International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET2293178

Spatial Panel Data Model of West Java's Regional Revenue

Tiyas Yulita
Politeknik Siber dan Sandi Negara
yulitatiyas@gmail.com

Article Info
Volume 9, Issue 4
Page Number : 25-31
Publication Issue :
July-August-2022
Article History
Accepted : 12 July 2022
Published: 19 July 2022

ABSTRACT

The spatial panel data model is the construction of a regression model that is used to explain the spatial dependence on panel data. Space dependence may apply between adjacent areas, as in the economic field. This should not be ignored because if the freedom between regions is not fulfilled. The spatial panel data model may be in the form of a SAR, SEM or GSM model. In this study, the spatial panel data model is used to model regional income in districts/cities in West Java, the results of the analysis obtained are that the SEM model with random effect is the best model because value of $R^2 - adj$ is 97.64%.
Keywords: Panel Data, Spatial Autocorrelation, Spatial Autoregression.

I. INTRODUCTION

Modeling in statistics was run to determine the relationship between the variables of interest to the investigator. It is not uncommon for data obtained from observations in several different locations at one time to be referred to as data across locations. In addition, if the reviewer creates observations for some time and then combines them with cross-site data, the existing data is referred to as panel data. If based on assumptions on component errors, panel data may have a fixed or random effect. If there is a constant impression, the error component of the existing regression model is a fixed parameter (location and time are determined by the investigator), whereas in a model with a random effect, the error component is a location and time impression determined randomly from the population. The problem that may arise from panel data is that there is a relationship between the location of the perception which is known as spatial dependence. This should not be ignored because if the independence between observations cannot be met, then spatial panel data modeling is required. The spatial panel data model is divided into two, namely the spatial lag model (spatial autoregressive model / SAR) where there is a spatial correlation in the bound variable and the spatial error model (spatial error model / SEM) where there is a spatial correlation. than that mistake.

In the economic field, especially in terms of regional income, when referring to Law Number 32 of 2014 concerning Regional Government, each region is given the authority to regulate its own region. In order for regional autonomy to be implemented to reduce the dependence of regional governments on the central government, especially in terms of finances, with the hope that each region is able to manage its finances freely, an appropriate basis is needed to increase the ability to explore. its own financial potential. Local government policies must be formed by taking into account the factors that can affect regional income, so that the utilization of existing resources in the region can be carried out optimally. However, it should be reminded that the economic condition of an area is always influenced by the surrounding area, such as the people of Bogor City who need supplies of food sources such as vegetables from the area in Bogor Regency which causes buying and selling transactions to take effect. applies automatically, so that the economic conditions of the two regions may affect each other.

Based on the background description, this study aims to analyze the factors that affect regional income in districts/cities in West Java Province and determine the impression of spatial dependence between regions. So that by knowing the factors that affect income, local governments can make a basis for increasing their income so that development for community welfare can be realized.

Model Spatial Data Panel

The spatial data panel model is able to capture the existence of spatial interactions in spatial units and mass as a whole. The following is a static model of panel data that includes bounded variable spatial lagging and spatial error autoregression.

$$y = \lambda(I_T \otimes W_N)y + X\beta + (i_T \otimes I_N)\mu + \rho(I_T \otimes W_N)\varepsilon + v \quad (1)$$

with

y : vector of concern for the size dependent variable $NT \times 1$

X : the matrix of concern for the size-independent variable $NT \times k$

I_T : identity matrix size T

W_N : space-weighted matrix with diagonal 0

λ : spatial autoregression parameter turned on y

i_T : unit of measure vector $T \times 1$

I_N : identity matrix size $N \times N$

μ : mass vector at the same attention with a certain impression (not spatial autocorrelation)

ρ : parameter of spatial autoregression in error ε

$v_{it} \sim IID\ (0, \sigma_v^2)\ dan\ \varepsilon_{it} \sim (0, \sigma_\varepsilon^2)$; N is the number of absorption locations, and T is the serial number of the time of observation[7].

