

COMPARISON OF ROBUST ESTIMATION ON MULTIPLE REGRESSION MODEL

Padrul Jana¹, Dedi Rosadi^{2*}, Epha Diana Supandi³

^{1,2} Department of Mathematics, FMIPA, Universitas Gadjah Mada
Bulaksumur, Sleman, Special Region of Yogyakarta, 55281, Indonesia.

¹ Department of Mathematics Education, FKIP, Universitas PGRI Yogyakarta
IKIP PGRI I Street Sonosewu No.117, Bantul, Special Region of Yogyakarta, 55182, Indonesia.

³ Department of Mathematics, FST, Universitas Islam Negeri Sunan Kalijaga Yogyakarta
Laksda Adisucipto Street, Sleman, Special Region of Yogyakarta, 55281, Indonesia.

Corresponding author's e-mail: *dedirosadi@ugm.ac.id

ABSTRACT

Article History:

Received: 19th January 2023

Revised: 26th April 2023

Accepted: 29th April 2023

Keywords:

Robust Regression;
M-Estimation;
MM-Estimation;
S-Estimation.

This study aimed to compare the robustness of the OLS method with a robust regression model on data that had outliers. The methods used on the robust regression model were M-estimation, MM-estimation, and S-estimation. The step taken was to check the characteristics of the data against outliers. Furthermore, the data were modeled with and without outliers using the OLS method and the M-, MM-, and S-estimations. The results were very different between the data with and without the outlier models in the OLS method. It was reflected in the intercept and standard error variables generated from the models. Meanwhile, the regression model with the M-, MM-, and S-estimations was quite stable and able to withstand the presence of outliers. Based on the three estimations that were robust against the outliers, the MM-estimation was the best candidate because, in addition to having a stable intercept parameter estimation, it also had the smallest standard error, which was 61.9 in the resulting model.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

How to cite this article:

P. Jana, D. Rosadi and E. D. Supandi., "COMPARISON OF ROBUST ESTIMATION ON MULTIPLE REGRESSION MODEL," *BAREKENG: J. Math. & App.*, vol. 17, iss. 2, pp. 0979-0988, June, 2023.

Copyright © 2023 Author(s)

Journal homepage: <https://ojs3.unpatti.ac.id/index.php/barekeng/>

Journal e-mail: barekeng.math@yahoo.com; barekeng_journal@mail.unpatti.ac.id

Research Article • Open Access

1. INTRODUCTION

Commonly, there are two variables in regression analysis, i.e., the dependent variable (Y) and the independent variable (X). The relationship between the dependent and the independent variables in regression analysis is represented through a mathematical model called the regression equation. The regression equation that states a linear relationship between one dependent variable (Y) and one independent variable (X) is called simple linear regression. Meanwhile, the regression equation that models a linear relationship between one dependent variable and more than one independent variable is known as multiple linear regression [1]. Regression analysis aims to estimate the regression coefficient in the regression model [2]. Some approaches to the method used are adjusted to the characteristics of the data that will be analyzed.

The characteristics or properties of the data are diverse; one of them is data with outliers [3], [4]. The problem that frequently arises in regression analysis is that one or more data points that are found are far from the general data pattern or are commonly called outliers. They usually arise due to errors in the measurement system, errors in the data input, measurement instrument errors, or an unusual event [5], [6]. However, it can provide information that cannot be delivered by other data points in some cases [7]–[11]. Practically, more special handling or methods are needed for data with outliers [12], [13].

Ordinary Least Square (OLS) is a parameter estimation method in regression modeling. This method is quite good to be used if the data do not have outliers [14]–[18]. However, if the assumptions are not met due to outliers, this method is not strong enough to model the regression [19]. Thus, we need another regression model in the estimation process. One of the robust models in the presence of outlier data is the robust regression model [20], [21]. According to Chen (2002), robust regression is a significant method for analyzing data contaminated by outliers [22]. Several estimation methods in the robust regression model are the M, S, and MM regression models. This method is selected because it has a high breakdown point (maximum number of outlier data that can be tolerated by a model) [23].

The main contribution of this paper is to explain the robustness of the robust regression method in the presence of outliers. This is important because the nature of the outliers in the data is random and exists in various conditions. This paper begins with checking outliers on the data used in the simulation [24]. Then, the regression model was formed and compared between the data with outliers and the data of which the outliers had been removed using the OLS method and the robust regression model. Furthermore, the robust method of the data with outliers was observed. The model criteria were seen from the stability of the estimated parameters and the standard error of the data with and without outliers.

2. RESEARCH METHODS

The research method contains explanations in the form of paragraphs about the research design or descriptions of the experimental settings, data sources, data collection techniques, and data analysis conducted by the researcher. This guide will explain writing headings. If your headers exceed one, use the second level of headings as below.

