On a Robust Subset Selection Procedure for the Slopes of Regression Equations[†]

Moon Sup Song* & Chang Hyuck Oh*

ABSTRACT

The problem of selection of a subset containing the largest of several slope parameters of regression equations is considered. The proposed selection procedure is based on the weighted median estimators for regression parameters and the median of rescaled absolute residuals for scale parameters. Those estimators are compared with the classical least squares estimators by a simulation study. A Monte Carlo comparison is also made between the new procedure based on the weighted median estimators and the procedure based on the least squares estimators. The results show that the proposed procedure is quite robust with respect to the heaviness of distribution tails.

1. Introduction

Consider a set of k simple linear regression equations

$$Y_{ij}=\alpha_i+\beta_i x_{ij}+e_{ij},\ \ j=1,\cdots,\ n\ ;\qquad i=1,\cdots,\ k \tag{1.1}$$

where α_i's and β_i's are unknown regression parameters, $x_{i1},\cdots,x_{in}$ are known constants, and $e_{i1},\cdots,\ e_{in}$ $(i=1,\cdots;k)$ are independent and identically distributed random variables with a continuous and symmetric density function f. It is also assumed that the e_{ij}'s have location and scale parameters 0 and σ, respectively. We are interested in selecting a subset containing the regression equation associated with the largest-slope parameter.

† This work was supported in part by the Research Fund of the Ministry of Education, Korean Government, in 1981. The authors are grateful to Dr. Woo-Chul Kim for his helpful suggestions throughout the preparation of this work.

* Department of Computer Science & Statistics, Seoul National University.

In the subset selection procedure it is usually required that for any given rule R the probability of a correct selection (CS) is at least a preassinged number P^*, i.e.,

$$\inf_{\beta} P(CS|R) \geq P^* \tag{1.2}$$

where $P^* \in \left(\frac{1}{k}, 1\right)$. Thus we need the information of the configuration of β_i's for which the P^* condition in (1.2) is satisfied. This configuration is called the least favorable configuration (LFC). For nonparametric selection procedures based on ranks, it is well known that the LFC is usually not given by the equi-parameter configuration (see Rizvi and Woodworth (1970)). Because of this difficulty the nonparametric procedures based on Hodges-Lehmann (H-L) type estimators are often suggested as robust procedures.

Assuming that the common variance σ^2 is known, Gupta and Huang (1977) proposed a nonparametric procedure for the selection of regression equation with the largest slope based on the H-L type estimators suggested by Adichie (1967). They used a robust procedure to estimate the regression parameter β, but not for the scale parameter σ. Thus, if we estimate σ by any classical method based on the normal theory, the robustness of the procedure may be lost.

In this paper we propose a robust selection procedure based on the weighted median estimator (WME). The WME is a H-L type estimator in estimating the regression parameter β, which has been studied by Scholz (1978) and Sievers (1978). To estimate the scale parameter σ, we use the median of rescaled absolute residuals, which is supposed to be quite robust in estimating scale parameters.

Section 2 deals with the definitions and properties of the WME of α and β. In Section 3 the small sample properties of the WME are investigated by a comparison with the LSE using a Monte Carlo experiment. The results show that the WME is much more robust than the LSE for heavy tailed distributions. A robust scale estimator based on rescaled absolute residuals is proposed and compared with the classical residual mean square estimator. Section 4 contains the formulations for selecting the "best" regression equation based on the WME and on the least squares estimator (LSE). We used a rather heuristic approach in making the subset selection rule based on the WME because of the complexity of the distributions. Section 5 consists of a Monte Carlo study to compare the selection procedures. The results also show that the WME procedure is quite robust with respect to the heaviness of distribution tails.

2. Regression Parameter Estimators

In this section we consider the simple linear regression model

$$Y_i=\alpha+\beta x_i+e_i,\ \ i=1,\cdots,\ n$$

instead of the model (1. 1). Without loss of generality we assume that $x_1\leq x_2\leq\cdots\leq x_n$ with

$$c_n{}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2>0.$$

The classical LSE of β and α are given by

$$\tilde{\beta}=\frac{\sum_1^n(x_i-\bar{x}_n)(Y_i-\bar{Y}_n)}{\sum_1^n(x_i-\bar{x}_n)^2}\quad\text{and}\quad\tilde{\alpha}=\bar{Y}_n-\tilde{\beta}\bar{x}_n,$$

respectively. Alternative H-L type estimators of β based on suitable rank tests are proposed by Theil (1950), Brown and Mood (1951), Adichie (1967), Sen (1968), Jaeckel (1972), Scholz(1978), Sievers(1978), Kildea(1981), and others. In this paper we use Scholz-Sievers procedure to estimate β.

