STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 7, No. 3, December 2025
2nd article, Pages: 292 – 307
DOI: https://doi.org/10.32493/sm.v7i3.54480

MODELING SPATIAL DEPENDENCE OF POVERTY BASED ON HUMAN DEVELOPMENT INDEX AND UNEMPLOYMENT RATE IN SOUTH SULAWESI PROVINCE 2024

Elisabeth Evelin Karuna[1,*], Mahrani[2], and Atiqah Azza El Darman[3]

[1,2,3]Public Administration Study Program, Faculty of Social Sciences and Law, University of Makassar, A.P. Pettarani Street, Makassar City, 90222, South Sulawesi, Indonesia

[*]Correspondance author: elisabeth.evelin@unm.ac.id

ABSTRACT

Poverty is one of the main problems in regional economic development that requires in depth statistical analysis to understand the factors that influence it. This study aims to model the spatial dependence of the percentage of poor people in South Sulawesi Province in 2024 by considering the influence of the Human Development Index (HDI) and the Open Unemployment Rate (OUR). The analysis was conducted using three spatial regression model approaches, namely the Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), and Spatial Autoregressive Moving Average Model (SARMA). The selection of the best model was based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values. The results of the analysis show that the SARMA model has the smallest AIC and BIC values (AIC = 102.6484 and BIC = 109.7167), so it was selected as the best model to explain the variation in poverty between regions in South Sulawesi. Parameter estimates in the SARMA model show that HDI has a significant negative effect on the percentage of poor people ($\beta = -0.3554$; $p_{value} = 0.0213$), while OUR has no significant effect ($X_2 = -0.1393$; $p_{value} = 0.6303$). In addition, there is spatial dependence in both the dependent variable ($\rho = -0.3029$) and the error component ($\lambda = 0.5572$), indicating the existence of interrelated poverty between regions. Thus, the results of this study confirm that an increase in HDI plays an important role in reducing poverty levels in South Sulawesi, and indicate the need for a development policy approach that considers spatial interrelationships between regions.

Keywords: Poverty, HDI, OUR, Spatial Regression, SARMA

ARTICLE INFO	
Submission received: 03 November 2025	Accepted: 30 December 2025
Revised: 28 December 2025	Published: 31 December 2025
Available on: https://doi.org/10.32493/sm.v7i3.54480	

1. INTRODUCTION

Poverty remains one of the complex socio-economic problems that continue to be faced by various regions in Indonesia. Although various poverty alleviation programs have been implemented by the government, regional disparities show that the distribution of poverty is not random but has a specific spatial pattern. This condition indicates the existence of spatial dependence, whereby the poverty level of a region is influenced by the conditions of the surrounding regions. Therefore, conventional statistical approaches such as OLS regression are often insufficient to accurately describe this relationship because they ignore the spatial effects that may arise between regions.

In the context of human development, the Human Development Index (HDI) and the

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 7, No. 3, December 2025
2nd article, Pages: 292 – 307
DOI: https://doi.org/10.32493/sm.v7i3.54480

open unemployment rate are two important indicators that are closely related to poverty. The HDI reflects a region's achievements in three main dimensions, namely longevity and healthy living; knowledge; and decent living standards (BPS, 2025). Meanwhile, the open unemployment rate is an indicator that shows the number of unemployed people in a community (Ardian, Syahputra, & Dermawan, 2022). Regions with low HDI and high unemployment rates generally have higher poverty rates (BPS, 2025). However, this relationship can vary between regions due to spatial influences, such as geographical proximity and economic linkages between districts/cities.

South Sulawesi Province is one of the provinces in eastern Indonesia with diverse socioeconomic characteristics. The city of Makassar, for example, shows a high level of human development and a relatively low poverty rate, while several districts in coastal and inland areas still face considerable development challenges (BPS, 2025). This pattern indicates the existence of spatial autocorrelation, whereby the poverty rate in a particular area does not stand alone but is influenced by conditions in the surrounding areas. Therefore, poverty analysis in South Sulawesi requires an approach that explicitly considers spatial aspects.

Several previous studies have examined this phenomenon. For example, Putra & Arka (2016) found that the open unemployment rate had a positive and significant effect on the poverty rate. Sari, Nasution, & Sihombing (2021) conducted an analysis of the factors affecting poverty in West Sumatra Province using spatial regression. The results showed that the best model was the SAR model.

On this background, this study aims to model the spatial dependence of poverty in South Sulawesi Province in 2024 by considering the influence of the Human Development Index (HDI) and the Open Unemployment Rate (OUR). The analysis was conducted using a spatial regression approach, namely the Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), and Spatial Autoregressive Moving Average (SARMA), to determine the extent to which spatial dependence plays a role in the distribution of poverty between regions. The results of this study are expected to provide a more comprehensive understanding of the spatial dynamics of poverty and serve as a basis for the formulation of more targeted and region-based development policies.

