BAREKENG: Journal of Mathematics and Its Applications
September 2024 Volume 18 Issue 3 Page 1573–1584
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss3pp1573-1584

ROBUST GEOGRAPHICALLY WEIGHTED REGRESSION WITH LEAST ABSOLUTE DEVIATION (LAD) ESTIMATION AND M-ESTIMATION ON GRDP OF WEST JAVA PROVINCE

Prizka Rismawati Arum[1*], Mohammad Ridwan[2], Ina Alfidayanti[3], Rochdi Wasono[4]

[1,3,4]Department of Statistics, Universitas Muhammadiyah Semarang
[2]Department of Accounting, Universitas Muhammadiyah Semarang
Jln. Raya Kedungmundu No. 18 Semarang, 50273, Indonesia

*Corresponding author's e-mail: * **prizka.rismawatiarum@unimus.ac.id***

ABSTRACT

Article History:

Received: 1st January 2023
Revised: 17th March 2024
Accepted: 23rd June 2024
Published: 1st September 2024

Keywords:

Gross Regional Domestic Product;
Least Absolute Deviation;
M-estimation;
Robust Geographically Weighted Regression.

Geographically Weighted Regression (GWR) is an analytical method for data that contains spatial heterogeneity effects. However, parameter estimation in the GWR model has a weakness, namely it is prone to outliers and can cause the parameter estimation to be biased. This can be overcome by the Robust Geographically Weighted Regression (RGWR) method which is more robust against the presence of outliers. This method is suitable for Gross Regional Domestic Product (GRDP) data in West Java Province, which contains outliers and also has spatial effects. The data used in this study are secondary data obtained from the Central Statistics Agency (BPS) of West Java Province. The purpose of this study is to compare the Robust Geographically Weighted Regression (RGWR) method with the Least Absolute Deviation (LAD) Estimation and M-estimation and also to find out the factors that affect the Gross Regional Domestic Product (GRDP) in West Java Province in 2021 based on the model resulting from. Selection of the best model is seen based on the value of the coefficient of determination (R^2) and Mean Squared of Error (MSE). The research results show that the Robust Geographically Weighted Regression (RGWR) method with M-estimation is much more effective in estimating the distribution of GRDP in West Java Province in 2021, seen from the larger coefficient of determination and the smaller Mean Square Error (MSE). The variables that have a significant influence on GRDP in West Java Province in 2021 are the variables of foreign investment and local income.

How to cite this article:

P. R. Arum, M. Ridwan, I. Alfidayanti and R. Wasono., "ROBUST GEOGRAPHICALLY WEIGHTED REGRESSION WITH LEAST ABSOLUTE DEVIATION (LAD) ESTIMATION AND M-ESTIMATION ON GRDP OF WEST JAVA PROVINCE," *BAREKENG: J. Math. & App.*, vol. 18, iss. 3, pp. 1573-1584, September, 2024.

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

National development aims to create a just, prosperous, competitive, advanced, and prosperous society. The central and regional governments have made various efforts to ensure the equitable development across various regions [1]. One important indicator to evaluate the success of regional development is economic growth. National development aims to create a just, prosperous, competitive, advanced and prosperous society [2] . Economic growth is a sustained change in the economic conditions of a country or region towards a better state over a specific period. An economy is considered to be developing when the level of economic activity in a particular country or region is higher than in the previous [3]. To measure the economic growth of a region during a specific period, one crucial indicator used is the Gross Regional Domestic Product (GRDP) [4]. GRDP is the total value added from goods and services produced by various production units in a region of a country during a specific period. The GRDP value can reflect the region's ability to manage and utilize available resources [5].

The Central Statistics Agency (BPS) notes that the national economic structure spatially is still dominated by provinces on the Java Island, contributing 57.89% to Indonesia's total Gross Domestic Product (GDP). West Java is one of the provinces on Java Island that contributes the highest Gross Regional Domestic Product (GRDP) to the GDP of Indonesia in 2021. The economy of West Java, based on the Gross Regional Domestic Product (GRDP) at current prices, reached Rp. 2.21 quadrillion throughout 2021. Meanwhile, when measured using the GRDP at constant prices (2010), the economy of West Java grew by 3.74% in 2021 compared to the previous year. This growth is significantly better than the contraction of 2.52% experienced in the previous year due to the Covid-19 pandemic [6]. However, when examining the contribution of Gross Regional Domestic Product (GRDP) in each regency/city, the distribution of GRDP in West Java is highly diverse, indicating unevenness.

From the data obtained for each observation, there are some observations that have values significantly higher or lower than the others. Data with unique and significantly deviating characteristics, appearing in the form of extreme values and deviating from the formed pattern, are called outliers [7]. The presence of outliers can have a significant impact on the estimation process of parameter estimates [8]. Outliers can cause an increase in data variability, leading to the mean value being unable to reflect the true value (bias), and they can also result in errors in decision-making and conclusions [9]. Additionally, data diversity can be influenced by various factors, both demographic and geographic. High GDRP figures in one region can also impact the GDRP figures in neighboring areas, and the location or region factor is suspected to have a spatial influence on GDRP figures in the West Java Province [10].