2.1 Multiple Linear Regression Model

Multiple linear regression is a statistical method that is useful in testing the effect of an independent variable on the dependent variable. The general form of the multiple linear regression equation is as follows [25], [26]:

$$Y_i = a + b_1X_{1i} + b_2X_{2i} + \dots + b_pX_{np} + \varepsilon_i, \varepsilon \sim N(0, \sigma^2) \quad (1)$$

Where:

Y = Dependent variable

a = Constant of a regression equation

b_i = Regression coefficient $i = 1, 2, 3, \dots, p$

$X_{1i} \dots, X_{np}$ = Independent variable $i = 1, 2, 3, \dots, n$

ε_i = Residual for the i^{th} observation

2.2 OLS Method

OLS is a regression method that minimizes the number of squared errors. The parameter estimation method used is the OLS method, which estimates the regression coefficient (β) by minimizing error. The parameter estimation is as follows:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (2)$$

Where $\hat{\beta}$ is a vector of the estimated parameters of size $(p + 1) \times 1$, X is a matrix of predictor variables of size $n \times (p + 1)$, and y is an observation vector of response variables of size $n \times 1$.

2.3 M-Estimation

M-estimation is an estimation of the robust regression that minimizes the ρ function, as follows [27]:

$$\hat{\beta}_M = \min_{\beta} \sum_{i=1}^n \rho(e_i) \quad (3)$$

$$\hat{\beta}_M = \min_{\beta} \sum_{i=1}^n \rho(y_i - \sum_{j=0}^k x_{ij} \beta_j) \quad (4)$$

Where:

$\hat{\beta}_M$ is the estimated beta of the M-method estimation results

ρ is a weighted representation of the residual

e_i is the i^{th} residual

Function ρ used is the Huber objective function by the following equation:

$$\rho(\varepsilon) = \begin{cases} \varepsilon^2; & -k \leq \varepsilon \leq k \\ 2k|\varepsilon| - k^2; & \varepsilon < -k \text{ atau } \varepsilon > k \end{cases}$$

Where $k = 1,5 \hat{\sigma}$. To estimate σ , use $\hat{\sigma} = 1,483 \text{ MAD}$, with MAD (Median Absolute Deviation) as the median of residue $|e_i|$ [14], [28], [29].

The steps of the robust regression model with M-estimation according to [14] are as follows:

1. Determine the residual of e_i .
2. Determine the Mean Absolute Deviation value, the value of $\hat{\sigma}$, and the value of k .
3. Compare the value of k with the initial residual value and the final residual value.
4. Determine the new estimated value by adding up the initial estimated value with the new residual.
5. Perform regression analysis again with the newly estimated value as the dependent variable.
6. Repeat until the n^{th} iteration and convergence is reached.

2.4 MM-Estimation

The first step is to find an estimator with the S-estimation, then to set the regression parameters using the M-estimation [30].

The MM-estimation is defined as:

$$\hat{\beta}_{MM} = \arg \min_{\beta} \sum_{i=1}^n \rho\left(\frac{e_i}{\hat{\sigma}_s}\right) = \arg \min_{\beta} \sum_{i=1}^n \rho\left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{\hat{\sigma}_s}\right) \quad (5)$$

Where $\hat{\beta}_{MM}$ is the estimated beta of the MM-method estimation results.

The steps of the robust regression model with MM-estimation according to [30], [31] are as follows:

1. Determine the initial estimator value of $\hat{\beta}_j$ and residue e_i
2. The residue obtained in the first step is used to determine the value of $\hat{\sigma}_s$ and calculate the initial weight of w_i
3. Residue in the first step and the value of $\hat{\sigma}_s$ in the second step is used in the initial iteration to estimate the Weighted Least Square to calculate the regression coefficient.
4. Calculate the new weight with the estimated scale from the initial iteration of Weighted Least Square.
5. Repeat these steps until the n^{th} iteration and $\sum_{i=1}^n |e_i^m|$ reaches convergence.

2.5 S-Estimation

The robust regression model with S-estimation can reach a breakdown point of up to 50%, meaning that half of the outliers can be overcome and have a good influence on other observations [14]. The S-estimation is defined as $\hat{\beta}_S = \min_{\beta} \hat{\sigma}_s(e_1, e_2, \dots, e_n)$, by determining the minimum scale of the $\hat{\sigma}_s$ robust estimation.

$$\min \sum \rho \left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{\hat{\sigma}_s} \right) \quad (6)$$

$$\text{Where } \hat{\sigma}_s = \sqrt{\frac{n \sum_{i=1}^n (e_i)^2 - (\sum_{i=1}^n e_i)^2}{n(n-1)}} \quad (7)$$

The weighting objective function minimizes the combined sum of the squared residuals and the absolute number of residues.