2. MATERIAL AND METHODS

This section will explain the data sources, variables used, and analysis methods applied in this study. The description covers the statistical and spatial techniques used to examine the relationship between poverty levels and the factors that influence them in South Sulawesi Province in 2024. The procedures described are expected to ensure the research and provide a strong and scientifically transparent basis for analysis.

2.1 Ordinary Least Squares Regression

Regression analysis is an analysis that serves to explore the relationship between variables and make accurate estimates based on that relationship (Chatterjee & Simonoff, 2013). In other words, regression analysis is a statistical method used to analyze the extent to which predictor variables influence response variables in a cause-and-effect relationship (Rinindah et al., 2025). The general form of a multiple linear regression model can be written as:

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 7, No. 3, December 2025
2nd article, Pages: 292 – 307
DOI: https://doi.org/10.32493/sm.v7i3.54480

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

where y is the dependent variable or response variable, while x_1, x_2,…, x_k are the independent variables.

Before estimating and interpreting regression models, it is important to test classical assumptions. These tests aim to ensure that regression coefficient estimates using the Ordinary Least Squares (OLS) method produce unbiased, efficient, and consistent estimators. Several assumptions that need to be tested include normality, autocorrelation, heteroscedasticity, and multicollinearity. The following is an explanation of each classical assumption test:

2.1.1 Normality Test

The Shapiro and Wilk (1965) test was originally limited to sample sizes of less than 50. The Shapiro-Wilk test is more recommended for small samples (n less than or equal to 50). This test is the preferred choice due to its good test power (Razali & Wah, 2011). The hypotheses used are H_0: Data is normally distributed vs H_1 : Data is normally distributed with the following test statistics:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2}$$

where $x_{(i)}$ is the i^{th}order statistic, $\bar{x}$ is the sample mean, $a_i = (a_1, a_2, \ldots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$, where $m = (m_1, m_2, \ldots, m_n)^T$ is the expected value of the order statistics of independent and identically distributed random variables drawn from the standard normal distribution, and V is the covariance matrix of these order statistics.

2.1.2 Autocorrelation Test

Various statistical tests can be used to detect the presence of autocorrelation. One widely used procedure is the test developed by Durbin and Watson, namely the Durbin Watson test. The hypotheses used are H_0 : $\rho = 0$ (No autocorrelation) vs H_1 : $\rho \neq 0$ (There is autocorrelation) with the following test statistics (Montgomery et al., 2012):

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} {e_t}^2} \tag{3}$$

This Durbin Watson value (d) will later be compared with two Durbin Watson table values, namely Durbin Upper (d_U) and Durbin Lower (d_L).

2.1.3 Heteroscedasticity Test

One test that can be used to test for heteroscedasticity is the Breusch-Pagan test. This test is used to test the null hypothesis that the error in a regression model is homoscedastic. Rejecting the null hypothesis indicates the presence of heteroscedasticity. The hypotheses used are H_0 : There is no heteroscedasticity vs. H_1 : There is heteroscedasticity with the following test statistics (Halunga et al., 2017):

$$BP_T = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij,T}^2 \tag{4}$$

when it is assumed that the cross-section units are independent and the residual variance is homogeneous over time (homoscedastic), then the statistical value $BP_T \to \chi_v^2$ with degrees of freedom equal to $v = \frac{1}{2}\, N(N - 1)$ for a fixed number of units N and a large time period T.

2.1.4 Multicollinearity Test

Multicollinearity indicates the presence of almost perfect linear dependence among independent variables (regressors). The presence of almost perfect linear dependence can greatly affect the ability to estimate regression coefficients. Therefore, to determine or detect the presence of multicollinearity, a test can be performed by looking at the Variance Inflation Factors (VIF) values. In general, the VIF for the jth regression coefficient can be written as follows (Montgomery et al., 2012):

$$VIF = \frac{1}{(1 - R_j^2)} \quad ; \; j = 1,2, \dots, k \tag{5}$$

where R_j^2 is the multiple coefficients of determination obtained from the regression of variable x_j against all other independent variables (regressors). VIF greater than 10 indicates a problem of multicollinearity.

