

A Comparative Study of M-Estimation and S-Estimation in Robust Regression Models: A Case Study on Crime Rates in Indonesia

Rahmat Hidayat¹, Ruliana^{2*}, Zakiyah Mar'ah³, Jumriani⁴

^{1,2,3,4} *Statistics Study Program, Universitas Negeri Makassar, Makassar, Indonesia*

*Corresponding author: ruliana.t@unm.ac.id

*Submission date: 31 July 2025, Revision: 12 August 2025, Accepted: 02 January 2026

ABSTRACT

Robust regression is a regression method used to handle outliers in statistical analysis. Robust regression consists of five estimation methods, namely M-estimation (Maximum Likelihood type), LMS estimation (Least Median of Squares), LTS estimation (Least Trimmed Squares), MM estimation (Method of Moments), and S-estimation (Scale). M-estimation is known for having the smallest variance among the estimators, with high efficiency reaching up to 95%, while S-estimation is based on the residual scale of M-estimation and is characterized by a high breakdown point of up to 50%. The 2022 crime rate data in Indonesia contains outliers. According to data from the Central Statistics Agency, there was a drastic increase in crime incidents in 2022, reaching 372,965 reported cases. Therefore, a comparison of robust regression estimation methods was conducted to obtain the best model that explains the factors influencing crime rates in Indonesia. This study employs robust regression using M-estimation and S-estimation with Tukey bisquare weighting. The dependent variable in this study is the number of crimes in Indonesia in 2022, while the independent variables include population density (x_1), open unemployment rate (x_2), number of poor people (x_3), mean years of schooling/MYS (x_4), and labor force participation rate/TPAK (x_5). The results indicate that S-estimation in robust regression provides the best performance among the methods analyzed.

KEYWORDS

Robust, M-estimation, S-estimation, Crimes, Outlier.

1. INTRODUCTION

Regression analysis is a statistical method used to determine the relationship between one or more independent variables (predictor variables) and a dependent variable (response variable). The objective is to model this relationship and measure the extent to which the independent variables can explain the variation in the dependent variable. In conducting regression analysis, the relationship between dependent and independent variables is commonly estimated using the Least Squares Method, also known as Ordinary Least Squares (OLS). This method is considered the best estimator because it is unbiased and has the minimum variance (efficient) among all linear unbiased estimators [1].

However, in practice, the OLS estimator can become biased and no longer efficient if its underlying assumptions are violated. One of the key assumptions is normality, which is not met when the data contain outliers data points that lie far from the general pattern of the dataset. According to Barnett (1981), as cited in [2, 3], outliers are observations that do not follow the general pattern of the data and are located far from the center of the data distribution.

Handling outliers in data can be done in several ways. First, by removing or deleting the i -th observation suspected to be an outlier. However, this approach is quite critical, as the data point may contain important information; therefore,

thorough evaluation is necessary before deciding to eliminate it. Second, by applying a data transformation. Third—and more recommended—is using robust regression. Robust regression is a regression method used to address outliers in statistical analysis. This method limits the influence of outliers by assigning smaller weights to data points identified as outliers [4, 5].

According to [6], robust regression consists of five estimation methods:

1. M-estimation (based on Maximum Likelihood),
2. LMS estimation (Least Median of Squares),
3. LTS estimation (Least Trimmed Squares),
4. MM-estimation (based on the Method of Moments), and
5. S-estimation (based on scale).

Each method has its own strengths and weaknesses. A good estimator is one that has a high breakdown point. Efficiency and breakdown point are used to describe the robustness of an estimation technique. The higher the efficiency and breakdown point of an estimator, the more resistant it is to outliers [7, 8].

M-estimation has the smallest variance among the estimators, with high efficiency of up to 95%, and is based on the Maximum Likelihood estimator [9, 10]. Meanwhile, S-estimation is based on the residual scale of M-estimation, and it is characterized by a high breakdown point of up to 50% [11, 12].

One study comparing robust regression methods was conducted by [13, 14], which found that S-estimation provided accurate estimates of the factors influencing the incidence of Dengue Hemorrhagic Fever (DHF), based on the smallest Mean Square Error (MSE) and the highest R^2 value [15].

2. LITERATURE REVIEW

2.1 Linear Regression Model

The general form of a linear regression model with k independent variables is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (1)$$

To obtain estimates from a regression model, several methods can be used. The Least Squares Method (LSM), also known as Ordinary Least Squares (OLS), is one of the most widely used methods for estimating regression parameters. OLS estimates the regression model by minimizing the Mean Square Error (MSE). In other words, the basic principle of the OLS method is to minimize the sum of the squared residuals [16]. The OLS method, which ensures that the sum of squared residuals is as small as possible, is explained as follows:

$$\sum e_i^2 = (Y_i - \hat{Y}_i)^2 \quad (2)$$

Furthermore, to assess the goodness-of-fit of the model, the coefficient of determination is used, with the following equation:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (3)$$

A regression model needs to be tested to determine whether it satisfies the underlying assumptions of linear regression. The key assumptions that must be met in linear regression are normality, homoscedasticity, no multicollinearity, and no autocorrelation.