Geographically Weighted Regression (GWR) is an analytical method for data containing spatial heterogeneity effects. However, parameter estimation in GWR has a weakness as it is susceptible to outliers, which can lead to bias in parameter estimation [11]. Dreper and Smith stated that the removal of outliers is not a wise action, as sometimes outliers can provide information not conveyed by other data points [12]. Therefore, the Robust Geographically Weighted Regression (RGWR) method is needed, which is more robust or resistant to the presence of outliers [13].

Robust Geographically Weighted Regression (RGWR) is a method that takes in to account outliers in the analyzed data. RGWR can be employed when the residual data is not normally distributed, and there are suspected outliers that may influence the model [14]. Therefore, this method is suitable for use with GDRP data for West Java Province, which contains outliers and exhibits spatial effects. RGWR has several estimations, including Least Trimmed Square (LTS), Least Median Square (LMS), Least Absolute Deviation (LAD), Method of Moment (MM-estimation), Maximum Likelihood Type (M-estimation), and Scale (S estimation [15].

One study conducted by Aqilah Salsabila [2021] on the Robust Geographically Weighted Regression Model with the Least Absolute Deviation (LAD) method in the case of Poverty in South Sulawesi Province found that the RGWR model with the Least Absolute Deviation (LAD) method is the most suitable for spatially diverse data containing outliers [16]. Another study by Puspitasari et al. [2021] on the Robust Regression Model for the Human Development Index in East Java with M-estimation concluded that the robust regression with M-estimation can produce convergent estimates using Tukey bi-square weights, with an R^2 value of 99.91%, and all independent variables significantly influence the Human Development Index in East [17].

Based on the description above, the Robust Geographically Weighted Regression method using the Least Absolute Deviation (LAD) and M-estimation can be employed to handle outliers, providing estimates that are close to the true values. Therefore, in this study, the researcher aims to compare the Robust Geographically Weighted Regression models using the Least Absolute Deviation (LAD) and M-estimation on the distribution of Gross Regional Domestic Product in each location of the districts/cities in West Java Province in 2021. Least Absolute Deviation (LAD) can automatically deal with outliers without having to detect data that is oulier and without having to reweight the residuals. LAD is the simplest estimation approach in non-parametric regression and can improve the robustness of the kernel function [18]. M-estimation is highly efficient and very good for estimating parameters affected by outliers on variable X. Parameter estimation using the M-estimation method is carried out through an iterative process to achieve convergent parameter estimate values [19]. This research was conducted to determine the factors that influence Gross Regional Domestic Product (GRDP) in West Java Province in 2021 based on the model produced, so that this research can be used as a basis for decision making by the West Java Government to determine effective policies or strategies to achieve national development goals.

2. RESEARCH METHODS

The data used in this study are secondary data obtained from the Central Statistics Agency (BPS) of West Java Province. There is a total of 26 data points used, representing the Gross Regional Domestic Product (GDRP) according to districts/cities in West Java Province for the year 2021. The response variable in this study is the Gross Regional Domestic Product ($\boldsymbol{Y}$), while the predictor variables assumed to have an influence on GDRP include: Foreign Direct Investment ($\boldsymbol{X_1}$), Domestic Direct Investment ($\boldsymbol{X_2}$), Regional Original Income ($\boldsymbol{X_3}$), Unemployment Rate ($\boldsymbol{X_4}$), and Population ($\boldsymbol{X_5}$).

2.1 Research Procedure

The sequence of steps conducted in this study is as follows:

1. Performing descriptive analysis on the Gross Regional Domestic Product (GDRP) data in West Java for the year 2021.
2. Conducting classical assumption tests.
   i. Autocorrelation Test
   Autocorrelation test can be conducted using the Durbin-Watson test, where the hypothesis for the Durbin-Watson test is defined as follows:

   H_0: no autocorrelation
   H_1: autocorrelation is present

   The Durbin-Watson test statistic is defined in the following form [20]:

$$d = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}(\hat{u}_t - \hat{u}_{it-1})^2}{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{u}_{it}^{\,2}} \tag{1}$$

   with:
   d = Durbin-Watson
   $\hat{u}_t$ = Residual value for t period
   $\hat{u}_{t-1}$ = Residual value for $(t-1)$ period

   ii. Heteroscedasticity Test
   Heteroscedasticity test can be performed using the Glejser test, where the absolute residual values of each response variable are regressed [21]. The hypotheses adopted for this test are as follows:

   H_0: No heteroscedasticity
   H_1: heteroscedasticity is present

   The assumption of heteroscedasticity is met if the p-value > 0.05, indicating that there is no heteroscedasticity in the model.

iii. Multicollinearity Test

Multicollinearity refers to the presence of a linear relationship between predictor variables in a regression model. The multicollinearity test is conducted with the following hypotheses:

H_0: no multicollinearity
H_1: multicollinearity is present

To detect multicollinearity, the Variance Inflation Factor (VIF) can be used. The VIF calculation can be performed using the following formula:

$$VIF_j = \frac{1}{1-R_j^2}; \text{with } j = 1,2,\dots,k \tag{2}$$

where R^2 is the coefficient of determination between the j-th independent variable and the remaining independent variables $(k-1)$. If the VIF value < 10, it can be concluded that there is no multicollinearity in the regression model [22].

iv. Normality Test

Normality testing is conducted using the Kolmogorov-Smirnov test or the Shapiro-Wilk test, with the following hypotheses [22]:

H_0: data follows a normal distribution
H_1: data does not follow a normal distribution

The decision for the test is to accept H_0 if the p-value > 0.05, indicating that the data has a normal distribution. Conversely, H_0 is rejected if the p-value < 0.05, indicating that the data is not normally distributed.

3. Detecting outliers through visualization using scatterplots and calculations using DfFITS. DfFITS can be implemented with the following equation [23]:

$$(\text{DfFITS})_i = t_i\left(\frac{hii}{1-h_{ii}}\right)^{\frac{1}{2}} \tag{3}$$

Data is identified as an outlier if the value of $|\text{DfFITS}| > 2\sqrt{\frac{p}{n}}$, where p is the number of parameters, and n is the number of observations.

4. Conducting a test for spatial heterogeneity using the Breusch-Pagan method.

Spatial heterogeneity testing aims to evaluate variability at each location observation point. In the context of spatial heterogeneity, there are significant differences in the variance of regression parameters [24]. The test statistic in Breusch-Pagan is given by [25]:

$$BP = \left(\frac{1}{2}\right) f^T \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T f - X^2{}_{(k)} \tag{4}$$

Here, the vector ff elements are obtained as : $f_i = \frac{e_i{}^2}{\sigma^2} - 1$
With:

e_i $= y_i - \hat{y}_i$
f $= (f_1, f_2, \dots, f_n)^T$
σ^2 = The variance of the variable y
$e_i{}^2$ = the squared residuals for the i-th observation location
Z = matrix of size $n \times (p+1)$ containing standardized vectors (z) for each observation.

The decision-making is as follows: H_0 is rejected if the value of $BP > X^2{}_{(k)}$ or the p-value < 0,05 indicating that there is a difference in variances across locations.

5. Selecting the optimal bandwidth.

The method of bandwidth selection is crucial for obtaining accurate kernel density estimates [26]. Mathematically, the Absolute Cross Validation (ACV) can be formulated as follows [18].

$$ACV\ (h) = \sum_{i=1}^{n} |y_i - \hat{y}_{\neq i}(h)| \tag{5}$$

With:

n	= the number of observation
i	$= 1,2,\dots,n$
y_i	= the i-th observation data
$\hat{y}_{\neq i}(h)$	= the estimated value of the i-th observation, obtained without involving the observation at the i-th location

6. Performing the analysis of Robust Geographically Weighted Regression on the Gross Regional Domestic Product (GDRP) data in West Java Province for the year 2021.

The equation of Robust Geographically Weighted Regression (RGWR) describes the relationship between the response and predictor variables at the local level in the context of spatial data. The general equation for RGWR can be written as follows:

$$y_i = \beta_{0i} + \beta_{1i}x_{1i} + \beta_{2i}x_{2i} + \dots + \beta_{pi}x_{pi} + e_i \tag{6}$$

With:

y_i	= the response variable at location i
$\beta_{0i}, \beta_{1i}, \beta_{2i}, \dots, \beta_{pi}$	= the local regression parameter (coefficient) at location i
$x_{1i}, x_{2i}, \dots, x_{pi}$	= the predictor values at location i
p	= the total number of predictor variables in the model
e_i	= the residual value at location i

In the RGWR model, the local regression parameters $(\beta_{0i}, \beta_{1i}, \beta_{2i}, \dots, \beta_{pi})$ are estimated individually for each location using robust estimation techniques that are resistant to outliers, such as M-estimation or Least Absolute Deviation (LAD). These estimations allow the model to capture variations in the relationships between variables across different locations.

i. Robust Geographically Weighted Regression with *Least Absolute Deviation* (LAD) Estimation.
The Least Absolute Deviation (LAD) estimation method can be performed by minimizing the total absolute residual values. The estimation procedure for regression coefficient values with the LAD method is based on the following function [27]:
Minimizing $\sum_{i=1}^{n}(\varepsilon_i^+ - \varepsilon_i^-)w_i(u_iv_i)$
With the constraint function:

$$\beta_0(u_iv_i) - \sum_{k=1}^{K}\beta_k(u_iv_i)X_{ik} + \varepsilon_i^+ - \varepsilon_i^- = y_i \text{ with } i = 1,2,\dots,n \tag{7}$$

where ε_i^+ and ε_i^- represent positive and negative deviations for observation i.
In LAD estimation, the Adaptive Gaussian Kernel weighting function is used and formulated as follows [28]:

$$w_j(u_i,v_i) = exp\left(-\left(\frac{d_{ij}}{h_i}\right)^2\right) \tag{8}$$

ii. Robust Geographically Weighted Regression with M-estimation
The Method of Moment (M-Estimation) aims to minimize the residual function ρ (objective function) of its residuals. The objective function in M-Estimation can be defined as follows [29]:

$$\sum_{i=1}^{n}\rho\,(u_i) = \min\sum_{i=1}^{n}\rho\left(\frac{e_i}{\hat{\sigma}MAD}\right) = \min\sum_{i=1}^{n}\rho\left(\frac{y_i\sum_{j=0}^{k}x_{ij}'\beta_j}{\hat{\sigma}MAD}\right) \tag{9}$$

with the estimated distance often formed from a linear combination of residuals, namely:

$$\hat{\sigma}MAD = \frac{median\,|e_i - median\,(e_i)|}{0{,}6745} \tag{10}$$

In M-estimation, the Ramsay weighting function is used and formulated as follows:

$$W(e) = \begin{cases} \exp(-c|e|), & |e| \leq \infty \\ 0, & |e| > \infty \end{cases} \tag{11}$$

With a tuning constant (c), which is a constant value associated with each Ramsay weight in the estimation process. The constant value (c) can influence the efficiency of robust regression techniques, and in Ramsay estimation, the value of c is set to 0.3 [30].

7. Testing the significance of model parameters using the F-test (Overall) and the T-test (Partial)
   i. F-test (Overall)
   The hypothesis for F-test can be formulated as follows:

   $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$
   H_1: at least $\beta_j \neq 0$ for $j = 0,1,\ldots,k$

   The decision is based on the value of F-statistic compared to the F-table. If the calculated F-value > critical F-value or if the p-value $< \alpha = 0.05$, then H_0 is rejected, indicating that at least one regression coefficient is not equal to zero.

   ii. T-test (Partial)
   The hypotheses used in this test are as follows:

   $H_0: \beta_j = 0$
   $H_1: \beta_j \neq 0$ for $j = 0,1,\ldots,k$

   The T-statistic is formulated as follows:

$$t_{statistics} = \frac{\hat{\beta}_j}{SE(\beta_j)} \tag{12}$$

   where $SE(\beta_j)$ is the standard error of β_j.
   The decision in the T-test is made by comparing the value $|t_{statistics}|$ with the critical t-value $t_{\left(1-\frac{\alpha}{2},n-k-1\right)}$ (where k is the number of parameters). If $|t_{statisics}| > t_{\left(1-\frac{\alpha}{2},n-k-1\right)}$ then reject H_0, indicating that there is an influence of the independent variable on the model.

8. Choosing the best estimation method is based on the values of the coefficient of determination (R^2) and Mean Squared Error (MSE).
   i. Coefficient of determination (R^2)
   R-squared allows for comparisons between multiple regression models. In some cases, comparing models with the highest R-squared can help choose the model that best fits the data [31]. Where the R-squared value can be calculated using the following equation [32]:

$$R^2 = \frac{\sum_{i=1}^{n}(\bar{Y}_i - Y)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{13}$$

   ii. Mean Squared Error (MSE)
   MSE provides an indication of how much error or deviation exists between the values predicted by a regression model and the actual values observed in the data. Models with lower MSE are considered better at estimating the true values. Here is the formula for calculating MSE (Mean Squared Error) [32]:

$$MSE = \frac{JKG}{n-k} \tag{14}$$

   and:

$$SSE = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{15}$$

   With SSE being the total sum of squares of the differences between the observed values and the predicted values in the regression

9. Model Interpretation

3. RESULTS AND DISCUSSION

3.1 General Overview of West Java Province's GRDP in 2021

Classification of data distribution is divided into 5 categories, namely very high, high, medium, low, and very low. The higher of GRDP value in an area is indicated by a darker color gradation.

