The Comparison of Outlier Detection in Multiple Linear Regression

Pimpan Amphanthong

***Abstract*— Four Outlier detection approaches in multiple linear regressions are reviewed, investigated and compared. The Monte Carlo simulation is based on the median absolute deviation (MAD) and the robustness standard deviation (RSD) criterion. The results of three and five regressors show that the MAD is better than the RSD for all situations. The DEFFIT$_i$ distance and the Mahalanobis distance (MD_i) are better than the others for all sample sizes with different percentages outliers in case of the X's outliers. For the Y's outliers, PRESS residual $(r_{(i)})$ and R-student (t_i) approaches are performed better than the others. For the both of X's and Y's outliers, the PRESS residual $(r_{(i)})$ and the Mahalanobis distance (MD_i) are better than the others for all sample sizes with different percentages outlier.**

***Index Terms*—Mean absolute deviation, Mean square error, Robustness standard deviation, Monte Carlo simulation, Multiple linear regression, Residuals.**

I. Introduction

The multiple regression models are widely used to study the relationship between the response variable and independent variables. A general multiple linear regression model is $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$ where $\underline{y}$ is an $n \times 1$ vector of observed values of the dependent variable, $X = \left[x_{i1}, x_{i2}, ..., x_{ip}\right]$, x_{ij} is an $n \times 1$ vector of the values of x_i , or regressors, $i = 1, 2, ..., n.$, $\underline{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\underline{\varepsilon}$ is an $n \times 1$ vector of errors with a traditional assumption of Gauss-Markov theorem is $\underline{\varepsilon} \sim N\left(0, \sigma^2 I\right)$. Various approaches to estimate unknown parameters of the model which have property as the best linear unbiased estimator (BLUE) , for example, the ordinary least squares (OLS) and the maximum likelihood estimator (MLE). However, in practice, $\underline{\varepsilon}$ are not always belonged the assumption, then the OLS and MLE may be arbitrarily bad. Furthermore, if outliers are exists in the model, then alternative approaches are needed. There are many authors have been studied and analyzed the multiple linear regression model when data has outliers (see [1], [2], [3] and [4]). According to the literatures (see[10], [11], [12], [13], [14], [15], [17], [18] and [19]), it's very important to know how to detect the outliers in multiple linear regression model and should be studied more carefully.

Manuscript received April 15, 2012. This work was supported in part by the Rajamankala University of Technology Suwanabhumi.

P. Amphanthong is with the Rajamankala Univeristy of Technology Suwanabhumi, Faculty of Science and Techonology, Department of Mathematics, Suphanburi Campus, 72130 Thailand, e-mail: pim_pimpan@hotmail.com.

This present paper, the author reviews four outlier detection approaches in multiple linear regression model and then compares theirs results by using the criterion which are called the median absolute deviation (MAD) and the robustness standard deviation (RSD).

II. Analytical approaches

In this paper, we review four outlier detection approaches in multiple linear regression model as many literatures used to identify the existence of outliers.

A. PRESS Residuals

The observations X_{ij}, $i = 1, 2, ..., n$ for each $j = 1, 2, ..., p$ are computed the prediction error where the fitted of the i^{th} are computed based on $n-1$ observations and deleted the i^{th} observed values. The PRESS residuals may be computed from the hat matrix and the residual as

$$r_{(i)} = e_i / (1 - h_{ii}) , \quad (1)$$

$i = 1, 2, ..., n$ where h_{ii} is the i^{th} diagonal element of $H = X(X'X)^{-1}X'$. If $r_{(i)} > 3$ then the i^{th} observation is identified as outliers (see [4]).

B. R-Student

A formal testing procedure for outliers detection based on R-student is given by

$$t_i = e_i / \sqrt{\hat{\sigma}^2_{(i)}(1 - h_{ii})} , \quad (2)$$

$i = 1, 2, ..., n$ where $|t_i| > t_{(\alpha/2n), n-(p-1)}$ indicates the existence outliers. (see [4]).