2.2 Robust Regression

Robust regression was developed by Rousseeuw and Leroy in 1987. It is a regression method used when the residuals are not normally distributed or when the dataset contains outliers that influence the model [17, 18]. Robust regression analyzes the entire dataset by assigning smaller weights to observations that exert a large influence. Outliers and influential observations can significantly affect the estimated values of the intercept, slope, R^2 , and S^2 .

Datasets containing outliers can be identified using several graphical methods, such as scatterplots, boxplots, leverage values, and DfFITS (Difference in Fitted Values) [19]. Using graphical methods, outliers can be detected when the residual plots (ZPRED vs. SRESID) display data points that deviate significantly from the overall data pattern. Furthermore, DfFITS values can be used to determine the extent to which the i -th observation influences the regression model in terms of the fitted values.

$$DfFITS = t_i \sqrt{\frac{h_{ii}}{1 - h_{ii}}} \quad (4)$$

$$h_{ii} = x_i'(X'X)^{-1}x_i \quad (5)$$

2.3 M-Estimation

M-estimation was introduced by Huber in 1964. M-estimation is an extension of the maximum likelihood estimation method and a robust regression technique that combines a simple approach in both computation and theory. The main principle of this method is to minimize the ρ -function (objective function) of the residuals. This objective function is used to define the weighting function in robust regression. One commonly used weighting function is the Tukey Bisquare function.

To obtain the M-estimation equation, the following equation must be solved:

$$y_i = \sum_{j=0}^k x_{ij}\beta_j + \varepsilon_i \quad (6)$$

$$\varepsilon_i = y_i - \sum_{j=0}^k x_{ij}\beta_j \quad (7)$$

$$\sum_{i=0}^k \rho(\varepsilon_i) = \sum \rho\left(y_i - \sum_{j=0}^k x_{ij}\beta_j\right) \quad (8)$$

$$\sum_{i=0}^k \rho\left(\frac{\varepsilon_i}{\sigma}\right) = \sum \rho\left(\frac{y_i - \sum_{j=0}^k x_{ij}\beta_j}{\sigma}\right) \quad (9)$$

The solution to equation(9) is obtained by taking the derivative with respect to β , resulting in:

$$\begin{aligned} \hat{\beta}_M &= \sum_{i=1}^n \rho'\left(\frac{y_i - \sum_{j=0}^k x_{ij}\beta}{\hat{\sigma}}\right) \\ &= 0 \end{aligned} \quad (10)$$

The matrix notation is obtained by solving the following equation:

$$y = x\beta_M + \varepsilon \quad (11)$$

$$\varepsilon = y - x\beta \quad (12)$$

By assigning a weight w to the error term ε , the following is obtained

$$\begin{aligned} \varepsilon'w\varepsilon &= (y - x\beta_M)'w(y - x\beta_M) \\ &= (y' - \beta_M'x')w(y - x\beta_M) \\ &= y'wy - y'wx\beta_M - \beta_M'x'wy + \beta_M'x'wx\beta_M \\ &= y'wy - 2\beta_M'x'wy + \beta_M'x'wx\beta_M \\ \frac{\partial \varepsilon'w\varepsilon}{\partial \beta} &= -2x'wy + 2(x'wx)\beta_M \end{aligned} \quad (13)$$

The partial derivative of equation (13), set equal to zero, becomes:

$$-2x'wy + 2x'wx\beta_M = 0 \quad (14)$$

$$x'wx\beta_M = x'wy \quad (15)$$

$$(x'wx)^{-1} (x'wx) \beta_M = (xwx)^{-1} x'wy \quad (16)$$

$$\hat{\beta}_M = (x'wx)^{-1} x'wy \quad (17)$$

The equation (13) is known as the Weighted Least Squares (WLS) equation, where w_i is a matrix with diagonal elements containing the weight values [11]

2.4 S-Estimation

This method is called S-estimation because it is based on a scale, specifically the standard deviation scale of the residuals. Similar to OLS, which minimizes the error variance, S-estimation minimizes the dispersion of the errors. S-estimation employs several weighting functions, including Welsch and Tukey Bisquare, which are used to generate weighted scale values through an iterative process until the resulting estimates converge. The Tukey Bisquare function is defined as follows:

$$\rho(u_i) = \begin{cases} \frac{u_i^2}{2} - \frac{u_i^4}{2c^2} + \frac{u_i^6}{6c^4}, & |u_i| \leq c \\ \frac{c^2}{6}, & |u_i| > c \end{cases} \quad (18)$$

Tukey Bisquare is influenced by a function,

$$\psi(u) = \begin{cases} u \left(1 - \left(\frac{u}{c}\right)^2\right)^2, & |u_i| \leq c \\ 0, & |u_i| > c \end{cases} \quad (19)$$

The weighting function is formulated as follows:

$$w(u) = \begin{cases} u \left(1 - \left(\frac{u}{c}\right)^2\right)^2, & |u_i| \leq c \\ 0, & |u_i| > c \end{cases} \quad (20)$$

With a value of $c = 1.547$. The value of c is a tuning constant that has been set to determine the robustness of a weight [20].