As in classical panel data, individual effects can be treated either permanently or randomly. In a model with a random effect, the assumption is that the unobserved effect of the individual is not correlated with other independent variables in the model, in this case $\mu_{it} \sim IID\ (0, \sigma_\mu^2)$, and error ε can be written as

$$\varepsilon = (I_T \otimes B_N^{-1})v \quad (2)$$

with $B_N = (I_N - \rho W_N)$, then

$$u = (i_T \otimes I_N)\mu + (I_T \otimes B_N^{-1})v \quad (3)$$

and variance covariance matric of ε is

$$\Omega_u = \sigma_\mu^2(i_T i_T^T \otimes I_N) + \sigma_v^2[I_T \otimes (B_N^T B_N)^{-1}] \quad (4)$$

Spatial autoregressive model / spatial lag (SAR) is obtained if the equation (1) value $\lambda \neq 0, \rho = 0$, and if $\lambda = 0, \rho \neq 0$ as *Spatial Error Model* (SEM), then if $\lambda \neq 0, \rho \neq 0$ as *General Spatial Model* (GSM).

Parameter estimation in the SAR and SEM models was carried out using the Maximum Likelihood method, the following is the log-likelihood function of the SAR model with a constant effect

$$\text{Log } L = -\frac{NT}{2}\ln(2\pi\sigma_\varepsilon^2) + T ln|I_N - \lambda W_N| - \frac{NT}{2\sigma_\varepsilon^2}e^T e \quad (5)$$

with $e = y - \lambda(I_T \otimes W_N)y - X\beta$ and $ln|I_N - \lambda W_N|$ as Jacobian Determinant.[3]

Log-likelihood function of SEM with fixed effect is

$$\text{Log } L = -\frac{NT}{2}\ln(2\pi\sigma_\varepsilon^2) + T ln|I_N - \rho W_N| - \frac{1}{2\sigma_\varepsilon^2}e^T[I_T \otimes (B_N^T B_N)]e \quad (6)$$

with $e = y - X\beta$. [3]

estimation of β and σ_ε^2 are

$$\beta = [X^T(I_T \otimes B_N^T B_N)X]^{-1}X^T(I_T \otimes B_N^T B_N)y \quad (7)$$

$$\sigma_\varepsilon^2 = \frac{e(\rho)^T e(\rho)}{NT} \quad (8)$$

$$\mu_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - x_{it}\beta) \quad (9)$$

Whereas log-likelihood from model with random effect is :

$$\text{Log } L\ (\beta, \sigma_e^2, \phi, \lambda, \rho) = -\frac{NT}{2}2\pi - \frac{NT}{2}\ln\sigma_v^2 + T ln|A| - \frac{1}{2}ln[T\phi I_T + (B_N^T B_N)^{-1}] + (T-1)ln|B_N| - \frac{1}{2\sigma_v^2}u^T\Sigma^{-1}u \quad (10)$$

with $\phi = \frac{\sigma_\mu^2}{\sigma_\varepsilon^2}$, $\bar{J}_T = \frac{J_T}{T}, E_T = I_T - \bar{J}_T, dan\ A_N = (I_N - \lambda W_N)$,

$$\Sigma = \phi(J_T \otimes I_N) + I_T \otimes (B_N^T B_N)^{-1}$$

$$\Sigma^{-1} = \bar{J}_T \otimes (T\phi I_N + (B_N^T B_N)^{-1})^{-1} + E_T \otimes B_N^T B_N$$

$$|\Sigma| = |T\phi I_N + (B_N^T B_N)^{-1}||(B_N^T B_N)^{-1}|^{T-1}$$

Parameter estimation is carried out by an iterative procedure until a convergent estimation result is obtained, with the formula

$$\beta = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Ay \qquad (11)$$

$$\sigma_v^2 = (Ay - X\beta)^T \Sigma^{-1} (Ay - X\beta)/NT \qquad (12)$$

II. METHODS AND MATERIAL

To perform modeling with spatial panel, data is required for each location and series of periods. The data from this study use secondary data sourced from the Statistics Indonesia of West Java which includes data from 26 districts and cities from 2009-2018[2] .Detail of data which are used in this methods can be viewed in Table 1.