C. DEFFITi Distance

The DEFFIT$_i$ is

$$DEFFIT_i = (h_{ii}^{1/2} e_i) / (\sigma_i (1 - h_{ii})) \quad , \quad (3)$$

$i = 1, 2, ..., n$. For each observation i compute or $(h_{ii} e_i)/(1 - h_{ii})$ which tells how much the predicted value $\hat{y}_i$, at the design point x_i would be affected if the i^{th} case were deleted. Belsley, Kuh and Welsch [5] suggested that any observation for which $|DEFFIT_i| > 2\sqrt{p/n}$ warrants attention for outliers.

D. Mahalanobis Distance

The measure of the leverage by means for MD_i (Mahalanobis distance) is

$$MD_i^2 = (\mu_i - \bar{\mu})\sigma^{2^{-1}}(\mu_i - \bar{\mu})' = (n-1)[h_{ii} - 1/n] , \quad (4)$$

$i = 1, 2, ..., n$ where $\bar{\mu} = 1/n(\sum_{i=1}^{n} \mu_i)$ and $\sigma^2 = 1/(n-1)*$

ISBN: 978-988-19251-3-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$\sum_{i=1}^{n}(\mu_i-\bar{\mu})'(\mu_i-\bar{\mu})$. If $MD_i^2 > \chi^2_{p-1,0.95}$ where $\chi^2_{p-1,0.95}$ is the 95th percentile of a chi-square distribution with $p-1$ degrees of freedom then there is an outlier (see [6]).

III. Main results

A. Criterions

There are many statistical values computed from the sample data that can be used to identify the existence of outliers. Most require different statistical criterion of the standard deviation (S.D.) of the residuals, $e_1, e_2, \ldots, e_n$, but the measure based on the mean squared error (MSE) is not robust, since it may be highly influenced by events of small probability. This paper, author uses the median absolute deviation which is denoted MAD (see [7], [16]) and the robustness standard deviation which is denoted RSD (see [8]), are defined as

$$MAD(e_i) = \frac{Med\left|e_i - Med(e_i)\right|}{0.6745}, \quad (5)$$

and

$$RSD(e_i) = 2.1Med\left\{\left|e_i\right|\right\}, \quad (6)$$

P. Amphanthong and P. Suwatee [1] studied the existences of outlier's detection in statistics and then comparison procedures in the multiple linear regression . They showed that Mahalanobis distance identifiers the presence of outliers more often than the others for small, medium and large sample sizes with different percentages outliers in the X-outliers and in both the X-Y outliers. The next best statistics for the detection are R-student and DEFFIT distance. As for the Y- outliers, R-student and PRESS residual perform better than the other approach.

B. Numerical Results

One thousand of data sets are generated from the model $y_i = \beta_0 + \beta_1 x_{i1} + \ldots + e_i, \quad i = 1,2,\ldots,n$ where all regression coefficients are fixed $\beta_j = 1$, for each $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,p$ and the errors are assumed to be independent.

The explanatory variables $x_{ij} \in R^{n\times p}$ are sampled independently from a $N(0,1)$. The sample data sets are generated under (p=3, p=5) regressors and the sample sizes are small sizes (n=10), medium sizes (n=20 and n=30), and large sizes (n=50 and n=100), with different percentage of outliers (10%, 20% and 30%).

The variation of four outlier detection approaches provide an indication of the sensitivity of them, then comparison of theirs' results by counting the number of times that each approaches can be identify outliers. The computations give the best of outlier detection approaches for different sample sizes and percentages of outlier with 1,000 replications. The results of four outlier detection approaches are as following;

Table 1: Comparisons of statistics' value of outlier detection by percentage of X's Outliers with three regressors.

SAMPLE SIZES	% OF OUTLIERS	$r_{(i)}$ (MAD)	$r_{(i)}$ (RSD)	$DEFFIT_i$ (MAD)	$DEFFIT_i$ (RSD)
10	10	0.970	0.968	1.000	1.000
	20	0.988	0.985	1.000	1.000
	30	0.901	0.880	1.000	1.000
20	10	0.976	0.972	1.000	1.000
	20	0.940	0.934	1.000	1.000
	30	0.991	0.982	1.000	0.995
30	10	0.917	0.919	1.000	1.000
	20	0.993	0.989	1.000	1.000
	30	1.000	1.000	0.998	0.988
50	10	0.989	0.987	1.000	1.000
	20	1.000	1.000	1.000	0.996
	30	1.000	1.000	0.999	0.971
100	10	1.000	1.000	1.000	1.000
	20	1.000	1.000	1.000	0.999
	30	1.000	1.000	1.000	0.967

Table 2: Comparisons of statistics' value of outlier detection by percentage of X's Outliers with three regressors (cont.).

