JSDS: JOURNAL OF STATISTICS AND DATA SCIENCE
VOLUME 3, No 2, October 2024
e-ISSN: 2828-9986
https://ejournal.unib.ac.id/index.php/jsds/index

Factors Affecting The Open Unemployment Rate in West Sumatra Province Using Spatial Autoregressive (SAR)

Clara Febby Adellia[1*], Devni Prima Sari[2]
[1] Department of Mathematics, Padang State University, Padang

* Corresponding Author: clarafebbyadellia@gmail.com

Article Info

Article History:
Received: August, 8 2024
Accepted: August, 30 2024
Available Online: October, 30 2024

Key Words:
SAR
Open Unemployment Rate
K Nearest Neighbor
AIC

Abstract

This paper proposes a Spatial Autoregressive (SAR) model to analyze the significant factors affecting the open unemployment rate in West Sumatra during 2023. The main advantage of the method is its ability to accurately capture spatial interactions between neighboring regions, such that it can provide a comprehensive understanding of regional unemployment patterns efficiently. By introducing the K Nearest Neighbor (KNN) weighting matrix and spatial lag parameter to the model, the effect of regional proximity on unemployment rates is more accurately captured. The viability of the SAR model is assessed by analyzing its ability to produce the lowest Akaike's Information Criterion (AIC) value, indicating its suitability for modeling regional unemployment patterns. The result indicates that the SAR model is more effective than the multiple linear regression model in capturing regional unemployment patterns, with an AIC value of 52.756. The factors that influence the open unemployment rate are gross regional domestic product, labor force participation rate and the percentage of poor people.

1. INTRODUCTION

Unemployment is one of the most serious problems in development in Indonesia, especially in West Sumatra. The indicator used to measure the high unemployment rate is the open unemployment rate [1]. The open unemployment rate is the percentage of the number of unemployed people to the total labor force [2]. The negative impact of an increase in the open unemployment rate on people's lives is a decrease in people's income and increase in the crime rate, which can hamper the stability and welfare of society [3]. The open unemployment rate in West Sumatra reached 5.94% and has not yet reached the 2020-2024 medium-term development plan target of 5.3%, placing this Province ranked in the top 8 nationally with the highest open unemployment rate.

Previous research on the open unemployment rate by Arizal (2019) employed panel regression analysis and found that the human development index significantly affects open unemployment rate [4]. However, this study did not explore spatial dependence between regions, which is crucial since the open unemployment rate in one region is likely influenced by neighboring regions, as explained by Tobler's law. Tobler's law states that everything in space is interrelated, but its influence decreases with distance [5]. To solve problems that contain spatial dependencies (dependence between a set of observations or locations) is Spatial Autoregressive (SAR).

Spatial Autoregressive (SAR) is a form of model that combines a simple regression model with a spatial lag on the response with a spatial lag on the response variable, this is explained by Anselin in his book. A fundamental component of spatial models is the spatial weight matrix, which reflects the relationship between one region and another. This model can illustrate how much influence a variable has as a factor in increasing unemployment and shows the relationship between a region and other neighboring regions [6]. The disadvantage of the SAR model is that SAR models sensitive to outliers in the data.

Research on the open unemployment rate has been conducted with spatial and non spatial approaches. Agustina (2022) used spatial regression to analyze the factors affecting the open unemployment rate in Indonesia, with the result that the SAR model is appropriate, where gross regional domestic product growth and labor force participation rate are significant to the open unemployment rate. Isnayanti (2017) used the OLS model in North Sumatra and found that education level has a positive effect on TPT [7]. Ningtias (2017) found that the spatial error model is the best for the open unemployment rate in East Java, with the percentage of poor population as a significant variable [8].

Therefore, an analysis method that considers the spatial aspect is needed in dealing with the problem of open unemployment rate. This research uses Spatial Autoregressive (SAR) using data from BPS that covers 19 districts/cities in West Sumatra Province. The dependent variables is the open unemployment rate and the independent variables include gross regional domestic product, labor force participation rate, percentage of poor population, highest level of education completed in high school, and expenditure per capita. Thus this research discusses the factors affecting the open unemployment rate in West Sumatra Province using Spatial Autoregressive (SAR).