TABLE I

VARIABLES

Variables	Explanation
Y	District/city local government revenue
X_1	Population
X_2	Percentage of working people
X_3	Average length of school

The stages of analysis carried out in spatial panel data modeling are as follows:

1. Data exploration to see the characteristics of the data in general.
2. Form a weighting matrix.
3. Using the Breush-Pagan test to determine the effect of location or time on panel data.
4. Testing with the Lagrange Multiplier (LM) test. The LM test has been developed to test the presence of random effects and correlations from serial or cross-sectional panel data models. Baltagi [1] derived a combined, marginal and conditional test for conditions of random influence and spatial correlation on the spatial model of panel data.
   a. The LM test is to find out whether there is a random effect on the data assuming there is no spatial autocorrelation, the hypothesis is

      $H_0 : \sigma_\mu^2 = 0$(asumsi $\lambda = 0$) $H_1 : \sigma_\mu^2 \neq 0$(asumsi $\lambda = 0$)

      Test-statistic

      $SLM_1 = \frac{LM_1 - E(LM_1)}{\sqrt{Var(LM_1)}}$,with $LM_1 = \sqrt{\frac{NT}{2(T-1)}} G^2; G = \frac{\bar{u}'(J_T \otimes I_N)\bar{u}}{\bar{u}'\bar{u}} - 1; u$ is error from Least Square Total

      Reject H_0 if statistics LM > $\chi^2_{(\alpha,1)}$, if there is no random effect, the next stage is 3.b, but if there is a random effect, then proceed to stage 3.c
   b. To find out whether there is a spatial effect on the model, the hypothesis is

      $H_0 : \lambda = 0(\ \sigma_\mu^2 = 0)$

      $H_1 : \lambda \neq 0(\ \sigma_\mu^2 = 0)$

      Test-statistics

      $SLM_2 = \frac{LM_2 - E(LM_2)}{\sqrt{Var(LM_2)}}$,with $LM_1 = \sqrt{\frac{N^2T}{b}} H^2; H = \frac{\bar{u}'(I_T \otimes (W+W')/2)\bar{u}}{\bar{u}'\bar{u}}, b = tr(W+W')^2/2.$

      Reject H_0 if LM > $\chi^2_{(\alpha,1)}$
   c. To find out whether there is a spatial effect with a random effect, the hypothesis is

      $H_0 : \lambda = 0(\ \sigma_\mu^2 \neq 0)$

      $H_1 : \lambda \neq 0$

      Test-statistics

$$LM_\lambda = \frac{\acute{D}(\lambda)^2}{\left[(T-1) + \frac{\hat{\sigma}_v^4}{\hat{\sigma}_1^4}\right] b}$$

with $\acute{D}(\lambda)^2 = \frac{1}{2}\acute{u}' \left[\frac{\hat{\sigma}_v^4}{\hat{\sigma}_1^4}\left(\bar{J}_T \otimes (W' + W)\right) + \frac{1}{\hat{\sigma}_v^4}\left(E_T \otimes (W' + W)\right)\right] \acute{u}.$ and, $\hat{\sigma}_1^4 = \frac{\acute{u}'(\bar{J}_T \otimes I_N)\acute{u}}{N}$,

$\hat{\sigma}_v^4 = \frac{\acute{u}'(E_T \otimes I_N)\acute{u}}{N(T-1)}$, $\acute{u}$ is error from least square methods.