Let T_β be the weighted rank statistic defined by

$$T_\beta=\sum_{1\leq i<j\leq n}w_{ij}\phi(Y_j-Y_i-\beta(x_j-x_i)),$$

where $\phi(\cdot)$ is an indicator function such that $\phi(t)=0$ or 1 according as $t\leq 0$ or $t>0$. Here the weights w_{ij} are non-negative, and $w_{ij}=0$ whenever $x_i=x_j$. Let $w_{..}$ be the sum of w_{ij}, *i.e.*, $w_{..}=\sum_{i<j}w_{ij}$. The H-L type estimator $\hat{\beta}$ of β based on T_β is given by

$$\hat{\beta}=(\hat{\beta}_U+\hat{\beta}_L)/2 \tag{2.1}$$

where

$$\hat{\beta}_U=\sup\{\beta : T_\beta\geq w_{..}/2\},$$

$$\hat{\beta}_L=\inf\{\beta : T_\beta\leq w_{..}/2\}.$$

The estimator $\hat{\beta}$ in (2. 1) is equivalent to those of Sievers (1978) and Scholz (1978). Note that for the special case $w_{ij}\equiv 1$, the distribution of T_β is the same as that of two-sample Wilcoxon statistic. But, we are in this paper interested in the weight $w_{ij}=x_j-x_i$, $i<j$, which is one of the optimal weights in the sense that it gives a minimum variance.

To estimate α simultaneously with β, we consider the rank statistic T_α defined by

$$T_\alpha=\sum_{i\le j}\psi(Y_i+Y_j-\hat{\beta}(x_i+x_j)-2\alpha),$$

and define the estimator $\hat{\alpha}$ of α by

$$\hat{\alpha}=(\hat{\alpha}_U+\hat{\alpha}_L)/2$$

where

$$\hat{\alpha}_U=\sup\{\alpha: T_\alpha\ge n(n+1)/4\},$$

$$\hat{\alpha}_L=\inf\{\alpha: T_\alpha\le n(n+1)/4\}.$$

The estimator $\hat{\alpha}$ is then given by

$$\hat{\alpha}=\operatorname*{med}_{i\le j}\frac{1}{2}\{Y_i+Y_j-\hat{\beta}(x_i+x_j)\}. \tag{2.3}$$

To find the explicit form of the $\hat{\beta}$ in (2.1) we consider $\binom{n}{2}$ pairwise slopes

$$S_{ij}=(Y_j-Y_i)/(x_j-x_i), \tag{2.4}$$

for $i<j$. Then T_β is a function of the slope S_{ij} since $\psi(Y_j-Y_i-\beta(x_j-x_i))=1$ whenever $S_{ij}>\beta$.

Consider the probability distribution on the $\binom{n}{2}$ points S_{ij} by assigning probability $w_{ij}/w_{..}$ to S_{ij}, $i<j$. Then $\hat{\beta}$ is the median of this probability distribution, and we call $\hat{\beta}$ the WME of β. Note that for the weights $w_{ij}\equiv 1$, $\hat{\beta}$ is the median of the slopes S_{ij} in (2.4), which was suggested by Theil (1950) and Sen(1968). For the weights $w_{ij}=x_j-x_i$, $\hat{\beta}$ is the estimator originally considered by Jaeckel(1972). We now present the invariance properties of $\hat{\alpha}$ and $\hat{\beta}$ in the following. The proof is similar to that of Adichie (1967, Lemma 4.1 & 4.2).

Theorem 1. Let $\hat{\beta}(Y)$ and $\hat{\alpha}(Y)$ be the estimators defined by (2.1) and (2.2), respectively. Then

i) $\hat{\beta}(Y+a+bx)=\hat{\beta}(Y)+b$

ii) $\hat{\alpha}(Y+a+bx)=\hat{\alpha}(Y)+a$

From this invariance property of $\hat{\beta}$, we can obtain the LFC in a subset selection procedure based on $\hat{\beta}$, which will be discussed in Section 3.

We now state the asymptotic distributions of $\hat{\alpha}$ and $\hat{\beta}$ in the following theorem. The proof is similar to that of Adichie (1967, Theorem 5.3), and is omitted here. See also Scholz (1978) and Sievers (1978) for the asymptotic normality of $\hat{\beta}$.

Theorem 2. Let $\hat{\alpha}$ and $\hat{\beta}$ be defined by (2.2) and (2.1), respectively. Let $w_{ij}=x_j-x_i$, $i<j$, and let $\bar{x}_n=\frac{1}{n}\sum_{i=1}^{n}x_i\to\bar{x}$ as $n\to\infty$. Assume that

i) $\int (f'/f)^2 f dy < \infty$,

ii) $\max_{1\le i\le n}(x_i-\bar{x}_n)^2/n \rightarrow 0$, as $n\rightarrow\infty$,

iii) $c_n^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2 \rightarrow c^2>0$, as $n\rightarrow\infty$.

Then $\sqrt{n}(\hat{\alpha}-\alpha,\ \hat{\beta}-\beta)$ has a limiting bivariate normal distribution with mean (0, 0) and covariance matrix $h^2(f)\Sigma$, where

$$h^2(f)=\frac{1}{12(\int f^2 dy)^2} \tag{2.5}$$

$$\Sigma=\begin{bmatrix} 1+\bar{x}^2/c^2 & -\bar{x}/c^2 \\ -\bar{x}/c^2 & 1/c^2 \end{bmatrix}$$

3. An Empirical Study on the Estimators

It is well known that the H-L type estimators discussed in Section 2 have desirable asymptotic properties. In this section we investigate the small sample behaviors of the WME. Moreover, to apply the estimators to the subet selection procedure we have to estimate the scale parameter σ. Thus not only the robustness of the slope estimators but also a robust scale estimator is our major concern. The scale estimator can be viewed more or less as a location estimation problem for $|r_i|$, where r_i is the residual of the i-th observation.