2. Theoretical Basic

2.1 Multiple Linear Regression

Multiple linear regression is a linear regression where the dependent variable (Y) is associated with at least two independent variables $\left(\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3, \dots, \mathrm{X}_\mathrm{p}\right)$. The general form of the multiple linear regression equation can be written as follows [9].

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \tag{1}$$

Where, y : dependent variable, β_0 : constant, $\beta_1, \beta_2, \dots, \beta_p$: regression coefficient value, $x_1, x_2, \dots, x_p$: independent variable, ε : error.

2.2 Spatial Weight Matrix

In modeling using spatial regression, the first thing to do is to determine the spatial weight matrix. The spatial weight matrix is a matrix that describes the relationship between a region and other regions. Matrix $\boldsymbol{W}$ is $n \times n$ matrix that shows how much influence one spatial observation unit has on another spatial observation units [10]. Information location can be known from two sources, namely relationship based on neighborhood and distance. Area linkages based on distance, the neighborhood relationship is determined by the distance between two areas [11]. The distance between locations can be determined using Euclid distance. Euclid distance is the shortest distance between two points in Euclidean space. Euclid distance can be obtained with the following formula :

$$d_{(i,j)} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \tag{2}$$

where, $d_{(i,j)}$: Euclidean distance between regions i and j, for $i, j = 1,2, \dots, n$, u_i : Latitude coordinate of the i-th region, u_j : Latitude coordinate of the j-th region, v_i : Longitude coordinate of the i-th region, v_j : Longitude coordinate of the j-th region.

In this study, the weight matrix that will be used is a k-location or k nearest neighbor spatial weight. The weight will be 1 for k nearest neighbors and 0 for those with no neighbors. The calculation of k nearest neighbors is done with the following steps: First calculate the Euclidian distance of location i to j; second sort the distances obtained; and third select the k locations with the closest distance as the optimum value obtained. Determining the value of k is first based on Moran's I statistics or Moran's index, the process is done in iteration. The value of k is selected based on the largest Moran's I value [12]. The weight matrix will be normalized so that each row will sum to 1, thus ensuring that the weight of each observation region is between 0 and 1. Normalization of weights by row is called row normalization.

$$\boldsymbol{W}_{ij} = \frac{\boldsymbol{W}_{ij}{}^*}{\sum_{j=1}^{n} \boldsymbol{W}_{ij}{}^*} \tag{3}$$

2.3 Moran's I Test

Moran's I is the most widely used method to calculate spatial autocorrelation globally. This method is used to detect the onset of spatial randomness that shows clustered patterns or trends in space [5].
Moran's index (Moran's I) can be formulated as follows:

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{W}_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{W}_{ij} \sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{4}$$

2.4 Spatial Autoregressive

Spatial regression model that combines linear regression with spatial lags in the response variable and shows the effect of spatial lags on the dependent variable using cross-section data. Spatial lag occurs when the observed values of the response variable at a certain location are correlated with the observed values of the response variable at surrounding locations. The SAR model is a linear regression model in which the response variable exhibits spatial correlation. The advantage of the SAR model is appropriate for spatial patterns with an area approach. The general form of the SAR model is as follows [5]:

$$\begin{gathered} \boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} \sim \boldsymbol{N}(0, \sigma^2 \boldsymbol{I}) \end{gathered} \tag{5}$$

Where, $\boldsymbol{y}$: vector of response variables of size $(n \times 1)$, ρ : spatial lag coefficient parameter of size (1×1), $\boldsymbol{W}$: spatial weighting matrix of size $(n \times n)$, $\boldsymbol{X}$: matrix of predictor variables of size $(n \times (p+1))$, $\boldsymbol{\beta}$: spatial regression coefficient parameter of size $(p+1) \times 1$, $\boldsymbol{\varepsilon}$: error vector of size $(n \times 1)$, $\boldsymbol{I}$: identity matrix of size $(n \times n)$.