S-estimation is defined as: $\hat{\beta}_s = \min_{\beta} \hat{\sigma}_s(e_1, e_2, \dots, e_n)$, by determining the estimator value ($\hat{\sigma}_s$) through minimization

$$\min \sum_{i=1}^n \rho \left(\frac{y_i \sum_{j=0}^k x_{ij} \beta}{\hat{\sigma}_s} \right) \quad (21)$$

by

$$\hat{\sigma}_s = \sqrt{\frac{1}{nK} \sum_{i=1}^n w_i e_i^2} \quad (22)$$

Value of K is 0.199, $w_i = w_{\sigma}(u_i) = \frac{\rho(u_i)}{u_i^2}$, and the initial estimate is selected

$$\hat{\sigma}_s = \frac{\text{median } |e_i - \text{median}(e_i)|}{0.6745} \quad (23)$$

where w_i is the weighting function of the Iteratively Reweighted Least Squares (IRLS).

$$w_i(u_i) = \frac{\psi(u_i)}{u_i} \begin{cases} u \left(1 - \left(\frac{u}{c}\right)^2\right)^2 \\ 0 \end{cases} \quad (24)$$

$$w_i(u_i) = \begin{cases} \left(1 - \left(\frac{u_i}{c}\right)^2\right)^2, & |u_i| \leq c \\ 0, & |u_i| > c \end{cases}$$

with $u_i = \frac{e_i}{\hat{\sigma}_s}$ and $c = 1,547$ [8]. Equation (24) can be solved using IRLS until convergence is achieved.

3. METHODOLOGY

This research employs a literature study and a quantitative approach, which involves reviewing relevant articles, journals, and books as reference materials, as well as analyzing numerical (quantitative) data using statistical methods, specifically robust regression. The case study analysis uses secondary data obtained from the publication of Statistics Indonesia (BPS), namely the 2022 Indonesian Crime Statistics, with the dependent variable being the number of reported crimes (Y), and the independent variables including population density (X_1), open unemployment rate (X_2), percentage of poor population (X_3), average length of schooling (X_4), and labor force participation rate (X_5).

The research procedure consists of several steps. First, classical assumption tests are conducted on the dataset [21, 22]. Second, regression parameter estimation is carried out using OLS (Ordinary Least Squares), and outliers in the model are identified using the DfFITS method, leverage values, and scatter plots. If the data are indicated to contain outliers, the analysis proceeds with robust regression using M-estimation and S-estimation methods.

3.1 M-estimation robust regression analysis

Parameter estimation using robust regression with M-estimation: The initial estimate of $\hat{\beta}$ is obtained using the least squares method, resulting in the estimated values Y_i .

1. Calculating the residual values $e_i = y_i - \hat{y}_i$
2. Calculating the standard deviation of the residuals (σ_i^2)
3. Calculating the robust residual values $u_i = \frac{e_i}{\sigma_i}$
4. Calculating the weighting function values (w_i)
5. Calculating the parameters $\hat{\beta}_M$ using the Weighted Least Squares (WLS) method with weights w_i .
6. Repeat steps 2 to 6 until the estimated value of $\hat{\beta}_M$ converges
7. Hypothesis testing is conducted to determine whether the independent variables have a significant effect on the dependent variable
8. Obtain the model based on robust regression using M-estimation

Parameter estimation using robust regression with S-estimation.

1. Calculating the parameters $\hat{\beta}^0$ with MKT.
2. Calculating the residual values $e_i = y_i - \hat{y}_i$.
3. Calculating the scale of the residuals ($\hat{\sigma}_s$)
4. Calculating the robust residuals $u_i = \frac{e_i}{\hat{\sigma}_i}$.
5. Calculating the weights (w_i)
6. Calculating the parameters $\hat{\beta}_s$ using the Weighted Least Squares (WLS) method / WLS (Weighted Least Squares) with weights w_i^0 .
7. Repeat steps one through six until a value is obtained $\hat{\beta}_s$. Convergence is achieved when the difference in values β_{sj}^{l+1} and β_{sj}^l approaches 0.
8. Hypothesis testing is conducted to determine whether the independent variables have a significant effect on the dependent variable.
9. Obtaining the model based on robust regression using S-estimation

10. The selection of the best regression model is based on the criteria of the highest R^2 value and the smallest Robust Residual Standard Error (RRSE).

