JSDS: JOURNAL OF STATISTICS AND DATA SCIENCE
VOLUME 2, No 2, October 2023
e-ISSN: 2828-9986
https://ejournal.unib.ac.id/index.php/jsds/index

Modeling the Open Unemployment Rate of Regency/City in West Java Province in 2021 using *Spatial Autoregresive Moving Average* and *Spatial Durbin Model*

Hermalia*[1], Cinta Rizki Oktarina[1], Eldo Aprianto [1], Nurul Hidayati[1]
[1] Department of Statistics, Faculty of Mathematics and Natural Sciences, Bengkulu University, Indonesia

* Corresponding Author: hermalialia3@gmail.com

Article Info

Article History:
Received: October 10, 2023
Accepted: December 6, 2023
Available Online: January 9, 2023

Key Words:
Spatial Regression
Spatial Autoregressive Moving Average
Spatial Durbin Model
Open Unemployment Rate
West Java

Abstract

The Open Unemployment Rate is an important indicator to see the non-absorption of labor by the labor market. According to statistic indonesia, in August 2021, the Open Unemployment Rate in Indonesia was 6.49% or around 9.1 million people a population of 50 million, West Java Province has a high unemployment rate, reaching 9.82%. When examined, the open unemployment rate in West Java tends to cluster higher in the west and lower in the east, indicating that there are spatial factors in the data. Therefore, an analysis was conducted involving the variables of Labor Force Participation Rate, Expected Years of Schooling, and Expenditure per capita as independent variables in measuring their influence on the Open Unemployment Rate, the methods used Spatial Autoregressive Moving Average and Spatial Durbin Model. The result shows that both methods are significant in all tests conducted, then the best method is chosen by comparing the AIC value, it is obtained that the best method in modeling the Open Unemployment Rate in West Java Province is the Spatial Durbin Model with Rsquared of 81.32%. Indicating that the independent variables have a significant effect of 81.32% while 18.68% is influenced by other variables not examined.

1. INTRODUCTION

According to statistic Indonesia, the open unemployment rate is an indicator used to measure labor that is not absorbed by the labor market and illustrates the underutilization of labor supply. The number of unemployed or Open Unemployment Rate in August 2021 reached 6.49 percent of the total labor force or equivalent to 9.10 million people. This figure is lower than the open unemployment rate in August 2020. West Java is one of the provinces with the largest population in Indonesia, reaching 50 million people, this certainly makes it a province with a fairly high number of unemployed people, which in 2021 amounted to 9.82%. With such a high percentage, it can have a negative impact on economic growth in Indonesia. Economic growth and unemployment have a close relationship, because the contribution of the working population in producing goods and services will be able to encourage the rate of economic growth. Conversely, quality economic growth is expected to absorb labor better. Okun's law states that there is a negative relationship between economic growth and unemployment. Gross Domestic Product (GDP) growth close to 3% will reduce unemployment by 1% [11].

Quoted from Kompas, West Java Province is the province that has the most companies and the area with the highest investment in Indonesia, which reaches more than 30%. This should be able to suppress the Open Unemployment Rate in the region, for this reason a tool is needed to see the gap in its suppression, one of which is to see what factors support the Open Unemployment Rate in West Java Province. Based on statistic Indonesia data, when examined, the Open Unemployment Rate in West Java Province tends to cluster higher in the western part and lower in the eastern part [14].

Therefore, an analysis is needed that is able to see the picture or pattern of distribution and the factors that influence the Open Unemployment Rate in West Java Province. Spatial regression analysis is a statistical method to look at data influenced by location or region. Tobler's first law of geography states that conditions at an observation location are closely related to conditions at other observation locations [8].

Spatial modeling can be divided into two, namely modeling with data types based on the point approach and data types based on the area approach. In addition, spatial analysis was developed with spatial influence at a higher level, namely Spatial Autoregressive Moving Average (SARMA). The SARMA model has a structure built through the Autoregressive Moving Average (ARMA) time series model. In the SARMA model there is a spatial weighting matrix as a substitute for the time effect in the ARMA time series model [10]. According to Kruk in [13], the advantage of this model is that spatial model parameters can be estimated for spatial relationships at a higher level.

SDM serves to produce regression estimates to see the relationship between variable X and variable Y, as well as based on the presence of spatial effects (spatial dependencies) in it. Based on the research that has been conducted, it can be seen that there are several predictor variables that significantly influence as well as the relationship between adjacent regions to the case under study (spatial dependency). However, each study does not explain the correlation or relationship pattern between observation areas based on spatial dependency. The Spatial Durbin Model (SDM) is a spatial model that considers the existence of spatial effects in the form of spatial lags not only on the response variable (Y) but also on the predictor variables (X).