The assumptions underlying the spatial autoregressive are :

1. Linearity
   The SAR model assumes a linear relationship between the dependent variable and the independent variables, as well as between the spatially lagged dependent variable and the dependent variable.
2. Satisfying the classical assumption tests
   The classical assumption tests are normality test, multicollinearity test, and homoscedasticity test.
3. Spatial dependence
   One of the primary assumptions of the SAR model is that there is spatial dependence in the data. The SAR model captures this spatial autocorrelation through a spatial lag in the response variable, represented by a spatial weights matrix (**W**), which defines the spatial relationships between different locations.
4. Correct specification of the spatial weights matrix
   The spatial weights matrix (**W**) must correctly represent the spatial structure of the data. This matrix defines the nature of spatial relationships between observations, such as which locations are considered neighbors and the weight of their influence.

In this study, parameter estimation of the spatial autoregressive model carried out using the Maximum Likelihood Estimation (MLE) method. Model parameter estimates are obtained by maximizing the likelihood function which is equivalent to maximizing the logarithm of the likelihood function in the equation likelihood function (L) in the following equation :

$$L = \frac{|\boldsymbol{I} - \rho \boldsymbol{W}|}{(2\pi)^{n/2} \sigma^n} exp \left[- \frac{(\boldsymbol{y} - \rho \boldsymbol{W} \boldsymbol{y} - \boldsymbol{X} \boldsymbol{\beta})^T (\boldsymbol{y} - \rho \boldsymbol{W} \boldsymbol{y} - \boldsymbol{X} \boldsymbol{\beta})}{2\sigma^2} \right]$$

$$\ln(L) = -\frac{n}{2} \ln(2\pi) - n \ln\sigma + \ln|\boldsymbol{I} - \rho \boldsymbol{W}| - \frac{(\boldsymbol{y} - \rho \boldsymbol{W} \boldsymbol{y} - \boldsymbol{X} \boldsymbol{\beta})^T (\boldsymbol{y} - \rho \boldsymbol{W} \boldsymbol{y} - \boldsymbol{X} \boldsymbol{\beta})}{2\sigma^2} \tag{6}$$

Parameter estimation $\sigma^2, \beta,$ dan ρ are obtained by maximizing the log likelihood function from the above equation. The first parameter that needs to be done is parameter β. The parameter estimation is found to be β is:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y} - (\boldsymbol{X}^T \boldsymbol{X})^{-1} \hat{\rho} \boldsymbol{W} \boldsymbol{y} \tag{7}$$

Estimator for σ^2 is obtained by deriving the log likelihood function against σ^2 as follows :

$$\hat{\sigma}^2 = \frac{\left(y - \rho W y - X\hat{\beta}\right)^T \left(y - \rho W y - X\hat{\beta}\right)}{n} \tag{8}$$

and for the estimator of the parameter ρ :

$$\hat{\rho} = (\boldsymbol{y}^T \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{y})^{-1} \boldsymbol{y}^T \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{y} \tag{9}$$

3. METHOD

This type of research is applied research and the data used is secondary data. The secondary data consists of data from 19 districts/cities in West Sumatra Province in 2023 obtained through the official publication of the Central Statistic Agency.

The steps taken in analyzing the data in this study are:

1. Data exploration
2. Multiple linear regression modeling with the following steps:
   a. Multiple linear regression parameter estimation based on (1)
   b. Simultaneous test of multiple linear regression models
   c. Partial test of multiple linear regression models
3. Classical assumption test
   a. Normality test using the Shapiro Wilk method
   b. Multicollinearity test
   c. Homoscedasticity test
4. Determining the spatial weight matrix with k nearest neighbor method
5. Spatial dependency test
   a. Moran's I test based on (4)
   b. Lagrange Multiplier (LM) Test
6. Spatial Autoregressive (SAR) modeling with the following steps:
   a. Calculate the parameter estimation of the SAR model based on (5)
   b. Testing the fit of the SAR model using the F test
   c. Testing the significance of SAR model parameters using the Wald test
7. Selection of the best model using Akaike's Information Criterion (AIC)

In its completion, the calculation of Spatial Autoregressive (SAR) uses the help of RStudio software.