The steps of the robust regression model with S-estimation according to [14] are as follows:

1. Determine the residual of e_i
2. Determine $\hat{\sigma}_s$
3. Determine $u_i = \frac{e_i}{\hat{\sigma}_s}$
4. Determine the weighting function
5. Determine $\hat{\beta}_S$ by using the Weighted Least Square with weighting w_i
6. Repeat these steps until the n^{th} iteration and $\hat{\beta}_S$ reaches convergence.

The data obtained are used for simulation by previously checking the outliers using the plotting data, boxplot, quartile values, and Mahalanobis distance and range to detect outliers. The criteria for univariate data are as follows:

$$Q_1 - \frac{3}{2} IQR < X < Q_3 + \frac{3}{2} IQR \quad (8)$$

2.6 Mahalanobis Distance

Measurement of the squared distance applied to detect outliers in multivariate data can use the following formula [32]:

$$d_{MD}^2(i) = (x_i - \bar{x})^T \Sigma^{-1} (x_i - \bar{x}) \quad (9)$$

Where:

$d_{MD}^2(i)$: Mahalanobis squared distance object at the i^{th} observation

x_i : Vector-object data at the i^{th} observation of sizes $p \times 1$

Σ : Covariance Matrix of size $(p \times p)$, where p is variable numbers

The i^{th} data are expressed as an outlier when $d_{MD}^2(i) > \chi_{p,1-\alpha}^2$.

Where $\chi_{p,1-\alpha}^2$ is the outlier limit with a probability of $1 - \alpha$ [33].

Data outside the interval are called outlier data. The next stage is to remove outliers in the data. After having and removing data with outliers, the regression parameters are estimated using the OLS and the estimations of M, MM, and S.

3. RESULTS AND DISCUSSION

3.1 Result

The data used in the simulation were student motivation data (X1), learning facilities (X2), and student learning outcomes in the cognitive domain (Y). Motivation and facilities were measured using a questionnaire with a score range of 0-30, while learning outcomes were measured by a test with a score range of 0-100. There were 20 students (respondents) in this study with detailed scores listed in Table 1.

Table 1. Table Name Motivation, Facilities, and Student Learning Outcomes Data

Resp	X1	X2	Y	Resp	X1	X2	Y
1	25	23	80	11	25	27	98
2	20	30	81	12	18	18	76
3	12	15	70	13	26	28	90
4	12	29	80	14	20	20	98
5	20	10	60	15	14	15	12
6	25	30	90	16	20	26	97
7	15	20	10	17	12	14	66

Resp	X1	X2	Y	Resp	X1	X2	Y
8	30	30	100	18	19	14	12
9	19	23	83	19	24	23	89
10	15	15	56	20	25	30	85

To see the entire data more easily, **Figure 1** presents the plotting data results.

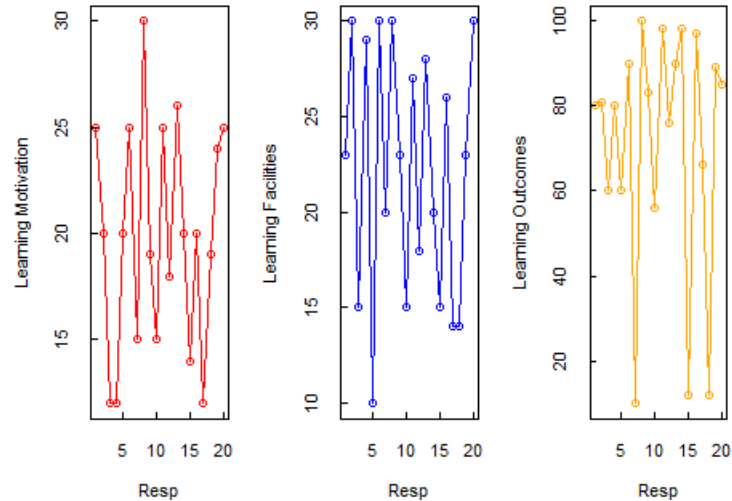


Figure 1. Plotting Data

Figure 1 shows that the data of the student learning outcome variables indicated outliers. Some outliers were far from the distribution of other data groups. To obtain the visualization of the data to be analyzed clearly, **Figure 2** presents boxplot data.

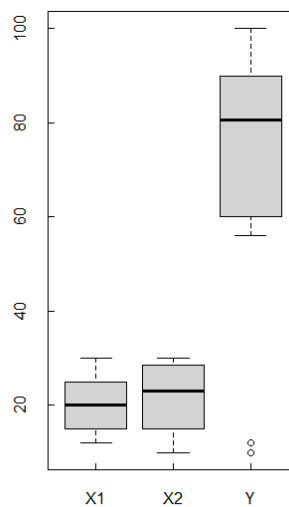


Figure 2. Boxplot Data

The results of the boxplot showed that there were outliers, notably in the student learning outcomes (Y) data. Then, the data with outliers were checked and validated using $Q_1 - \frac{3}{2}IQR < \text{data} < Q_3 + \frac{3}{2}IQR$. The results are shown in **Table 2**.