2. THEORETICAL REVIEW

2.1 Multiple Linear Regressions

Multiple linear regression is a statistical technique used to determine the relationship model of one response variable (*Y*) involving more than one predictor variable up to p predictor variables where the number of p is less than the number of observations (*n*). The multiple regression model is as follows:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_{i2} X_{i2} + \cdots + \beta_p X_{i_p} \varepsilon_i ; i = 1,2, \ldots, n \quad (1)$$

Where Y_i is the value of the dependent variable in the i-th observation, $\beta_0, \beta_1, \ldots, \beta_p$ are parameters of unknown value, $X_{i1}, X_{i1}, \ldots X_{ip}$ are the values of the independent variables of the i-th observation and ε_i are random errors and normally distributed with zero mean and variance σ^2 [12].

2.2 Spatial Weight Matrix

The spatial weight matrix is a matrix that states the relationship of the observation area. In this study, the spatial weight matrix used is the queen spatial weight matrix. The spatial queen weighting matrix defines for regions that are adjacent or the corner points meet with the region of interest while for other regions [6]. In regression analysis, sometimes there is data that has spatial effects that require handling using spatial analysis. A very important thing in spatial analysis is the spatial weight matrix. The spatial weighting matrix is used to determine the weight between observed locations based on the neighboring relationship between locations. Neighborliness can be defined in several ways, namely Rook contiguity is the observation area is determined based on the sides that intersect and the angle is not taken into account, Bishop contiguity is the observation area is determined based on the corners that intersect and the sides are not taken into account, Queen contiguity is the observation area is determined based on the sides that intersect and the angle is also taken into account [16].

One way to determine the existence of spatial dependencies between locations is to conduct a spatial autocorrelation test using Moran's I statistics. Spatial autocorrelation is an estimate of the correlation between observed values related to the location of the same variable. If there is a systematic pattern in the distribution of a variable, then there is spatial autocorrelation. Determine whether there is spatial autocorrelation between locations, a spatial autocorrelation test using Moran's I can be performed. Moran's index is a local statistical test to see the value of spatial autocorrelation, usually widely used in testing to see spatial autocorrelation or identify a location of spatial clustering. The value of this index ranges between -1 and 1 [9].

2.3 Lagrange Multiplier

LM (Lagrange Multiplier) test is a test to determine whether the model has spatial effects or not. Lagrange Multiplier (LM) which in this test, the residual value is obtained from least squares and the spatial weight matrix count used is W. LM test form [8]. In the Lagrange Multiplier (LM) Test, there are three types of models that can be generated, namely SEM, and SAR.

1. LM lag test for Spatial Autoregressive (SAR)
   Model The SAR model is formed if the value of ρ ≠ 0 dan λ = 0, so that the following general form is obtained [7]:

$$\boldsymbol{Y} = \rho\boldsymbol{W} + \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2}$$
$$\boldsymbol{\varepsilon} = N(0, \sigma^2 I)$$

$\boldsymbol{Y}$: Vector of dependent variables of size $n \times l$
ρ: Coefficient of spatial lag dependent variable
$\boldsymbol{W}$: Spatial weight matrix
$\boldsymbol{X}$: Matrix of independent variables of size $n \times p$
$\boldsymbol{\beta}$: Parameter that reflects the effect of the dependent variable on variations in the independent variable
$\boldsymbol{\varepsilon}$: Error model

2. LM error test for Spatial Error Model (SEM)
   The SEM model is formed if the value of $\lambda \neq 0$ dan $\rho = 0$, so that the following general form is obtained [7]:

$$u = (\boldsymbol{I} - \lambda\boldsymbol{W})^{-1}\varepsilon \tag{3}$$
$$\boldsymbol{Y} = \boldsymbol{X\beta} + (I - \lambda\boldsymbol{W})^{-1}$$

$\boldsymbol{Y}$: Vector of dependent variables of size $n \times l$
λ: Parameter which is the spatial correlation error coefficient
$\boldsymbol{W}$: Spatial weight matrix
$\boldsymbol{X}$: Matrix of independent variables of size $n \times p$
$\boldsymbol{\beta}$: Parameter that reflects the effect of the dependent variable on variations in the independent variable
$\boldsymbol{\varepsilon}$: Error model