After testing the classical assumptions, the next step is to test the significance of the parameters, both simultaneously and partially. The regression parameter significance test is used to determine whether there is a linear or significant relationship between the dependent variable y and one of the independent variables (regressors) x_1, x_2,…, x_k. Simultaneous testing (F-test) can be performed with the hypothesis used being H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ vs H_1: At least $\beta_j \neq 0, j = 1, \dots, k$ with the test statistic used being (Montgomery et al., 2012):

$$F = \frac{SS_R/k}{SS_{Res}/(n - k - 1)} = \frac{MSR}{MSE} \tag{6}$$

This statistic follows an F distribution with degrees of freedom $(k, n - k - 1)$. Rejection of the null hypothesis (H_0) indicates that at least one explanatory variable (x_1, x_2,…, x_k) contributes significantly to the model. Meanwhile, for the partial test (t-test), the hypotheses used are H_0: $\beta_j = 0$ vs H_1: $\beta_j \neq 0, j = 1, \dots, k$ with the test statistic used being (Montgomery et al., 2012):

$$t = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \tag{7}$$

where the null hypothesis will be rejected if the test statistic value $|t| > t_{\frac{\alpha}{2}, n-k-1}$.

2.2 Spatial Autocorrelation Test (Moran's Index)

Starting from a concise mathematical form, we can deduce a clear new relationship for spatial autocorrelation analysis. Suppose there are n regional units, such as cities or districts, in a geographical area. To measure the level of spatial interdependence between these regions, the Moran Index is used. By considering variables that have been adjusted to population size and a spatially weighted matrix that has been globally normalized, the Moran Index value can be expressed in the following equation (Chen, 2023):

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 7, No. 3, December 2025
2nd article, Pages: 292 – 307
DOI: https://doi.org/10.32493/sm.v7i3.54480

$$I = z^T W z \tag{8}$$

where I refers to the Moran Index, $\mathbf{z} = [z_1, z_2, \dots, z_n]^T$ is a vector of standardized measures with z-scores, $W = [w_{ij}]]$ is an $n \times n$ symmetric and globally normalized spatial weight matrix, and the superscript T indicates the transposition of the matrix or vector $(i, j = 1, 2, \dots, n)$. Moran's I values range from -1 to +1. Negative values indicate negative spatial autocorrelation, while positive values indicate positive spatial autocorrelation. Values close to zero indicate that the spatial pattern is random or that there is no spatial correlation between regions (Das & Ghosh, 2016). To test Moran's I, the following hypothesis is first determined: H_0: there is no spatial autocorrelation vs H_1: there is spatial autocorrelation, with the global Moran's I test statistic value standardized to a Z value, which can be calculated using the following formula (Mathur, 2015):

$$Z = \frac{I - E(I)}{\sqrt{V(I)}} \tag{9}$$

where $E(I)$ is the expected value of Moran's I and $V(I)$ is the variance of Moran's value. Reject H_0 if the Z value is greater than 1.96 or less than –1.96 or $p_{value} < \alpha$, which means there is significant spatial autocorrelation.

2.3 Spatial Regression

Spatial analysis is used to analyze data related to geographical locations or the position of an area on the earth's surface, taking into account the relationships and patterns that emerge between these areas (Aswi et al., 2021). In classical linear regression models, spatial dependence can be incorporated through two main approaches: first, as an additional explanatory variable representing the dependent variable with spatial lag (Wy), or second, through an error component that contains a spatial dependence structure (Anselin, 1999).

2.3.1 Lagrange Multiplier (LM)

The LM test is used to test for the presence of spatial dependence in the model and is formulated in a specific form according to the relationship between the residuals and the spatial structure represented by the spatial weight matrix. The Lagrange Multiplier (LM) test can be written separately for spatial dependence on lags and errors. Lagrange Multiplier Lag, the hypothesis used is $H_0 : \rho = 0$ (no spatial dependence on lag) vs $H_1 : \rho \neq 0$ (there is spatial dependence on lag) with the test statistics as follows (LeSage, 1998):

$$LM_{lag} = \frac{\left[e'WY/(\frac{ee'}{N})\right]^2}{D} \tag{10}$$

with $D = [\{(WX\beta)'(I - X(X'X)^{-1}X')(WX\beta)\}/\sigma^2] + \mathrm{tr}(\mathrm{W'W} + \mathrm{WW})$. Reject H_0 if the value of $LM_{lag} > \chi^2{}_{\alpha(1)}$. Meanwhile, for Lagrange Multiplier Error, the hypotheses used are $H_0 : \lambda = 0$ (no spatial dependence on error) vs $H_1 : \lambda \neq 0$ (there is spatial dependence on error) with the following test statistics (LeSage, 1998):

$$LM_{err} = \frac{\left[e'We/(\frac{ee'}{n})\right]^2}{\mathrm{tr}(\mathrm{W'W} + \mathrm{WW})} \tag{11}$$

where H_0 will be rejected if the value of $LM_{err} > \chi^2{}_{\alpha(1)}$. If significant at lag, then the model formed is a Spatial Autoregressive Model (SAR). Meanwhile, if significant at error, then the model formed is a Spatial Error Model (SEM).

