

ROBUST SPATIAL REGRESSION MODEL ON ORIGINAL LOCAL GOVERNMENT REVENUE IN JAVA 2017 *

Winda Chairani Mastuti¹, Anik Djuraidah^{2‡}, and Erfiani³

¹Department of Statistics, IPB University, windachairani1@gmail.com

² Department of Statistics, IPB University, anikdjuraidah@apps.ipb.ac.id

³ Department of Statistics, IPB University, erfiani@apps.ipb.ac.id

‡corresponding author

Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)

Vol 4 No 1 (2019), 68 - 79

Copyright © 2019 Winda Chairani Mastuti, Anik Djuraidah, and Erfiani. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium provided the original work is properly cited.

Abstract

Spatial regression measures the relationship between response and explanatory variables in the regression model considering spatial effects. Detecting and accommodating outliers is an important step in the regression analysis. Several methods can detect outliers in spatial regression. One of these methods is generating a score test statistics to identify outliers in the spatial autoregressive (SAR) model. This research applies a robust spatial autoregressive (RSAR) model with S- estimator to the Original Local Government Revenue (OLGR) data. The RSAR model with the 4-nearest neighbor weighting matrix is the best model produced in this study. The coefficient of the RSAR model gives a more relevant result. Median absolute deviation (MdAD) and median absolute percentage error (MdAPE) values in the RSAR model with 4-nearest neighbor give smaller results than the SAR model.

Keywords: *outlier, mean shift, OLGR, robust spatial regression, SAR, score test.*

* Received Oct 2019; Accepted Nov 2019; Published online on Feb 2020

1. Introduction

Spatial regression is an analysis of the statistical process that measures the relationship between response and explanatory variables in the regression model considering spatial influence. According to Anselin (1988), spatial influence between locations can be caused by spatial dependencies and spatial heterogeneity. Spatial dependence is the dependency between an area and its neighbour. Spatial dependencies indicate that the area of their neighbour's influence causes the subject is plural. There are some models in spatial regression, Spatial Autoregressive (SAR), which has spatial dependencies on the response variable was the focus of this study.

According to Choi (2009), outliers in the regression analysis affect parameter estimators. Outliers produce a large variety of residuals. Outliers also affect the testing of parameter hypotheses. Detecting and accommodating outlier is important because regression with outliers will give wrong conclusions. Haining (1993) studied that spatial outliers are observations different from others by considering the location element. Outliers can be seen on a plot between the response variable with the weighted response. Shekhar et al. (2003) explain outliers as a spatial object with variable values significantly different from their neighbour.

There are several methods to detect spatial outliers. Jin et al. (2016) propose score test statistics to detect spatial outlier. Score test statistics derived from the mean shift outlier model (MS) and variance-weight. In this study, the outlier detection method based on score test statistics on the MS model modified from the SAR model with adding a shifting parameter. This method detects outliers following the spatial regression model obtained. Score test statistics also contain spatial parameter from the model and also spatial weight.

Handling outliers in spatial regression can be approached by several methods. Some methods that can be used are M-estimator, least median square estimator, least trimmed squares estimator, method of moment, and S-estimator. According to Harini et al. (2019) among the several methods, the S-estimator has the highest breakdown point value reaching 50%. Tharmaratnam & Claeskens (2013) studied S estimator method can overcome outlier data on the response variables and explanatory variables. Susanti et al. (2014) compared robust regression for corn production data in Indonesia, and the S-estimator is the best estimator because it has a minimum AIC value. In this research, S-estimator will be used in SAR model and now we can say it, robust spatial autoregressive model (RSAR). The RSAR model will be applied to Original Local Government Revenue (OLGR) Java in 2017.