4. RESULTS AND DISCUSSION

4.1. Data exploration

The first stage of research data exploration is the use of descriptive statistics which aims to obtain information and create an overview of the research data consisting of 19 districts/cities in West Sumatra Province. This descriptive statistics will be conducted on the data of the dependent variable (Y), namely the open unemployment rate. The independent variables used consist of gross regional domestic product (X_1), labor force participation rate (X_2), percentage of poor population (X_3), highest level of education completed in high school (X_4), and expenditure per capita (X_5) in West Sumatra Province in 2023.

Table 1 is descriptive statistics for dependent and independent variables, it can be seen the open unemployment rate (Y) in districts/cities in West Sumatra has an average of 5.11 percent, with a minimum value of 1.33 percent and a maximum value of 10.86 percent. Gross regional domestic product (X_1) has an average of 9700.52 billion rupiah, with a minimum value of 2747.1 billion rupiah and a maximum value of 47185.1 billion rupiah. The labor force participation rate (X_2) has an average of 69.24 percent, with a minimum value of 64.42 percent and a maximum value of 72.61 percent. The percentage of poor people (X_3) has an average of 5.9 percent, with a minimum value of

2.27 percent and a maximum value of 13.72 percent. The highest level of education completed by senior high school (X_4) has an average of 33.77 percent, with a minimum value of 24.78 percent and a maximum value of 45.46 percent. Per capita expenditure (X_5) has an average of 11261.21 thousand rupiah/person/year, with a minimum value of 6891 thousand rupiah/person/year and a maximum value of 15089 thousand rupiah/person/year.

Table 1. Descriptive Statistic

Variable	Mean	Standard Deviation	Minimum	Maximum
Y	5.11	1.86	1.33	10.86
X_1	9700.52	985.00	2747.10	47185.10
X_2	69.24	2.59	64.42	72.61
X_3	5.90	2.39	2.27	13.72
X_4	33.77	6.36	24.78	45.46
X_5	11261.21	1937.00	6891.00	15089.00

4.2. Multiple linear regression modeling

a. Multiple Linear Regression Parameter Estimation

Multiple linear regression was used to evaluate the relationship between the open unemployment rate and several independent variables that were considered influential.

Table 2. Multiple Linear Regression Parameter Estimation

Variable	Estimation	Standard Error	T-Value	$p-value$
(*Intercept*)	16.900000	13.600000	1.24	0.236000
X_1	0.000149	0.000029	5.17	0.000179
X_2	-0.137000	0.147000	-0.93	0.369000
X_3	-0.412000	0.209000	-1.97	0.071000
X_4	-0.028000	0.067800	-0.41	0.687000
X_5	-0.000033	-0.000257	-0.13	0.901000

Table 2 shows that X_1 and X_3 are significant, with $p-value$ smaller than the significant level $\alpha = 10\%$. The multiple linear regression model, formed based on the estimation results shown in Table 2, is as follows:

$$\hat{y} = 16.9 + 0.000149X_1 - 0.412X_3$$

b. Simultaneous Test of Multiple Linear Regression Models

The simultaneous parameter test is a test to determine whether all predictor variables that enter the model have a significant effect together on the model using F test with the hypothesis :

$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ (model parameters are not significant)
$H_1: \beta_j \neq 0$, for at least $j = 1,2,3, \dots, k$ (model parameters are significant)

Table 3. F-Test

Variable	F-Value	$p-value$
Regression	10.66	0.0003089

The F Test results show that the multiple linear regression model is significant, with a p-value of 0.0003089 which is smaller than the significant level α=10%. So it can be concluded that there is at least one predictor variable that has a significant effect on the factors affecting the unemployment rate in West Sumatra Province in 2023.

The R^2 value of the model is 80.39% which means that 80.39% of the total variation in y is explained by x and there is still 19.61% more variation in y that cannot be explained by the multiple linear regression model. The remaining 19.61% may be due to other factors that failed to be accounted for in the model.

c. Partial Test of Multiple Linear Regression Models

Partial test of the model can be done using the T test. The T test is used to test each predictor variable individually in the model with the hypothesis.