However, the variance of r_i depends on the design point x_i. It may be desirable to rescale each r_i so that the variance of r_i is approximately σ^2 in the normal case. When the LSE are used, the standard deviation of r_i is $\delta_i\sigma$ where δ_i is given by

$$\delta_i^2=\frac{n-1}{n}-\frac{(x_i-\bar{x}_n)^2}{\sum(x_i-\bar{x}_n)^2}$$

We thus, as a heuristic estimator, employ the following scale estimator:

$$\hat{\sigma}=1.48\ \mathrm{med}\{|r_i^*|\}, \tag{3.1}$$

where the regression equation is fitted by WME, r_i^* is the rescaled r_i, *i.e.*, $r_i^*=r_i/\delta_i$, and the factor 1.48 makes $\hat{\sigma}$ an approximately unbiased estimator of σ in normal case. Holland and Welsch (1977) discussed the other aspect of scale estimator.

In normal case σ^2 is usually estimated by the residual mean square

$$s^2=\Sigma r_i^2/(n-2), \tag{3.2}$$

using LSE to fit the regression equation.

In order to investigate the small sample properties of WME and $\hat{\sigma}$ we performed Monte Carlo experiments with pseudo random variables for uniform, normal, double exponential, SLASH (normal/uniform), and Cauchy distributions. In each case the location and scale parameters used are 0 and 1, respectively. The observed values are simulated from the model $Y_i=\alpha+\beta x_i+e_i$, $i=1,\cdots,n$, with $\alpha=0$, $\beta=1$, and $x_i=i$, $i=1,\cdots,n$, for $n=5$, 10, and 15. In each case 1000 observations are generated. The uniform random numbers were generated by using the intrinsic subroutine RANDU in PDP 11, and converted to normal variates by standardizing sum of twelve random numbers. The inverse integral transformation was applied to generated double exponential and Cauchy samples. All computations in this paper are carried out on PDP 11.

Table 1 summarized the results of the Monte Carlo study about WME ($\hat{\alpha}$ and $\hat{\beta}$) and LSE ($\tilde{\alpha}$ and $\tilde{\beta}$). The values are empirical means and variances (appeared in parentheses) of estimators. The relative efficiency (R.E.) of the two estimators is computed as an inverse ratio of empirical variances.

For the uniform, normal, and double exponential distributions, the LSE and WME are compatible. But the relative efficiency of the WME with respect to the LSE in SLASH and Cauchy case is tremendously increased. The estimated values for the WME are also very close to the true values, but not for the LSE.

The results of the empirical study for the performance of $\hat{\sigma}$ in (3.1) and s in (3.2) are presented in Table 2. From the table we can see that $\hat{\sigma}$ is much more robust than s. The relative efficiency of the two estimators, denoted by R.E. $(\hat{\sigma}, s)$ in the table, is again computed as an inverse ratio of empirical variances. The values of R.E. $(\hat{\sigma}, s)$ are uniformly less than one up to double exponential, but significantly large for very heavy tailed distributions.

Table 1 Empirical Means, Variances*, and Relative Efficiencies of the Estimators of Regression Parameters

(true value: $\beta=1$, $\alpha=0$)

	n	$\tilde{\beta}$(LSE)	$\hat{\beta}$(WME)	R.E.$(\hat{\beta}, \tilde{\beta})$
	5	1.01(0.08)	1.01(0.08)	0.94
Uniform	10	1.00(0.01)	1.00(0.01)	0.86
	15	1.00(0.00)	1.00(0.00)	0.86
	5	0.99(0.10)	0.98(0.12)	0.94
Normal	10	1.00(0.01)	1.00(0.01)	0.87
	15	1.00(0.00)	1.00(0.00)	0.92

Table 1 (Continued)

	n	$\tilde{\beta}$(LSE)	$\hat{\beta}$(WME)	R.E.$(\hat{\beta}, \tilde{\beta})$
	5	1.02(0.22)	1.03(0.24)	0.91
Double Exp	10	1.00(0.03)	1.00(0.02)	1.15
	15	1.01(0.01)	1.00(0.01)	1.19
	5	0.56(817.)	0.71(336.)	3.
SLASH	10	1.16(12.7)	1.00(0.18)	73.
	15	0.73(16.0)	0.99(0.04)	385.
	5	1.51(374.)	1.44(94.5)	4.
Cauchy	10	0.89(339.)	1.00(0.14)	2348.
	15	1.70(197.)	1.00(0.03)	7915.
	n	$\tilde{\alpha}$(LSE)	$\hat{\alpha}$(WME)	R.E.$(\hat{\alpha}, \tilde{\alpha})$
	5	−0.03(0.83)	−0.03(0.90)	0.92
Uniform	10	−0.02(0.38)	−0.01(0.45)	0.85
	15	0.03(0.23)	0.03(0.27)	0.85
	5	0.06(1.09)	0.08(1.22)	0.89
Normal	10	−0.00(0.45)	−0.00(0.51)	0.88
	15	0.01(0.30)	0.01(0.32)	0.93
	5	−0.06(2.39)	−0.07(2.60)	0.92
Double Exp.	10	−0.01(1.02)	−0.02(0.87)	1.18
	15	−0.02(0.56)	−0.01(0.47)	1.20
	5	0.66(12639)	0.79(4819)	3.
SLASH	10	0.62(339.)	0.02(6.62)	51.
	15	2.40(1866.)	0.02(3.06)	610.
	5	0.14(5055.)	−0.97(505.)	10.
Cauchy	10	−2.08(11227)	0.02(6.05)	1855.
	15	−3.71(5758.)	0.00(1.94)	2973.