Original Local Government Revenue (OLGR) is one important indicator to determine the condition of regional improvement. The OLGR can evaluate regional growth based on regional autonomy. Java, which is the centre of the economic and industrial, has a higher income than the other islands. Java also has the most populous compared to other islands. In 2018, five provinces in Java are the top five highest OLGR groups in Indonesia, according to the Central Bureau of Statistics (BPS). Jakarta Province dominates Java regions by having the highest OLGR. OLGR conditions in Java are also diverse which allow for several values that have different values from other regions.

This study aims to compare the SAR and RSAR model. The outlier in SAR model will be detected using score test method. The goodness of fit is compared with

calculating the median absolute deviation (MdAD) and Median Absolute Percentage Error (MdAPE).

2. Methodology

2.1 Spatial Regression

General regression models that have spatial dependency effects can be written as follows (Elhorst, 2014):

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}\mathbf{y} + \boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \mathbf{u} \\ \mathbf{u} &= \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \end{aligned} \quad (1)$$

with $|\lambda|$ and $|\rho| < 1$ assuming $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$, \mathbf{y} is the response vector, \mathbf{X} is the explanatory variable matrix. $\boldsymbol{\beta}, \rho, \lambda, \boldsymbol{\theta}$ is the parameter to be predicted, \mathbf{W} is the spatial weighting matrix. \mathbf{u} is the error vector assumed to contain autocorrelation. $\boldsymbol{\varepsilon}$ is an error vector that is no autocorrelated, ρ is the spatial lag autoregression coefficient, λ is spatial error autoregressive coefficient, $\boldsymbol{\theta}$ is a spatial Durbin coefficient, and \mathbf{I} is the identity matrix. In general model, $\mathbf{W}\mathbf{y}$ denotes the interaction effect between the response variables, $\mathbf{W}\mathbf{X}$ symbolises the interaction effect between explanatory variables, and $\mathbf{W}\mathbf{u}$ expresses the interaction effect between the error.

There are several models used in spatial regression, that is:

- i) If $\rho \neq 0$ and $\lambda = 0$, the model becomes the Spatial Autoregressive (SAR).
- ii) If $\lambda \neq 0$ and $\rho = 0$, the model becomes a Spatial Error Model (SEM).
- iii) If $\rho \neq 0$ and $\lambda \neq 0$, then the model will become a Generalized Spatial Model (GSM).

2.2 Mean Shift Outlier Model

The score test statistic (ST) calculation is similar to the likelihood ratio test in hypothesis testing. ST values are calculated to test outliers based on MS model. Modifications to the MS model for the SAR model are as follows:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \boldsymbol{\beta}\mathbf{X} + d_i \boldsymbol{\gamma} + \boldsymbol{\varepsilon} \quad (2)$$

where $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$, $\boldsymbol{\gamma}$ is a parameter to detect outliers. The model is directed to test outliers in the observation. Outliers testing is carried out with the hypothesis as follows:

$$H_0 : \boldsymbol{\gamma} = 0 \text{ dan } H_1 : \boldsymbol{\gamma} \neq 0$$

The statistical test for the hypothesis is the score test (SC) as follows:

$$\begin{aligned} SC_i &= \left(\frac{\hat{e}_i}{\hat{\sigma} \sqrt{1 - p_{ii}}} \right)^2 \left(\frac{q_{ii}}{q_{ii} - \hat{b}_i^2 / \hat{c}} \right) \\ &= \hat{t}_i^2 \left(\frac{q_{ii}}{q_{ii} - \hat{b}_i^2 / \hat{c}} \right) \end{aligned} \quad (3)$$