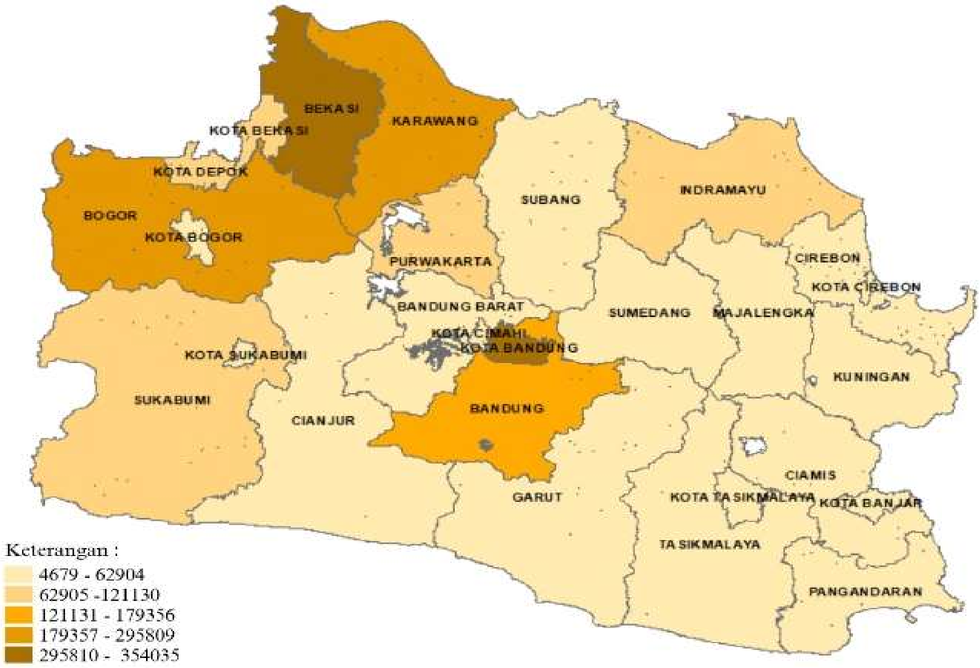

Figure 1. Gross Regional Domestic Product in West Java Province in 2021

Based **Figure 1** illustrates the distribution of Gross Regional Domestic Product (GDRP) contributions by Regency/City in West Java Province in 2021. Regencies/Cities with very high GDRP contributions, such as Bekasi Regency and Bandung City, are highlighted in dark brown color, indicating GDRP contributions ranging from 295,810 to 354,035 billion rupiahs. This is followed by surrounding areas, including Karawang Regency, Bogor Regency, and Bandung Regency, with GDRP contributions ranging from 121,131 to 179,356 billion rupiahs, marked in light brown color. Other Regencies/Cities highlighted in light orange color fall into the category of low GDRP contributions, ranging from 4,679 to 62,904 billion rupiahs.

3.2 Classical Assumption Test

1. Autocorrelation Test

Table 1. Durbin-Watson Test Results

dL	4-dL	DW	dU	4-dU	Decision
1.004	2.996	1.918	1.861	2.139	Accept H_0

Based on **Table 1**, the Durbin-Watson value is 1.918, which will be compared with the dL (1.004) and dU (1.861) values obtained from the Durbin-Watson table with $n = 27$ and $k = 5$. From **Table 1**, it is found that dU (1.861) < DW (1.918) < 4-dU (2.139) so it can be concluded that Ho is accepted, indicating no autocorrelation between residuals.

2. Heteroscedasticity Test

Table 2. Heteroscedasticity Test Result

	P-value	**Decision**
Heteroscedasticity	0.300 > 0.05	Accept H_0

Based on Table 2, the p-value (0.300) > 0.05. Therefore, it can be concluded that H_0 is accepted, indicating that there is no heteroscedasticity in the model.

3. Multicollinearity Test

Table 3. VIF Test Result

Variable	**VIF**
X_1	1.7832
X_2	2.4083
X_3	8.1935
X_4	1.1239
X_5	6.1567

From Table 3, it is known that all predictor variables (X_1, X_2, X_3, X_4, and X_5) have VIF values $<$ 10. It can be concluded that there is no multicollinearity in the model.

4. Normality Test

Table 4. Normality Test Result

	P-value	**Decision**
Shapiro-Wilk	0,02 < 0,05	Reject H_0

Based on Table 4, it is known that the p-value is 0.02, which is smaller than $\alpha = 0.05$. This means that H_0 is rejected, indicating that the data is not normally distributed.

3.3 Detecting Outlier

Based on the normality test results, it is known that the data is not normally distributed, indicating a potential presence of outliers. At this stage, outlier detection will be carried out using scatterplots visualization.

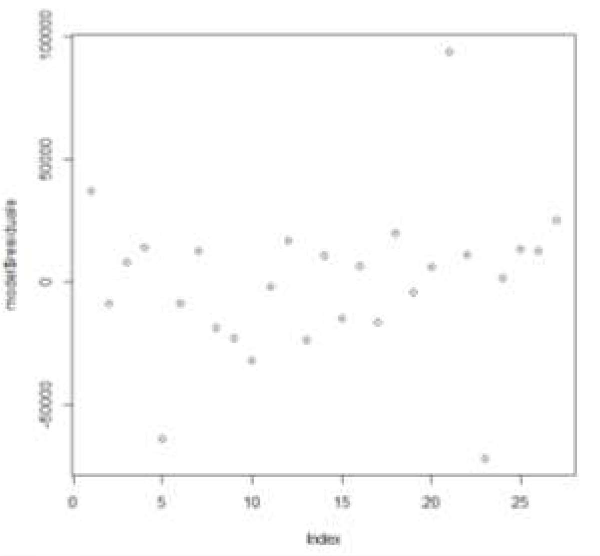

Figure 2. Scatter Plot of Residual Data

Figure 2 shows that there are several data points that are isolated or far from the spread of other data. This indicates that these data points are outliers. However, the data plot does not provide specific information about which data points are considered outliers. Therefore, another method is needed to detect the presence of outliers.

In addition to the scatterplots, outlier detection can be performed using the DfFITS method. A data point is considered an outlier if the cut-off value $|\text{DfFITS}| > 2\sqrt{\frac{p}{n}} = 2\sqrt{\frac{6}{37}} = 0.942$ where p represents all parameters, including the intercept.