SAMPLE SIZES	% OF OUTLIERS	MD_i (MAD)	MD_i (RSD)	t_i (MAD)	t_i (RSD)
10	10	1.000	0.969	0.460	0.688
	20	0.995	0.887	0.631	0.789
	30	0.975	0.312	0.513	0.795
20	10	1.000	1.000	0.641	0.717
	20	1.000	1.000	0.781	0.951
	30	1.000	0.994	0.968	0.998
30	10	1.000	1.000	0.655	0.621
	20	1.000	1.000	0.979	0.991
	30	1.000	0.998	0.999	1.000
50	10	1.000	1.000	0.960	0.915
	20	1.000	1.000	1.000	0.996
	30	1.000	1.000	1.000	1.000
100	10	1.000	1.000	1.000	0.981
	20	1.000	1.000	1.000	1.000
	30	1.000	1.000	1.000	1.000

It can be seen from table 1 to 2, the best criterion is median absolute deviation (MAD), and the best of X's outlier detection approaches are $DEFFIT_i$ and MD_i. Their performances are highest values of outlier detection (1.000) for all sample sizes and percentage of outliers. Furthermore, the performance of $r_{(i)}$ and t_i are high for large sample sizes and all percentage of outliers [Fig. 1(a)].

ISBN: 978-988-19251-3-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table 3: Comparisons of statistics' value of outlier detection by percentage of Y's Outliers with three regressors.

SAMPLE SIZES	% OF OUTLIERS	$r_{(i)}$ (MAD)	$r_{(i)}$ (RSD)	$DEFFIT_i$ (MAD)	$DEFFIT_i$ (RSD)
10	10	0.996	0.994	0.013	0.010
	20	1.000	1.000	0.002	0.002
	30	1.000	1.000	0.000	0.000
20	10	1.000	0.999	0.000	0.000
	20	1.000	1.000	0.000	0.000
	30	1.000	1.000	0.000	0.000
30	10	1.000	1.000	0.000	0.000
	20	1.000	1.000	0.000	0.000
	30	1.000	1.000	0.000	0.000
50	10	1.000	1.000	0.000	0.000
	20	1.000	1.000	0.000	0.000
	30	1.000	1.000	0.000	0.000
100	10	1.000	1.000	0.000	0.000
	20	1.000	1.000	0.000	0.000
	30	1.000	1.000	0.000	0.000

Table 4: Comparisons of statistics' value of outlier detection by percentage of Y's Outliers with three regressors (cont.).

SAMPLE SIZES	% OF OUTLIERS	MD_i (MAD)	MD_i (RSD)	t_i (MAD)	t_i (RSD)
10	10	0.012	0.000	0.448	0.010
	20	0.021	0.000	0.104	0.422
	30	0.031	0.000	0.015	0.750
20	10	0.072	0.014	0.962	0.137
	20	0.134	0.028	0.634	0.457
	30	0.186	0.039	0.241	0.742
30	10	0.107	0.041	1.000	0.166
	20	0.221	0.075	0.943	0.451
	30	0.319	0.104	0.671	0.740
50	10	0.214	0.077	1.000	0.160
	20	0.381	0.138	1.000	0.468
	30	0.518	0.208	0.994	0.721
100	10	0.379	0.151	1.000	0.173
	20	0.624	0.292	1.000	0.481
	30	0.787	0.413	1.000	0.747

It can be seen from table 3 to 4, the best criterion is median absolute deviation (MAD), and the best of Y's outliers detection approach is $r_{(i)}$, for all percentage of outliers and sample sizes. Furthermore, the performance of t_i is better than the $DEFFIT_i$ and MD_i in medium (n=30) and large sample sizes for all percentage of outliers [Fig. 1(b)].

Table 5: Comparisons of statistics' value of outlier detection by percentage of both X's and Y's Outliers with three regressors.