with \hat{e}_i is i -th element of the vector $\hat{e} = \mathbf{y} - \rho \mathbf{W} \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}$, \hat{b}_i is the i -th element of the vector $\hat{\mathbf{b}} = \mathbf{Q} \hat{\mathbf{V}} \mathbf{X} \hat{\boldsymbol{\beta}}$, with $\mathbf{Q} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\hat{\mathbf{V}} = \mathbf{W}(\mathbf{I} - \rho \mathbf{W})^{-1}$, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \rho \mathbf{W})\mathbf{y}$. $q_{ii} = 1 - p_{ii}$ and p_{ii} are diagonal elements of the matrix $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. $\hat{c} = \hat{\sigma}^2 \left(\text{tr}(\hat{\mathbf{V}}\hat{\mathbf{V}}' + \hat{\mathbf{V}}^2) - 2 \text{tr}^2 \left(\frac{\hat{\mathbf{V}}}{n} \right) \right) + \hat{\mathbf{b}}'\hat{\mathbf{b}}$ where $\hat{\sigma}^2$ estimate based on the likelihood function. So $\hat{\sigma}^2 = \frac{1}{n}(\hat{\mathbf{e}}'\hat{\mathbf{e}})$. Score test, SC_i , will follow the distribution $\chi^2_{(1)}$. An observation is detected outlier if the value of SC_i is greater than $\chi^2_{(1)}$.

2.3 S-estimator

Rousseeuw & Yohai (1984) proposed the parameter estimator with the S-estimator. The S-estimator is based on the residual scale. The S-estimator method corrects weaknesses from the M-estimator. The S-estimator method uses residual standard deviations to overcome the w(Salibian-Barrera & Yohai, 2006) weaknesses of the median method. Salibian-Barrera & Yohai (2006) studied that S-estimator is defined by $\hat{\beta}_s = \min_{\beta} \hat{\sigma}_s(e_1, e_2, \dots, e_n)$. It minimize the estimator scale C with:

$$\min \sum_{i=1}^n \rho \left(\frac{y_i - \sum_{j=1}^n x_{ij} \beta_j}{\hat{\sigma}_s} \right)$$

with

$$\hat{\sigma}_s = \sqrt{\frac{1}{nK} \sum_{i=1}^n \theta_i e_i^2} \tag{4}$$

$K = 0.199$, $\theta_i = \theta_{\sigma}(u_i) = \frac{\rho'(u_i)}{u_i^2}$ and ρ are objective functions that are used to find the weight in robust regression. One objective function ρ used is Tukey's biweight, namely:

$$\rho(u_i) = \begin{cases} \frac{u_i^2}{2} - \frac{u_i^2}{2c^2} - \frac{u_i^2}{6c^4}, & |u_i| \leq c \\ \frac{c^2}{6}, & |u_i| > c \end{cases} \tag{5}$$

where $u_i = \frac{e_i}{\hat{\sigma}_s}$ and $c = 1.547$. The initial $\hat{\sigma}$ for iteration is:

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

the $\hat{\beta}$ estimation can be solved by finding the derivative of β_j with $j = 0, 1, 2, \dots, k$:

$$\sum_{i=1}^n x_{ij} \psi \left(\frac{y_i - \sum_{i=1}^n x_{ij} \beta_j}{\hat{\sigma}_s} \right) = 0$$

with ψ derived from uncton ρ as follows:

$$\psi(u_i) = \rho'(u_i) = \begin{cases} u_i \left[1 - \left(\frac{u_i}{c} \right)^2 \right]^2, & |u_i| \leq c \\ 0, & |u_i| > c \end{cases} \quad (6)$$

Furthermore, the S-estimator function is obtained based on $\theta_i = \theta_\sigma(u_i) = \frac{\rho'(u_i)}{u_i^2}$ so that $\theta_i(u_i)$ is an iteratively reweighted least-squares (IRLS) weight function as follows:

$$\theta_i(u_i) = \begin{cases} \left[1 - \left(\frac{u_i}{c} \right)^2 \right]^2, & |u_i| \leq c \\ 0, & |u_i| > c \end{cases} \quad (7)$$

The weighting matrix θ is a symmetrical matrix of size n based on the weighting function θ_i in equation 7. The matrix element θ is $\theta_{ij} = \theta_i(u_i)$ for $i = j$ dan 0 for the other. This weighting matrix is different from the spatial weighting matrix W . The values in the matrix θ indicate that the i -th observation that matches $k = \theta_i$ is an outlier or not, so a large θ_i value indicates that the observation is an outlier observation. The weighting matrix θ will be entered into the spatial regression equation together with the weighting matrix W . So the definition of u_i is adjusted based on the spatial regression model used.