*Empirical variances appeared in parentheses

Table 2 Empirical Means, Variances*, and Relative Efficiencies of the Estimators of Scale Parameter

	n	s	$\hat{\sigma}$	R.E.$(\hat{\sigma}, s)$
	5	0.80(0.09)	0.94(0.32)	0.27
Uniform	10	0.85(0.03)	0.96(0.14)	0.18
	15	0.85(0.01)	1.00(0.09)	0.17

Table 2 (Continued)

	n	s	$\hat{\sigma}$	R.E.$(\hat{\sigma}, s)$
	5	0.93(0.15)	1.08(0.47)	0.32
Normal	10	0.98(0.06)	1.00(0.16)	0.37
	15	0.98(0.04)	0.99(0.11)	0.32
	5	1.22(0.45)	1.30(0.94)	0.48
Double Exp.	10	1.35(0.24)	1.11(0.30)	0.85
	15	1.37(0.15)	1.10(0.19)	0.82
	5	13.70(3661.)	8.64(1776.)	2.
SLASH	10	14.69(5587.)	2.70(2.93)	1905.
	15	21.94(8561.)	2.54(1.44)	5957.
	5	12.11(5984.)	4.96(497.)	12.
Cauchy	10	20.10(37447.)	2.02(2.63)	14246.
	15	22.21(38474.)	1.76(0.85)	45368.

*Empirical variances appeared in parantheses

4. Subset Selection Procedures for the Largest Slope

We consider again the set of k regression equations in (1.1)

$$Y_{ij}=\alpha_i+\beta_i x_{ij}+e_{ij}, \quad j=1,\cdots,\ n\ ; \quad i=1,\cdots,\ k.$$

Here we assume that the regularity conditions in Theorem 2 hold for each $i=1,\cdots,\ k$. We again use the weights $w_{lij}=x_{lj}-x_{li}$, $l=1,\cdots,\ k$. Let

$$c_{in}^2=\frac{1}{n}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{in})^2 \quad \text{with} \quad \bar{x}_{in}=\frac{1}{n}\sum_{j=1}^{n} x_{ij}, \tag{4.1}$$

and assume that $c_{in}^2 \to c_i^2 > 0$ as $n\to\infty$, for $i=1,\cdots,\ k$.

Let $\beta_{[1]}\leq\beta_{[2]}\leq\cdots\leq\beta_{[k]}$ denote the ordered β_i's, and let $c^2_{(in)}$, $c^2_{(i)}$, and $\pi_{(i)}$ be the values of c_{in}^2, c_i^2, and the regression equation associated with $\beta_{[i]}$ respectively. Here we are interested in selecting a subset which contains the "best" population associated with $\beta_{[k]}$.

Assuming that the common variance σ^2 is known, Gupta and Huang (1977) have suggested the following rule:

R : Select π_i if and only if

$$\hat{\beta}_i^* \geq \max_{1\leq j\leq k}\left(\hat{\beta}_j^* - \frac{d\sigma}{\sqrt{n}}\sqrt{c_i^{-2}+c_j^{-2}}\right), \tag{4.2}$$

where $\hat{\beta}_i{}^*$ is the Adichie estimator for β_i with Wilcoxon scores or normal scores, and $d=d(k, P^*, n)>0$ is chosen to satisfy the P^*-condition. For conservative values of d they suggested the inequality

$$P^* \geq \int_{-\infty}^{\infty} \Phi^{k-1}\left(\frac{d^*\sigma-\alpha_1 u}{\sqrt{1-\alpha_1{}^2}}\right) d\Phi(u)$$

where $\sigma_1^2=c_{[1]}^2/(c_{[1]}^2+c_{[k]}^2)$ with $c_{[1]}\leq c_{[2]}\leq\cdots\leq c_{[k]}$, $d^*=\sqrt{0.864}d$ or d according as Wilcoxon scores or normal scores are used, and $\Phi(\cdot)$ is the cdf of standard normal. (Note that we use the notation σ for both standard deviation and scale parameter without confusion.)