2.3.2 Spatial Autoregressive Model (SAR)

The Spatial Autoregressive Model (SAR) is also known as the spatial lag model and is used when the focus of research is to assess the existence and strength of spatial interactions between regions. This model describes the existence of substantive spatial dependence, namely direct relationships between regions that are reflected in the spatial model structure (Anselin, 1999). The SAR model is formed when the value of $\rho \neq 0$ and $\lambda = 0$. Formally, the SAR model can be expressed by the following equation (LeSage, 1998):

$$y = \rho W y + X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2 I_n) \quad (12)$$

where y is an $n \times 1$ vector containing the dependent variable, X is an $\times k$ data matrix containing the explanatory variables, and W is a known spatial weight matrix, usually in the form of a first-order contiguity matrix. The parameter ρ (rho) is the predictor coefficient of the spatial lag model, while the parameter β (beta) represents the effect of explanatory variables on the variation in the dependent variable y.

2.3.3 Spatial Error Model (SEM)

Spatial dependence that arises in regression error components is known as the Spatial Error Model (SEM). This model is appropriate when the purpose of the analysis is to correct for the effects of spatial autocorrelation, which can cause bias due to the use of spatially related data, regardless of whether the main model is spatial or not (Anselin, 1999). The spatial error model is formed when $\rho = 0$ and $\lambda \neq 0$. The SEM model can be expressed as follows (LeSage, 1998):

$$y = X\beta + u \quad (13)$$

$$U = \lambda W u + \varepsilon, \varepsilon \sim N(0, \sigma^2 I_n) \quad (14)$$

Vector y contains $n \times 1$ vectors representing dependent variables, while X is an $n \times k$ data matrix containing explanatory (independent) variables. W is a known spatial weight matrix, and the parameter λ (lambda) is a coefficient that describes the spatial correlation in the error component, similar to the serial autocorrelation problem in time series models. Meanwhile, the parameter β (beta) reflects the influence of explanatory variables on the variation in the dependent variable y.

2.4 Types and Sources of Data

The data used in this study is secondary data obtained from the official website of the Central Statistics Agency (BPS) of South Sulawesi Province. The data used covers several key socioeconomic variables, namely the Percentage of Poor Population, Human Development Index (HDI), and Open Unemployment Rate (OUR) by district/city in South Sulawesi Province in 2024.

2.5 Research Variables

This study uses two types of variables, namely dependent variables and independent variables, each of which plays a role in explaining the relationship between socioeconomic factors in South Sulawesi Province. The dependent variable in this study is the Percentage of Poor Population (Y). The independent variables in this study consist of: Human

StatMat
Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 7, No. 3, December 2025
2nd article, Pages: 292 – 307
DOI: https://doi.org/10.32493/sm.v7i3.54480

Development Index (X_1) and Open Unemployment Rate (X_2).

2.6 Analysis Steps

The following are the stages of analysis conducted using spatial regression.

1. **Descriptive Statistics**
   This stage aims to provide an overview of the characteristics of the research variable values in each district/city in South Sulawesi Province. Descriptive analysis is used to identify data distribution, minimum and maximum values, and the average value of each variable studied.
2. **Correlation Analysis**
   This stage aims to determine whether there is a correlation between the variables used, as an initial indication for further analysis.
3. **Visualization and Exploration of Spatial Data**
   This stage aims to understand the geographical patterns of each research variable before conducting a more in-depth statistical analysis. This visualization is carried out by mapping the data into thematic maps based on the administrative boundaries of districts/cities in South Sulawesi Province.
4. **OLS Regression Model**
   The initial stage was conducted by forming a classical linear regression model (Ordinary Least Squares / OLS) to identify the basic relationship between the dependent variable (percentage of poor population) and the independent variables (Human Development Index and Open Unemployment Rate). This model became the basis for comparison for the next spatial model.
5. **Spatial Autocorrelation Test (Moran's Index)**
   This stage aims to detect spatial dependencies between regions. The test is conducted using Global Moran's I, which measures whether the distribution pattern of poverty in districts/cities in South Sulawesi Province is random, clustered, or dispersed.
6. **Lagrange Multiplier Test**
   After detecting spatial dependence, a Lagrange Multiplier test is conducted to determine the appropriate spatial model type, namely SAR if the dependence occurs in the dependent variable, or SEM if the dependence appears in the error component, or both SARMA.
7. **Spatial Regression Model**
   Based on the LM test results, spatial models are constructed using the SAR, SEM, or SARMA approaches.
8. **Selection of the Best Model**
   The final stage is to determine the best model based on the log-likelihood value, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The model with the smallest AIC and BIC values is selected as the most appropriate model to explain poverty variation in South Sulawesi Province.

3. RESULTS AND DISCUSSION

In this section, the results of the analysis will be presented systematically, starting from descriptive statistics, correlations between variables, to the spatial regression model used.