2.4 Spatial Regression

Spatial regression is a regression method used for spatial data types or data that have spatial effects. The spatial effect consists of two types: spatial dependency and spatial heterogenity. Spatial dependency means that observations at location i depend on other observations at location j, $j \neq i$. While spatial heterogeneity occurs due to random location effects, namely the difference between one location and another. The basis for the development of spatial regression methods is the classical linear regression method (multiple linear regression). The development is based on the influence of place or spatial on the data being analyzed. According to Miller in [6] revealed in the first Law of Geography, that everything is interconnected with one another, but something close has more influence than something far away [8]

The general spatial regression model can be written as follows [7]:

$$y = \rho \mathrm{Wy} + \mathrm{X}\beta + \mathrm{u}$$
$$\mathrm{u} = \lambda \mathrm{Wu} + \varepsilon$$
$$\varepsilon \sim N(0, \sigma_\varepsilon^2 \mathrm{I_n})$$

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{4}$$

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & \cdots & w_{1n} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & w_{n3} & \cdots & w_{nn} \end{bmatrix}$$

From the general spatial regression model equation can be formed several other models as follows [8]:

1. If $\rho = 0$ and $\lambda = 0$, it is called a classical linear regression model with the equation formed is:

$$y = X\beta + \varepsilon \tag{5}$$

2. If $\rho \neq 0$ and $\lambda = 0$ it is called spatial autoregressive model (SAR) regression with the equation formed is:

$$y = \rho W y + X\beta + \varepsilon \tag{6}$$

3. If $\rho = 0$ and $\lambda \neq 0$ it is called *spatial error model* (SEM) regression with the equation formed is:

$$\begin{gathered} y = X\beta + u \\ u = \lambda W u + \varepsilon \end{gathered} \tag{7}$$

4. if $\rho \neq 0$ and $\lambda \neq 0$ it is called *spatial autoregressive moving average* (SARMA) regression with the equation formed is:

$$\begin{gathered} y = \rho W y + X\beta + u \\ u = \lambda W u_i + \varepsilon \end{gathered} \tag{8}$$

2.5 *Spatial Autoregressive Moving Average*

SARMA is a spatial analysis that assumes that there is a spatial influence on the dependent variable and the residuals. The SARMA model is a spatial analog of a class of ARMA time series models with a spatial weighting matrix as a substitute for the time factor in the ARMA time series model. The SARMA model is developed into a SARMA regression model characterized by the addition of independent variables to the SARMA model. The general SARMA regression model is as follows [10]:

$$\widehat{y_i} = \beta_0 + \rho \sum_{j=1, i \neq j}^{n} w_{ij} y_i + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \lambda \sum_{j=1, i \neq j}^{n} w_{ij} u_i \tag{9}$$

$\widehat{y_i}$: Dependent variables
ρ: Coefficient of spatial lag dependent variable
λ: Parameter which is the spatial correlation error coefficient
w_{ij}: Spatial weight matrix
X: Independent variables
β: Parameter that reflects the effect of the dependent variable on variations in the independent variable

2.6 Spatial Durbin Model

SDM is modeling in spatial regression analysis that is used to analyze the factors that influence a problem in terms of social, economic and health. SDM serves to produce regression estimates to see the relationship between variable X and variable Y, as well as based on the existence of spatial effects (spatial dependencies) in it. Based on the research that has been conducted, it can be seen that there are several predictor variables that significantly influence as well as the relationship between adjacent regions to the case under study (spatial dependency). However, each study does not explain the correlation or relationship pattern between observation areas based on spatial dependency. The Spatial Durbin Model (SDM) is a spatial model that considers the existence of spatial effects in the form of spatial lags not only on the response variable (Y) but also on the predictor variable (X). The SDM equation can be seen as follows [17]

$$\boldsymbol{Y} = \rho \boldsymbol{W} \boldsymbol{Y} + \beta_0 + \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{W} \boldsymbol{X} \theta + \boldsymbol{\varepsilon} \tag{10}$$

$\boldsymbol{Y}$: Vector of dependent variables of size $n \times l$
ρ: Coefficient of spatial lag dependent variable
$\boldsymbol{W}$: Spatial weight matrix

$\boldsymbol{X}$: Matrix of independent variables of size $n \times p$
$\boldsymbol{\beta}$: Parameter that reflects the effect of the dependent variable on variations in the independent variable
θ:Independent variable on spatial lag
$\boldsymbol{\varepsilon}$: Error model
β_0: Intercept