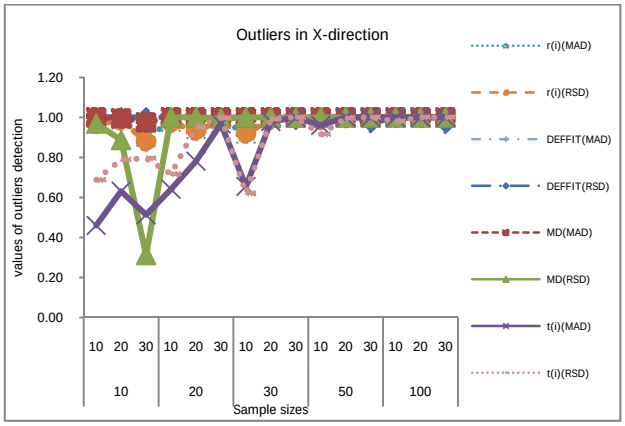
SAMPLE SIZES	% OF OUTLIERS	$r_{(i)}$ (MAD)	$r_{(i)}$ (RSD)	$DEFFIT_i$ (MAD)	$DEFFIT_i$ (RSD)
10	10	0.994	0.993	0.998	0.998
	20	1.000	1.000	0.968	0.965
	30	0.998	0.998	0.652	0.694
20	10	1.000	1.000	0.964	0.963
	20	1.000	1.000	0.372	0.534
	30	1.000	1.000	0.014	0.121
30	10	1.000	1.000	0.872	0.920
	20	1.000	1.000	0.059	0.270
	30	1.000	1.000	0.000	0.063
50	10	1.000	1.000	0.439	0.705
	20	1.000	1.000	0.003	0.158
	30	1.000	1.000	0.000	0.040
100	10	1.000	1.000	0.044	0.510
	20	1.000	1.000	0.000	0.144
	30	1.000	1.000	0.000	0.016

Table 6: Comparisons of statistics' value of outlier detection by percentage of both X's and Y's Outliers with three regressors (cont.).

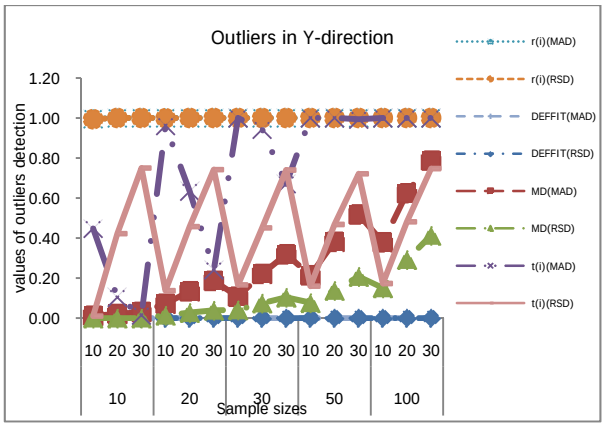
SAMPLE SIZES	% OF OUTLIERS	MD_i (MAD)	MD_i (RSD)	t_i (MAD)	t_i (RSD)
10	10	1.000	0.969	0.717	0.356
	20	0.995	0.887	0.832	0.353
	30	0.975	0.312	0.474	0.052
20	10	1.000	1.000	0.978	0.231
	20	1.000	1.000	0.999	0.189
	30	1.000	0.994	0.903	0.418
30	10	1.000	1.000	0.995	0.031
	20	1.000	1.000	1.000	0.264
	30	1.000	0.998	0.991	0.472
50	10	1.000	1.000	1.000	0.089
	20	1.000	1.000	1.000	0.284
	30	1.000	1.000	1.000	0.484
100	10	1.000	1.000	1.000	0.120
	20	1.000	1.000	1.000	0.296
	30	1.000	1.000	1.000	0.506

It can be seen from table 5 to 6, the best criterion is median absolute deviation (MAD) and the best of both X's and Y's outlier detection approaches are $r_{(i)}$ and MD_i. The performances of $r_{(i)}$ and MD_i approaches are highest values of the detection outliers (1.000) in all sample sizes and percentage of outliers. Furthermore, the performance of t_i is better than the $DEFFIT_i$ for large sample sizes and all percentage of outliers [Fig. 1(c)].