2.4 Data

This study uses OLGR of regencies/cities in Java 2017. Data were obtained based on BPS publications in 2017. Table 1 presents the variables used in this study.

2.5 Data Analysis Procedure

Data analysis uses software R 3.4.3. The steps of analysis are as follows:

1. Exploring data using descriptive statistical analysis on Java OLGR data.
2. Calculating VIF values to see multicollinearity on explanatory variables.
3. Calculating the spatial weighting matrix (W) with the queen contiguity matrix, k-nearest neighbour, inverse distance weight, and exponential weight.
4. Test the spatial association with the Moran index.
5. Test the effect of spatial heterogeneity using the Breusch-Pagan test.
6. Test the effects of spatial dependencies with the Lagrange Multiplier (LM) test.
7. Test outlier data following score test from MS model.
8. Modelling the robust spatial regression (RSAR) model with S-estimator.
9. Comparing the SAR and RSAR models with MdAD and MdAPE.

Table 1: List of variables in the study.

Variable	Description
Y	Original Local Government Revenue (OLGR) (Billion)
X_1	Total population (Million)
X_2	General Allocation Fund (GAF) (Billion)
X_3	Gross Regional Domestic Product (GRDP) (Billion)
X_4	Number of restaurants (units)
X_5	Jumlah hotel (units)

3. Results

3.1 Data Exploration

There are six provinces in Java Island. Five of them have the highest original local government revenue in Indonesia. The distribution of OLGR is described by grouping it into four categories: low, medium, high, and very high. Figure 1 shows the regions have the highest amount of OLGR that is more than Rp. 469 billion, there are South Jakarta, Central Jakarta, Surabaya City, and others. Regions with the lowest OLGR are Banyuwangi, Pandeglang, Cianjur, Garut, and other regions that have OLGR less than Rp. 267 billion. Area and its neighbour mostly have similar OLGR. The similarity of colours in nearby regions can also show that there are spatial autocorrelations between neighbouring areas.

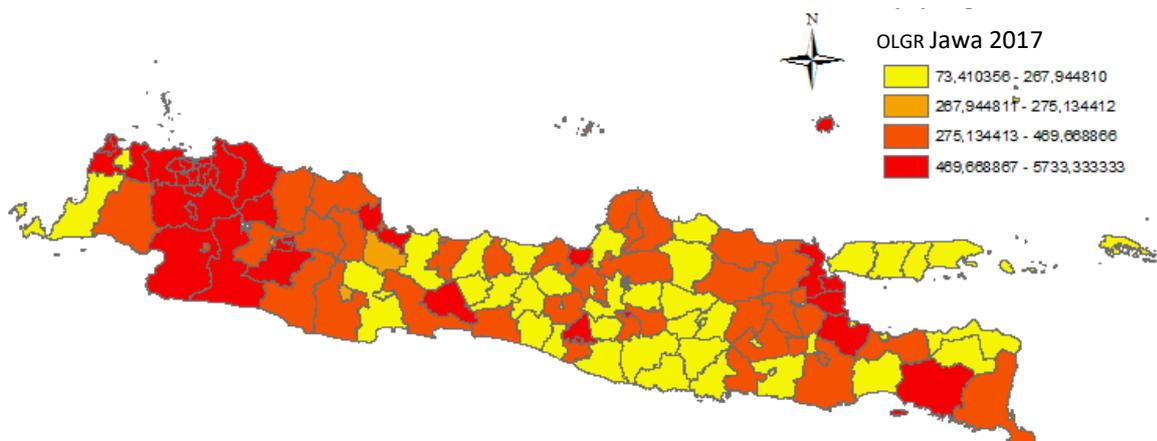


Figure 1: Java OLGR Distribution Map for 2017 (no scale map).