Table 5. DfFITS

No	|DfFITS|		Decision
1	1.442		
4	1.628		
15	1.125	> 0.942	Outlier Data
21	5.861		
23	2.144		

Table 5 shows that there are several data points with |DfFITS| > 0.942, namely observations 1, 4, 15, 21, and 23. This indicates that these observations are outliers in both variables X and Y.

3.4 Spatial Heterogenity

Table 6. Spatial Heterogenity Result

Breusch-Pagan	Df	*P-Value*
20.085	5	0.001

Table 6 shows a p-value of 0.001, which is less than $\alpha = 0.05$. Therefore, it can be concluded that H_0 is rejected, indicating that there is a difference in variance for each location or the presence of spatial heterogeneity effects.

3.5 Robust Geographically Weighted Regression

1. Least Absolute Deviation (LAD) Estimation

The summary results of the RGWR model parameter estimates with the Least Absolute Deviation estimation obtained using the R software are presented in the following **Table 7**:

Table 7. Summary of RGWR Model Parameter with LAD Estimation

Parameter	Min	Mean	Median	Max
β_0	-0.0346	-0.0046	-0.0086	0.0702
β_1	0.3542	0.3812	0.3833	0.3925
β_2	0.0298	0.2092	0.2016	0.2598
β_3	0.3478	0.4724	0.4524	1.0426
β_4	0.0180	0.0272	0.0198	0.0373
β_5	-0.4501	-0.0726	-0.0341	-0.0036

Based on **Table 7**, it is known that the variables that have a positive impact on the response variable (GDRP) are foreign investment, domestic investment, local revenue, and the open unemployment rate. Meanwhile, the population variable has a negative impact on the response variable (GDRP). Below, we will display one of the models, namely the local model for Bekasi Regency:

$$\hat{Y}_i = 0.0214 + 0.3705X_{1i} + 0.2016X_{2i} + 0.4523X_{3i} + 0.0180X_{4i} - 0.0036X_{5i}$$

2. M-estimation

The summary results of the RGWR model parameter estimates with the M-estimation obtained using the R software are presented in the following **Table 8**:

Table 8. Summary of RGWR Model Parameter with M-estimation

Parameter	Min	Mean	Median	Max
β_0	-0.0082	-0.0039	-0.0041	0.0021
β_1	0.3886	0.4581	0.4578	0.5244
β_2	0.0262	0.0797	0.0797	0.1103
β_3	0.5094	0.6148	0.6107	0.8443
β_4	0.0575	0.0808	0.0814	0.0975
β_5	-0.2527	-0.0809	-0.0789	0.0098

Based on Table 8, it can be seen that the variables that have a positive impact on the response variable (GDRP) are almost the same as the LAD estimates, namely foreign investment, domestic investment, local revenue, and the open unemployment rate. Meanwhile, the population variable has a negative impact on the response variable (GDRP). Below, we will display one of the models, namely the model for Bekasi Regency:

$$\hat{Y}_i = -0.0045 + 0.4708X_{1i} + 0.0652X_{2i} + 0.6001X_{3i} + 0.0864X_{4i} - 0.0590X_{5i}$$

3.6 Parameter Significance Test

1. F-test

The F test is used to determine whether the predictor variables together have an effect on the response variable. Table 9 below is a summary of the results of the significance test of the RGWR model parameters with LAD estimation and M-estimation:

Table 9. F-test Result

	F-test		
	$\lvert F_{statistics} \rvert$	p-value	Decision
LAD Estimation	0.8956	0.02	Reject H_0
M-estimation	2.2712	0.05	Reject H_0

Based on Table 9, it is known that the p-value of LAD estimation and M-estimation is 0.02 and 0.05, respectively, which is $\leq \alpha$ (0.05), so H_0 is rejected. This means that the parameters generated are diverse or different in each location, so the RGWR model is appropriate and suitable for use.

2. T-test

The T test was carried out to determine whether the predictor variables significantly influenced the response variable. In Table 10, the results of the T test for the RGWR model with LAD estimation and M-estimation for locations in Bekasi Regency are presented:

Table 10. T-test Result

	Parameter	Estimated Value	T-test		
			$\lvert t_{hitung} \rvert$	t_{table}	Decision
Estimation LAD	Intercept	0.0214	3.0724	2.0518	**Reject H_0**
	β_1	0.3705	7.6873		**Reject H_0**
	β_2	0.2016	3.5930		**Reject H_0**
	β_3	0.4523	4.3700		**Reject H_0**
	β_4	0.0180	0.4675		Accept H_0
	β_5	-0.0036	0.0409		Accept H_0
Estimation M	Intercept	-0.0045	-0.3416	2.0893	**Reject H_0**
	β_1	0.4708	4.3788		**Reject H_0**
	β_2	0.0652	0.5465		Accept H_0
	β_3	0.6001	3.0947		**Reject H_0**
	β_4	0.0864	1.1752		Accept H_0
	β_5	-0.0590	-0.3477		Accept H_0

Based on Table 10 in the RGWR model with LAD estimation, it can be seen that variables X_1, X_2, and X_3 have values of $|t_{statistics}| > t_{table}$ (2.0518), so it can be concluded that H_0 is rejected, meaning that the variables of foreign investment (X_1), domestic investment (X_2), and regional original income (X_3) significantly influence the variable GDRP (Y). On the other hand, the variables of open unemployment rate (X_4) and population (X_5) are not significant because they have values of $|t_{statistics}| > t_{table}$ (2.0518).

