left-skewed distributions such as $GF(0.5, 6)$ and $GF(0.2, 4)$.

<Table 3> ARE(ms, rs) for $GF(0.5, 6)$ distribution

r \ s	0.1	0.5	1.0	1.5	2.0	2.5	3.0
0.1	0.155	0.128	0.050	0.019	0.008	0.004	0.002
0.5	0.089	0.094	0.047	0.025	0.015	0.010	0.008
1.0	0.117	0.112	0.064	0.040	0.028	0.023	0.021
1.5	0.161	0.142	0.087	0.059	0.045	0.039	0.037
2.0	0.211	0.176	0.113	0.080	0.064	0.058	0.055
2.5	0.264	0.211	0.138	0.102	0.085	0.077	0.075
3.0	0.317	0.245	0.164	0.123	0.105	0.097	0.095

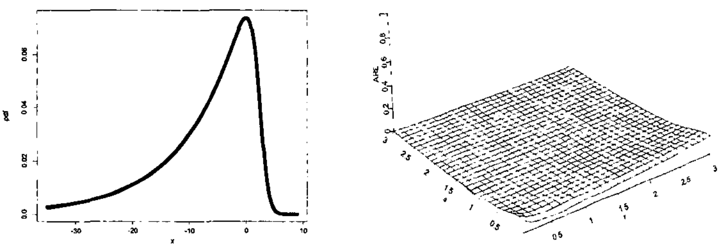


<Figure 3> pdf and ARE(ms, rs) for $GF(0.5, 6)$ distribution

<Table 4> ARE(ms, rs) for $GF(0.2, 4)$ distribution

r \ s	0.1	0.5	1.0	1.5	2.0	2.5	3.0
0.1	0.096	0.062	0.016	0.005	0.001	0.001	0.001
0.5	0.045	0.042	0.016	0.007	0.004	0.003	0.002
1.0	0.055	0.048	0.022	0.012	0.008	0.007	0.006
1.5	0.074	0.061	0.031	0.019	0.014	0.013	0.012
2.0	0.098	0.077	0.041	0.027	0.022	0.020	0.020
2.5	0.126	0.094	0.053	0.037	0.031	0.029	0.029
3.0	0.155	0.112	0.066	0.047	0.040	0.039	0.039

pdf ARE(ms, rs)



<Figure 4> pdf and ARE(ms, rs) for $GF(0.2, 4)$ distribution

As we compare the left-skewed distributions with the right-skewed distributions, similar results can be observed. <Table 3>, <Figure 3> and <Table 4>, <Figure 4> show that our procedure has higher efficiency than the McKean and Sievers scores for all values of r and s. Further for the left-skewed distributions with low negative skewness, our proposed score generating function provides much improved efficiency over the McKean and Sievers scores for all values of r and s.

5. Conclusions

In this paper we derived the asymptotic relative efficiency, ARE(ms, rs), of our score function with respect to the McKean and Sievers scores for the asymmetric distributions by using the generalized F distributions. We thoroughly explored our new score generating function that provides much improvement over the McKean and Sievers scores. The result shows that our proposed score generating function provides severely improved efficiency over the McKean and Sievers scores for all values of r and s under asymmetric distributions which we commonly encounter in practice.

References

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

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

[3] Choi, Y.H. (2004a). Asymptotic Relative Efficiency for New Score Functions in Rank Regression Models. *The Korean Journal of Applied Statistics*, Vol. 17, 269-280.

[4] Choi, Y.H. (2004b). Asymptotic Relative Efficiency for New Score Functions in the Generalized F Distribution. *The Korean Communications in Statistics*, Vol. 11, 435-446.

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

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

[7] McKean, J.W. and Sievers, G.L. (1989). Rank Scores Suitable for Analyses of Linear Models under Asymmetric Error Distributions. *Technometrics*, Vol. 31, 207-218.

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

[9] 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*, Vol. 29, 233-244.

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

[Received May 2006, Accepted June 2006]