$H_0: \beta_j = 0$ (model parameter is not significant)
$H_1: \beta_j \neq 0$, with $j = 1,2, \ldots, k$ (model parameters are significant)

The partial test results of multiple linear regression models shown in Table 2 indicate that, with a significant level of α=10%, there are two predictor variables that have a significant effect on the level of open unemployment in West Sumatra Province in 2023, namely X_1 (Gross Regional Domestic Product) and X_3 (Percentage of Poor Population).

4.3. Classical Assumption Test

a. Normality test using the Shapiro Wilk method

To find out whether the data is normally distributed or not, the Shapiro Wilk test can be done with the hypothesis [13]:

H_0: residuals are normally distributed
H_1: residuals are not normally distributed

Based on the computational results with RStudio software, the test result results is obtained with a $p - value$ of 0.4657, which is greater than the significant level α = 10%. Thus, it can be concluded that the residuals are normally distributed.

b. Multicollinearity Test

Multicollinearity test aims to see if there is a correlation between predictor variables. This test uses the *Variance Inflation Factor* (VIF) to measure the level of multicollinearity [14].

Table 4. Multicollinearity Test

Variable	VIF
X_1	1.556
X_2	2.756
X_3	4.767
X_4	3.550
X_5	4.713

At the Table 5, it can be seen that he VIF values for all predictor variables are less than 10, so it can be concluded that there is no multicollinearity in the predictor variables in the model.

c. Homoscedasticity Test

The homoscedasticity test aims to ensure that the residual variance of the regression model is constant. Homoscedasticity is an important condition in linear regression, because heteroscedasticity (residual variance that is not constant) can lead to inefficient estimation results and unclear conclusions. The hypothesis in the homoscedasticity test is as follows :

$H_0: {\sigma_1}^2 = {\sigma_2}^2 = \ldots = {\sigma_n}^2$ (Homoscedasticity)
H_1: there is at least one ${\sigma_i}^2 \neq {\sigma_j}^2$ where $i \neq j$ (Heteroscedasticity)

Table 5. Homoscedasticity Test

BP	DF	$p - value$	Description
2.3527	5	0.7985	Accept H_0

The homoscedasticity test results in Table 6 using the BP test show that the $p - value$ of 0.7985 is greater than the significant level $\alpha = 10\%$. This indicates that H_0 is accepted, which means there is no heteroscedasticity problem in the model. In other words, the residual variance is constant and the condition of homoscedasticity is met.

4.4. Determining the spatial weight matrix with k nearest neighbor method

A spatial weighting matrix is a matrix that describes the relationship between a region and other regions. In this research, the spatial weighting matrix used is the K-Nearest Neighbors (KNN) matrix with a matrix size of 19×19. Calculating the k nearest neighbors is done with the following steps: First, calculate the Euclidean distance of location

i to j; second, sort the distances obtained and last, select the k locations with the closest distance as the optimum value. In this study, experiments to determine the value of k were conducted for k = 1, 2, 3, 4, and 5. The greater the value of k, the smaller the Moran's I value obtained. Based on the experimental results, the largest Moran's I value is 0.148 with k = 3, which indicates that the optimum k value is 3.

Figure 1. K Nearest Neighbor Spatial Weighting with K=3

For example, if we look at Agam Regency (1) in Figure 1, this region is connected to three other nearby regions, namely Lima Puluh Kota Regency (4), Bukittinggi City (13) and Padang Pariaman Regency (5). This relationship shows that based on distance calculations, these three regions are the closest neighbors of Agam Regency (1).

4.5. Spatial Dependency Test

a. Moran's I Test

This method is used to detect the onset of spatial randomness that shows clustered patterns or trends in space. The Morans's I test is used to assist in selecting an appropriate spatial model. To calculate the Moran's I index using equation (4) and the hypothesis :

$H_0: I = 0$ (no dependency between locations)
$H_1: I \neq 0$ (there are dependencies between locations)

Table 6. Moran's I Test

Variable	Moran's I Value	$p - value$	Conclusion
Y	-0.1984	0.8398	Accept H_0
X_1	-0.1670	0.7810	Accept H_0
X_2	0.1487	0.0776	Reject H_0
X_3	0.0711	0.1891	Accept H_0
X_4	-0.1085	0.6437	Accept H_0
X_5	-0.0836	0.5775	Accept H_0

