Spatial Autoregressive Models with Heteroskedasticity Disturbance using Generalized Method of Moments

Rohimatul Anwar[*], Anik Djuraidah[*], Aji Hamim Wigena [*]

[*] Department of Statistics, IPB University

DOI: 10.29322/IJSRP.9.07.2019.p91105
http://dx.doi.org/10.29322/IJSRP.9.07.2019.p91105

Abstract- Spatial regression is one of the statistical methods that has problems of spatial dependency and heteroskedasticity. Spatial autoregressive regression (SAR) concerns only to the dependence on lag. The estimation of SAR parameters containing heteroskedasticity using the maximum likelihood estimation (MLE) method gives biased and inconsistent. The alternative method is generalized method of moments (GMM). GMM uses a combination of linear and quadratic moment functions simultaneously, so that the computation is easier than that of MLE. This study is to develop SAR model with heteroskedasticity disturbances using the GMM. The model is evaluated based on residual variance and pseudo R^2. Furthermore, this method is applied to the Java's Gross Regional Domestic Product (GRDP) data on 2017. The results showed that the district minimum wage and local revenue were significantly influence to the Java's GRDP data in 2017. This model provides pseudo R^2 value of 75.3% which means it is good enough to illustrate the diversity of Java's GRDP in 2017.

Index Terms: Heteroskedasticity, spatial autoregressive, maximum likelihood, generalized method of moments.

I. Introduction

The spatial dependence and spatial heteroskedasticity are problems in spatial data [1]. Lesage [2] stated that spatial dependence can be described in regression models, such as autoregressive response, error, predictor, or combination the variable. Models with dependencies in response are called spatial autoregressive models (SAR). Fotheringham [3]stated that spatial heteroskedasticity can be described using geographically weighted regression (GWR).

Ord [4] considered the maximum likelihood (ML) for the estimation of the regression model. Kelejian and Prucha [5] extended that the MLE estimator is inconsistent in heteroskedasticity disturbances. Anselin [1] introduced the two-stage least squares (S2SLS) method. Kelejian and Prucha [6] introduced the generalized method of moments (GMM). GMM does not require a distribution assumption of the disturbance and computationally easier than the ML [7].

The results of Kelejian and Prucha's research [5] showed that the estimation method is valid if the assumption of errors is stochastic and identical normal. However, heteroscedacity can occur in aggregation data. In this case heteroskedacity originates from a data averaging process with many different observations at the time of aggregation [6]. Kelejian and Prucha [7] develop the GMM method into robust form that has been proven to be consistent if there is heteroskedacity. Combination of linear and quadratic in moment functions are simultaneously assumed by GMM.

In general, spatial regression modeling only considers one of the spatial effects and uses the ML estimation method. Gross regional domestic product (GRDP) data using the SAR model has been conducted by Hikmah [8] and Yulita [9] has heteroskedasticity using the GWR model. GRDP data is economic data that reflects the economic development of a region. Bivand [10] suggested that economic data always showed spatial patterns. In addition, GRDP publications in Indonesia published by the BPS in 2014 showed that Java had a high level of GDP distribution compared to outside Java. This study would handle heteroscedastic problems in the SAR model using GMM. The purpose of this study is modelling the Java 2017 GRDP data using GMM.

II. MATERIALS AND METHODS

2.1 SAR Model with GMM Approach

In this study, the following SAR model specificiation is considered $\mathbf{y} = \rho\mathbf{W}\,\mathbf{y}+\mathbf{X\beta}+\boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, $\mathbf{y}$ is the n×1 vector of dependent variable, $\mathbf{X}$ is the n×k matrix predictor, $\boldsymbol{\beta}$ is the k×1 vector of regression coefficient parameter, $\mathbf{W}$ is the n×n spatial weight matrix, $\boldsymbol{\varepsilon}$ is the n×1 vector of disturbances (or innovations), and ρ is the coefficient autoregressive spatial lag [1].