The rule R in (4.2) is not useful for practical purpose in two aspects. Firstly the scale parameter σ should in most cases be estimated from the data. Secondly if the usual pooled sample estimator s^2 to estimate σ^2 is used, then the procedure may be too conservative to discriminate bad populations. The numerical solution of the Adichie estimator is not so simple as that of the WME. Considering these aspects, we propose the following rule:

R_1 : Select π_i if and only if

$$\hat{\beta}_i\geq\max_{1\leq j\leq k}(\hat{\beta}_j-\frac{d_1\hat{\sigma}}{\sqrt{n}}\sqrt{c_{in}{}^{-2}+c_{jn}{}^{-2}}), \qquad (4.3)$$

where $d_1=d_1(k, P^*, n, c_1, \cdots, c_k)>0$ is determined so that the P^*-condition is satisfied and $\hat{\sigma}$ is the pooled sample estimate of σ defined by $1.48\text{med}\ \{|r_{ij}{}^*|,\ j=1,\cdots,\ n\ ;\ i=1,\cdots,\ k\}$. To obtain the exact values of d_1 we have to know the exact distribution of $\hat{\beta}$ and $\hat{\sigma}$. We therefore try to find approximate values of d_1, assuming that the error term has a normal distribution.

Using the stochastic ordering property of the distribution of $\hat{\beta}_i$ from Theorem 1, it can be easily shown that the infimum of the probablity of CS occurs when $\beta_1=\cdots=\beta_k=\beta$. We thus have

$$\begin{aligned}&\inf P(CS|R_1)\\&=\inf P\{\hat{\beta}_{(k)}\geq\max_{1\leq i\leq k-1}(\hat{\beta}_{(i)}-\frac{d_1\hat{\sigma}}{\sqrt{n}}\sqrt{c_{(kn)}{}^{-2}+c_{(in)}{}^{-2}})\ |\beta_1=\cdots=\beta_k=\beta\},\end{aligned}$$

where the infimum is over β and all permutations of $(c_1, \cdots, c_k)$, and $\hat{\beta}_{(i)}$ is the estimator of $\beta_{(i)}$. Using the same argument as in Gupta and Huang (1977), we have for large n,

$$\inf P(CS|R_1)$$

$$=\inf P\{Y_j \leq \frac{d_1\hat{\sigma}}{h(f)},\ j=1,\cdots,\ k-1 \mid \beta_1=\cdots=\beta_k\}$$

where

$$Y_j=\frac{\sqrt{n}(\hat{\beta}_{(j)}-\hat{\beta}_{(k)})}{\sqrt{h^2(f)(c_{(jn)}{}^{-2}+c_{(kn)}{}^{-2})}}$$

and $h^2(f)$ is defined by (2.5). Note that under the assumption of normality, $h^2(f)=\sigma^2\pi/3$. From Theorem 2 the asymptotic distributions of Y_j's are standard normal with correlation matrix $\{\rho_{ij}\}$ defined by $\rho_{ij}=1$ or $\alpha_i\alpha_j$ as $i=j$ or $i\neq j$, where

$$\alpha_i=[1+\frac{c_{(k)}{}^2}{c_{(i)}{}^2}]^{-1/2},\ i=1,\cdots,\ k-1.$$

Therefore from Gupta and Huang (1976) we have for large n,

$$\inf P(CS|R_1)$$

$$\approx\int_0^\infty\int_{-\infty}^\infty\prod_{j=1}^{k-1}\Phi\left(\frac{d_1{}^*\hat{\sigma}^*-\alpha_j u}{\sqrt{1-\alpha_j{}^2}}\right)\,d\Phi(u)dQ_\nu{}^*(\hat{\sigma}^*), \tag{4.4}$$

where $d_1{}^*=\sqrt{3/\pi}d_1$, $\hat{\sigma}^*=\hat{\sigma}/\sigma$, and $Q_\nu{}^*(\hat{\sigma}^*)$ denotes the cdf of $\hat{\sigma}^*$.

We now consider the selection rule based on the LSE, which can be given by

R_2 : Select π_i if and only if

$$\tilde{\beta}_i\geq\max_{1\leq j\leq k}(\tilde{\beta}_j-\frac{d_2 s}{\sqrt{n}}\sqrt{c_{in}{}^{-2}+c_{jn}{}^{-2}}), \tag{4.5}$$

where $d_2=d_2(k,P^*,c_1,\cdots,\ c_n)$ is determined to satisfy the P^*-condition and s^2 is the usual pooled sample unbiased estimator for σ^2. Then by the same argument as above, we can obtain

$$\inf P(CS|R_2)$$

$$=\int_0^\infty\int_{-\infty}^\infty\prod_{j=1}^{k-1}\Phi\left(\frac{d_2 s-\alpha_j u}{\sqrt{1-\alpha_j{}^2}}\right)\,d\Phi(u)dQ_\nu(s), \tag{4.6}$$

where $Q_\nu(s)$ is the cdf of $\chi_\nu/\sqrt{\nu}$ variate with $\nu=k(n-2)$. For special values of α_j we can find tables of d_2. For example when $\alpha_j=1/\sqrt{2}$, $j=1,\cdots,\ k$, Gupta and Sobel (1957) provides tables of $\sqrt{2}d_2$ for $P^*=0.75,\ 0.90,\ 0.95,\ 0.975,\ 0.99$. For $\alpha_1{}^2=\cdots=\alpha_{k-1}{}^2=\rho$, we may use the tables for multivariate t distribution in Krishnaiah and Armitage (1966).