Figure 2 shows a Java OLRG boxplot. Based on the boxplot, we can see that the distribution of data right skewness. The shape of the boxplot is not symmetrical and has many outliers. This outlier has a value far from the other values. Boxplot is one of the explorative ways to detect outliers on the response variable. Regions that are outliers based on this boxplot are West Jakarta, South Jakarta, Bandung City, Bogor Regency, and some other regions. Areas with high OLRG are areas with industry and also big cities that get income from buildings and taxes.

Multicollinearity of explanatory variables calculates from the VIF value. The VIF value in this study shows in Table 2. All VIF values less than 10. These results indicate that there is no multicollinearity in the explanatory variables.

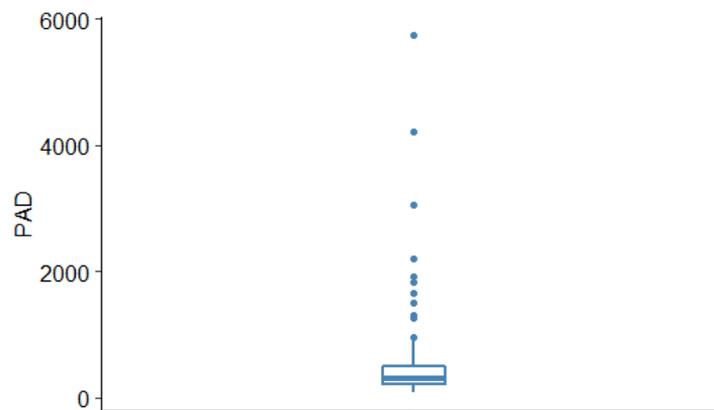


Figure 2: Boxplot Java OLRG 2017.

Table 2: VIF values of explanatory variables.

Variable	VIF
X_1	2.61
X_2	2.26
X_3	2.58
X_4	2.26
X_5	1.09

Moran index shows spatial autocorrelation in data. The Moran index test results show in Table 3. Based on the test results in Table 3, the Moran index is significant at the 5% level for all spatial weight matrix. This result shows that there is positive Moran index value ($I > 0$) in the Java OLRG data 2017. These results indicate that the OLRG value in the data has a positive spatial autocorrelation. These results also indicate that neighbouring regions have similar OLRG values (in groups).

3.2 Spatial Regression

There are two kinds of testing spatial regression effect, spatial heterogeneity and spatial dependencies. In this study, Breusch Pagan (BP) test is performed to see

spatial heterogeneity. The resulting BP value is 8.42, with a p-value of 0.13. The p-value is higher than the significant level $\alpha = 5\%$, so this data has homogeneous variance. Spatial regression models can be used on Java OLGR 2017.

Table 3: Moran index test results.

Weighting Matrix	Moran index
Queen Contiguity	0.59 **
Exponential	0.25 **
4-nearest neighbor	0.74 **
Invers Distance	0.33 **

Note: ** significant at 5% level

We use the LM test to select a spatial regression model. LM test is used to test the spatial dependencies on a model that in SAR, SEM, or GSM. A summary of the LM test results can be seen in Table 4. The test results for SAR, SEM, and GSM are significant at 5% significance level for exponential weight matrix, k-nearest neighbour, and inverse distance weight. SAR, SEM, and GSM models are not significant in queen contiguity matrix. The results of this test indicate that the model to be used in this study is the autoregressive spatial model (SAR).

Table 4: p-value Lagrange Multiplier (LM) Test.