Kelejian and Prucha [7] motivated to control spatial autocorrelations in the model at a tractable level also implies that the model and then reduced form of the model becomes feasible as $\mathbf{Wy} = \mathbf{WS}^{-1}\mathbf{X\beta} + \mathbf{WS}^{-1}\boldsymbol{\varepsilon}$ where $\mathbf{S=(I} - \rho\mathbf{W)}$ and the variable $\mathbf{Wy}$ is known as the spatial lag of the dependent variable. Lee and Liu [11] define for $\mathbf{Q}$ a matrix built from functions $\mathbf{W}$ and $\mathbf{X}$. So, $\mathbf{G}= \mathbf{WS}^{-1}$ and then $\mathbf{Q}=(\mathbf{GX\beta}, \mathbf{X})$ is a non-stochastic part that forms a population moment function $\mathbf{Q}'\boldsymbol{\varepsilon}$. Let $\mathbf{P}$ be a n×n matrix with $\mathrm{tr}(\mathbf{P})=\mathrm{tr}\big(\mathbf{G}\text{-Diag}(\mathbf{G})\big)=0$. $\boldsymbol{\varepsilon}'\,\mathbf{P}_j\boldsymbol{\varepsilon}$ from orthogonal forms is obtained the moment function as follows:

$$E(\mathbf{Q}'\boldsymbol{\varepsilon}) = \mathbf{Q}'E(\boldsymbol{\varepsilon})=\mathbf{0}_{(k+1)\times 1} \tag{1}$$

$$\begin{aligned} E(\boldsymbol{\varepsilon}'\,\mathbf{P}_j\,\boldsymbol{\varepsilon}) &= E\left(\mathrm{tr}\left(\boldsymbol{\varepsilon}'\,\mathbf{P}_j\,\boldsymbol{\varepsilon}\right)\right) = E\left(\mathrm{tr}\left(\mathbf{P}_j\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'\right)\right) \\ &= \mathrm{tr}\left(\mathbf{P}_j E(\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}')\right) = \mathrm{tr}\left(\mathbf{P}_j\boldsymbol{\Sigma}\right)=0 \end{aligned} \tag{2}$$

Q is linear and $\mathbf{P}_j$ is quadratic in $\boldsymbol{\varepsilon}$ as the moment function. Matrix quadratic moments $\mathbf{P}_j$ asymptotically efficient as GMM estimator [11]. The parameter moment function of the spatial model simultaneously is the combination of linear and quadratics moment functions as follows:

$$\mathbf{g_n}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{Q}'\boldsymbol{\varepsilon}(\boldsymbol{\theta}) \\ \boldsymbol{\varepsilon}(\boldsymbol{\theta})'\,\mathbf{P}_j\boldsymbol{\varepsilon}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} (\mathbf{GX\beta},\mathbf{X})'\boldsymbol{\varepsilon}(\boldsymbol{\theta}) \\ \boldsymbol{\varepsilon}(\boldsymbol{\theta})'\big(\mathbf{G}\text{-Diag}(\mathbf{G})\big)\boldsymbol{\varepsilon}(\boldsymbol{\theta}) \end{pmatrix} \quad (3)$$

Suppose that $\boldsymbol{\Omega}$ is a matrix of moment functions. $\boldsymbol{\Omega}$ consist variance and covariance that are linear and quadratic[1)] in $\boldsymbol{\varepsilon}$.

$$\boldsymbol{\Omega}=\mathrm{E}[\mathbf{g_n}(\boldsymbol{\theta_0})\mathbf{g_n}'(\boldsymbol{\theta_0})] = \begin{pmatrix} \mathrm{tr}\left(\boldsymbol{\Sigma}\mathbf{P}^*\left(\mathbf{P}^{*\prime}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{P}^*\right)\right) & \mathbf{0}_{1\times(k+1)} \\ \mathbf{0}_{1\times(k+1)} & \mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q} \end{pmatrix} \quad (4)$$

[2)] $\boldsymbol{\Sigma}=$ diag $(\sigma_1^2,....,\sigma_n^2)$ and $\mathbf{P}^* = (\mathbf{G}\text{-Diag}(\mathbf{G}))$, the parameters of the spatial model are simultaneously suspected by the formula of a combination of linear and quadratic in moment functions by GMM approach. GMM robust estimator specification is considered:

$$\hat{\boldsymbol{\theta}}= \begin{pmatrix} \hat{\rho} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = \mathrm{argmin}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\ \mathbf{g_n}'(\boldsymbol{\theta})\ \hat{\boldsymbol{\Omega}}^{-1}\mathbf{g_n}(\boldsymbol{\theta}) \quad (5)$$

Where $\hat{\boldsymbol{\Omega}}$ is a consistent estimator for $\boldsymbol{\Omega}$.