At Table 7 shows the results of the Moran's I test using a K-NN weighting matrix with k=3 and shows that the variable that has spatial autocorrelation between locations is variable X_2, namely the labor force participation rate with a $p - value = 0.0776$ significant at $\alpha = 10\%$. This indicates that there is a spatial clustering or linkage pattern in the labor force participation rate variable among neighboring regions. In contrast, variables other than labor force participation rate have insignificant Moran's I values. This means that there is no significant spatial autocorrelation for these variables.

b. Lagrange Multiplier Test

The Lagrange Multiplier test is conducted to determine whether there is spatial dependence in the regression model. This test helps determine the most suitable spatial model based on two different types of spatial dependence. If LM_{lag} is significant then the appropriate model is SAR, if LM_{error} s significant then the appropriate model is SEM and if both are significant then the appropriate model is SARMA (Spatial Autoregressive Moving Average). If both are not significant, then the appropriate model is multiple linear regression analysis. The SAR hypothesis:

$H_0: \rho = 0$ (no spatial lag dependence)
$H_1: \rho \neq 0$ (there is spatial lag dependence)

Table 7. Lagrange Multiplier Test

Spatial Dependence Test	$p-value$	Decision
LM_{lag}	0.0446	Reject H_0
LM_{error}	0.5449	Accept H_0

Table 8 shows the results of the Lagrange Multiplier test using the K-NN weighting matrix showing that the p-value on LM_{lag} 0.0446 which is significant at $\alpha = 10\%$. This means there is sufficient evidence to reject H_0 and indicates the presence of spatial lag dependence in the data. In contrast, the $p-value$ of LM_{error} is 0.5449 which is not significant at $\alpha = 10\%$, so the null hypothesis for spatial dependence in the error term cannot be rejected. Thus, these results indicate that the Spatial Autoregressive (SAR) model is the most appropriate model for this analysis, as it can handle the spatial dependence in the dependent variable detected by LM_{lag}.

4.6. Spatial Autoregressive Modeling

a. Calculate the parameter estimation of the SAR

Parameter estimation in the SAR model aims to determine the effect of independent variables on the dependent variable by considering spatial effects, namely the influence of the value of the dependent variable in adjacent locations (spatial lag).

Table 8. Parameter Estimation of The SAR

Parameter	Estimation	Z value	$p-value$
$\hat{\rho}$	-0.583690	-2.892	0.015456
$\widehat{\beta_0}$	28.690000	2.9073	0.003645
$\widehat{\beta_1}$	0.000149	7.4553	8.97e-14
$\widehat{\beta_2}$	-0.248710	-2.3620	0.018178
$\widehat{\beta_3}$	-0.416120	-2.8814	0.003959
$\widehat{\beta_4}$	-0.023274	-0.4979	0.618554
$\widehat{\beta_5}$	-0.000132	-0.7452	0.456143

Table 9 shows the parameter estimation results of the Spatial Autoregressive (SAR) model using a k-nn weighting matrix with $k = 3$. Variables that have a $p-value$ less than $\alpha = 10\%$ mean that the independent variable has a significant effect on the dependent variable. It can be seen that there are 3 independent variables that affect the dependent variable, namely gross regional domestic product (X_1), labor force participation rate (X_2) and percentage of poor people (X_3). Meanwhile, the other independent variables have no significant effect on the open unemployment rate. The SAR equation model formed is as follows :

$$\hat{y}_i = -0{,}58369 \sum_{j=1, i\neq j}^{19} W_{ij} y_j + 28{,}69 + 0{,}000149 X_{1i} - -0{,}24871 X_{2i} - 0{,}41612 X_{3i}$$

b. Testing the fit of the SAR model using the F test

The SAR model fit test can be performed using the F test with the hypothesis :

$H_0: \beta_j = \rho = 0$ (Multiple linear regression model is better)

$H_1: \beta_j = \rho \neq 0$ and there is at least one $\beta_j \neq 0, j = 1,2, \dots, k$ (SAR Model is better)