Test Statistics	Weighting Matrix			
	Queen Contiguity	Exponential	k-nearest neighbour	Inverse distance
LMerr	0.94	0.80	0.41	0.94
LMLag	0.30	0.01**	0.01 **	0.002**
RLMerr	0.38	0.27	0.23	0.14
RLMLag	0.17	0.01**	0.001 **	6.7×10^{-4} **
SARMA (GSM)	0.39	0.01**	0.002 **	3×10^{-3} **

Note: ** significant at 5% level

Coefficient of the spatial regression parameters for the SAR model shown in Table 5. The results of the SAR model show that the GAF variable is significant and has a negative effect on the model with four types of weights used. The number of hotels does not have a significant effect on the four types of weights used, but if seen from the sign on the coefficient has a negative sign on the queen contiguity weights and exponential weights. Other variables have a positive influence on OLGR and significant at 5% level. Purwaningsih (2011) has studied that GAF has a positive influence on OLGR and the population have negative effect on OLGR because GAF is a component in OLGR, and the population can decrease OLGR in the area. The results of the parameter estimator for the SAR model don't follow existing research.

One of the problems in regression is outlier which can change result in parameter estimation. Detection of spatial outliers in this study uses ST as a statistical test based on the MS model. The area detected as an outlier is an area with very high OLGR. This region with very high OLGR also has neighbours with low OLGR values. Areas detected as outliers for several types of weights used are presented in Table 6.

Table 5: SAR model parameters estimator.

Parameters	Weighting Matrix			
	Queen Contiguity	Exponential	4-nearest neighbour	Inverse Distance
ρ	0.12	0.28 **	0.30 **	0.44 **
β_0	805.97 **	654.04 **	433.33**	453.61**
β_1	0.46**	0.40 **	0.28 **	0.34**
β_2	-1.27**	-1.22 **	-0.79**	-1.03**
β_3	6.21**	6.62 **	5.66**	6.18**
β_4	1.12**	1.15 **	1.09**	1.09**
β_5	-0.17	-1.60×10^{-2}	1.70×10^{-2}	0.06

Note: ** significant at 5% level

Table 6: Outliers based on several weights.

Weighting Matrix	Outlier Area
Queen Contiguity	West Jakarta, South Jakarta, Jakarta, Central Jakarta, East Jakarta, North Jakarta, Thousand Island, Bogor, Surabaya
Exponential	West Jakarta, South Jakarta, Jakarta, Central Jakarta, Bogor, Surabaya, East Jakarta, North Jakarta, Thousand Island
4-nearest neighbor	West Jakarta, South Jakarta, Jakarta, Central Jakarta, Bogor, Surabaya, East Jakarta, North Jakarta, Thousand Island Bandung City, Depok
Invers Distance	West Jakarta, South Jakarta, Jakarta, Central Jakarta, Bogor, Surabaya, East Jakarta, North Jakarta, Thousand Island, Bandung City, Depok

3.3 Robust Spatial Autoregressive Model

Robust spatial autoregressive modelling begins with weighting using the S-estimator. The value of e_i is calculated first, and then the weight is calculated for the data. Table 7 presents parameter estimators for the RSAR model. The results of the RSAR model parameter estimators differ when compared to the SAR estimator in Table 5. Changes

do not only differ in the coefficient values for the parameter estimators. The coefficient also changes from positive to negative or vice versa. The coefficients of the RSAR model have more appropriate signs based on existing research. Tests at 5% significance level showed the results that the population variable had significant and negative effects on all types of weights. These results are following Purwaningsih (2011) research that the more the number of population, government spending will increase and will decrease OLGR. The number of hotels is significant and has a positive effect except for queen contiguity weight. This result is also suitable because hotel tax will be one component in the OLGR. Other variables are significant and have a positive effect on OLGR Jawa in 2017.

Table 7: RSAR model parameters estimator.