Reveals that what is meant by unemployment is the number of workers in the economy who are actively seeking work, but have not yet obtained it. Unemployment is often a problem in the economy because with unemployment, the productivity and income of the community will decrease, which can lead to poverty and other social problems such as crime. Unemployment has a tendency to increase every year. This is certainly a big challenge for the Indonesian government and also local governments considering that one of the indicators of successful development is being able to lift poverty and reduce unemployment significantly so that social problems do not occur. Efforts to reduce unemployment are to use an economic development plan that includes a more democratic employment plan regarding the rights to choose a job, job opportunities, job location according to ability, and willingness to work without discrimination [15].

3. METHOD

The data source in this study is secondary data obtained from the official bps.go.id website. This study will use cross section data with observations of districts / cities in West Java province with 27 districts / cities in 2021. The research variables used consist of 4 independent variables (labor force participation rate, expected years of schooling and per capita expenditure) and 1 response variable, namely the open unemployment rate. stages of data analysis as follows:

1. Exploring the data
2. Modeling with multiple regression least squares method
3. Testing the classical assumptions of the least squares method regression model (normality test of residuals, multicollinearity test, heteroscedasticity test and autocorrelation test)
4. Determine the weight matrix (queen contiguity)
5. Testing for spatial dependency (moran's scatterplot and moran's index)
6. Lagrange multiplier test (lagrange multiplier error and lagrange multiplier lag)
7. Robust lagrange multiplier test (robust lagrange multiplier error and robust lagrange multiplier lag)
8. Parameter estimation
   - Spatial Autoregressive Moving Average Model
   - Spatial Durbin Model
9. Parameter assumption test (SARMA and SDM)
10. Best model selection (akaike information criterion and residual standard error)
11. Model Interpretation

4. RESULTS AND DISCUSSION

4.1 Testing Spatial Effects

The following are each district/city of West Java with the number of neighbors according to the intersection of sides and angles:

Table 1. Neighborliness of West Java Districts/Cities

No.	District/City	Neighbors	Description
1	Bandung	2 6 8 11 15 24 26	West Bandung, Cianjur, Bandung City, Cimahi City, Subang, Sumedang
2	West Bandung	1 6 11 15 23 24	Bandung. Cianjur, Bandung City, Cimahi City, Purwakarta, Subang
3	Bekasi	4 10 13	Bogor, Karawang, Bekasi City
4	Bogor	3 6 10 13 14 17 23 25	Bekasi, Cianjur, Karawang, Bekasi City, Bogor City, Depok City, Purwakarta, Sukabumi
5	Ciamis	12 19 20 21 22 27	Banjar City, Tasikmalaya City, Kuningan, Majalengka, Tasikmalaya
6	Cianjur	1 2 4 8 23 25	Bandung, West Bandung, Bogor, Garut, Purwakarta, Sukabumi

7	Cirebon	9 16 20 21	Indramayu, Cirebon City, Kuningan, Majalengka
8	Garut	1 6 26 27	Bandung, Cianjur, Sumedang, Tasikmalaya
9	Indramayu	7 21 24 26	Cirebon, Majalengka, Subang, Sumedang
10	Karawang	3 4 23 24	Bekasi, Bogor, Purwakarta, Subang
11	Bandung City	1 2 15	Bandung, West Bandung, Cimahi City
12	Banjar City	5	Ciamis
13	Bekasi City	3 4 17	Bekasi, Bogor, Depok City
14	Bogor City	4	Bogor
15	Cimahi City	1 2 11	Bandung, West Bandung, Bandung City
16	Cirebon City	7	Cirebon
17	Depok City	4 13	Bogor, Bekasi City
18	Sukabumi City	25	Sukabumi
19	Tasikmalaya City	5 27	Ciamis, Tasikmalaya
20	Kuningan	5 7 21	Ciamis, Cirebon, Majalengka
21	Majalengka	5 7 9 20 26 27	Ciamis, Cirebon, Indramayu, Kuningan, Sumedang, Tasikmalaya
22	Pangandaran	5 27	Ciamis, Tasikmalaya
23	Purwakarta	2 4 6 10 24	West Bandung, Bogor, Cianjur, Karawang, Subang
24	Subang	1 2 9 10 23 26	Bandung, West Bandung, Indramayu, Karawang, Purwakarta, Sumedang
25	Sukabumi	4 6 18	Bogor, Cianjur, Sukabumi City
26	Sumedang	1 8 9 21 24 27	Bandung, Garut, Indramayyu, Majalengka, Subang, Tasikmalaya
27	Tasikmalaya	5 8 19 21 22 26	Ciamis, Garut, Tasikmalaya City, Majalengka, Pangandaran, Sumedang