Now we may intuitively consider the use of d_2 in (4.6) for $d_1{}^*$ in (4.4). Because of the behavior of $\hat{\sigma}$ which was discussed in section 3, the use of d_2 for $d_1{}^*$ may make the rule R_1 fail to meet the P^*-condition. But it is expected that the number of non-best regression equations in the selected subset is considerably decreased when there are significant differences in β_i's. The results of a small sample Monte Carlo study in the next section demonstrate this point.

5. An Empirical Study on the Procedures

In this section we present a Monte Carlo study comparing the two selection procedures, the LSE procedure based on the rule R_2 and the WME procedure based on the rule R_1, in their ability to select the best equation and also to eliminate non-best equations. Since the selection rule R_1 is adjusted to be approximately equivalent to the rule R_2 under the assumption of normality, it is expected that the two rules are much alike in their behavior for medium tailed distributions. For short tailed distributions the WME is less efficient than the LSE as seen in Section 3. We therefore expect that the rule R_1 is less efficient than the rule R_2 in their ability to eliminate non-best equations. But the reverse is expected to hold for heavy tailed distributions, even though the rule R_1 fails to satisfy the P^*-condition.

In our Monte Carole study we compared the two rules on the uniform $U(-1.5,\ 1.5)$, standard normal, double exponential, and Cauchy distributions. The random variates are generated as those in Section 3. To investigate the performance of the two rules we considered the case when the slope parameters are equally spaced, *i.e.* $\beta_i=\beta_0+(i-1)\delta\sigma$, $i=1, \cdots, k$, where $\delta\geq 0$ is a given constant and σ is the standard deviation of each distribution. (For Cauchy distribution, $\sigma=2$ is used just for convenience.) The parameters used in our simulation are $k=5$, $n=10$, $\alpha_i=0$ $(i=1, \cdots, k)$, $\beta_0=1$, and $x_{ij}=i$ for $j=1, \cdots, n$; $i=1, \cdots, k$. The values of δ chosen in this simulation study are $\delta c^*=0.0,\ 0.5,\ 2.0$, and 4.0, where $c^*=\Sigma(x_i-\bar{x})^2$.

Table 3 presents the number of times that each equation is selected for the configuration $(\beta_0, \beta_0+\delta\sigma, \cdots, \beta_0+(k-1)\delta\sigma)$ in 500 replications. The entries in Table 3 are based on separate simulations for each value of $\delta c^*=0.0,\ 0.5,\ 2.0$, and 4.0. The average of each column divided by 500 can be interpreted as the sample estimate of the expected proportion of the selected equations, which is denoted by EP in the table. The values of EP are rounded at the third decimal place.

Note that the values of EP for $\delta=0.0$ in Table 3 is the empirical P^*, *i.e.* the empirical probability of CS for LFC. For the short tailed uniform distribution there are no significant differences in the values of empirical P^* between the two rules. For the normal case empirical P^* values of the rule R_1 are slightly lower than those of the rule R_2. The maximum difference is about 0.04 appeared at $P^*=0.75$. In the cases of double exponential and Cauchy distributions, the empirical P^* values of the rule R_1 show that the P^*-condition is significantly violated. The maximum difference is about 0.11 occurred at $P^*=0.75$ in Cauchy.

In terms of the probability of CS for the cases $\delta c^* \geq 0.5$, there are no significant differences between the two rules.

We now investigate the efficiency of the two rules in terms of the size of the selected subset. As a measure of "goodness" of a subset selection procedure, we use the ratio of the expected number populations $E(S)$ to the probability of a correct selection $P(CS)$, *i.e.* $E(S)/P(CS)$. (See, for example, McDonald (1977) for the definition of "goodness" of a procedure.) A rule R is said to be "better" than a rule R^* if the ratio for R is less than the corresponding ratio for R^*. We may thus define the relative efficiency of the procedure R_1 relative to the procedure R_2 as an inverse ratio of the measures of goodness, *i.e.*

$$e(R_1, R_2) = \frac{E(S|R_2)}{E(S|R_1)} \times \frac{P(CS|R_1)}{P(CS|R_2)} \text{ .}$$

Note that this relative efficiency depends on the number of equations k, and the bounds are

$$1/k \leq e(R_1, R_2) \leq k.$$

For example, when R_1 always selects only the best equation and R_2 selects all the equations, the value of $e(R_1, R_2)$ is k. The empirical relative efficiencies of the rule R_1 relative to the rule R_2 are computed from Table 3 and summarized in Table 4.

For the uniform distribution the empirical relative efficiencies of R_1 relative to R_2 are less than 1 except one place, but in most cases greater than 0.9. In normal case the efficiencies of the two rules are almost the same. For the double exponential distributi on the empirical $e(R_1, R_2)$ is uniformly greater than 1, but in most cases less than 1.2. For the Cauchy distribution which has very heavy tails, the efficiency of the rule R_1 is much higher than that of the rule R_2 as expected. The maximum value of the empirical $e(R_1, R_2)$ is 2.652, which is a significant improvement in its efficiency considering that the upper bound of $e(R_1, R_2)$ is 5 in this empirical study.

As a conclusion, the WME procedure based on the rule R_1 performs significantly better than the LSE procedure based on the rule R_2 for heavy tailed distributions.

Table 3 also shows that the WME procedure is quite robust in terms of the underlying distributions.