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 7, No. 3, December 2025
2nd article, Pages: 292 – 307
DOI: https://doi.org/10.32493/sm.v7i3.54480

The discussion focuses on the interpretation of key findings, variable distribution patterns, and the influence of independent factors on dependent variables.

3.1 Descriptive Statistics

Descriptive statistics were performed as an initial step to provide an overview of the data used in this study. This descriptive stage aimed to examine the distribution, mean values, and spread of data for each variable, thereby providing an initial understanding of the socioeconomic conditions of the districts/cities prior to further analysis. The descriptive statistics results are presented in Table 1 below to facilitate interpretation.

Table 1. Statistics Descriptive.

Variable	Min	Q1	Median	Mean	Q3	Max	Skewness	Kurtosis
Y	4.97	6.82	8.29	8.60	10.79	12.41	0.07	1.77
X_1	69.45	72.17	73.93	74.49	74.97	85.23	1.55	5.22
X_2	1.51	2.37	3.23	3.71	4.19	9.71	1.52	5.14

The descriptive statistics in Table 1 provide information that the dependent variable Y has a mean of 8.60 with a median of 8.29, indicating a relatively symmetrical data distribution and a distribution that is not too elongated. For the independent variables, the Human Development Index (X_1) has a mean of 74.49 and a median of 73.93 with a right-skewed distribution and a fairly high peak, indicating that some districts/cities have a Human Development Index above the average. Meanwhile, the open unemployment rate (X_2) has an average of 3.71 and a median of 3.23, also showing a positively skewed distribution, so that some regions have higher unemployment rates than the average. Overall, the dependent variable data is relatively symmetrical, while the independent variables tend to have a right-skewed distribution. Next, the dependency structure or relationship between variables will be examined using Pearson's Correlation.

3.2 Correlation Analysis

Correlation can be defined as the degree of relationship or connection between two variables (Asuero et al., 2006). In this study, Pearson's correlation analysis was used to determine whether there was a relationship between variable Y and each of variables X_1 and X_2, which was a preliminary step that needed to be taken before proceeding to regression modeling. The results of the correlation analysis are presented in Table 2 below:

Table 2. Pearson Correlation.

Variable Pair	Coefficient Correlation	Statistics Test (t)	df	p_{value}
Y & X_1	-0.5395	-3.0051	22	0.0065
Y & X_2	-0.3975	-2.0318	22	0.0544

Based on Table 2, the Pearson correlation results show that the dependent variable Y has a negative relationship with both independent variables. The correlation between the Percentage of Poor Population (Y) and HDI (X_1) is -0.540, indicating that the higher the HDI value, the lower the Y value tends to be, and this relationship is statistically significant. Meanwhile, the correlation between Y and the OUR (X_2) is -0.398, indicating a negative relationship that is close to significant, so that an increase in TPT tends to be associated with a decrease in the value of Y, although the effect is weaker than that of X_1. Overall, these

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 7, No. 3, December 2025
2nd article, Pages: 292 – 307
DOI: https://doi.org/10.32493/sm.v7i3.54480

results confirm a negative relationship between the dependent and independent variables, with HDI having a stronger influence on Y than OUR. This confirms that these three variables are interrelated. Therefore, these three variables will be used for further analysis, namely Spatial Regression.

3.3 Visualization and Spatial Exploration

Before conducting further analysis, particularly spatial regression, it is important to examine the characteristics of data distribution patterns using thematic map visualizations of the variables used, particularly the percentage of poor people. The results of spatial visualization and exploration can be seen in Figure 1:

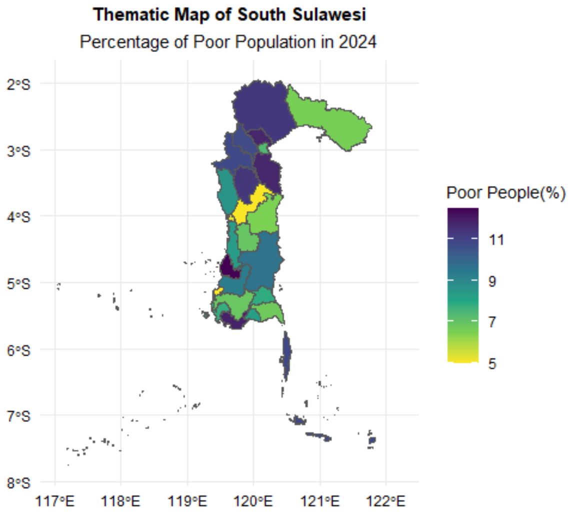

Figure 1. Thematic Map of Poverty Rate in Regencies/Cities of South Sulawesi.