Identifying the significant factors influencing the crime rate in Indonesia based on the best robust regression estimation.

4. RESULT & DISCUSSION

The results of the descriptive analysis of the research data are presented in **Table 1**. Based on **Table 1**, the descriptive analysis shows that the highest number of reported crimes is 51,905, with a mean of 10,967.55. The average values for population density, open unemployment rate, percentage of poor population, average years of schooling, and labor force participation rate are 751.14; 4.96; 10.29; 9.24; and 68.64, respectively.

Table 1. Descriptive Analysis

Variables	Min	Max	Mean	Std Deviation
Crime Rate	1220	51905	10967.55	12698.21
Population Density	10	16158	751.1471	2751.22
Open Unemployment Rate	2.34	8.31	4.9662	1.6002
Poor Population	4.53	26.80	10.2901	5.28564
Average Years of Schooling	7.31	11.30	9.2468	0.81736
Labor Force Participation Rate	63.08	77.75	68.6438	3.64053

The standard deviation for the number of crimes is 12,698.21. Population density has a standard deviation of 2,751.22; the open unemployment rate has a standard deviation of 1.6; the percentage of poor population has a standard deviation of 5.28; the average years of schooling has a standard deviation of 0.81; and the labor force participation rate has a standard deviation of 3.64. The parameter estimation using the Ordinary Least Squares (OLS) method yields the following regression equation:

$$\hat{y} = -1.075 \times 10^{-16} + 0.3853X_1 + 0.4492X_2 - 0.02271X_3 - 0.2918X_4 + 0.1434X_5$$

Before conducting further analysis, a simultaneous test using the F-test is performed to determine whether the independent variables have a significant effect on the dependent variable. Hypotheses:

H_0 : There is no effect of the independent variables on the dependent variable.

H_1 : There is an effect of the independent variables on the dependent variable.

Based on the research results, the p-value obtained is 0.09989, which fails to reject H_0 . This indicates that there is no significant effect of the independent variables on the dependent variable.

4.1 Classical Assumption Test

Normality assumption testing was conducted using the Shapiro-Wilk test with the following hypotheses:

H_0 : The residuals are normally distributed

H_1 : The residuals are not normally distributed The analysis produced a Shapiro-Wilk test statistic of 0.80628 with a p-value of 0.00003398, which is less than the 5% significance level. Therefore, H_0 is rejected, indicating that the residuals are not normally distributed.

Multicollinearity detection was performed by examining the Variance Inflation Factor (VIF) values, using the criterion that VIF values should be less than 10. Table 2 shows that there is no multicollinearity problem among the independent variables, as all VIF values are below 10. Autocorrelation testing was carried out using the Durbin-Watson (DW) test, with the following

Table 2. Variance Inflation Factor (VIF) Value

Variable	X_1	X_2	X_3	X_4	X_5
VIF	1.255086	1.656784	1.354112	1.894815	1.693125

hypotheses:

H_0 : There is no autocorrelation in the regression model

H_0 : There is autocorrelation in the regression model

Table 3 shows that the calculated Durbin-Watson test statistic (d_{count}) is 1.94, which is greater than the upper bound $dU = 1.81$. This can be written as $dU < d_{count} < (4 - dU)$, with a p-value of $0.4103 > 5\%$. Therefore, H_0 is accepted, indicating that there is no autocorrelation problem in the regression model. The heteroscedasticity assumption was tested using

Table 3. Durbin-Watson Test Results

	d_{count}	$d_{L;0.05}$	$d_{U;0.05}$	$4 - d_{L;0.05}$	$4 - d_{U;0.05}$
Value	1.94	1.15	1.81	2.85	2.19

the Breusch-Pagan (BP) test, with the decision rule to reject H_0 if the p-value $< \alpha = 5\%$. The hypotheses are as follows:
 H_0 : There is no heteroscedasticity in the regression model
 H_1 : There is heteroscedasticity in the regression model
 The BP test produced a test statistic value of 11.016 with a p-value of $0.05107 > 5\%$. Thus, it can be concluded that the model fails to reject H_0 , meaning that there is no heteroscedasticity problem in the regression model.

4.2 Outlier Detection

Outlier detection was performed using a scatter plot. The scatter plot of the research data with the residual values is shown in **Figure 1**.