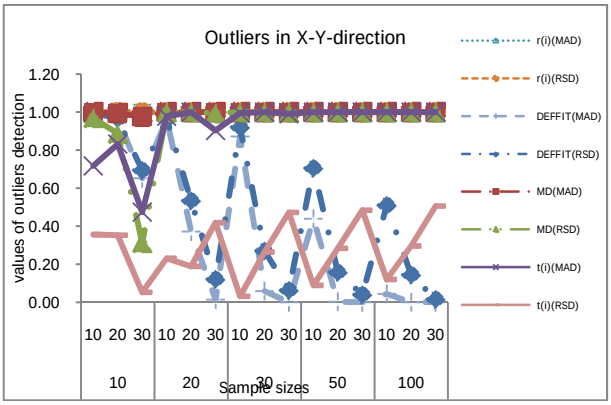
ISBN: 978-988-19251-3-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)



(a)



(b)



(c)

Figure.1 A Comparison of Statistics' value of outlier detection by Sample Sizes with Three Regressors.(a) X's Outliers; (b) Y's Outliers; (c) Both X's and Y's Outliers.

Furthermore, we compare the results of five regressors. The computations give the best of outlier detection approaches for different sample sizes and percentages of outlier with 1,000 replications, the results are as following;

Table 7: Comparisons of statistics' value of outlier detection by percentage of X's Outliers with five regressors.

SAMPLE SIZES	% OF OUTLIERS	$r_{(i)}$ (MAD)	$r_{(i)}$ (RSD)	$DEFFIT_i$ (MAD)	$DEFFIT_i$ (RSD)
10	10	0.996	0.996	1.000	1.000
	20	1.000	0.999	1.000	1.000
	30	1.000	1.000	1.000	1.000
20	10	1.000	1.000	1.000	1.000
	20	0.996	0.996	1.000	1.000
	30	0.982	0.977	1.000	1.000
30	10	1.000	1.000	1.000	1.000
	20	0.988	0.987	1.000	1.000
	30	1.000	0.999	1.000	1.000
50	10	0.999	0.997	1.000	1.000
	20	1.000	1.000	1.000	1.000
	30	1.000	1.000	1.000	0.998
100	10	1.000	1.000	1.000	1.000
	20	1.000	1.000	1.000	1.000
	30	1.000	1.000	1.000	0.992

Table 8: Comparisons of statistics' value of outlier detection by percentage of X's Outliers with five regressors (cont.).

SAMPLE SIZES	% OF OUTLIERS	MD_i (MAD)	MD_i (RSD)	t_i (MAD)	t_i (RSD)
10	10	1.000	1.000	0.584	0.606
	20	1.000	1.000	0.814	0.733
	30	1.000	0.998	0.907	0.763
20	10	1.000	1.000	0.756	0.747
	20	1.000	1.000	0.860	0.793
	30	1.000	1.000	0.867	0.972
30	10	1.000	1.000	0.862	0.775
	20	1.000	1.000	0.896	0.939
	30	1.000	1.000	0.995	0.999
50	10	1.000	1.000	0.863	0.663
	20	1.000	1.000	0.999	0.994
	30	1.000	1.000	1.000	1.000
100	10	1.000	1.000	1.000	0.973
	20	1.000	1.000	1.000	1.000
	30	1.000	1.000	1.000	1.000

It can be seen from table 7 to 8, the best criterion is median absolute deviation (MAD). The performance of MD_i and $DEFFIT_i$ are highest values of detection outlier (1.000) for all sample sizes and percentage of outliers.

ISBN: 978-988-19251-3-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table 9: Comparisons of statistics' value of outlier detection by percentage of Y's outliers with five regressors.

SAMPLE SIZES	% OF OUTLIERS	$r_{(i)}$ (MAD)	$r_{(i)}$ (RSD)	$DEFFIT_i$ (MAD)	$DEFFIT_i$ (RSD)
10	10	0.998	0.996	0.041	0.034
	20	1.000	1.000	0.007	0.007
	30	1.000	1.000	0.002	0.003
20	10	1.000	0.999	0.001	0.000
	20	1.000	1.000	0.000	0.000
	30	1.000	1.000	0.000	0.000
30	10	1.000	1.000	0.000	0.000
	20	1.000	1.000	0.000	0.000
	30	1.000	1.000	0.000	0.000
50	10	1.000	1.000	0.000	0.000
	20	1.000	1.000	0.000	0.000
	30	1.000	1.000	0.000	0.000
100	10	1.000	1.000	0.000	0.000
	20	1.000	1.000	0.000	0.000
	30	1.000	1.000	0.000	0.000

Table 10: Comparisons of statistics' value of outlier detection by percentage of Y's Outliers with five regressors (cont.).

