International Journal of Engineering & Technology, 8 (1.9) (2019) 114-116

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Parameter Estimate for Spatial Lag Regression Model with Outlier

[1]Sri Harini[*], [2]Siscaviyana Sheppy, [3]Marita Siti Nurmala Sari, [4]Purhadi

1Lecturer of Mathematics Department, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang
2Student of Mathematics Department, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang
3Graduate Student of Statistics Department, Faculty of Mathematics and Natural Sciences, Universitras Brawijaya
4Lecturer of Statistics Department, Institut Teknologi Sepuluh Nopember Surabaya
**Corresponding author E-mail: sriharini@mat.uin-malang.ac.id*

Abstract

Spatial lag regression is the result of linear regression model development that considers the effect of spatial data to the dependent variable, the Spatial Autoregressive Model. In the model of spatial lag regression model, it is often found there is an outlier that affects the created model. One of the methods to detect the outlier from the spatial regression model is by using the S estimator model. The S estimator method is a method that is used to determine the outlier by minimizing the objective function and the function of the number of error square. The result of the study shows the interpreting parameter $\hat{\beta}^{m+1} = (X^T {\omega_i}^m X)^{-1} X^T {\omega_i}^m (I - \rho W_1) y$ that bears biased characteristic.

Keywords: *spatial; outlier, estimator; error, bias*

1. Introduction

Spatial data analyze is a data analyzing that containing location information. In spatial data, the observation of a particular location often depends on the observation of other neighboring locations Cressie (1991). It is said that spatial data is one of kind dependent data, where data on a location is influenced by measurement of another location. In consequence, when the spatial data is processed using linear regression analysis with Ordinary Least Square (OLS), it will result in an inappropriate data. This is because, in linear regression analysis with OLS, it is assumed that the error variable (homoscedasticity) is constant and there should be no dependence between errors (autocorrelation) in each location of the observation (LeSage, J.P, 1994).

How to analyze the spatial data is considered an important issue in the science of statistics. There have been a number of statisticians who write references and resources that deal with spatial problems, especially those for shifting (non-stationary) spatial data. Anselin (2003) observes the spatial uni-variant regression model that contains lag (spatial lag regression), error (spatial error regression), and the combination between spatial lag regression with spatial error regression. Spatial lag regression is a linear regression model that considers the effect of space on the error. Leung, Mei and Zhang (2000) estimates the model parameter of Geographically Weighted Regression (GWR) by considering location factor as a weighing factor. Zhang and Land Gove (2005) determines the autocorrelation value using the OLS method, Linear Mixed Model (MLM), Generalized Additive Model (GAM), and GWR. Mennis (2006) formulates estimating parameter of the GWR model by utilizing MLE method. For the development theory Harini, Purhadi, Mashuri and Sunaryo (2012) determines the statistical test for Multivariate Geographically Weighted Regression (MGWR) model by utilizing Maximum Likelihood Ratio Test. This process is a result of statistical test developed from the GWR model developed by Fotheringham, Brundson and Charlton (2002).

One problem that is faced in spatial data analysis is the occurrence of Outlier. The existence of the outlier creates a biased estimated parameter. In order to solve the problem, a robust and resistant method is needed so that when a small portion of change occurs, it will not affect the estimated value. The resistant estimation will not be affected by size of data changes.

A robust method to estimate the regression can be approached using the M-estimator (Maximum Likelihood Type), LMS-Estimator(Least Median Square), LTS-Estimator (Least Trimmed Squares), MM-Estimator (Method of Moment), and S-Estimator (Scale). The S-estimator is a robust estimation that provides the highest breakdown point of 50%. The breakdown point is the smallest fraction from data contaminated with outlier, which can render the estimator useless (Montgomery, Peck, and Vining, 2006).

The aim of the research is to determine the estimating parameter for spatial lag regression model in the outlier by utilizing the S-estimator method. This study refers to the research conducted by Chen (2002) that has successfully applied the estimation method in a robust regression for fixed and random data.

2. Methods

The method used to determine estimating parameter of on the model for spatial regression is by determining parameter $\hat{\beta}^0$ as an initial estimator and β parameter using the Iteratively Reweighted Least Square (IRLS) to minimize the objective function and the function of the error square.