Testing the fit of the SAR model using the F Test it can be seen with rho (ρ), the $p - value$ of 0.015456 indicates that the model is significant with $\alpha = 10\%$. So it can be concluded that H_0 is rejected and the SAR model formed is appropriate.

c. Parameter Estimation of Spatial Autoregressive (SAR) on significant variables

In the initial analysis, the SAR model was estimated using all independent variables in the study. The initial estimation results show that some independent variables do not have a significant effect on the dependent variable. Based on the significance test, the insignificant variables were identified and removed from the model. This estimation aims to obtain more precise model parameters and avoid variables that do not contribute significantly to the influence of the open unemployment rate variable [15].

Table 9. Parameter Estimation of Spatial Autoregressive (SAR) on significant variables

Parameter	Estimasi	Z value	p-value
$\hat{\rho}$	-0.54432	-2.6066	0.022732
$\widehat{\beta_0}$	19.846	3.6035	0.000314
$\widehat{\beta_1}$	0.000138	8.1668	2.220e-16
$\widehat{\beta_2}$	-0.16598	-2.3418	0.019189
$\widehat{\beta_3}$	-0.28925	-3.9652	0.000073

Table 10 shows the he parameter estimation results of the Spatial Autoregressive (SAR) model using significant variables and k-nn weighting matrix with k=3 as follows :

$$\hat{y}_i = -0,54432 \sum_{j=1, i\neq j}^{19} W_{ij} y_j + 19.846 + 0.000138 X_{1i} - 0.16598 X_{2i} - 0.28925 X_{3i}$$

From the SAR model obtained, there is one variable that has a positive effect, namely gross regional domestic product (X_1). Which means that with the increase of gross regional domestic product, the open unemployment rate in West Sumatra Province in 2023 will also increase. Meanwhile, the variables of labor force participation rate and the percentage of poor people have a negative effect on the open unemployment rate. If the labor force participation rate and the percentage of poor people increase, there will be a decrease in the value of the open unemployment rate in West Sumatra Province in 2023. The result of the parameter ρ (rho) shows that a decrease in the influence of the area surrounding a district/city will decrease the percentage of open unemployment rate in a district/city by -0.54432.

d. Testing the significance of SAR model parameters using the Wald test

To determine whether the parameters in the SAR model are statistically significant, a parameter significance test is conducted using the Wald test. This test can be done with hypothesis :

$H_0: \hat{\beta}_j = 0$ (Parameteris not significant)

$H_1: \hat{\beta}_j \neq 0$ (Parameter is significant

Table 10. Wald Test

Wald	$p - value$	Description
6.7946	0.0091433	Reject H_0

The Wald test results presented in Table 11 show a Wald statistical value of 6.7946 with a p-value of 0.0091433. Since the p-value is smaller than the significance level α=10%, the null hypothesis is rejected. This means that the parameters tested in the SAR model are statistically significant and contribute to explaining the variation in the open unemployment rate in West Sumatra Province. Thus, the variables associated with these parameters do have a significant influence on the dependent variable in the SAR model.

4.7. Selection of the best model using Akaike's Information Criterion (AIC)

Akaike Information Criterion (AIC) is a measure used to compare statistical models, where a model with a lower AIC value is considered better because it shows a balance between the accuracy and complexity of the model.

Table 11. AIC Value

Model	AIC
Multiple linear regression	55.945
SAR	52.756

Based on the AIC values displayed in Table 12, it can be seen that the SAR model has a lower AIC value of 52.756 compared to the multiple linear regression model with AIC of 55.945. The lower AIC value in the SAR model indicates that this model is better at explaining data variations by considering the complexity of the model. This means that the SAR model is more appropriate to use in this study because it is able to capture the relationship between the variables studied, especially with the spatial linkage factor between regions.