SAMPLE SIZES	% OF OUTLIERS	MD_i (MAD)	MD_i (RSD)	t_i (MAD)	t_i (RSD)
10	10	0.091	0.000	0.323	0.013
	20	0.180	0.000	0.063	0.500
	30	0.262	0.000	0.011	0.808
20	10	0.316	0.113	0.906	0.169
	20	0.524	0.204	0.478	0.554
	30	0.694	0.286	0.148	0.835
30	10	0.429	0.190	0.999	0.199
	20	0.710	0.382	0.884	0.552
	30	0.869	0.532	0.518	0.848
50	10	0.675	0.347	1.000	0.210
	20	0.889	0.580	1.000	0.570
	30	0.961	0.721	0.982	0.821
100	10	0.880	0.581	1.000	0.197
	20	0.990	0.852	1.000	0.551
	30	0.999	0.936	1.000	0.820

From table 9 to 10, the best criterion is median absolute deviation (MAD). The best of Y's outliers detection is $r_{(i)}$, its' performance are good for all sample sizes and percentages of outliers. Furthermore, the performance of t_i is better than the $DEFFIT_i$ and MD_i for large sample sizes and all percentage of outliers**.**

Table 11: Comparisons of statistics' value of outlier detection by percentage of both X's and Y's Outliers with five regressors.

SAMPLE SIZES	% OF OUTLIERS	$r_{(i)}$ (MAD)	$r_{(i)}$ (RSD)	$DEFFIT_i$ (MAD)	$DEFFIT_i$ (RSD)
10	10	0.998	0.998	1.000	1.000
	20	1.000	1.000	1.000	1.000
	30	1.000	1.000	0.999	0.999
20	10	1.000	1.000	1.000	1.000
	20	1.000	1.000	0.991	0.988
	30	1.000	1.000	0.559	0.698
30	10	1.000	1.000	0.999	0.999
	20	1.000	1.000	0.730	0.879
	30	1.000	1.000	0.034	0.229
50	10	1.000	1.000	0.995	0.996
	20	1.000	1.000	0.026	0.359
	30	1.000	1.000	0.000	0.060
100	10	1.000	1.000	0.300	0.813
	20	1.000	1.000	0.000	0.193
	30	1.000	1.000	0.000	0.024

Table 12. Comparisons of statistics' value of outlier detection by percentage of both X's and Y's Outliers with five regressors (cont.).

SAMPLE SIZES	% OF OUTLIERS	MD_i (MAD)	MD_i (RSD)	t_i (MAD)	t_i (RSD)
10	10	1.000	1.000	0.729	0.394
	20	1.000	1.000	0.917	0.529
	30	1.000	0.998	0.960	0.518
20	10	1.000	1.000	0.964	0.396
	20	1.000	1.000	0.993	0.364
	30	1.000	1.000	0.998	0.264
30	10	1.000	1.000	0.998	0.318
	20	1.000	1.000	1.000	0.178
	30	1.000	1.000	1.000	0.414
50	10	1.000	1.000	0.999	0.055
	20	1.000	1.000	1.000	0.236
	30	1.000	1.000	1.000	0.441
100	10	1.000	1.000	1.000	0.112
	20	1.000	1.000	1.000	0.314
	30	1.000	1.000	1.000	0.507

From table 11 to 12, the best criterion is median absolute deviation (MAD). The best outlier detection approaches in both X's and Y's outliers are $r_{(i)}$ and MD_i. The performance of $r_{(i)}$ and MD_i are highest values of detection outlier (1.000) for all sample sizes and percentage of outliers. Furthermore, the performance of t_i is better than the $DEFFIT_i$ for large sample sizes and all percentage of outliers, the performance of $DEFFIT_i$ is better than t_i for small and medium sizes (n=30) and all percentage of outliers.