2.2 Data

The data used is secondary data from the Indonesian Central Bureau of Statistics. The data is data based on the constant price of 2010 as the response variable. The response and independent variable are presented in table 1.

Table 1: The variables used this the study

Variables	Unit
GRDP at the constant price based on year 2010 (Y)	Trillion Rupiah
Amount of Labor (X_1)	Thousand Souls
District minimum wage (X_2)	Million Rupiah
Locally generated revenue (X_3)	Million Rupiah
Human development index (X_4)	Percent

2.3 Data Analysis Procedure

Data Analysis used software R Studio 3.5.2 for Java GRDP data on 2017. The steps of analysis will be by the following:

1. Exploring GRDP data in Java in 2017.
2. Testing heteroscedastic variance errors using the Pagan Breusch (BP) test statistic [1].
3. Calculating the Moran's Index to spatial dependence [1].
4. Identify of spatial models with LM test includes spatial influence test on lag (SAR), error (SEM), GSM, robust LM_{lag} and robust LM_{error}[1].
5. Estimating model parameters by GMM approach.

[1)] Appendix 1 can be used to derive Ω matrices in this section
[2)] Appendix 2

III. RESULTS

3.1 Exploring GRDP data in Java in 2017

Gross regional domestic product (GRDP) is the amount of gross domestic product (GDP) of a region. Economic growth can be used as a macroeconomic parameter, both on a national scale and regional scale that reflects the economic condition of a country or region. GRDP can describe the ability of a region to manage natural resources. The average GRDP in each province in Java on 2017 is presented in Figure 1.

The distribution of GRDP-values in Java is presented in Figure 2. Figure 2 describe GRDP-values based on grouping by quartile. They are 3.91-17.8, 17.8-30, 30-59.5, and 59.5-591. Districts / cities with high GRDP-values (> 59.5 trillion rupiahs) are cities in DKI Jakarta, Bekasi, and Surabaya provinces. Districts / cities that have a low GRDP-value (<3.91 trillion rupiah) including Banjar, Blitar, and others. The pattern of the distribution of GRDP-values in districts / cities in Java is presented in Figure 2 without scale.

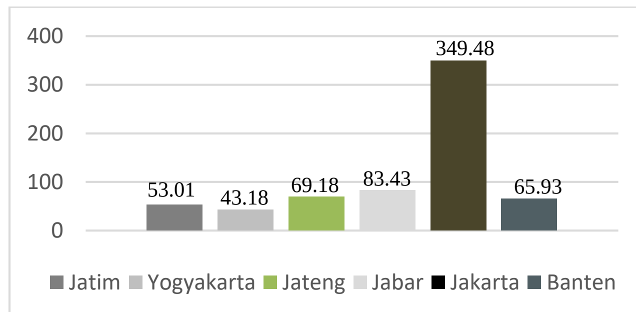

Figure 1: The average Java's GRDP in each province in 2017

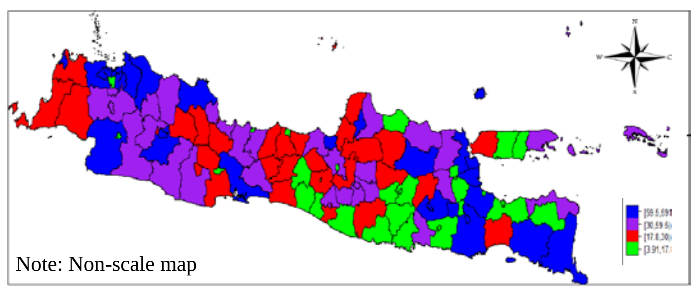

Figure 2: Map of the distribution of the Java's GRDP in 2017

GDP produced by a region, population, as well as social conditions, the surrounding area, and the welfare of the people that differ between regencies / cities, geographical condition of an area can cause the difference of GRDP. The regions are the center of government and economic centers tend to produce high GRDP-values such as cities in DKI Jakarta, Surabaya, and Bandung. The Moran scatter plot is used to detect a location that is outlier in other locations.