Spatial autocorrelation is an estimate of the correlation between observed values related to the location of the same variable. If there is a systematic pattern in the distribution of a variable, then there is spatial autocorrelation. One way to determine the presence of spatial dependencies between locations is to conduct a spatial autocorrelation test using Moran's I statistic *Moran's I* [9]. In the test with H_0 there is no spatial autocorrelation in the open poverty level data between districts/cities in West Java province which are located close to each other and H_1 is the opposite which means there is spatial autocorrelation in the open poverty level data between districts/cities in West Java province which are located close to each other. The rejection criterion is to reject H_0 if the p-value is smaller than α with H_0 spatial autocorrelation does not exist and H_1 spatial autocorrelation exists

Table 2. Spatial Autocorrelation Test

Parameters	Significance		Decision
	I	P-value	
Open Unemployment Rate (Y)	0.608	1.58×10^{-6}	Reject H_0
Labor Force Participation Rate (X_1)	0.260	0.014	Reject H_0
Expected Years of School (X_2)	0.193	0.051	Reject H_0
Per Capita Expenditure (X_3)	0.299	0.006	Reject H_0

The results of the Moran index test state that each independent variable and the dependent variable have a $p_{value} < 10\%$. So that for each variable there is positive spatial autocorrelation or there is no relationship in the data between districts / cities in West Java Province which are located close to the real level of 10%.

4.2 Lagrange Multiplier

There are three spatial dependency models, namely spatial autoregressive model (SAR), spatial error model (SEM), and spatial autoregressive moving average (SARMA). In testing with H_0 which means there is no spatial error effect in the model and H_1 there is a spatial error effect in the model. then with the rejection criteria, reject ho if the p-value is smaller than α with H_0 no spatial lag effect in the model and H_1 there is a spatial lag effect in the model.

Table 3. LM Test

Parameters	Significance		Decision
	Parameter Values	P-value	
LM_{err}	0.42	0.5138	Accept H_0
LM_{lag}	4.31	0.0377	Reject H_0
$SARMA$	11.47	0.0032	Reject H_0

Lagrange multiplier test results in the model for LM_error **(**H_0 is accepted, which means $\lambda = 0$ or there is no spatial error effect in the model with a real level of 10%). LM_lag (H_0 is rejected, which means $\lambda \neq 0$ or there is no spatial error effect in the model with a real level of 10%), and LM_SARMA(H_0 is rejected, which menas $\rho, \lambda \neq 0$ or there is no spatial error effect in the model with a real level of 10%).

4.3 Robust Lagrange Multiplier

Robust Lagrange Multiplier is a modification of the Lagrange multiplier used to overcome local misspecification. Therefore, a modification of the LM test on the spatial dipendency model is carried out to overcome the problem of local misspecification based on OLS estimation called the robust lagrange multiplier. In testing with H_0 which means there is no lag in the model and H_1 means there is a spatial lag effect in the model. Then, using the rejection criteria, reject H_0 if the p-value is smaller than α with H_0 no spatial lag effect in the model and H_1 there is a spatial lag effect in the model.

Table 4. Robust LM Test

Parameters	Significance		Desicion
	Parameter values	P-value	
RLM_{err}	7.15	0.0074	Reject H_0
RLM_{lag}	11.05	0.0008	Reject H_0

Lagrange multiplier test results in the model for RLM_{err} (H_0 is rejected, which means $\rho \neq 0$ and $\lambda \neq 0$ or there is a spatial error effect in the model with a real level of 10%) and RLM_lag (H_0 is rejected, which means $\rho \neq 0$ and $\lambda \neq 0$ or there is a spatial lag effect in the model with a real level of 10%).

4.4 Parameter Estimation of Spatial Autoregressive Moving Average

The spatial dependency test results show that there is a spatial dependency of lag and error which can then be applied to the spatial autoregressive moving average method to estimate fit model. Parameter estimation of the SARMA model is done using the maximum likelihood method with the results presented in Table 5.