Table 2. Summary of Learning Outcomes Variables

Parameter	Value
Minimum	10.00
Q1	60.00
Q2	80.50
Mean	71.15
Q3	90.00
Maximum	100.00

Parameter	Value
$Q_1 - \frac{3}{2}IQR$	15.00
$Q_3 + \frac{3}{2}IQR$	135.00

Table 2 confirms that there was data below 15.00. This means that all values of the learning outcomes variable in the category below 15.00 were detected as outliers. Meanwhile, if we take a look closely, data with outliers were found in the 7th, 14th, and 19th respondents. The causes of outliers could vary, such as recording errors, measuring instrument errors, and so on. Furthermore, the data were analyzed using the OLS method, M-estimation, MM-estimation, and S-estimation to see the accuracy and robustness of the outlier method.

3.2 Detection of outliers with Mahalanobis Distance

In the Mahalanobis distance calculation, the mean and variance-covariance matrix were calculated firstly, thus obtaining:

$$\text{Mean} = \bar{x} = \begin{bmatrix} 19.80 \\ 22.00 \\ 71.65 \end{bmatrix} \text{ while the variance-covariance matrix is as follows:}$$

$$\Sigma = \begin{bmatrix} 28.16842 & 20.84211 & 82.66316 \\ 20.84211 & 43.57895 & 118.31579 \\ 82.66316 & 118.31579 & 826.02895 \end{bmatrix} \text{ and}$$

$$\Sigma^{-1} = \begin{bmatrix} 0.059335398 & -0.020055946 & -0.003065167 \\ -0.020055946 & 0.044327846 & -0.004342216 \\ -0.003065167 & -0.004342216 & 0.002139306 \end{bmatrix}$$

Thus, the Mahalanobis distance of the data was obtained as follows:

Table 3. Mahalanobis Distance

i	$d_{MD}(i)$	$d_{MD}^2(i)$	i	$d_{MD}(i)$	$d_{MD}^2(i)$
1	1.2506389	1.564098	11	1.1709325	1.371083
2	2.3011408	5.295249	12	0.8522773	0.726377
3	3.4185425	11.68643	13	1.4512414	2.106102
4	8.0129607	64.20754	14	2.1064609	4.437178
5	5.5724039	31.05169	15	4.4043829	19.39859
6	1.6332782	2.667598	16	1.1426137	1.305566
7	6.4053564	41.02859	17	3.3495576	11.21954
8	3.7141777	13.79512	18	5.7934046	33.56354
9	0.3470779	0.120463	19	0.9691214	0.939196
10	1.3034666	1.699025	20	1.8009641	3.243472

Then, it was compared to the value of $\chi_{p,1-\alpha}^2$ with $p = 20$ and $\alpha = 5\%$. It obtained a value of $\chi_{20,0.95}^2 = 10.8508$. The i^{th} data were declared as an outlier if $d_{MD}^2(i) > \chi_{p,1-\alpha}^2$. Therefore, the 3rd, 4th, 5th, 7th, 8th, 17th, and 18th observation data were detected as outliers.

3.3 Ordinary Least Square (OLS) Method

The OLS method is the most commonly used in the regression. The multiple regression analysis with the OLS method was presented using two data in this discussion, namely data with and without outliers. The analysis results are presented in **Table 4** below.

Table 4. Comparison of the OLS Method with And Without Outliers

	With Outlier			Without Outlier		
	Intercept	X1	X2	Intercept	X1	X2
Estimate	-4.9277	1.5615	2.0527	34.8248	0.9859	1.1609
Std. Error	21.0053	1.2017	0.9662	9.1318	0.4874	0.4012
Pr (> t)	0.8173	0.2112	0.0486	0.0019	0.0627	0.0118
Adjusted R-squared	0.4003			0.6245		
P-value	0.005034			0.000413		

	With Outlier			Without Outlier		
	Intercept	X1	X2	Intercept	X1	X2
Residual Normality Test	p-value = 0.02866 (Unfulfilled)			p-value = 0.2768 (Fulfilled)		
Autocorrelation	p-value = 0.9072 (Fulfilled)			p-value = 0.161 (Fulfilled)		
Residual Homogeneity	p-value = 0.09092 (Fulfilled)			p-value = 0.897 (Fulfilled)		
Multicollinearity	Vif < 10 (Fulfilled)			Vif < 10 (Fulfilled)		
Residual Standard Error	22.35			8.654		
Mean Square Error	424.5			178.7		

Based on **Table 4** using the OLS method, it can be seen that there were differences in the analysis results between the data with and without outliers. In the Intercept section, it can be seen that the data with outliers had a value of -4.9277 which was very different from the data without outliers of which the value was 34.8248. It gave a contrast-interpretation. The data with outliers had a quite high standard error, namely 21.0053, while the data without outliers was 9.1318. If we look at the mean squared error (MSE), there was a significant difference between the data with and without outliers, indicating that the Ordinary Least Square (OLS) method was not a robust method for data with outliers.