Reject H_0 if $LM_\lambda > \chi^2_{(\alpha,1)}$

5. Hausman test is done by comparing estimators with random and fixed effects and testing whether there is a random effect on the data. Mult and Pfaffermayr (2011) developed the Hausman test in the spatial case, with the formula $H = NT(\hat{\theta}_{FGLS} - \hat{\theta}_W)^T(\hat{\Sigma}_W - \hat{\Sigma}_{FGLS})^{-1}(\hat{\theta}_{FGLS} - \hat{\theta}_W)$

With $\hat{\theta}_{FGLS}$ dan $\hat{\theta}_W$ is *Generalized Least Square* (GLS) and estimation with fixed effect, whereas $\hat{\Sigma}_W$ dan $\hat{\Sigma}_{FGLS}$ is coefficients of estimation from matric variance-covariance, with $(H_0)$is model has random effect. Reject H_0 if $H > \chi^2_{(\alpha,k)}$; k is number variables in model[7]

1. Analyze spatial panel data using the maximum likelihood estimation method
2. Estimation of panel data parameters, AR / MA if there is a random effect.
3. Selection of the best model

III.RESULTS AND DISCUSSION

Data Exploration

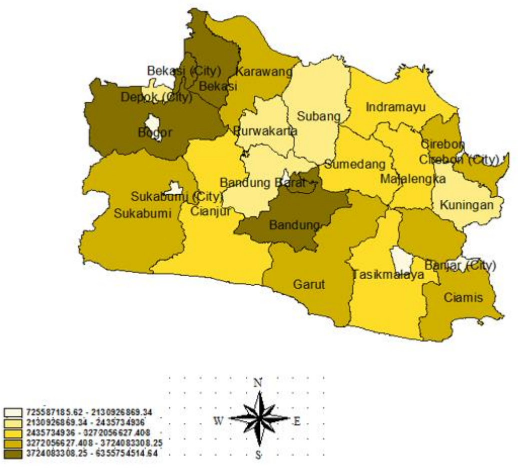

Figure 1: Map of Regional Income in Regency/City of West Java

From Figure 1, it can be seen that the region with high income, namely Bandung, Bogor, Bekasi, and Bekasi City, with a minimum regional income in 2016. In several regions, relatively the same color was found, so it is suspected that there is a spatial influence between regions.

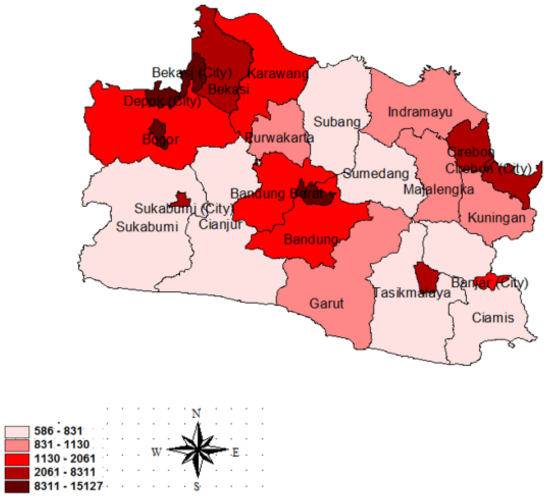

Figure 2: Population Density Map in Regency/City of West Java

From Figure 2 regarding population density, urban areas tend to be densely populated, such as Bogor City, Bandung City, Depok City, and Bekasi City, which affects the surrounding population density, such as Bandung City, which is densely populated, directly intersects with Bandung and Bandung Barat whose population is also quite dense.

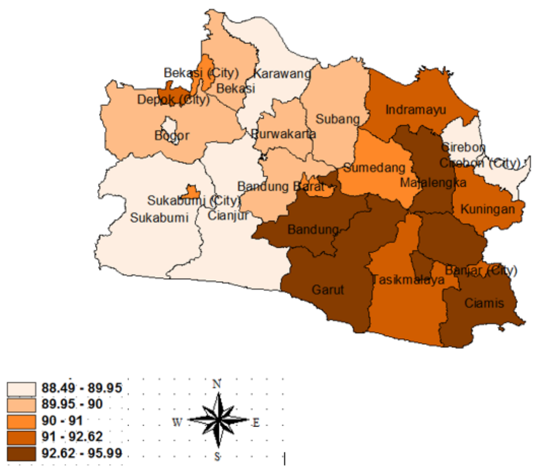

Figure 3: Map of Percentage of Population Working People in

Regency/City of West Java

From Figure 3, it is known that the area with the highest percentage of the working population is Garut, Bandung, Majalengka, Ciamis, and Tasikmalaya City. However, the percentage in each region is not much different from other regions, only around 88.49% - 95.99%.