In Figure 3 quadrant I (top right) shows districts/cities with positive autocorrelation, because the java's GDRP values of districts/cities and surrounded by surrounding areas are high, such as Sidoarjo, Gresik and Surabaya. Quadrant II (top left) shows districts/cities with negative autocorrelation, because the Java's GDRP values of districts/cities is low and surrounded areas are high, such as South Jakarta City.

ISSN 2250-3153

Quadrant III (lower left) shows districts/cities with positive autocorrelation, because the Java's GDRP values of districts/cities and surrounded by surrounding areas are low, such as in Depok Regency, Bogor Regency, and Bogor City. Quadrant IV shows districts/cities with negative autocorrelation, because the Java's GDRP values of districts/cities is high and surrounded areas are low, such as Kediri City, Bekasi Regency, Sukabumi Regency, East Jakarta City and West Jakarta City are the outlier.

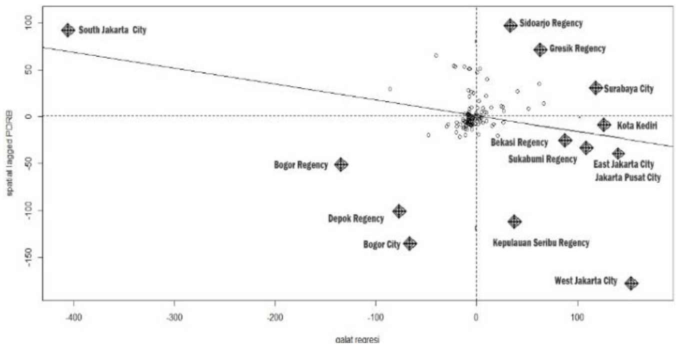

Figure 3: Moran scatter plot

Correlation between regional original income and district minimum wages is 0.63. Correlation human development index and number of workers is negative correlation. Chatterjee [12] stated the value of the variance inflation factor (VIF) greater than 10 indicates that the data has a multicollinearity problem. Table 2 shows that variable independent has not a VIF value of more than 10.

Table 2: VIF values for each independent variable

Variables	VIF
x1	1.35
x2	1.86
x3	2.02
x4	1.38

3.2 *Spatial Effect Test*

BP test produces a BP-value of 35.79 with a p-value of $3.32x10^{-6}$. This shows that the homogeneity of spatial diversity is not fulfilled or there is a problem with heteroskedasticity in the data. The Moran's index is 0.395 with a p-value of 5.05×10^{-7} which indicates that there is a positive spatial autocorrelation (I>0) between GRDP in each neighboring district / city at the 5% level and expectation value from the Moran's index is -0.00847 or $I<E(I)$ then a random pattern.

3.3 *Testing Heteroskedasticity Variance Errors*

The Lagrange Multiplier (LM) test is conducted to test the effect of spatial dependence in response. Table 3 shows the results of the spatial comparison test with the LM test. LM test obtained the spatial dependence value on the response (SAR) of 6.678 with a p-value of 0.0097. Thus, it can be concluded that there is a spatial dependence on the response with significant at level of 0.01. In addition, the value of dependence in spatial error model (SEM) and in the response and error (GSM) results are less than 0.05, so that it can also be modeled with the SEM or GSM model.

Table 3: The results of *Lagrange Multiplier* (LM) test

Model	Statistics LM test	*P-value*
LM lag (SAR)	6.6777	0.0097**
LM error (SEM)	3.9592	0.0466*
SARMA (GSM)	6.7488	0.0342*

*significant at 5% level
**significant at 1% level

3.4 *Parameter Estimation SAR Model by GMM Approach*

The parameter estimates in the SAR model are presented in Table 4. The independent variable that has a significant influence on GDP in Java in 2017 is the district minimum (X_2) and local income (X_3). The estimated parameters, residual variance, pseudo R^2 SAR models use the GMM estimation presented in Table 4.

Table 4: Estimated parameters, residual variance, pseudo R^2

Coefficient	Parameter	Estimates
Rho	ρ	-0.14
x1	β_1	0.01
x2	β_2	19.86*
x3	β_3	0.07**
x4	β_4	-0.74
Residual variance		53.89
Pseudo R^2		0.75

*significant at 5% level
**significant at 1% level

Based on Table 4, increasing one unit of district minimum wages will cause increasing 19.86 of GRDP and increasing one unit of regional original income will cause increasing 0.07 of GRDP. Thus, heteroskedasticity handling produces a model that is simpler than the model with GWR.