Parameter	Weighting Matrix			
	Queen Contiguity	Exponential	4-nearest neighbour	Inverse Distance
ρ	0.82**	0.87**	0.78**	0.9**
β_0	-105.27**	-290.17**	-133.01**	-361.84**
β_1	-0.02**	-0.03	-0.02**	-0.04**
β_2	0.13**	0.17**	0.14**	0.10**
β_3	1.01**	0.49**	0.97**	1.68**
β_4	0.01**	0.08**	0.26**	2.6×10^{-3} **
β_5	-0.17**	0.79**	0.38**	0.89**

Note: ** significant at 5% level

3.4 The Goodness of fit MdAD and MdAPE

The goodness of fit model is compared using the MdAD value and MdAPE value. The model has better estimator when the MdAD and MdAPE values are smaller. A comparison of the two values is presented in Table 8. The results in Table 8 shows that the RSAR model with the 4-nearest neighbour weight matrix gives better results than SAR model because MdAD and MdAPE are smaller than SAR model.

4 Conclusion

Detection of outliers using the MS model is more accurate because it contains spatial parameters. This test is only for SAR model. The robust spatial autoregressive model (RSAR) can be used to modelling data that contain outliers. The application of Java OLGR data 2017 shows that the RSAR model with the 4-nearest neighbour produces a model with a better accuracy value, MdAD value is 87.9 and MdAPE value is 27.94. Handling outliers and selecting the weighting matrix is important for the results of parameter estimation in the spatial regression model.

Table 8: Goodness of Fit of Model.

Weighting Matrix	Goodness of fit			
	MdAD		MdAPE	
	SAR	RSAR	SAR	RSAR
Queen Contiguity	190.74	132.10	43.08	51.36
Exponential	190.73	135.59	43.08	53.81
4-nearest neighbor	177.63	87.97	47.88	27.94
Inverse Distance	197.31	653.83	52.66	217.71

References

- Anselin, L. (1988). *Spatial econometrics: methods and models*. Dordrecht (NL): Kluwer Academic Publisher.
- Choi, S.-W. (2009). The effect of outliers on regression analysis: regime type and foreign direct investment. *Quarterly Journal of Political Science*, 4(2): 153–165.
- Elhorst, J. P. (2014). *Spatial econometrics: from cross-sectional data to spatial panels* (Vol. 479). New York (US): Springer.
- Haining, R. (1993). *Spatial data analysis in the social and environmental sciences*. Cambridge University Press.
- Harini, S., Sheppy, S., Sari, M. S. N., & Purhadi. (2019). Parameter estimate for spatial lag regression model with outlier. *International Journal of Engineering & Technology*, 8(1.9): 114–116.
- Jin, L., Dai, X., Shi, A., & Shi, L. (2016). Detection of outliers in mixed regressive-spatial autoregressive models. *Communications in Statistics-Theory and Methods*, 45(17): 5179–5192.
- Purwaningsih, E. (2011). *Analisis faktor yang mempengaruhi pendapatan asli daerah di Kabupaten Sragen 1999-2008* [tesis]. Surakarta(ID): Universitas Sebelas Maret.
- Rousseeuw, P., & Yohai, V. (1984). Robust regression by means of S-estimators. In *Robust and nonlinear time series analysis* (pp. 256–272). <https://doi.org/10.1007/978-1-4615-7821-5-15>.
- Salibian-Barrera, M., & Yohai, V. J. (2006). A fast algorithm for S-regression estimates. *Journal of Computational and Graphical Statistics*, 15(2): 414–427.
- Shekhar, S., Lu, C.-T., & Zhang, P. (2003). A unified approach to detecting spatial outliers. *GeoInformatica*, 7(2): 139–166.

Susanti, Y., Pratiwi, H., & others. (2014). M estimation, S estimation, and MM estimation in robust regression. *International Journal of Pure and Applied Mathematics*, 91(3): 349–360.

Tharmaratnam, K., & Claeskens, G. (2013). A comparison of robust versions of the AIC based on M-, S-and MM-estimators. *Statistics*, 47(1): 216–235.