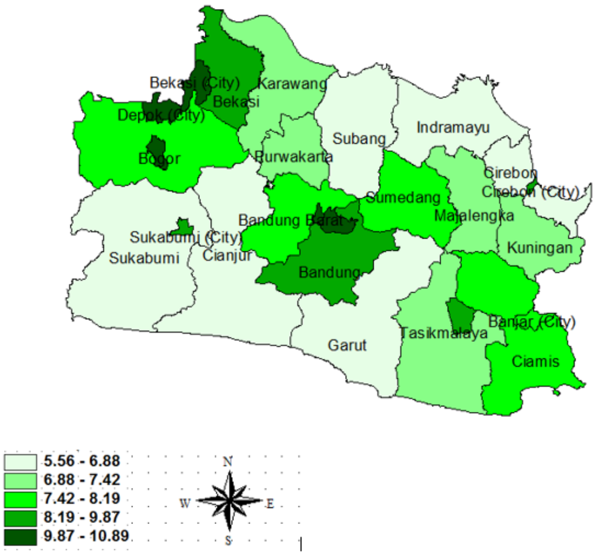

Figure 3: Map of Average Years of Schooling in Districts/Cities of West Java

From Figure 4, it can be seen that urban residents tend to have more than 8 years of education. In Bandung City, Bogor City, Depok City and Bekasi City, the population has an education of more than 9 years or up to college.

The weighting matrix used in this analysis is obtained by the queen contiguity method which defines the relationship of intersections between locations where the side-angle intersection defines $bobot_{ij} = 1$ for the location that is side by side (common side), or the corner point (common vertex) meets the location of concern, $bobot_{ij} = 0$ for other locations [5].

Weighted

Figure 5: Map of Regency/City of West Java

For example in Sukabumi, by looking at the map of the province of West Java in Figure 5, the neighbor locations that are in direct contact with the district. Sukabumi, namely Cianjur, Sukabumi City, and Bogor are given a weight of 1. After all neighboring areas have been recorded, then standardization is carried out on the rows. The same is done for each district and city so that a weighting matrix can be formed.

Breusch-Pagan Test

TABLE II

BREUSCH-PAGAN TEST

Effect	χ^2**(df)**	**p-value**
Time and individual	390,1 (2)	< 0,000
Time	90,7 (1)	< 0,000
Individual	299,4 (1)	< 0,000

If a significance rate of 5% is used, then the decision is:

1. Time and individual effect

P-value < 5%, then reject Ho which means there is sufficient evidence to state that there is at least one effect (time or individual) on the model.

2. Time effect

P-value < 5%, then reject Ho, which means there is sufficient evidence to state that there is an effect of time on the model.

3. Individual effect

P-value <5%, then reject Ho, which means there is sufficient evidence to state that there is an individual effect on the model.

From the results above, it can be concluded that there is a significant effect which caused by time and individuals on local government revenue data in districts/cities of West Java Province. So it is very possible if there is a spatial effect between the observation locations.

Lagrange Multiplier (LM) test

TABLE III

LAGRANGE MULTIPLIER TEST

LM-test	H_0	p-value
SLM_1	$\sigma_\mu^2 = 0(\lambda = 0)$	< 0,000
SLM_2	$\lambda = 0(\sigma_\mu^2 = 0)$	< 0,000
LM_λ	$\lambda = 0(\sigma_\mu^2 \neq 0)$	< 0,000

From the results of the LM test in Table 3, information is obtained that based on the test, p-value < 5%, so it can be concluded that there is a random effect on the data, while based on the SLM_2test, the p-value <5%, so there is a spatial autocorrelation in the data. Then from the test results LM_λ, p-value < 5%, the test results show that there is a spatial autocorrelation with a random effect on the data. So that the next District/City Government Revenue data in West Java will be modeled using the spatial panel data method with random effects. The random effect that exists can occur because time has a significant effect on the data.