3. Results and Discussions

Spatial lag regression model with the response variable is modeled as a linear combination from the area coinciding with that is influenced by predictor variable. The spatial lag model is formulated as follows:

$$y = \rho W_1 y + X\beta + \varepsilon_i \quad (1)$$

If outlier that is assumed by θ occurs in (1), the estimated spatial lag model that contains the outlier can be formulated as follows:

$$\theta\varepsilon_i = \theta(I - \rho W_1)y - \theta X\beta \quad (2)$$

In order to estimate the β parameter by minimizing the objective function (by minimizing the θ residual), the spatial regression model is formulated as follows:

$$\sum_{i=1}^{n} \theta(\varepsilon_i) = \sum_{i=1}^{n} \theta((I - \rho W_1)y - \theta X\beta) \quad (3)$$

Based on (3), the function of the sum squared error containing the outlier can be formulated as follows:

$$SSE = \left(\theta((I - \rho W_1)y - X\beta)\right)^T \left(\theta((I - \rho W_1)y - X\beta)\right)$$
$$= y^T(I - \rho W_1)^T\theta(I - \rho W_1)y - 2\beta^T X^T\theta(I - \rho W_1)y + \beta^T X^T \theta X\beta \quad (4)$$

In order to minimize (4), the first derivation from $\varepsilon^T\theta\varepsilon$ to β^T has to be found. Then the formulation can be listed as follows:

$$\hat{\beta}_{OLS} = (X^T\theta X)^{-1}X^T\theta(I - \rho W_1)y \quad (5)$$

After β is obtained as shown by (5), the initial error obtained from the Ordinary Least Square (OLS) can be determined as follows:

$$\varepsilon_i = (I - \rho W_1)y - X\hat{\beta}_{OLS} \quad (6)$$

Next, the equation from (6) is used to obtain the θ value, which functions as an outlier parameter from the spatial regression model by assuming $\theta = \psi$ as the influence function. Therefore, the equation in (6) can be rendered as:

$$\hat{\beta}_{OLS} = (X^T\psi X)^{-1}X^T\psi(I - \rho W_1)y \quad (7)$$

With its influence function as follows :

$$\boldsymbol{\omega}_i = \frac{\psi(\varepsilon_i{}^*)}{\varepsilon_i{}^*} \quad \text{with } \varepsilon_i{}^* = \frac{\varepsilon_i}{\hat{\sigma}} \quad (8)$$

The value of $\varepsilon_i{}^*$ is a standardised error to standard deviation $(\hat{\sigma})$. In order to obtain εi *, the standard deviation of $\hat{\sigma}$ is computed. According to Rousseeuw and Yohai (1984), the $\hat{\sigma}$ value can be obtained utilizing the following equation:

$$\hat{\sigma}_s = \frac{MAD\ (x)}{0{,}6745}\ untuk\ iterasi = 1$$

$$\hat{\sigma}_s = \sqrt{\frac{1}{nK}\sum_{i=1}^{n} \boldsymbol{\omega}_i \varepsilon_i{}^2}\ untuk\ iterasi > 1$$

If $\psi = \boldsymbol{\omega}_i$ then, (1.8) can be rendered as :

$$\hat{\beta} = (X^T \boldsymbol{\omega}_i X)^{-1} X^T \boldsymbol{\omega}_i (I - \rho W_1) y \quad (9)$$

with ω_i is a weighted matrix $n \times n$ where the diagonal elements contain weighed $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \dots, \boldsymbol{\omega}_n$. The equation for this is known as Weighted Least Square (WLS). In this study, the weighted function that is used is the bisquare Tukey weighted function.

If the $\boldsymbol{\omega}_i$ function is not linear, the estimated parameter can be processed utilizing Iteratively Reqeighted Least Square (IRLS) (Alma, O.G, 2011).