IV. CONCLUSION

GMM can be used to analyze data containing heteroskedasticity in SAR modeling. The application of GMM methods in GRDP data in Java on 2017 provide to significant variable that influence GRDP. Those are district minimum wage (X_2) and regional original income (X_3) with pseudo R^2 is 0.753 and residual variance of 53.89.

ACKNOWLEDGEMENTS

This work is fully supported by Kemenristek DIKTI (Kementerian Riset Teknologi dan Pendidikan Tinggi) of Indonesia.

APPENDIX

1. Suppose that Ω is a matrix of moment functions consisting of variance and covariance that are linear and quadratic in $\boldsymbol{\varepsilon}$. For a $\mathbf{G}$ quadrilateral matrix, set Diag $(\mathbf{G})=(g_{11},...,g_{nn})'$ 'is a vector of diagonal matrix elements $\mathbf{G}$.

$$\begin{aligned}\boldsymbol{\Omega} &= \mathrm{var}(\mathbf{g}_n(\boldsymbol{\theta}_0))\\ &= \mathrm{E}[\mathbf{g}_n{}^2(\boldsymbol{\theta}_0)]-(\mathrm{E}[\mathbf{g}_n(\boldsymbol{\theta}_0)])^2\\ &= \mathrm{E}[\mathbf{g}_n(\boldsymbol{\theta}_0)\mathbf{g}_n'(\boldsymbol{\theta}_0)]-0\end{aligned}$$

$= E[\mathbf{g}_n(\boldsymbol{\theta}_0)\mathbf{g}_n'(\boldsymbol{\theta}_0)]$

2. Let $\mathbf{X}_n$, $\mathbf{Y}_n$ and $\mathbf{Z}_n$ be nxn matrices. $\mathbf{X}_n$ and $\mathbf{Y}_n$ have zero diagonal elements, and $\mathbf{Z}_n$ has uniformly bounded row and column sums in absolute value. Assume $\boldsymbol{\Sigma}=$ diag $(\sigma_1^2,\ldots,\sigma_n^2)$.

a. $$E(\boldsymbol{\varepsilon}'_n \mathbf{X}'_n\boldsymbol{\varepsilon}'_n.\boldsymbol{\varepsilon}'_n \mathbf{Y}'_n\boldsymbol{\varepsilon}'_n) = \Sigma_{i=1}^{n}\Sigma_{j=1}^{n} \mathbf{x}_{n,ij}(\mathbf{y}_{n,ij} + \mathbf{x}_{n,ji})\boldsymbol{\sigma}_{ni}^2\boldsymbol{\sigma}_{nj}^2$$
$$= \text{tr}\,(\boldsymbol{\Sigma}_n\mathbf{X}_n(\mathbf{Y}'_n\boldsymbol{\Sigma}_n + \mathbf{Y}_n\boldsymbol{\Sigma}_n))$$

b. $$E(\varepsilon_n\boldsymbol{\Sigma}_n\boldsymbol{\varepsilon})^2 = \Sigma_{i=1}^{n}\mathbf{z}_{n,ii}^2[E(\boldsymbol{\varepsilon}_{ni}^2) - 3\,\boldsymbol{\sigma}_{ni}^4] + (\Sigma_{i=1}^{n}\mathbf{z}_{n,ii}^2\boldsymbol{\sigma}_{ni}^2)$$
$$+ \Sigma_{i=1}^{n}\Sigma_{j=1}^{n}\mathbf{c}_{n,ij}^2(\mathbf{z}_{n,ij}\mathbf{z}_{n,ji})\boldsymbol{\sigma}_{ni}^2\boldsymbol{\sigma}_{nj}^2$$
$$= \Sigma_{i=1}^{n}\mathbf{z}_{n,ii}^2[E(\boldsymbol{\varepsilon}_{ni}^4) - 3\,\boldsymbol{\sigma}_{ni}^4] + \text{tr}^2(\Sigma_n\mathbf{Z}_n)$$
$$+ \text{tr}\,(\boldsymbol{\Sigma}_n Z_n\mathbf{Z}'_n\boldsymbol{\Sigma}_n + \boldsymbol{\Sigma}_n\mathbf{Z}_n\boldsymbol{\Sigma}_n\mathbf{Z}_n)$$