Hausman test

By using a significance level of 5%, from the Hausman test results is p-value > 0.05, there is not enough evidence to reject Ho, so the spatial panel data model follows the random effect model.

Spatial Panel Data Model

Then the data will be modeled with the Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), and General Spatial Model (GSM) with random effects. The results of the estimated parameters for the three models are given in Table 4.

TABLE IV

ESTIMATED PARAMETERS

Model	Panel Model	SAR	SEM	GSM
Parameter				
$\hat{\beta}_1$	0,24 ***	0,06	0,014	0,012
$\hat{\beta}_2$	144,10***	44,21**	38,871 **	28,852 *
$\hat{\beta}_3$	-546,64**	-8,06	21,682	-25,246
Intercept	-7506,15**	-3161,7*	-1753,6	800,76
λ	-	0,67***	-	-0,54 ***
ρ	-	-	0,742**	0,870***
ϕ	-	7,089**	6,253***	8,074 **
θ	0,7737	-	-	-
The goodness of fit				
R^2-adj	21,23%	92,79%	97,64%	93,62%
AIC		3183,117	3177,008	3156,785

with : $^{***}(\alpha = 0{,}001), ^{**}(\alpha = 0{,}01), ^{*}(\alpha = 0{,}05)$

From Table 4, it can be seen that based on R^2-adj, the panel data model with random effect is not very good when used in the estimation, which means that the regional income of districts/cities in West Java can be explained by the factors contained in the panel data model only 21.23%, this may be due to the spatial effect between regions. Meanwhile, from the results of spatial panel data modeling, the SEM model with random effect was chosen as the best model because it produces the smallest MSE value, with R^2-adj of 97.64%, although the AIC value generated by the SEM model is greater than GSM, but not too much different. The model formed is

$$\hat{y}_{it} = 0{,}01438x_{it1} + 38{,}8710x_{it2} - 8{,}0698x_{it3} + \mu_i + 0{,}7423\sum_{j=1}^{26} w_{ij}\phi_{jt}$$

In the SEM model with random effects, it is known that the variable that has a significant and positive value is the percentage of the working population (X_2), so that if the local government wants to increase its regional income, the government must increase employment opportunities to provide greater opportunities for residents to be able to work. The potential of human resources in the area will be optimally empowered, such as through cooperation with investors to develop the natural potential in the area which will automatically absorb labor in the area, so that local government revenues can increase through tax levies.

IV.CONCLUSION

Based on the results and discussion, panel data modeling of regional income in districts/cities in West Java province by adding spatial elements gives better results than the classic panel data model. The SEM panel model with random effect is better at explaining diversity than the SAR or GSM model with an R^2-adj value obtained of 97.64%. While the variable that has a significant positive effect is the percentage of the working people (X_2).

V. REFERENCES

[1]. Baltagi, B.H. 2005. Econometrics Analysis of Panel Data Third Edition. John Wiley & Sons. Chicester, England.

[2]. BPS. Jawa Barat Dalam Angka Tahun 2010-2018. Bps.jabar.go.id.

[3]. Elhorst, J.P. 2009. Spatial Panel Data Models : Handbook of Applied Spatial Analysis. New York: Springer.

[4]. Elhorst, J.P. 2014. Spatial Panel Data Models : Spatial Ecoometrics From Cross-Sectional Data to Spatial Panel, Ch.3, New York : Springer.

[5]. LeSage, J.P. 1999. The Theory and Practice of Spatial Econometrics.Ohio: Department of Econometric. University of Toledo 6Millo, G., Piras, G., 2012. Splm:Spatial Panel data Models in R. Journal of Statistical Software. Vol 47, Issue 1.

[6]. Mult J, Pfaffermayr M. 2011. The Hausman test in a Cliff and Ord Panel Model. Econometrics Journal, 14. 48-76

Cite this article as :

Tiyas Yulita, "Spatial Panel Data Model of West Java's Regional Revenue ", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 4, pp. 25-31, July-August 2022. Available at doi : https://doi.org/10.32628/IJSRSET2293178

Journal URL : https://ijsrset.com/IJSRSET2293178