In the RGWR model with M-estimation, it can be seen that only variables X_1 and X_3 have values of $|t_{statistics}| > t_{table}$ (2.0893). Thus, H_0 is rejected, indicating that the variables of foreign investment (X_1) and regional original income (X_3) significantly influence the variable GDRP (Y). On the other hand, the variables of domestic investment (X_2), open unemployment rate (X_4), and population (X_5) are not significant because they have values of $|t_{statistics}| < t_{table}$ (2.0893).

3.7 Selection of Best Model

Compare the R-squared value and the Mean Square Error (MSE) value to determine the best estimation method. A R-squared value close to 1 means that the regression model is able to explain most of the variation in the model. Meanwhile, a lower MSE value indicates that the regression model has better implications, because it produces model predictions that match the values in the actual data. Table 11 below is a comparison of R-squared and MSE in the RGWR model with LAD estimates and M estimates:

Table 11. Comparison of R^2 and MSE Values

Parameter	Estimate LAD	Estimate M
R^2	0,9271	**0,9285**
MSE	0,8581	**0,0044**

Based on Table 11, it can be seen that the M-estimation produces the highest R-squared value, which is 0.9285. Meanwhile, the smallest MSE value is also produced by the M-estimation, which is 0.0044. Therefore, it can be concluded that the use of the RGWR method with M-estimation is the most optimal and effective robust estimation approach in estimating regression parameters for Gross Regional Domestic Product data in Regencies/Cities in West Java Province in 2021.

The results of the model validation process using the T-test (partial) indicate the formulation of the Robust Geographically Weighted Regression model with M-estimation as follows:

$$\hat{Y}_i = -0,0045 + 0,4708X_{1i} + 0,6001X_{3i}$$

The model above has an R-squared value of 0.9285, indicating that 92.85% of the variation in the response variable can be explained by the predictor variables, while the remaining is influenced by other factors outside the study. With a Mean Squared Error (MSE) value of 0.0044, it means that the formed model is very good.

4. CONCLUSIONS

The most effective and optimal method for handling outliers in the Gross Regional Domestic Product (GDRP) data for Districts/Cities in West Java Province in 2021 is the Robust Geographically Weighted Regression Model with M-estimation. This model produces the highest R-squared value and the lowest MSE compared to Least Absolute Deviation (LAD) Estimation. Based on the formed model, the significant factors influencing the Gross Regional Domestic Product (GDRP) in the Districts/Cities of West Java Province in 2021 are the variables of foreign direct investment (X_1) and local income (X_3). Future researchers are expected to be able to develop M-estimation methods with other weight such as Huber weight, Hampel weight and Tukey weight

REFERENCES

[1] R. P. Gautama, I. Fitriani, F. Naza, and P. R. Arum, “Identifying Factors that Influence Life Expectancy in Central Java Using Spatial Regression Models,” vol. 16, no. 2, pp. 605–613, 2023.

[2] M. L. Jhingan, “Ekonomi Pembangunan dan Perencanaan, terjemahan D. Guritno.” Jakarta: PT. Raja Grafindo Persada, 2003.

[3] R. Okriatama, “Penerapan Analisis Error Correction Model Dalam Menentukan Faktor-Faktor Yang Mempengaruhi Pdrb Provinsi Dki Jakarta,” 2021.

[4] K. N. Annisa, I. M. Nur, and P. R. Arum, “Pemodelan Spatial Seemingly Unrelated Regression (S-SUR) Pada Produk Domestik Regional Bruto Sektor Unggulan Di Indonesia Spatial Seemingly Unrelated Regression (S-SUR) Modeling on Gross Regional Domestic Product of Superior Sectors in Indonesia Unive,” vol. 3, pp. 293–301, 2020.

[5] S. Alfiani and P. R. Arum, “Pemodelan Pertumbuhan Ekonomi di Jawa Barat Menggunakan Metode Geographically Weighted Panel Regression Universitas Muhammadiyah Semarang,” vol. 15, no. 2, pp. 219–227, 2022.

[6] BPS, *Produk Domestik Regional Bruto Kabupaten/Kota di Indonesia*. BPS, 2022.

[7] P. J. Rousseeuw and A. M. Leroy, “Robust Regression and Outlier Detection. New York: John Wiley& Sons.” Inc, 1987.

[8] E. Oktaviana, P. R. Arum, and M. Al Haris, “Pemodelan Spatial Autoregressive Quantile Regression (SARQR) Menggunakan Pembobot Queen Contiguity Pada Kasus Stunting Balita di Indonesia Spatial Autoregressive Quantile Regression (SARQR) Modeling Using Queen Contiguity Weights in Toddler Stunting C,” vol. 6, pp. 411–425, 2023.