Table 5. SARMA Parameter Estimation

Parameters	Partial Test		Decision	Parameter
	Z_{count}	P-value		Estimation
Intercept	4.94	7.61×10^{-7}	Reject H_0	38.6694
Labor Force Participation Rate (X_1)	5.24	1.56×10^{-7}	Reject H_0	-0.3814
Expected Years of Schooling (X_2)	2.73	0.0061	Reject H_0	-0.9605
Per Capita Expenditure (X_3)	2.74	0.0060	Reject H_0	0.0003
ρ	2.73	0.0062	Reject H_0	0.4053
λ	3.22	0.0012	Reject H_0	-0.76578

In the table 5, the partial test is obtained for each variable, it can be seen that all independent variables used have a significant influence on the open unemployment rate. Based on all independent variables, the relationship with the dependent variable, namely the open unemployment rate, can be known directly by looking at the parameter estimation value for each variable. The spatial autoregressive moving average model is obtained as follows:

$$\hat{y}_i = 38.66 + 0.405 \sum_{j=1, i \neq j}^{n} \boldsymbol{w_{ij} y_i} - 0.381X_1 - 0.96X_2 + 0.00038X_3 + \boldsymbol{u}$$

$$\boldsymbol{u_i} = -0.765 \sum_{j=1, i \neq j}^{n} \boldsymbol{w_{ij} u_i} + \boldsymbol{\varepsilon_i}$$

The data model equation can be described by the regression coefficient of the labor force participation rate variable (X_1) of -0.381 which means that assuming other variables are constant, then every one unit increase in the labor force participation rate will decrease the open unemployment rate in a region by 38.1 percent. Because the coefficient value is negative, there is a negative relationship between the labor force participation rate and the open unemployment rate. The greater the labor force participation rate, the smaller the open unemployment rate.

For example, ρ value is divided into the neighborhood of the area), u value itself is the zero neighbor error and substitute the original data of each variable, the following is the SARMA model for Bekasi district:

$$\widehat{y_3} = 38.66 + 0.135y_4 + 0.135y_{10} + 0.135y_{13} - 0.381X_1 - 0.96X_2 + 0.00038X_3 - 0.255u_4 - 0.255u_{10} - 0.255u_{13}$$

$$\widehat{y_3} = 38.66 + 0.135(12.22) + 0.135(11.83) + 0.135(10.88) - 0.381(65.87) - 0.96(13.10) + 0.00038(11341) - 0.255(0.84) - 0.255(0.33) - 0.255(-0.28)$$

$$\hat{y}_{bekasi} = 9.785$$

Table 6. Significance of SARMA Model

Parameters	Significance Value		Decision
	Stat-test	P-value	
Residual Normality Test	0.968	0.561	Accept H_0
Residual Diversity Test	1.645	0.649	Accept H_0
Residual Autocorrelation Test	0.564	0.572	Accept H_0
Residual Randomness Test	11	0.2377	Accept H_0

Furthermore, it is necessary to test the residual assumptions to find out that the residuals of the SARMA model are independent or there is no autocorrelation, follow a normal distribution and the diversity is homogeneous. Based on the Table 6, we can know that all assumptions have been met, namely that the residuals of the SARMA model are independent or there is no autocorrelation, follow a normal distribution and the diversity is homogeneous.

4.5 Parameter Estimation of Spatial Durbin Model

The spatial dependency test results show that there is a spatial dependency of lag and error which can then be applied to the spatial durbin model method to estimate the right model. Parameter estimation of the SARMA model is done using the maximum likelihood method with the results presented in the table below.

Table 7. Parameter Estimation of SDM

Parameters	Partial Test		Decision	Parameter Estimation
	Z-count	P-value		
Intercept	6.98	2.7×10^{-12}	Reject H_0	101.25
Labor Force Participation Rate (X_1)	6.26	3.6×10^{-10}	Reject H_0	-0.381
Expected Years of Schooling (X_2)	2.15	0.03080	Reject H_0	-0.774
Per Capita Expenditure (X_3)	2.03	0.04221	Reject H_0	0.00028
Lag_X_1	3.54	0.00038	Reject H_0	-0.653
Lag_X_2	2.69	0.00704	Reject H_0	-1.463
Lag_X_3	3.16	0.00157	Reject H_0	0.00072
ρ	4.31	0.03789	Reject H_0	-0.6635