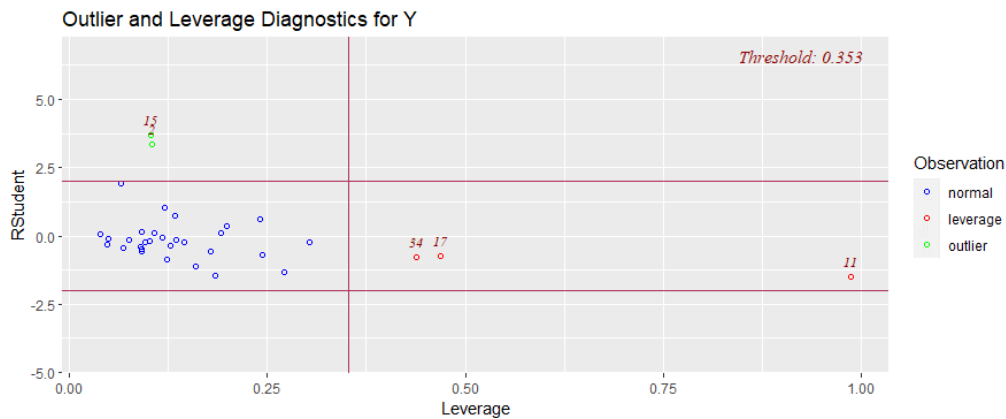


Figure 1. Outlier detection with scatter plot

Outlier detection using DFFITS and leverage values was carried out based on the following hypotheses:

H_0 : The i -th outlier has no influence, for $i = 1, 2, 3, \dots, 34$

H_1 : The i -th outlier has an influence

The critical value for the DFFITS test is:

$$2\sqrt{\frac{p}{n}} = 2\sqrt{\frac{6}{34}} = 0.840168$$

The critical value for leverage is commonly determined using the formula:

$$\frac{2k}{n} = \frac{2(6)}{34} = 0.35294$$

The decision criteria for detecting outliers based on leverage and DFFITS are as follows:

$$DfFITS = \begin{cases} \leq 0.840168, & H_0 \text{ is accepted} \\ > 0.840168, & H_0 \text{ is rejected} \end{cases}$$

$$laverage = \begin{cases} \leq 0.35294, & H_0 \text{ is accepted} \\ > 0.35294, & H_0 \text{ is accepted} \end{cases}$$

Based on the test results using DfFITS and leverage values, observations no. 2 and 15 were identified as outliers with DfFITS

Table 4. Results of DfFITS and Leverage value

No.	DfFITS	—DfFITS—	Leverage Value	No.	DfFITS	—DfFITS—	Leverage Value
1	0.05981041	0.05981041	0.19188254	18	-0.07516922	0.07516922	0.09613486
2	1.15455917	1.15455917	0.10490605	19	-0.26089706	0.26089706	0.17909318
3	-0.1742828	0.17428281	0.09189666	20	-0.47760720	0.47760720	0.15974338
4	0.29447895	0.29447895	0.13372429	21	-0.12431468	0.12431468	0.09006634
5	-0.0694600	0.06946003	0.04826661	22	-0.15085539	0.15085539	0.09215854
6	0.01144230	0.01144230	0.03942619	23	-0.06986197	0.06986197	0.10134932
7	-0.0374396	0.03743966	0.07590662	24	-0.11446367	0.11446367	0.06899333
8	-0.0203452	0.02034524	0.04996837	25	-0.02444052	0.02444052	0.11822087
9	-0.3308593	0.33085931	0.12341487	26	0.05091332	0.05091332	0.09187560
10	-0.8163807	0.81638070	0.27159482	27	0.50520111	0.50520111	0.06571517
11	-12.959972	12.95997230	0.98706883	28	0.04024438	0.04024438	0.10809596
12	0.33577992	0.33577992	0.24058048	29	-0.09446688	0.09446688	0.14472318
13	0.38212133	0.38212133	0.12095524	30	-0.13758532	0.13758532	0.12815413
14	0.18125086	0.18125086	0.19881832	31	-0.39224554	0.39224554	0.24419498
15	1.25649677	1.25649677	0.10361864	32	-0.06028124	0.06028124	0.13558054
16	-0.6996338	0.69963380	0.18422154	33	-0.15203672	0.15203672	0.30323429
17	-0.6858801	0.68588012	0.46844789	34	-0.68398123	0.68398123	0.43805235

values of 1.15455917 and 1.25649677, respectively. Meanwhile, observations no. 11, 17, and 34 were identified as outliers in the leverage test with values of 0.98706883, 0.46844789, and 0.43805235, respectively. These values exceed the critical threshold, leading to the rejection of H_0 and indicating that these observations are influential outliers in the regression model. Since outliers are present in the dataset, robust regression analysis will be conducted using M-estimation and S-estimation methods.