Table 3 Number of Times Selected for the Equally Spaced Configuration $\beta_i=\beta_0+(i-1)\delta\sigma$ in 500 Replications

	R_1					R_2				
i \ P^*	.75	.90	.95	.975	.99	.75	.90	.95	.975	.99
					Uniform with $\delta c^*=0.0$					
1	355	440	475	488	495	380	450	469	481	487
2	369	443	478	489	495	386	448	470	483	491
3	367	446	469	481	495	383	448	476	482	491
4	356	443	467	488	491	366	447	469	476	486
5	371	439	472	486	498	385	446	476	485	493
EP	.73	.88	.94	.97	.99	.76	.90	.94	.96	.98
					Uniform with $\delta c^*=0.5$					
1	156	265	346	387	440	189	325	382	416	448
2	231	348	410	444	477	262	381	428	453	468
3	316	421	450	469	486	348	434	453	472	485
4	392	464	480	487	497	401	463	480	490	497
5	467	494	500	500	500	465	492	493	498	500
EP	.63	.80	.87	.91	.96	.67	.84	.89	.93	.96
					Uniform with $\delta c^*=2.0$					
1	0	0	0	0	1	0	0	1	2	9
2	0	2	3	10	35	2	6	25	53	115
3	43	90	137	209	270	63	147	211	275	331
4	229	327	375	420	457	271	358	407	434	469
5	500	500	500	500	500	499	500	500	500	500
EP	.31	.37	.41	.46	.51	.33	.40	.46	.51	.60
					Uniform with $\delta c^*=4.0$					
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	2	0	0	0	3	13
4	30	83	125	186	258	58	144	202	249	316
5	500	500	500	500	500	500	500	500	500	500
EP	.21	.23	.25	.27	.30	.22	.26	.28	.30	.33

Table 3 (Continued)

i \ P^*	R_1 .75	.90	.95	.975	.99	R_2 .75	.90	.95	.975	.99
				Normal with $\delta c^*=0.0$						
1	372	455	480	491	493	357	438	471	485	491
2	371	459	481	491	497	358	436	472	483	490
3	373	460	480	492	498	357	435	468	481	491
4	365	439	474	489	496	350	432	465	482	488
5	390	452	476	487	496	362	449	466	481	486
EP	.75	.91	.96	.98	.99	.71	.88	.94	.96	.98
				Normal with $\delta c^*=0.5$						
1	163	284	351	403	450	153	273	334	391	432
2	246	358	412	454	477	240	344	388	440	467
3	318	427	454	479	493	313	403	435	462	489
4	401	451	474	488	497	379	453	468	481	492
5	463	489	497	498	500	464	484	489	494	498
EP	.64	.80	.88	.93	.97	.62	.78	.85	.91	.95
				Normal with $\delta c^*=2.0$						
1	0	0	0	1	1	0	0	0	1	5
2	0	2	7	23	45	0	3	15	30	50
3	25	83	135	192	271	27	92	143	200	271
4	232	334	397	435	460	226	336	376	408	445
5	497	499	500	500	500	496	498	498	499	500
EP	.30	.37	.42	.46	.51	.30	.37	.41	.46	.51
				Normal with $\delta c^*=4.0$						
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	1	2	4
4	38	88	135	185	254	41	83	137	184	250
5	500	500	500	500	500	500	500	500	500	500
EP	.22	.24	.25	.27	.30	.22	.23	.26	.27	.30
				Double Exponential with $\delta c^*=0.0$						
1	363	445	479	488	494	325	403	438	461	480
2	378	442	473	484	495	347	414	445	463	479
3	375	452	471	482	489	345	404	448	465	478
4	381	447	474	489	494	343	414	447	464	475
5	375	448	472	481	493	346	411	446	465	477
EP	.75	.89	.95	.97	.99	.68	.82	.90	.93	.96

Table 3 (Continued)

	R_1					R_2				
i \ P^*	.75	.90	.95	.975	.99	.75	.90	.95	.975	.99
				Double Exponential with $\delta c^*=0.5$						
1	146	271	343	390	433	95	189	244	296	350
2	213	342	398	445	466	160	264	325	372	411
3	306	413	456	475	488	250	343	394	428	451
4	395	456	478	488	495	335	419	450	468	485
5	470	488	496	496	499	453	497	491	494	495
EP	.61	.79	.81	.92	.95	.52	.68	.76	.82	.88
				Double Exponential with $\delta c^*=2.0$						
1	0	0	0	1	6	0	0	0	0	0
2	0	3	10	22	48	0	2	5	10	24
3	32	79	127	176	260	16	40	62	91	136
4	236	360	401	438	464	165	252	314	359	402
5	496	498	500	500	500	496	499	500	500	500
EP	.31	.38	.42	.45	.51	.27	.32	.35	.38	.42
				Double Exponential with $\delta c^*=4.0$						
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	2	0	0	0	0	0
4	19	73	129	194	256	11	37	63	98	138
5	500	500	500	500	500	500	500	500	500	500
EP	.21	.23	.25	.28	.30	.20	.21	.23	.24	.26
				Cauchy with $\delta c^*=0.0$						
1	353	424	455	467	486	306	366	393	421	438
2	355	417	444	462	487	311	364	399	427	452
3	372	426	459	470	488	303	379	416	441	450
4	365	427	459	472	488	302	389	414	430	448
5	362	421	455	466	489	308	369	411	428	446
EP	.72	.85	.91	.93	.97	.61	.75	.81	.86	.89
				Cauchy with $\delta c^*=0.5$						
1	340	409	438	471	488	131	225	267	311	362
2	359	413	441	465	486	203	281	314	358	389
3	374	431	459	473	490	259	343	385	410	439
4	402	443	470	478	492	350	404	427	447	468
5	412	454	472	483	498	401	442	460	468	474
EP	.75	.86	.91	.95	.98	.54	.68	.74	.80	.85