3.4 Robust Regression Model with M-Estimation

We will present the application of a robust regression model using a bisquare weighting with M-estimation in this section. This aimed to test whether the M-estimation can be robust to the outlier data. The results of the analysis using M-estimation are presented in **Table 5** below.

Table 5. Robust Regression With M-Estimation

	With Outlier			Without Outlier		
	Intercept	X1	X2	Intercept	X1	X2
Estimate	33.5227	0.9520	1.2133	33.1399	0.9437	1.2267
Std. Error	8.9422	0.5116	0.4113	9.3720	0.5003	0.4118
t-value	3.7488	1.8609	2.9498	3.5361	1.8864	2.9789
Mean Square Error	546.5			62,87		

The robust regression model with M-estimation showed no significant difference between the data with and without outliers. The M-estimation carried out reached convergence in the 7th iteration. The data intercept parameter with an outlier of 33.5227 could survive even by removing the outlier, which was 33.1399. On the standard error and others, they had almost adjacent values. However, if we looked at the MSE robust regression method with M-estimation on the data with and without outliers, the difference was quite far. The data with outliers had an MSE of 546.5, while those without outliers had an MSE of 62.87.

3.5 Robust Regression Model with MM-Estimation

Furthermore, a robust regression method with MM-estimation was applied. The iterations required to reach convergence were seven iterations on the data with outliers in the MM-estimation. The results obtained are presented in **Table 6** below.

Table 6. Robust Regression With MM-Estimation

	With Outlier			Without Outlier		
	Intercept	X1	X2	Intercept	X1	X2
Estimate	34.2759	0.9704	1.1841	33.9391	0.9619	1.1975
Std. Error	6.2774	0.2281	0.2855	6.1486	0.2229	0.2771
Pr (> t)	4.23e-05	0.000535	0.000674	7.55e-05	0.000714	0.000703
Adjusted R-squared	0.6238			0.6396		
Residual Standard Error	11.13			62,87		
Mean Square Error	546.5			61,9		

The robust regression model with MM-estimation showed no significant difference between the data with and without outliers. The MM-estimation carried out reached convergence in the 8th iteration on the data without outliers. The intercept parameter data with an outlier of 34.2759 was quite able to survive even by

removing the outliers, which was 33.9391. In terms of the standard errors and others, both the data with and without outliers had quite competitive values. However, the difference was quite far when using the MSE robust regression method with MM-estimation on the data with and without outliers. The data with outliers had an MSE value of 546.5, while those without outliers had an MSE value of 61.9.

3.6 Robust Regression Model with S-Estimation

The convergence was achieved in Iteratively Reweighted Least Squares (IRWLS) in the application of the S-method. The results obtained are presented in **Table 7** below.

Table 7. Robust Regression With S-Estimation

	With Outlier			Without Outlier		
	Intercept	X1	X2	Intercept	X1	X2
Estimate	33.0059	0.9251	1.2083	34.0982	0.8858	1.1869
Std. Error	11.7776	0.6008	0.5105	17.2207	0.8586	0.7351
Pr (> t)	0.0122	0.1420	0.0301	0.0677	0.3197	0.1287
Adjusted R-squared		0.8069			0.8296	
Mean Square Error		531,7			66,61	

3.7 Discussion

The results of the comparison of the regression model with the OLS and the M, MM, S estimations are described in this section. Then, a robust model to the presence of outliers was determined. The following **Table 8** presents a summary of each method.