The thematic map of the percentage of poor people in South Sulawesi Province in 2024 in Figure 1 shows variations in poverty levels between districts/cities, ranging from 4.97% to 12.41%. Areas colored yellow to light green, such as Makassar City, Parepare City, and Sidenreng Rappang, have the lowest poverty rates below 7%. Meanwhile, areas colored dark green to purple, such as Pangkajene Islands, Jeneponto, Luwu, and North Luwu, show higher poverty rates above 11%. This pattern indicates that regions with better access to economic opportunities, education, and infrastructure tend to have lower poverty rates than inland and mountainous areas. Therefore, spatial analysis is needed to understand the geographical and socioeconomic factors that influence the distribution of poverty, so that poverty alleviation policies can be more targeted according to the characteristics of each region.

3.4 OLS Regression

To determine the relationship or influence between variables, especially those with spatial correlation aspects, a statistical method called spatial regression is required. However, before performing spatial regression analysis, ordinary linear regression (OLS) analysis is first performed to see the initial relationship between the dependent and independent variables without considering spatial aspects. This analysis serves to identify variables that have a significant effect and to assess the suitability of the model in general.

3.4.1 OLS Regression Classical Assumption Test

Before interpreting the results of the OLS regression model estimation, a classical assumption test is first conducted to ensure that the model meets the Best Linear Unbiased Estimator (BLUE) criteria. This test aims to examine whether the data and regression model are free from violations of basic assumptions so that the analysis results are reliable and statistically valid. This test consists of four tests, namely normality, autocorrelation, heteroscedasticity, and multicollinearity. The test results are presented in Table 3:

Table 3. OLS Regression Classical Assumption Test.

Types of Tests	Test Statistics	Test Statistics Value	p_{value}	Conclusion
Normality Test (Shapiro-Wilk)	W	0.9609	0.4561	There are no normality issues
Autocorrelation Test (Durbin-Watson)	DW	2.2449	0.6758	There is no autocorrelation
Heteroscedasticity Test (Breusch-Pagan)	BP	3.4549	0.1777	There is no heteroscedasticity
Multicollinearity Test (VIF)	VIF(X_1, X_2)	(2.63268, 2.63268) < 10		There is no multicollinearity

All results of classical assumption tests in Table 3 show that the OLS regression model has met the assumptions of normality, homoscedasticity, no autocorrelation, and no multicollinearity, so that the model is suitable for further analysis.

3.4.2 Estimation of OLS Regression Parameters

This stage is for estimating the parameter values of the regression model that will be formed after fulfilling the classical assumptions. The estimated parameter values or regression coefficients in OLS analysis serve to show the direction and magnitude of the influence of each independent variable on the dependent variable. The results of the parameter estimation and the parameter significance test are presented in Table 4.

Table 4. Parameter Estimation and t-Test (Partial Test).

	Estimate	Std. Error	Statistics Test (t_{value})	p_{value}
(Intercept)	37.1766	13.4341	2.7673	0.0115
X_1	-0.3879	0.1937	-2.0024	0.0583
X_2	0.0848	0.3505	0.2418	0.8113

The results in Table 4 show that the constant value obtained is 37.1766, indicating the average value of the dependent variable when all independent variables are zero. Variable X_1 has an estimated coefficient of -0.3879, which has a negative effect on the dependent variable and is statistically significant at a significance level of 6% ($\alpha = 0.06$), but is not yet significant at a level of 5% ($\alpha = 0.05$). Meanwhile, variable X_2 has a coefficient of 0.0848, which has a positive but statistically insignificant effect on the dependent variable. The results of the simultaneous parameter significance test will be presented again below.

Table 5. F-test (Simultaneous Test).

F-statistic	df_1	df_2	p_{value}
4.3513	2	21	0.0262

The results in Table 5 provide information on the F-statistic value of 4.3513 with degrees of

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 7, No. 3, December 2025
2nd article, Pages: 292 – 307
DOI: https://doi.org/10.32493/sm.v7i3.54480

freedom (df_1 = 2, df_2 = 21) and p_{value}= 0.0262, indicating that the regression model is significant at a significance level of 0.05. This means that simultaneously, variables X_1 and X_2 have a significant effect on the dependent variable, so that the OLS model used is considered feasible and statistically significant in explaining the variation in the dependent variable. After the OLS regression model has been declared to have met all the classical assumptions, the next step is to perform regression modeling. This stage aims to analyze the relationship between the independent and dependent variables and measure the extent of the influence of each explanatory variable on the variable being studied. The resulting regression model is as follows:

$$\hat{y} = 37.1766 - 0.3879x_1 + 0.0848x_2$$

This regression model provides a coefficient of determination (R^2) value of 0.293 or 29.3%, which means that the Human Development Index (X_1) and Open Unemployment Rate (X_2) together can explain 29.3% of the variation in the Percentage of Poor Population (Y). Meanwhile, the remaining percentage is explained by other factors outside the model. After conducting OLS regression analysis, the next step is to perform spatial analysis by initially identifying whether the data in this study has a correlation or relationship from a spatial aspect. Therefore, the initial step that needs to be taken in spatial regression modeling is a spatial dependency test, which will be discussed in the following subchapter.