5. CONCLUSION

Based on the research results that have been, obtained, it can be concluded as follows :

1. From the results of modelling the factors affecting the open unemployment rate in West Sumatra Province using Spatial Autoregressive (SAR) with k-nn weighting matrix, it can be concluded that the Spatial Autoregressive (SAR) model is better and more appropriate to use than multiple linear regression, because SAR has a smaller AIC value than the AIC value of multiple linear regression, which is 52.756. The SAR model equation obtained is as follows:

$$\hat{y}_i = -0{,}54432 \sum_{j=1,i\neq j}^{19} W_{ij}y_j + 19{,}846 + 0{,}000138X_{1i} - 0{,}16598X_{2i} - 0{,}28925X_{3i}$$

   where the gross regional domestic product variable (X_1) has a positive and significant relationship to the open unemployment rate in West Sumatra Province. Meanwhile, the labour force participation rate variable (X_2) and the percentage of poor people (X_3) have a negative and significant effect on the open unemployment rate in West Sumatra Province.
2. Factors that influence the open unemployment rate in West Sumatra Province using Spatial Autoregressive (SAR) based on the SAR model that has been done are gross regional domestic product, labour force participation rate and percentage of poor population.

REFERENCES

[1] W. Gessan Septiyanto and E. Tusianti, "Analisis Spasial Tingkat Pengangguran Terbuka di Provinsi Jawa Barat," *Jurnal Ekonomi Indonesia* , vol. 9, 119–131, 2020.

[2] Badan Pusat Statistik, *Provinsi Sumatera Barat dalam Angka 2023.* BPS Provinsi Sumatera Barat, 2023.

[3] R. Ardian, U. Sultan, A. Tirtayasa, M. Syahputra, and D. Dermawan, "Pengaruh Pertumbuhan Ekonomi terhadap Tingkat Pengangguran Terbuka di Indonesia," *Jurnal Ekonomi, Bisnis dan Manajemen*, vol. 1, no. 3, 2022.

[4] M. Arizal and M. Marwan, "Pengaruh Produk Domestik Regional Bruto dan Indeks Pembangunan Manusia terhadap Tingkat Pengangguran Terbuka di Provinsi Sumatera Barat," *Jurnal EcoGen*, vol. 2, no. 3, pp. 433-442, 2019.

[5] L. Anselin, *Spatial Econometrics: Methods and Models*, vol. 4. in Studies in Operational Regional Science, vol. 4. Dordrecht: Springer Netherlands, 1988. doi: 10.1007/978-94-015-7799-1.

[6] J. LeSage and R. Kelley Pace, “Introduction to Spatial Econometrics,” 2009.

[7] N. D. I. A. Ritonga, “Analisis Faktor-Faktor yang Mempengaruhi Tingkat Pengangguran Di Provinsi Sumatera Utara Tahun 1978-2014 Dengan Metode Ordinary Least Square,” *Karismatika: Kumpulan Artikel Ilmiah, Informatika, Statistik, Matematika dan Aplikasi*, Vol 3, No 2 , 2017.

[8] I. P. Ningtias and S. P. Rahayu, “Pemodelan Faktor-faktor yang Mempengaruhi Tingkat Pengangguran Terbuka di Provinsi Jawa Timur Tahun 2015 Menggunakan Regresi Spasial,” *Jurnal Sains dan Seni ITS*, vol. 6, no. 2, 2017.

[9] R. K. Sembiring, *Analisis Regresi Edisi Kedua*. Bandung: ITB, 1995.

[10] D. Novitasari and L. Khikmah, "Penerapan Model Regresi Spasial pada Indeks Pembangunan Manusia (IPM) di Jawa Tengah," *Jurnal Statistika*, vol. 19, no. 2, pp. 123-134, 2019.

[11] R. Kosfeld, *Spatial Econometric*. Germany: University of Kassel, 2006.

[12] I. Gede Nyoman Mindra Jaya and B. Tantular, “Optimalisasi Matrik Bobot Spasial Berdasarkan K-Nearest Neighbor dalam Spasial Lag Model,” 2017.

[13] T. Cahyono, *Statistik Uji Normalitas*. Purwokerto: Yayasan Sanitarian Banyumas, 2015.

[14] D. N. Gujarati and D. C. Porter, *Basic Econometrics*, 5th ed. Americas: McGraw-Hill, 2009.

[15] H. Yasin, A. R. Hakim, and W. Budi, *Regresi Spasial (Aplikasi dengan R)*. Wade Group, 2020.