Table 8. Comparison of the OLS-Method, M-Estimation, MM-Estimation, And S-Estimation

		Methods			
		OLS	M	MM	S
Data with outlier	Intercept	-4.9277	33.5227	34.2759	33.0059
	Std. Error	21.0053	8.9422	6.2774	11.7776
	Adjusted R-squared	0.4003	-	0.6238	0.8069
	Mean Square Error	424.5	546.5	546.5	531,7
Data without outlier	Intercept	34.8248	33.1399	33.9391	34.0982
	Std. Error	9.1318	9.3720	6.1486	17.2207
	Adjusted R-squared	0.6245	-	0.6396	0.8296
	Mean Square Error	178.7	62,87	61,9	66,61

In the OLS method, there was a significant difference between the data with and without outliers. The difference between the intercept and standard error models was quite significant, indicating that multiple linear regression with the OLS method was weak against the outlier data. This is in line with research from [10], [34]–[36]. The robust regression model with M-estimation was quite reliable in overcoming the outlier data. It can be seen from the intercept of both the data with and without outliers of 33.5227 and 33.1399, which were not significantly different. In terms of the standard error, the data with outliers had an error of 8.9422 and those without outliers of 9.3720. The resulting standard error did not differ significantly when compared to the OLS method. This result is certainly in line with [37]. The regression model with MM-estimation was also quite reliable in overcoming the outliers. Even the MM method resulted in the smallest standard error compared to the other models tested. Thus, it can be said that the MM-model was the best estimation candidate in this case study. The reliability of the MM method in overcoming outliers has also been confirmed by research conducted by [38], [39]. Meanwhile, the robust regression model on the S-estimation was also quite reliable in overcoming outliers, in which it generated a higher standard error than the M-estimation and MM-estimation for both the data with and without outliers. However, this result is slightly different from the research conducted by [23], [40] in their case study, showing that the S-estimation was better than the MM-estimation. In the data without outliers, the MM-estimation method was still the best method with an MSE value of 61.9, which was smaller than the other models. Therefore, the regression equation with the best estimation was $\hat{Y} = 33.0059 + 0.9251X_1 + 1.2083X_2$ for the data with outliers and $\hat{Y} = 33.9391 + 0.9619X_1 + 1.1975X_2$ for the data without outliers.

4. CONCLUSIONS

In modeling the regression, it is necessary to first know the characteristics of the data to be used. By understanding the data characteristics, it is easier to determine the most suitable method. For data without outliers, the OLS method is quite reliable in modeling regression. However, for data with outliers, the OLS method is not yet capable of modeling it. Therefore, a robust regression model with M, MM, and S estimations is needed. These estimations are quite reliable in modeling data with outliers. In this study, the regression model with MM-estimation was better than the M and S estimations. MM-estimation is better in modeling robust regression on data with and without outliers. The open issues in future research are related to the addition of fuzzy to the robust regression model in the M, MM, and S estimations or others.