3.5 Spatial Dependency Test

Moran's Index (Moran's I) is used to measure or test the level of spatial dependence or correlation between attribute values of geographic units, both in the form of polygons and points (Aswi et al., 2021). This test is conducted to measure the extent to which the value of a variable in one area correlates with the value of the same variable in the surrounding area. If the results are significant, it indicates that there is a spatial effect that needs to be taken into account in the regression model. The results of the Moran Index test are presented in Table 6:

Table 6. Moran's Index Test Results.

Moran I Statistic	Expectation	Variance	Z_{value}	p_{value}	Conclusion
0.2411	-0.0435	0.0228	1.8842	0.0298	There is spatial autocorrelation

Table 6 provides information that there are indications of statistically significant positive spatial autocorrelation. This means that areas with similar values, such as high poverty, tend to be geographically close to each other. These results can also be seen in the form of the following Moran's Index plot.

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 7, No. 3, December 2025
2nd article, Pages: 292 – 307
DOI: https://doi.org/10.32493/sm.v7i3.54480

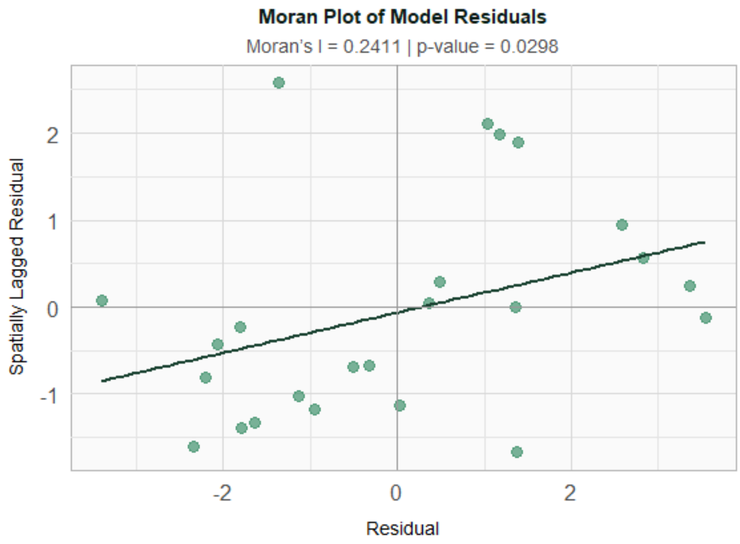

Figure 2. Moran's Index Plot.

The information provided by Table 6 and Figure 2 leads to the conclusion that there are indications of significant spatial autocorrelation. Therefore, further analysis can be carried out using spatial regression.

3.6 Spatial Regression

Spatial regression techniques are applied because this method is able to consider spatial interrelationships between regions. Thus, the parameter estimates obtained are more accurate and reliable, provided that the spatial structure as represented in the covariance matrix can be modeled accurately (Fortin et al., 2012). This section describes the application of spatial regression to analyze the effect of independent variables on dependent variables by considering the interrelationships between regions. Before determining the appropriate model, a Lagrange Multiplier (LM) test was conducted to detect spatial autocorrelation. The results of this test formed the basis for model selection, namely the Spatial Autoregressive Model (SAR) if dependence occurred in the dependent variable or the Spatial Error Model (SEM) if dependence appeared in the error component. The results of the LM test are presented in Table 7:

Table 7. Lagrange Multiplier Test Results.

Model	Test Statistics Value	df	p_{value}
RSerr	2.8384	1	0.0920
RSlag	0.4052	1	0.5244
adjRSerr	7.7387	1	0.0054
adjRSlag	5.3054	1	0.0213
SARMA	8.1439	2	0.0170

The results in Table 7 show that based on the test results on the RSerr and RSlag models, the $p_{value} > \alpha = 0.05$, so H_0 is accepted. These results indicate that there is no significant spatial dependence in either the lag or the error. However, after conducting a standardized (adjusted) LM test for both the adjRSerr and adjRSlag models, it was found that for adjRSerr, the $p_{value} = 0.0054 < \alpha = 0.05$, and for adjRSlag, the $p_{value} = 0.0213 < \alpha = 0.05$.

StatMat
Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 7, No. 3, December 2025
2nd article, Pages: 292 – 307
DOI: https://doi.org/10.32493/sm.v7i3.54480

These results lead to the conclusion that H_0 is rejected, which means that there is significant spatial dependence at the 5% significance level for both lag and error, so that the models formed are SAR and SEM. In addition, in the SARMA model, the $p_{value} = 0.0170 < \alpha = 0.05$, so H_0 is rejected. This result also shows that the SARMA model is significant, thus strengthening the indication of spatial dependence in the model. Therefore, there are three spatial regression models formed in this study, namely SAR, SEM, and SARMA.