4.3 M-Estimation Robust Regression

The results of parameter estimation with robust M-estimation regression are presented in **Table 5**. The results of parameter

Table 5. M-estimation robust regression

No.	y_i	$\hat{y}_i$	ε_i	$ MAD $	$\hat{\sigma}_M$	u_i	w_i
1	-0.06541	-0.16964	0.10423	0.28986	0.42975	0.24254	0.99465
2	2.5663	0.03722	2.52907	0.28986	0.42975	5.885	0
3	-0.25803	0.23255	-0.49057	0.28986	0.42975	-1.14154	0.88479
4	0.11194	-0.54044	0.65238	0.28986	0.42975	1.51805	0.80104
5	-0.44168	-0.15782	-0.28386	0.28986	0.42975	-0.66053	0.96064
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
34	-0.26646	0.27634	-0.542803	0.28986	0.42975	-1.26307	0.85992

estimation with robust M-estimation regression for iteration 1 are obtained:

$$\begin{aligned} \hat{\beta}_m^1 &= [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5] \\ &= [-0.19734968, 0.40087083, 0.3947812, 0.03325691, -0.25710976, 0.011110205] \end{aligned}$$

So that the robust regression equation for the M-estimate is obtained at iteration 1.

$$\hat{y}^{(1)} = -0,197 + 0,400X_1 + 0,339X_2 + 0,0331X_3 - 0,257X_4 + 0,011X_5 \quad (25)$$

The results of the first iteration have not converged because the difference between the OLS estimation model and iteration 1 is not close to 0, so it will be continued with iteration 2 using the calculations in iteration 1 and so on until convergence is achieved, namely the model in the m -th iteration is the same as the model in the m -th iteration + 1. Based on the iteration results, it is known that the iteration has converged in the 8th iteration because the difference in values in the 7th iteration and the 8-th iteration is equal to zero, so that the M-estimation equation is obtained:

$$\hat{y}_M = -0,338 + 0,380X_1 + 0,067X_2 + 0,046X_3 - 0,048X_4 - 0,007X_5 \quad (26)$$

With $R^2 = 0.6643$, $R^2(adj) = 0.6043$, and Robust Residual Standard Error (RRSE) value of 0.3177, the interpretation of the equation is that every additional unit of population density index (X_1), open unemployment rate (X_2), and number of poor people (X_3) will increase the number of crimes by 0.380, 0.067, and 0.046, respectively. Every additional unit of average years of schooling (X_4) and labor force participation rate (X_5) will decrease the number of crimes by 0.048 and 0.007, respectively. A partial test (t-test) was then conducted to determine whether each independent variable had a significant effect on the dependent variable. The results are presented in Table 6. Table 6 explains that the constant value has a value of $p\text{-value} < 0.05$, which

Table 6. Partial test of independent variables of crime prediction model with M estimation

	Coefficient	t-count	p-value	Conclusion
Constant	-0.33845	-6.457	5.41×10^{-7}	Significant
X_1	0.38036	6.915	1.62×10^{-7}	Significant
X_2	0.06792	0.917	0.367	not significant
X_3	0.04641	0.824	0.417	not significant
X_4	-0.04866	-0.646	0.524	not significant
X_5	-0.00746	-0.113	0.911	not significant

means that at least one independent variable influences the dependent variable. It can be seen that the variable $p\text{-value} < 0.05$ affects the dependent variable X_1 , namely population density has a significant effect on the number of crimes in Indonesia because it has a value of $p\text{-value} < 0.05$.

4.4 S-Estimation Robust Regression

The results of parameter estimation using robust S-estimation regression are presented in Table 7. The results of parameter

Table 7. S-estimation robust regression

No.	y_i	$\hat{y}_i$	ϵ_i	ϵ_i^2	σ_s	u_i	w_i
1	-0.06541	-0.29432	0.22891	0.05239	0.30887	0.74108	0.5937
2	2.5663	-0.13368	2.69998	7.28991	0.30887	8.74101	0
3	-0.25803	0.03203	-0.2901	0.08413	0.30887	-0.9391	0.3988
4	0.11194	-0.62339	0.73533	0.54070	0.30887	2.38057	0
5	-0.44168	-0.30536	-0.13632	0.01858	0.30887	-0.4413	0.8438
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
34	-0.26646	0.03952	-0.30598	0.09362	0.30887	-0.9906	0.34806

estimation using robust M-estimation regression for iteration 1 are obtained:

$$\begin{aligned} \hat{\beta}_s^1 &= [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5] \\ &= [-0.23629450, 0.37273039, 0.34395281, 0.03482291, -0.22148747, 0.01823836] \end{aligned}$$

So that the robust regression equation for the M-estimate is obtained at iteration 1.

$$\hat{y}^{(1)} = -0,236 + 0,372X_1 + 0,343X_2 + 0,034X_3 - 0,221X_4 + 0,018X_5 \quad (27)$$

The equation above is the robust regression model for the S-estimate for iteration 1. At iteration 1, the equation has not converged, so it will continue to the next iteration. The iteration will stop when the difference in the j values of $\hat{\beta}_{i+1}$ and $\hat{\beta}_{ji}$ approaches zero. The robust regression equation that converges with the S-estimate is

$$\hat{y}_S = -0,462 + 0,411X_1 + 0,021X_2 + 0,017X_3 - 0,008X_4 + 0,083X_5 \quad (28)$$

With $R^2 = 0.9493$, $R^2(adj) = 0.9402$, and a Robust Residual Standard Error (RRSE) value of 0.2996. The interpretation of the above equation is that each unit increase in the population density index (X_1), the open unemployment rate (X_2), the percentage of the poor (X_3), and the labor force participation rate (X_5) will increase the number of crimes in Indonesia by 0.411; 0.021; 0.017 and 0.083, respectively. Meanwhile, each unit increase in the average number of years of schooling (X_4) will decrease the number of crimes in Indonesia by 0.00862 if the population density index (X_1), the open unemployment rate (X_2), the percentage of the poor (X_3), and the labor force participation rate (X_5) are fixed.