In this iteration, the $\boldsymbol{\omega}_i$ value will change in every iteration, hence the obtained value of $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$. For the parameter with *m* value, it is the sum of the parameter that should be estimated with the initial estimation of $\hat{\beta}^0$ as follows:

$$\hat{\beta}^0 = (X^T X)^{-1} X^T (I - \rho W_1) y \quad (10)$$

with the $\boldsymbol{\omega}_i^{\ 0}$ as the initial weighted matrix $n \times n$ containing $\boldsymbol{\omega}_1^{\ 0}, \boldsymbol{\omega}_2^{\ 0}, \boldsymbol{\omega}_3^{\ 0} \dots \boldsymbol{\omega}_n^{\ 0}$, the next estimator can be formulated as follows:

$$\hat{\beta}^1 = (X^T \boldsymbol{\omega}_i^{\ 0} X)^{-1} X^T \boldsymbol{\omega}_i^{\ 0} (I - \rho W_1) y \quad (11)$$

Next, the $\widehat{\boldsymbol{\beta}}$ iteration is repeatedly until the following is obtained:

$$\hat{\beta}^{m+1} = (X^T \boldsymbol{\omega}_i^{\ m} X)^{-1} X^T \boldsymbol{\omega}_i^{\ m} (I - \rho W_1) y \quad (12)$$

The aforementioned equation is conducted repeatedly until the convergent estimator is obtained. This is when the difference between $\hat{\beta}^{m+1}$ and $\hat{\beta}^m$approaches 0, with *m* value as the sum of iteration. The result of the analysis shows that estimator β bears biased characteristics, where $E\left(\hat{\beta}^{m+1}\right) \neq \beta$.

4. Conclusions

From the result of the analysis, the parameter estimate for the spatial lag regression model for the outlier data using the S estimator can be obtained by using the equation $\hat{\beta}^{m+1} = (X^T \boldsymbol{\omega}_i^{\ m} X)^{-1} X^T \boldsymbol{\omega}_i^{\ m} (I - \rho W_1) y$, with $E\left(\hat{\beta}^{m+1}\right) \neq \beta$ which bears a biased characteristic.

Acknowledgment

The writers are grateful for the Research Sub Directorate, Society Service and Scientific Publications General Directorate of Islamic Higher Education for funding for scientific publication in 2015.

References

[1] Alma, O.G. (2011). Comparison of Robust Regression Methods in Linear Regression .International Journal Contemp. Math. Sciences, Vol. 6, no. 9, page 409 - 421.

[2] Anselin, L. (2003). An Introduction to Spatial Regression Analysis in R ,http://www,sal,uiuc,edu/stuff/stuff-sum/pdf/spdepintro,pdf Access at 6th Januari 2007.

[3] Cressie, N.A.C. (1991). Statistics for Spatial Data, Revised ed, John Wiley and Sons, New York.

[4] Chen, C. (2002). Robust Regression and Outlier Detection with the ROBUSTREG Procedure. Paper Statistics and Data Analysis, SUGI 27, Hal.265-267.

[5] Fotheringham, A.S, Brundson,C and Charlton, M. (2002). Geographically Weighted Regression: the analysis of spatially varying relationships, John Wiley & Sons Ltd, England.

[6] Harini, S., Purhadi, Mashuri, M and Sunaryo, S. (2012). Statistical Test for Multivariate Geographically Weighted Regression Model Using the Method of Maximum Likelihood Ratio Test. International Journal of Applied Mathematics & Statistics, Vol. 29, Issue Number 5, page.110-115.

[7] LeSage, J.P.(1994). Regression Analysis of Spatial Data, Departemen of Economics, University of Toledo.

[8] Leung, Y., Mei, C.L., and Zhang, W.X. (2000). Statistic Tests for Spatial Non-Stationarity Based on the Geographically Weighted Regression Model, Journal Environment and Planning A, Vol. 32, page.9-32.

[9] Mennis, J. (2006). Mapping the Results of Geographically Weighted Regression, TheCartographic Journal, Vol. 43, No. 2, page. 171–179.

[10] Montgomery, D.C., Peck, E.A., and Vining, G.G. (2006). Introduction to Linier Regression Analysis. 4th Ed. Canada: John Wiley & Sons.

[11] Rousseeuw, P.J. and Yohai, V.J. (1984). Robust Regression by Mean of S-Estimators, Robust and Nonlinear Time Series. eds. J. Franke, W. Hardle, and D. Martin, Lecture Notes in Statistics, 26, 256-272, Berlin: Springer-Verlag.

[12] Zhang and Land Gove, J.H. (2005). Spatial Assessment of Model Errors from Four Regression Techniques, Journal Forest Science, Vol. 51, No. 4, page. 334-346.