[9] Y. L. Sukestiyarno and A. Agoestanto, “Batasan prasyarat uji normalitas dan uji homogenitas pada model regresi linear,” *Unnes Journal of Mathematics*, vol. 6, no. 2, pp. 168–177, 2017.

[10] R. S. Wirawan, “Analisis Faktor-Faktor yang Mempengaruhi PDRB di Provinsi Jawa Barat Tahun 2015-2020,” 2021.

[11] D. Permata Sari, “Robust Geographycally Weighted Rergression dengan Metode Least Absolute Deviation Pada Kasus Penyebaran Covid-19 di Indonesia.” 2021.

[12] N. R. Draper, H. Smith, and B. Sumantri, *Analisis regresi terapan*. PT Gramedia Pustaka Utama, 1992.

[13] Febawanti, “Pemodelan Robust Geographycally Weighted Rergression (RGWR) Pada Data yang Mengandung Pencilan (Studi Kasus Indeks Pembangunan Manusia Provinsi Jawa Timur Tahun 2015),” 2017.

[14] N. Herawati, K. Nisa, and E. Setiawan, “Analisis Ketegaran Regresi Robust Terhadap Letak Pencilan: Studi Perbandingan,” *Bulletin of Mathematics*, vol. 3, no. 01, pp. 49–60, 2011.

[15] C. Chen, “Robust Regression and Outlier Detection with the Robustreg Procedure,” in *Proceedings of the Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference*, 2002.

[16] A. S. Rahman, “Estimasi Model Robust Geographically Weighted Regression dengan Metode Least Absolute Deviation.” Universitas Hasanuddin, 2021.

[17] M. T. S. Puspitasari, Y. Susanti, and S. S. Handajani, “Model Regresi Robust untuk Indeks Pembangunan Manusia di Jawa Timur dengan Estimasi M,” in *PRISMA, Prosiding Seminar Nasional Matematika*, 2021, pp. 659–665.

[18] F. T. Wang and D. W. Scott, “The L 1 method for robust nonparametric regression,” *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 65–76, 1994.

[19] F. Astuti, R. D. Bekti, A. A. Br Keliat, and T. I. Sebo, “Analisis Produksi Padi dI Indonesia Menggunakan Model Regresi Robust Estimasi M, Estimasi S dan Estimasi MM,” *Jurnal Teknologi Technoscientia*, vol. 16, no. 1, pp. 33–40, Sep. 2023, doi: 10.34151/technoscientia.v16i1.4513.

[20] D. N. Gujarati and D. C. Porter, “Basic econometrics (ed.),” *Singapore: McGrew Hill Book Co*, 2003.

[21] I. Ghozali, “Aplikasi Analisis Multivariate dengan program SPSS . semarang: undip,” 2009.

[22] M. S. Dahlan, *Regresi Linier Disertai Praktik dengan SPSS*. Jakarta: Epidemiologi Indonesia, 2012.

[23] Soemartini, “Pencilan (Outlier),” *Universitas Padjadjaran ,Jatinangor*, 2007.

[24] C. A. Ayuwida, P. R. Arum, and M. Al Haris, “Model Seemingly Unrelated Regression Pada Data Kemiskinan Jawa Timur Menggunakan Matriks Pembobot Queen Contiguity Dan Rook Contiguity,” vol. 9, no. 1, pp. 64–68, 2021.

[25] L. Anselin, *Spatial econometrics: methods and models*, vol. 4. Springer Science & Business Media, 1988.

[26] D. L. Sogen, P. R. Arum, and R. Wasono, “Pemodelan Geographically Weighted Poisson Regression (GWPR) Dengan Pembobot Fixed Bisquare Kernel Dan Adaptive Bisquare Kernel Pada,” vol. 6, pp. 371–387, 2023.

[27] D. Birkes and Y. Dodge, *Alternative methods of regression*. John Wiley & Sons, 2011.

[28] Y. Leung, C.-L. Mei, and W.-X. Zhang, “Statistical tests for spatial nonstationarity based on the geographically weighted regression model,” *Environment and Planning A*, vol. 32, no. 1, pp. 9–32, 2000.

[29] J. Fox and S. Weisberg, *An R companion to applied regression*. Sage publications, 2011.

[30] D. C. Montgomery, E. A. Peck, and G. G. Vining, “Introduction to Linear Analysis.” Tinker, NA, DE Mather, TK Blake, KG Briggs, TM Choo, 1982.

[31] A. O. Winarta, P. R. Arum, I. M. Nur, and E. A. Purnomo, *Modelling Gender-Based Human Development Index in Indonesia Using Spatial Seemingly Unrelated Regression (S-SUR)*. Atlantis Press SARL, 2023. doi: 10.2991/978-2-38476-176-0.

[32] R. K. Sembiring, “Analisis Regresi Edisi Ke-2.” Proyek Pengembangan Staf dan Sarana Perguruan Tinggi. Bandung: ITB, 2003.