i \ P^*	R_1 .75	R_1 .90	R_1 .95	R_1 .975	R_1 .99	R_2 .75	R_2 .90	R_2 .95	R_2 .975	R_2 .99
				Cauchy with $\delta c^*=2.0$						
1	161	237	295	332	372	4	5	5	13	15
2	210	303	348	390	428	6	14	21	33	52
3	282	359	409	435	471	40	81	114	157	188
4	372	436	461	476	492	192	283	322	366	397
5	449	473	483	490	499	486	492	493	493	496
EP	.59	.72	.80	.85	.90	.29	.35	.38	.42	.46
				Cauchy with $\delta c^*=4.0$						
1	96	139	174	205	239	1	1	1	1	1
2	128	193	233	264	306	3	3	3	3	4
3	207	284	324	353	377	4	6	9	11	21
4	328	392	420	451	466	46	88	126	161	200
5	462	479	487	491	498	492	494	495	496	498
EP	.49	.59	.66	.71	.75	.22	.24	.25	.27	.30

Table 4 Empirical Relative Efficiencies $e(R_1, R_2)$ based on 500 Replications

δc^*	P^* 0.75	0.90	0.95	0.975	0.99
		Uniform			
0.5	0.934	0.947	0.964	0.978	1.001
2.0	0.923	0.909	0.887	0.901	0.886
4.0	0.950	0.905	0.890	0.904	0.914
		Normal			
0.5	1.029	1.016	1.018	1.016	1.012
2.0	1.005	0.986	1.003	1.009	1.005
4.0	0.994	1.009	0.995	0.999	1.000
		Double Exponenential			
0.5	1.140	1.141	1.129	1.110	1.078
2.0	1.129	1.188	1.178	1.184	1.203
4.0	1.016	1.067	1.117	1.161	1.188
		Cauchy			
0.5	1.367	1.235	1.199	1.152	1.096
2.0	2.192	2.149	2.133	2.011	1.959
4.0	2.381	2.590	2.626	2.652	2.605

REFERENCES

(1) Adichie, J.N. (1967). "Estimates of reqression parameters based on rank tests," Ann. Math. Statist. 38, 894-904.

(2) Brown, G.W. and Mood, A.M. (1951). "On median tests for linear hypotheses," Proc. Second Berkeley Symp. Math. Statist. Prob., Univ. of Calif. 159-166,

(3) Gupta, S.S. and Huang, D.Y. (1976). "Subset selection procedures for the means and variances of normal populations: unequal sample sizes case," Sankhyā Ser. B 38, 112-128.

(4) Gupta, S.S. and Huang, D.Y. (1977). "Nonparametric selection procedures for selecting the regression equations with largest slope," J. Chinese Statist. Assoc. 15, 5717-5724.

(5) Gupta, S.S. and Sobel, M.(1957). "On a statistic which arises in selection and ranking problems," Ann. Math. Statist. 28, 957-967.

(6) Hollander, P.W. and Welsch, R.E. (1977). "Robust regression using iteratively reweighted least-squares," Commun. Statist.-Theor. Meth. A6, 813-827.

(7) Jacekel, L.A. (1972). "Estimating regression coefficients by minimizing the dispersion of the residuals," Ann. Math. Statist. 43, 1449-1458.

(8) Kildea, D.G. (1981). "Brown-Mood type median estimators for simple regression models," Ann. Statist. 9, 438-442.

(9) Krishnaiah, P.R. and Armitage, J. V.(1966). "Tables for multivariate t distribution," Sankhyā Ser. B 28, 31-56.

(10) McDonald, G.C. (1977). "Subset selection procedures based on sample ranges," Commun. Statist.-Theor. Meth. A6 1055-1079.

(11) Rizvi, M.H. and Woodworth, G.G. (1970). "On selection procedures based on ranks: counterexamples concerning least favorable configurations," Ann. Math. Statist. 41, 1942-1951.

(12) Scholz, F.W. (1978). "Weighted median regression estimates," Ann. Statist, 6, 603-609.

(13) Sen, P.K. (1968). "Estimates of the regression coefficients based on Kendall's tau,„ J. Amer. Statist. Assoc. 63, 1379-1389.

(14) Sievers, G.L. (1978). "Weighted rank statistic for simple linear regression," J. Amer. Statist. Assoc. 73, 628-631.

(15) Theil, H. (1950). "A rank-invariant method of linear and polynomial regression analysis, I, II, and III," Nederl. Akad. Wetensch. Proc. 53, 386-392, 521-525, and 1397-1412.