ISBN: 978-988-19251-3-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

IV. Conclusions

The Monte Carlo simulation shows the performance of four outlier detection approaches in multiple linear regression. We use the MAD and RSD as the criterions, which MAD is better than RSD for all situations. We get the same agreement for three and five regressors, the $DEFFIT_i$ distance and the Mahalanobis distance (MD_i) are better than the others for all sample sizes with different percentages outliers in case of the X's outliers. The PRESS residual ($r_{(i)}$) and R-student (t_i) approaches are performed better than the others in the case of Y's outliers. The PRESS residual ($r_{(i)}$) and the Mahalanobis distance (MD_i) are better than the others for all sample sizes with different percentages outlier in the case of both of the X's and Y's outliers. Furthermore, we have seen that the performance of t_i is better for large sample sizes and all percentage of outliers in all cases of the outliers.

References

[1] P. Ampanthong and P. Suwattee, "A Comparative Study of Outlier Detection Procedures in Multiple Linear Regression (Periodical style—Submitted for publication)," IMECS 2009 submitted for publication.

[2] A. C. Atkinson, "In: Plots, Tronasformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis," Oxford : Clarendon Press, 1985.

[3] A. S. Hadi and J. S. Simonoff, "Procedures for the Identification of Multiple Outliers in Linear Models," J. Ammer. Statist. Assoc. Vol. 88, 1993, pp. 1264-1272.

[4] P. J. Huber, "Robust Statistic," New York : John Wiley & Sons, 1981.

[5] D. A. Belsley and R. E. S. Welsch, "Regression Diagnostics : Identifying Influential Data and Source of Collinearity," New York: John Wiley & Sons, 1980.

[6] P. J. Rousseeuw and A. M. Leroy, "Robust Regression and Outlier Detection," New York : John Wiley & Sons, 1987.

[7] A. S. Kosinsk, "A procedure for the detection of multivariate outliers," Computational Statistics and Data Analysis. Vol. 29, 1998, pp. 2145-2161.

[8] J. O. Ramsay, "A Comparative Study of Several Robust Estimates of Slope, Intercept, and Scale in Linear Regression," Journal of the American Statistical Association. Vol. 72, 1977, pp. 608-615.

[9] Y. Jiazhong, "A Monte Carlo Comparison of Several High Breakdown and Efficient Estimator," Computational Statistics & Data analysis. Vol. 30, 1999, pp. 205-219.

[10] H. P. Lopuhaa and P. J. Rousseeuw, "Breakdown Point of Affine Equivariant Estimators of Maultivariate Location and Covariance Matrice," Technical Report, Faculty of Mathematics and Informatics, Netherlands: Delft University of Technology, 1987.

[11] D. C. Montgomery, E. A. Peck and G.G. Vining, "Introduction to Linear Regression Analysis," 3rd ed. New York: John Wiley & Sons, 2003.

[12] V. Barnett and T. Lewis, "Outliers in Statistical Data," 3rd ed. UK : Wiley, Chicester, 1994.

[13] D. Birkes and Y. Dodge, "Alternative Methods of Regression," New York : John Wiley & Sons, 1993.

[14] A. S. Hadi, "A Modification of a Method for the Detection of Outliers in Multivariate Samples," J. Roy. Statist. Soc. Ser B . Vol. 56, 1994, pp. 393-396.

[15] T. P. Ryan, "Modern Regression Methods," New York: John Wiley & Sons, 1997.

[16] P. J. Rousseeuw, "Least Median of Squares Regression," J. Ammer. Statist. Assoc. Vol. 79, 1984, pp. 871-880.

[17] P. J. Rousseeuw and K. V. Driessen, "A Fast Algorithm for the Minimum Covariance Determinant Estimator," Technometrics, Vol. 41, 1999, pp. 212-223.

[18] J. W. Wisnowski, D. C. Montgomery and J. R. Simpson, "A Comparative Analysis of Multiple Outlier Detection Procedures in the Linear Regression Model," Computational Statistics and Data Analysis. Vol. 6, 2001, pp. 351-382.

[19] J. You, "A Monte Carlo comparison of several high breakdown and efficient estimators," Computational Statistics and Data Analysis. Vol. 30, 1999, pp. 205-219.

ISBN: 978-988-19251-3-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)