REFERENCES

- [1] W. Wilsen, W. Rahayu, and V. M. Santi, "Penerapan Imputasi Ganda dengan Metode Predictive Mean Matching (PMM) untuk Pendugaan Data Hilang pada Model Regresi Linear," *J. Stat. dan Apl.*, vol. 2, no. 1, pp. 12–20, 2018, doi: 10.21009/jsa.02102.
- [2] M. B. Rahman and E. Widodo, "Perbandingan Metode Regresi Robust Estimasi Least Trimmed Square, Estimasi Scale, dan Estimasi Method Of Moment," in *Prosiding Seminar Nasional Matematika*, 2018, vol. 1, pp. 426–433. [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/19689>
- [3] A. T. Tholaby MS and A. Djunaidy, "Untuk Memperbaiki Kinerja Fuzzy Time Series Pada Peramalan Harga Saham Bursa Efek," *J. Ilm. Mikrotek*, vol. 2, no. 4, 2017.
- [4] G. S. Na, D. Kim, and H. Yu, "DILOF: Effective and memory efficient local outlier detection in data streams," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 1993–2002, 2018, doi: 10.1145/3219819.3220022.
- [5] S. Indra, D. Vionanda, and R. Sriningsih, "Pendeteksian Data Pencilan dan Pengamatan Berpengaruh pada Beberapa Kasus Data Menggunakan Metode Diagnostik," *UNP J. Math.*, vol. 1, no. 2, pp. 67–74, 2013, [Online]. Available: <http://ejournal.unp.ac.id/students/index.php/mat/article/view/1265/942>
- [6] C. Chen and L.-M. Liu, "Joint Estimation of Model Parameters and Outlier Effects in Time Series," *J. Am. Stat. Assoc.*, vol. 88, no. 421, pp. 284–297, 1993, doi: 10.1080/01621459.1993.10594321.
- [7] S. Candraningtyas, D. Safitri, and D. Ispriyanti, "Regresi Robust MM-Estimator Untuk Penanganan Pencilan Pada Regresi Linier Berganda," *J. Gaussian*, vol. 2, no. 4, pp. 395–404, 2013, doi: <https://doi.org/10.14710/j.gauss.v2i4.3806>.
- [8] Z. A. Bakar, R. Mohamad, A. Ahmad, and M. M. Deris, "A comparative study for outlier detection techniques in data mining," in *2006 IEEE Conference on Cybernetics and Intelligent Systems*, 2006, pp. 1–6. doi: 10.1109/ICCIS.2006.252287.
- [9] J. A. S. Almeida, L. M. S. Barbosa, A. A. C. C. Pais, and S. J. Formosinho, "Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering," *Chemom. Intell. Lab. Syst.*, vol. 87, no. 2, pp. 208–217, 2007, doi: 10.1016/j.chemolab.2007.01.005.
- [10] A. Shodiqin, A. N. Aini, and M. R. Rubowo, "Perbanding Dua Metode Regresi Robust yakni Metode Least Trimmed Squares (LTS) dengan metode Estimator-MM (Estmasi-MM) (Studi Kasus Data Ujian Tulis Masuk Terhadap Hasil IPK Mahasiswa UPGRIS)," *J. Ilm. Teknosains*, vol. 4, no. 1, p. 35, 2018, doi: 10.26877/jitek.v4i1.2403.
- [11] A. S. Hadi and J. S. Simonoff, "Procedures for the identification of multiple outliers in linear models," *J. Am. Stat. Assoc.*, vol. 88, no. 424, pp. 1264–1272, 1993, doi: 10.1080/01621459.1993.10476407.
- [12] G. Barbato, E. M. Barini, G. Genta, and R. Levi, "Features and performance of some outlier detection methods," *J. Appl. Stat.*, vol. 38, no. 10, pp. 2133–2149, 2011, doi: 10.1080/02664763.2010.545119.
- [13] L. C. Chang, D. K. Jones, and C. Pierpaoli, "RESTORE: Robust estimation of tensors by outlier rejection," *Magn. Reson. Med.*, vol. 53, no. 5, pp. 1088–1095, 2005, doi: 10.1002/mrm.20426.
- [14] N. A. Atamia, Y. Susanti, and S. S. Handajani, "Analisis Perbandingan Regresi Robust Estimasi-M Huber dan Estimasi-S dalam Mengatasi Outlier," in *PRISMA, Prosiding Seminar Nasional Matematika*, 2015, vol. 4, pp. 673–679.
- [15] S. Sunaryo and T. H. Siagian, "Mengatasi Masalah Multikolinearitas Dan Outlier Dengan Pendekatan Robpca," *J. Mat. Saint dan Teknol.*, vol. 12, no. 1, pp. 1–10, 2011.
- [16] G. A. M. Srinadi, "Pengaruh Outlier Terhadap Estimator Parameter Regresi dan Metode Regresi Robust," *Pros. Konf. Nas. Mat. XVII - 2014*, no. June 2014, pp. 1259–1266, 2014.
- [17] A. R. Fadhilah, A. Fitrianto, and I. M. Sumertajaya, "OUTLIER IDENTIFICATION ON PENALIZED SPLINE REGRESSION MODELING FOR POVERTY GAP INDEX IN JAVA," vol. 16, no. 4, pp. 1231–1240, 2022.
- [18] Y. Susanti, H. Pratiwi, H. Sulistijowati, and T. Liana, "M Estimation, s estimation, and mm estimation in robust regression," *Int. J. Pure Appl. Math.*, vol. 91, no. 3, pp. 349–360, 2014, doi: 10.12732/ijpam.v91i3.7.
- [19] D. Cahyawati, H. Tabuji, and R. Abdiati, "Efektivitas Metode Regresi Robust Penduga Welsch dalam Mengatasi Pencilan pada Pemodelan Regresi Linear Berganda," *J. Penelit. Sains*, vol. 12, no. 1, pp. 1–7, 2009, doi: <https://doi.org/10.26554/jps.v12i1.182>.
- [20] D. Blatná, "Application of Robust Regression and Bootstrap in Productivity Analysis of GERD Variable in EU27," *Stat. Stat. Econ. J.*, vol. Vol. 94, no. 2, pp. 62–76, 2014, [Online]. Available: <https://pdfs.semanticscholar.org/d648/592c98bc63164ac6f6989e1af5c525a06817.pdf>
- [21] P. J. Rousseeuw and M. Hubert, "Robust statistics for outlier detection," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 1, no. 1, pp. 73–79, 2011, doi: 10.1002/widm.2.
- [22] C. Chen, "Robust Regression and Outlier Detection with the ROBUSTREG Procedure," in *SAS Institute Inc.*, 2002, pp. 265–27.
- [23] A. Semar, F. Virgantari, and H. Wijayanti, "Perbandingan Estimasi S (Scale) Dan Estimasi Mm (Method of Moment) Pada Model Regresi Robust Dengan Data Pencilan," *Statmat J. Stat. Dan Mat.*, vol. 2, no. 1, p. 21, 2020, doi:

- 10.32493/sm.v2i1.4207.
- [24] J. Ha, S. Seok, and J. S. Lee, "A precise ranking method for outlier detection," *Inf. Sci. (Ny)*, vol. 324, pp. 88–107, 2015, doi: 10.1016/j.ins.2015.06.030.
- [25] A. A.-F. N. Wahyudin, A. Primajaya, and A. S. Y. Irawan, "Penerapan Algoritma Regresi Linear Berganda Pada Estimasi Penjualan Mobil Astra Isuzu," *Techno.Com*, vol. 19, no. 4, pp. 364–374, 2020, doi: 10.33633/tc.v19i4.3834.
- [26] J. Grosz, "Identification of Influential Points in a Linear Regression Model," *Stat. Stat. Econ. J.*, vol. 48, no. 1, pp. 71–77, 2011.
- [27] D. C. Montgomery, P. E. A., and G. G. Vining, *Introduction To Linear Regression Analysis, 5nd Edition*. New York: John Wiley & Sons, Inc., 2006.
- [28] L. J. Sinay and M. W. Talakua, "Pemodelan Harga Saham Indeks Lq45 Menggunakan Regresi Linier Robust M-Estimator: Huber Dan Bisquare," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 9, no. 1, pp. 51–61, 2014, doi: 10.30598/barekengvol9iss1pp51-61.
- [29] D. Q. F. de Menezes, D. M. Prata, A. R. Secchi, and J. C. Pinto, "A review on robust M-estimators for regression analysis," *Comput. Chem. Eng.*, vol. 147, p. 107254, 2021, doi: 10.1016/j.compchemeng.2021.107254.
- [30] A. Z. S. W. R. Hendra Perdana, "Analisis Regresi Robust Estimasi-Mm Dalam Mengatasi Pencilan Pada Regresi Linear Berganda," *Bimaster Bul. Ilm. Mat. Stat. dan Ter.*, vol. 9, no. 1, pp. 123–128, 2020, doi: 10.26418/bbimst.v9i1.38666.
- [31] V. J. Yohai, *High Breakdown Point and High Efficiency Robust Estimates for Regression*. Annals of Statistics, 1987.
- [32] C. Lays, O. Klein, Y. Dominicy, and C. Ley, "Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance," *J. Exp. Soc. Psychol.*, vol. 74, no. September 2017, pp. 150–156, 2018, doi: 10.1016/j.jesp.2017.09.011.
- [33] M. A. Nahdliyah, T. Widiharah, and A. Prahutama, "Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index," *J. Gaussian*, vol. 8, no. 2, pp. 161–170, 2019.
- [34] I. F. Maharani, N. Satyahadewi, and D. Kusnandar, "Metode Ordinary Least Squares Dan Least Trimmed Squares Dalam Mengestimasi Parameter Regresi Ketika Terdapat Outlier," *Bul. Ilm. Mat. Stat. dan Ter.*, vol. 03, no. 3, pp. 163–168, 2014, doi: <http://dx.doi.org/10.26418/bbimst.v3i03.7350>.
- [35] F. Daniel, "Mengatasi Pencilan Pada Pemodelan Regresi Linear Berganda Dengan Metode Regresi Robust Penaksir Lms," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 13, no. 3, pp. 145–156, 2019, doi: 10.30598/barekengvol13iss3pp145-156ar884.
- [36] A. Fitrianto and S. H. Xin, "Comparisons Between Robust Regression Approaches in the Presence of Outliers and High Leverage Points," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 1, pp. 243–252, 2022, doi: 10.30598/barekengvol16iss1pp241-250.
- [37] A. D. Deria, A. Hoyyi, and M. Mustafid, "Regresi Robust Estimasi-M Dengan Pembobot Andrew, Pembobot Ramsay Dan Pembobot Welsch Menggunakan Software R," *J. Gaussian*, vol. 8, no. 3, pp. 377–388, 2019, doi: 10.14710/j.gauss.v8i3.26682.
- [38] N. Nurdin, Raupong, and A. Islamiyati, "Penggunaan Regresi Robust Pada Data Yang Mengandung Pencilan Dengan Metode Momen," *Mat. Stat. dan Komputasi*, vol. 10, no. 2, p. 115, 2014, [Online]. Available: <http://journal.unhas.ac.id/index.php/jmsk/article/download/3418/1955>
- [39] R. Finger, "Investigating the performance of different estimation techniques for crop yield data analysis in crop insurance applications," *Agric. Econ. (United Kingdom)*, vol. 44, no. 2, pp. 217–230, 2013, doi: 10.1111/agec.12005.
- [40] J. W. Wisnowski, D. C. Montgomery, and J. R. Simpson, "A comparative analysis of multiple outlier detection procedures in the linear regression model," *Comput. Stat. Data Anal.*, vol. 36, no. 3, pp. 351–382, 2001, doi: 10.1016/S0167-9473(00)00042-6.