c. $$\text{Var}(\boldsymbol{\varepsilon}_n\mathbf{Z}_n\varepsilon) = \Sigma_{i=1}^{n}\mathbf{z}_{n,ii}^2[[E(\boldsymbol{\varepsilon}_{ni}^4) - 3\,\boldsymbol{\sigma}_{ni}^4]$$
$$+\Sigma_{i=1}^{n}\Sigma_{j=1}^{n}\,\mathbf{z}_{n,ij}(\mathbf{z}_{n,ij}\mathbf{z}_{n,ji})\boldsymbol{\sigma}_{ni}^2\boldsymbol{\sigma}_{nj}^2$$
$$=\Sigma_{i=1}^{n}\mathbf{z}_{n,ii}^2[E(\boldsymbol{\varepsilon}_{ni}^4) - 3\,\boldsymbol{\sigma}_{ni}^4]$$
$$+\text{tr}\,(\boldsymbol{\Sigma}_n\mathbf{Z}_n\mathbf{Z}'_n\boldsymbol{\Sigma}_n + \boldsymbol{\Sigma}_n\mathbf{Z}_n\boldsymbol{\Sigma}_n\mathbf{Z}_n)$$

REFERENCES

[1] B. S. Yandell and L. Anselin, *Spatial Econometrics: Methods and Models.*, vol. 85, no. 411. 2006.

[2] J. LeSage, "Spatial Econometrics Toolbox," *A companion to Theor. Econom.*, 1999.

[3] Fotheringham AS, Brunsdon C, Charlton M. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. England (GB): John Wiley & Sons, LTD. 2002.

[4] K. Ord, "Estimation methods for models of spatial interaction," *J. Am. Stat. Assoc.*, vol. 70, no. 349, pp. 120–126, 1975.

[5] H. H. Kelejian and I. R. Prucha, "A generalized moments estimator for the autoregressive parameter in a spatial model," *Int. Econ. Rev. (Philadelphia).*, vol. 40, no. 2, pp. 509–533, 1999.

[6] H. H. Kelejian and I. R. Prucha, "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances," *J. Econom.*, vol. 157, no. 1, pp. 53–67, 2010.

[7] O. Doğan and S. Taşpinar, "Spatial autoregressive models with unknown heteroskedasticity: A comparison of Bayesian and robust GMM approach," *Reg. Sci. Urban Econ.*, vol. 45, no. 1, pp. 1–21, 2014.

[8] Hikmah,Nur. Statistika, F. Matematika dan Ilmu Pengetahuan Alam, "Produk Domestik Regional Bruto Kabupaten / Kota Jawa Barat Dengan Spasial Data Panel," 2012.

[9] I. Yulita, "Pemodelan Regresi Komponen Utama Dan Lasso Terboboti Geografis," pp. 0–10, 2016.

[10] R. S. Bivand and D. W. S. Wong, "Comparing implementations of global and local indicators of spatial association," *Test*, vol. 27, no. 3, pp. 716–748, 2018.

[11] L. Lee, "Efficient GMM Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances," *Science (80-.).*, 2006.

[12] A. Chatterjee, *110 Regression Analysis Example*. 4th ed. New York (US): J Wiley.2006.

.

AUTHORS

First Author – Rohimatul Anwar S. Si, college student, Department of Statistics, Faculty of Mathematics and Natural Sciences (FMIPA), IPB University, Bogor,16680, Indonesia, and rohimatul_anwar@apps.ipb.ac.id

Second Author – Dr. Anik Djuraidah, MS, Lecturer, Departement of Statistics, Faculty of Mathematics and Natural Sciences (FMIPA), IPB University, Bogor,16680, Indonesia, and anikdjuraidah@apps.ipb.ac.id

Third Author – Prof. Dr. Aji Hamim Wigena, M.Sc, Lecturer, Department of Statistics, Faculty of Mathematics and Natural Sciences (FMIPA), IPB University, Bogor,16680, Indonesia, and aji_hw@apps.ipb.ac.id

Correspondence Author - Dr. Anik Djuraidah, MS, Lecturer, Department of Statistics, Faculty of Mathematics and Natural Sciences (FMIPA), IPB University, Bogor,16680, Indonesia, and anikdjuraidah@apps.ipb.ac.id