In the table 7, partial test is obtained for each variable, it can be seen that all independent variables used have a significant influence on the open unemployment rate. Based on all independent variables, the relationship with the dependent variable, namely the open unemployment rate, can be known directly by looking at the parameter estimation value for each variable. The spatial durbin model is obtained as follows:

$$\hat{y}_i = -0.6635 \sum_{j=1, i \neq j}^{n} \boldsymbol{w_{ij} y_i} + 101.25 - 0.381X_1 - 0.774X_2 + 0.00028X_3 - 0.653 \sum_{j=1}^{n} w_{ij}\theta_{1j} - 1.46 \sum_{j=1}^{n} w_{ij}\theta_{2j} + 0.00072 \sum_{j=1}^{n} w_{ij}\theta_{3j}$$

The data model equation can be described by the regression coefficient of the labor force participation rate variable (X_1) of -0.381 which means that assuming other variables are constant, then every one unit increase in the labor force participation rate will decrease the open unemployment rate in a region by 38.1 percent. Because the coefficient value is negative, there is a negative relationship between the labor force participation rate and the open unemployment rate. The greater the labor force participation rate, the smaller the open unemployment rate.

Based on the model obtained, we can know that $\rho = -0.6635$ which has a significant effect, meaning that there is an influence of the level of open unemployment in the i-th district/city that is close to the level of open unemployment in the surrounding district/city. The model obtained is a general model for 27 districts/municipalities in West Java, for each district/municipality the model is adjusted to the weights that have been determined for each district/municipality observed. For example, the following is the SDM model for Bekasi district:

$$\begin{gathered}\widehat{\boldsymbol{y_3}} = -0.2211(12.22) - 0.2211(11.83) - 0.2211(11.79) + 101.25 - 0.381(65.87) \\ - 0.774(13.10) + 0.00028(11341) - 0.22(62.55) - 0.22(64.19) \\ - 0.22(61.77) - 0.49(12.49) - 0.49(12.10) - 0.49(13.42) \\ + 0.00024(10410) + 0.00024(11522) + 0.00024(11716) \\ \widehat{\boldsymbol{y}}_{bekasi} = 6.3858\end{gathered}$$

Table 8. Signification of SDM Model

Parameters	Signifikation Value		Decision
	Stat-test	P-value	
Residual Normality Test	0.989	0.993	Accept H_0
Residual Diversity Test	3.784	0.705	Accept H_0
Residual Autocorrelation Test	0.566	0.571	Accept H_0
Residual Randomness Test	15	0.841	Accept H_0

Next, it is necessary to test the residual assumptions to find out that the residuals of the HR model are independent or there is no autocorrelation, follow a normal distribution and the diversity is homogeneous. Based on based on table 8, we can see that all assumptions have been met, namely that the residuals of the HR model are independent or there is no autocorrelation, follow a normal distribution and their diversity is homogeneous.

4.6 Best Model Selection

The selection of the best model between the SARMA and SDM models aims to determine which model is better to apply to the open unemployment rate in West Java. The criteria for the goodness of fit is comparing the AIC value and residual standard error.

Table 9. Best Model Selection

	Classical Regression	SARMA	SDM
AIC	105.66	99.15	95.05
Residual Standard Error	1.541	1.132	0.912
Rho		0.405	−0.663
Lamda		−0.765	
Residual Normality Test	√	√	√
Residual Diversity Test	√	√	√
Residual Autocorrelation Test	√	√	√
Residual Randomness Test	√	√	√

Based on table 9 after choosing the best model based on the AIC value and residual standard error. From the results of the comparison of all these models, it can be seen that classical regression, SARMA and SDM are the best models. The model that has the smallest AIC value and residual standard error is the HR model. So it can be concluded that the 2021 open unemployment rate data in West Java can be modeled with SDM which obtains the smallest AIC value and residual standard error to minimize errors.

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}} = -0.6635 \sum_{j=1, i\neq j}^{n} \boldsymbol{w_{ij}y_i} + 101.25 - 0.381X_1 - 0.774X_2 + 0.00028X_3 - 0.653\sum_{j=1}^{n} w_{ij}\theta_{1j} - 1.46\sum_{j=1}^{n} w_{ij}\theta_{2j} + 0.00072\sum_{j=1}^{n} w_{ij}\theta_{3j}$$

Based on the results of parameter estimation using the SDM method, the $R^2_{adjusted}$ value is 81.32% which means that the estimated model is able to explain the diversity of the open unemployment rate in West Java, 81.32% of which is explained by the variables of labor force participation rate, expected years of schooling, per capita expenditure, and 18.68% is explained by other variables not studied.