Next, a partial test (t-test) is conducted to determine whether each independent variable has a significant effect on the dependent variable. Table 8 presents a partial test of the independent variables of the crime prediction model using S-estimation. **Table 8** explains that the constant value has a p-value < 0.05 , indicating that at least one independent variable influences the

Table 8. Partial Test of Independent Variables of Crime Prediction Model with S Estimation

	Coefficient	t-count	p-value	Conclusion
Constant	-0.46255	-13.897	4.33×10^{-14}	Significant
X_1	0.41112	13.518	8.54×10^{-14}	Significant
X_2	0.02108	0.479	0.635	not significant
X_3	0.01756	0.535	0.597	not significant
X_4	-0.00862	-0.194	0.847	not significant
X_5	-0.08354	1.971	0.058	not significant

dependent variable. Therefore, it can be seen that the variable with a p-value < 0.05 influences the dependent variable. X_1 , population density, has a significant effect on the number of crimes in Indonesia because it has a p-value < 0.05 .

The comparison of the robust M-estimation and S-estimation regression models uses two criteria: the coefficient of determination (R^2) and the Robust Residual Standard Error (RRSE). The best model will have the largest R^2 value and the smallest RRSE value. Table 9 shows the comparison of the R^2 and RRSE values. **Table 9** shows that the estimation results

Table 9. Comparison of R^2 and RRSE values

Method	R^2	RRSE
M-estimation	0,6643	0,3177
S-estimation	0,9493	0,2996

using S-estimation provide a larger R^2 value and a smaller RRSE value than the robust M-estimation regression model, thus indicating that the S-estimation model is the better model. The convergent robust regression equation of S-estimates shows that there is one independent variable, namely population density, which has a significant effect on the number of crimes in Indonesia with a value of p-value of 8.54×10^{-14} . The following is a population density model (X_1) with S-estimates.

$$\hat{y} = -0.462 + 0.411X_1 \quad (29)$$

The equation above shows that if population density (X_1) increases by one unit, it will increase the number of crimes in Indonesia by 0.411. The R^2 value = 0.9493 shows that the total variation in Y of 94.93% is explained by the variable X_1 , while 5.98% is explained by other variables outside the observation.

5. CONCLUSION

The results of the robust regression model equation that converged with Tukey Bisquare weights on the 2022 crime rate data in Indonesia are:

$$\hat{Y}_M = -0.338 + 0.380X_1 + 0.067X_2 + 0.046X_3 - 0.048X_4 - 0.007X_5$$

Robust regression model S-estimation converges:

$$\hat{Y}_S = -0.462 + 0.411X_1 + 0.021X_2 + 0.017X_3 - 0.008X_4 + 0.083X_5$$

The results of a comparison of the robust regression models, M-estimation and S-estimation, using Tukey Bisquare Weights on crime data in Indonesia, based on the criteria of the largest R^2 value and the smallest RRSE, showed that the robust regression method, S-estimation, produced the best robust regression equation model with an $R^2 = 94.93\%$ and an RRSE value of 0.2996. The factors influencing the number of crimes in Indonesia in 2022, based on the robust regression model, S-estimation, as the best robust regression model, were population density. This means that if population density increases, the number of crimes in Indonesia will also increase.