5. CONCLUSION

Based on assumption testing, the data on the open unemployment rate in West Java in 2021 uses two models, namely, spatial autoregressive moving average and spatial durbin model. After going through several stages of analysis, it is found that the spatial regression model using the spatial durbin model is the best model. This can be seen based on the smallest AIC value of 95.054 and obtaining an $R^2_{adj} = 81.32\%$ value. The spatial durbin model produces several variables that have a significant effect on the open unemployment rate in West Java in 2021. The following is the global model of the spatial durbin model:

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}} = -0.6635 \sum_{j=1, i\neq j}^{n} \boldsymbol{w_{ij}y_i} + 101.25 - 0.381X_1 - 0.774X_2 + 0.00028X_3 - 0.653\sum_{j=1}^{n} w_{ij}\theta_{1j} - 1.46\sum_{j=1}^{n} w_{ij}\theta_{2j} + 0.00072\sum_{j=1}^{n} w_{ij}\theta_{3j}$$

The general factors that influence the open unemployment rate are minimum wage, labor force, economic growth, and education level. Meanwhile, the factors that influence the open unemployment rate based on the spatial effect (spatial dependency) are labor force participation rate, expected years of schooling, and per capita expenditure.

REFERENCES

[1] Badan Pusat Statistik, Pengeluaran Perkapita Provinsi Jawa Barat, Jawa Barat: BPS Provinsi Jawa Barat 2021.

[2] Badan Pusat Statistik, Tingkat Partisipasi Angkatan Kerja Provinsi Jawa Barat, Jawa Barat: BPS Provinsi Jawa Barat, 2021.

[3] Badan Pusat Statistik, Tingkat Pengangguran Terbuka Provinsi Jawa Barat, Jawa Barat: BPS Provinsi Jawa Barat, 2021.

[4] Badan Pusat Statistik, Angka Harapan Lama Sekolah Indonesia, Indonesia: Badan Pusat Statistik Indonesia, 2022.

[5] H. J. Miller. "Tobler's first law and spatial analysis". Annals of the association of American geographers, 94, 2, 284-289, 2004

[6] J. Lee, & S. D. Wong, Statistical Analysis with Arcview GIS. New York: John Willey & Sons. Inc, 2001.

[7] J. P. Lesage, Theory and Practice of Spatial Econometrics. United States: University of Toledo, 1999.

[8] L. Anselin, Spatial Econometrics: Methods and Models. Netherlands: Kluwer Academic Publishers, 1988.

[9] M. F. Goodchild, Spatial Autocorrelation, Norwich: Geobooks, 1989.

[10] M. N. Widyastuti, I. G. A. M. Srinadi, & M. Susilawati, "Pemodelan Jumlah Kasus Pneumonia Balita Di Jawa Timur Menggunakan Regresi Spatial Autoregressive Moving Average". E-Jurnal Matematika, 8, 3, 236-245, 8, 2019.

[11] M. Prachowny. "Okun's Law: Theoretical Foundations and Revised Estimates". Review of Economics and Statistics, 75, 2, 331–36, 5, 1993.

[12] R. Cahyandari, & N. Hisani, "Model Regresi Linier Berganda Menggunakan Penaksir Parameter Regresi Robust M-Estimator (Studi Kasus: Produksi Padi di Provinsi Jawa Barat Tahun 2009)". JURNAL ISTEK, 6, 1-2, 85-92, 6, 2012.

[13] R. V. D. Kruk, "A general Spatial ARMA Model: Theory and Application", Europian Regional Science Association Conference (ERSA), 2002.

[14] S. Rani, Tingkat Pengangguran di Jawa Barat Capai 2,13 Juta Orang, Tertinggi City Bogor, [10 November 2022] Available at: https://amp.kompas.com/bandung/read/2022/11/10/120339478/tingkat-pengangguran-di-jawa-barat-capai-213-juta-orang-tertinggi-City-bogor

[15] S. Sukirno, Makroekonomi Teori Pengantar, Edisi Ketiga, Jakarta: Raja Grafindo Persada, 2004

[16] T. Wuryandari., A. Hoyyi., D. S. Kusumawardani., & D. Rahmawati, "Identifikasi Autokorelasi Spasial pada Jumlah pengangguran Di Jawa Tengah Menggunakan Indeks Moran". Media Stat, 7, 1, 1-10, 6, 2014.

[17] Z. Feng, & W. Chen, "Environmental Regulation, Green Innovation, and Industrial Green Development: an Empirical Analysis Based on the Spatial Durbin Model", Sustainability (Switzerland), 10, 1, 1-22, 1, 2018.