REFERENCES

- [1] S. Syilfi, D. Ispriyanti, and D. Safitri, "Analisis Regresi Linier Piecewise Dua Segmen," *Jurnal Gaussian*, vol. 1, no. 1, pp. 219–228, Oct. 2012.
- [2] S. Soemartini, *Pencilan (Outlier)*. Universitas Padjajaran, 2007.
- [3] H. Lainun, G. M. Tinungki, and A. Amran, "Perbandingan Penduga M, S, dan MM pada Regresi Linier dalam Menangani Keberadaan Outlier," *Jurnal Matematika Statistika dan Komputasi*, vol. 15, no. 1, pp. 88–96, Jul. 2018.
- [4] N. Herawati, K. Nisa, and E. Setiawan, "Analisis Ketegaran Regresi Robust Terhadap Letak Pencilan: Studi Perbandingan," *Bulletin of Mathematics*, vol. 3, no. 1, pp. 49–60, 2011.
- [5] N. A. Atamia, Y. Susanti, and S. S. Handajani, "Perbandingan Analisis Regresi Robust Estimasi-S dan Estimasi-M dengan Pembobot Huber dalam Mengatasi Outlier," in *Prosiding Seminar Nasional Matematika*, vol. 4. PRISMA, 2021, pp. 673–679.
- [6] C. Chen, "Robust Regression and Outlier Detection with the ROBUSTREG Procedure," in *SUGI Conference Proceedings*. SAS Communities, Jan. 2002, pp. 265–278.
- [7] A. M. Maziyah, "Penggunaan Regresi Robust Dengan Estimasi-S Dan Estimasi-Mm Dalam Pengembangan Sistem Pendukung Keputusan Guna Memprediksi Tingkat Produksi Padi," Undergraduate thesis, Institut Teknologi Sepuluh November, 1995.
- [8] Z. Setiarini and E. Listyani, "Analisis Regresi Robust Estimasi-S Menggunakan Pembobot Welsch Dan Tukey Bisquare," *Jurnal Matematika*, vol. 6, no. 1, pp. 48–55, 2017.
- [9] V. J. Yohai, "High Breakdown-Point and High Efficiency Robust Estimates for Regression," *Ann. Statist.*, vol. 15, no. 2, Jun. 1987.
- [10] A. A. Dewayanti and E. Widodo, "Perbandingan Metode Estimasi M Dan Estimasi Mm (Methode Of Moment) pada Regresi Robust," in *Prosiding Konferensi Nasional Penelitian Matematika dan Pembelajarannya*. Surakarta, Indonesia: Program Studi Pendidikan Matematika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Muhammadiyah Surakarta, Mar. 2016, pp. 751–758.
- [11] M. Mardiana, "Perbandingan Model Regresi Robust Estimasi M, Estimasi S dan Estimasi MM pada Faktor yang Mempengaruhi Angka Kejadian Demam Berdarah Dengue di Provinsi Jawa Timur Tahun 2017," Thesis, Universitas Airlangga, Surabaya, Indonesia, Jun. 2019.

- [12] S. Suyono, *Analisis Regresi untuk Penelitian*. deepublish, Aug. 2015.
- [13] H. Ihsan, W. Sanusi, and N. Nurfadillah, “Estimasi Parameter Regresi Linear Pada Kasus Data Outlier Menggunakan Metode Estimasi Method Of Moment,” *JMathCoS*, vol. 1, no. 1, p. 38, May 2019. [Online]. Available: <https://ojs.unm.ac.id/JMathCoS/article/view/9176>
- [14] F. Hidayatulloh, D. Yuniarti, and S. Wahyuningsih, “Regresi Robust dengan Metode Estimasi,” in *Prosiding UNINDRA*. Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Indraprasta PGRI, 2015, pp. 163–170.
- [15] A. Widarjono, *Ekonometrika: Teori dan Aplikasi untuk Ekonomi dan Bisnis*. Yogyakarta Ekonesia, 2007.
- [16] T. Tarno, “Estimasi Model Regresi Linier dengan Metode Median Kuadrat Terkecil,” *Jurnal Sains & Matematika (JSM)*, vol. 15, no. 2, pp. 69–72, Apr. 2007.
- [17] P. R. Sihombing, S. Suryadiningrat, D. A. Sunarjo, and Y. P. A. C. Yuda, “Identifikasi Data Outlier (Pencilan) dan Kenormalan Data Pada Data Univariat serta Alternatif Penyelesaiannya,” *JESI*, vol. 2, no. 3, pp. 307–316, Jan. 2023. [Online]. Available: <https://jurnaljesi.com/index.php/jurnaljesi/article/view/112>
- [18] Y. Susanti, H. Pratiwi, S. Sulistijowati H., and T. Liana, “M Estimation, S Estimation, and MM Estimation in Robust Regression,” *Int. J. of Pure and Appl. Math.*, vol. 91, no. 3, Mar. 2014. [Online]. Available: <http://www.ijpam.eu/contents/2014-91-3/7/>
- [19] S. Wahyuningsih and D. Yuniarti, “Regresi Robust Linear Sederhana dengan Menggunakan Estimasi MM (Method of Moment) Simple Linear Regression Robust with MM,” *Estimation*, vol. 6, pp. 179–186, 2015.
- [20] N. R. Draper and H. Smith, *Applied Regression Analysis*, 1st ed., ser. Wiley Series in Probability and Statistics. Wiley, Apr. 1998. [Online]. Available: <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118625590>
- [21] “Robust Regression,” in *Dictionary of Statistics & Methodology*. 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc., 2005. [Online]. Available: <https://methods.sagepub.com/reference/dictionary-of-statistics-methodology/n1702.xml>
- [22] Y. Susanti, S. P. Hasih, and H. Sulistijowati, “Optimasi Model Regresi robust untuk Memprediksi Produksi Kedelai di Indonesia,” in *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika*. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Yogyakarta, Dec. 2013, pp. 253–262.