After conducting the Lagrange Multiplier (LM) test, the next step is to form a model from the three types of spatial regression produced. However, before determining the final model, it is necessary to retest the classical assumptions in the three models, namely SAR, SEM, and SARMA, with the aim of ensuring that the selected model meets the basic assumptions of regression and provides valid and reliable estimation results. The following presents the results of testing the classical assumptions for the three models.

Table 8. Spatial Regression Classical Assumption Test.

Types of Tests	Spatial Regression Model	Test Statistics	Test Statistics Value	p_{value}	Conclusion
Normality Test (Shapiro-Wilk)	SAR	W	0.9669	0.5903	There are no normality issues
	SEM		0.9490	0.2581	There are no normality issues
	SARMA		0.9296	0.0957	There are no normality issues
Autocorrelation Test (Moran's I)	SAR	Z	1.5562	0.0598	There is no autocorrelation
	SEM		0.5835	0.2798	There is no autocorrelation
	SARMA		0.3290	0.3711	There is no autocorrelation
Heteroscedasticity Test (Breusch-Pagan)	SAR	BP	2.2503	0.3246	There is no heteroscedasticity
	SEM		4.8144	0.0901	There is no heteroscedasticity
	SARMA		3.4801	0.1755	There is no heteroscedasticity

The results of the classical assumption test in Table 8 show that the three spatial regression models, SAR, SEM, and SARMA, have met all the basic regression assumptions. Based on the normality test (Shapiro-Wilk), the $p_{value} > 0.05$, indicating that the residuals are normally distributed. Furthermore, the results of the autocorrelation test (Moran's I) also show a $p_{value} > 0.05$, meaning that there is no spatial autocorrelation in the model residuals. Finally, the results of the heteroscedasticity test (Breusch-Pagan) indicate $p_{value} > 0.05$ for all models, indicating no heteroscedasticity. Thus, the three models have met the classical assumptions and are suitable for further analysis.

After all classical assumptions are fulfilled in the three spatial regression models, the next step is to perform spatial regression modeling to determine the effect of independent variables on dependent variables, taking into account the spatial effects between regions. This modeling is carried out in three main forms, namely the Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), and Spatial Autoregressive Moving Average Model (SARMA). The estimation results from each model will be compared to determine the best

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 7, No. 3, December 2025
2nd article, Pages: 292 – 307
DOI: https://doi.org/10.32493/sm.v7i3.54480

model that is most suitable for explaining the variation in poverty levels between districts/cities in South Sulawesi Province. The estimation results from the three models are presented in Table 9:

Table 9. Spatial Regression Parameter Estimation Results.

Spatial Regression Model	Variable	Estimate	Std. Error	Z_{value}	p_{value}	Spatial Dependency Parameters	
						Rho (ρ)	Lamda (λ)
SAR	(Intercept)	36.9500	12.5545	2.9432	0.0032	0.0880	
	X_1	-0.3931	0.1794	-2.1911	0.0284		
	X_2	0.04851	0.3244	0.1495	0.8811		
SEM	(Intercept)	42.4539	12.038	3.5268	0.0004		0.3653
	X_1	-0.4458	0.1755	-2.5403	0.0111		
	X_2	-0.1494	0.3340	-0.4473	0.6546		
SARMA	(Intercept)	38.3135	10.3407	3.7051	0.0002	-0.3029	0.5572
	X_1	-0.3554	0.15433	-2.3027	0.0213		
	X_2	-0.1393	0.2895	-0.4813	0.6303		

Based on the parameter estimation results in Table 9, the three spatial models tested, SAR, SEM, and SARMA, show a relatively consistent effect of independent variables on the poverty rate of districts/cities in South Sulawesi Province in 2024. In all three models, variable X_1 (Human Development Index) had a negative and significant effect on poverty levels (p_{value} < 0.05). This indicates that the higher the HDI of a region, the lower its poverty level tends to be. Conversely, variable X_2 (Open Unemployment Rate) has no significant effect (p_{value} > 0.05) in all models, indicating that variations in unemployment between regions do not have a direct and significant effect on poverty levels in this spatial context. In terms of spatial dependence, the ρ (rho) parameter in the SAR model has a positive value (0.0880), indicating a weak but unidirectional spatial relationship between regions, meaning that regions with high poverty rates tend to be close to other regions that also have high poverty rates. Meanwhile, the λ (lambda) parameter in the SEM (0.3653) and SARMA (0.5572) models is positive and quite large, indicating that spatial effects are more dominant through the error component rather than directly through the dependent variable. Overall, these results reinforce that spatial aspects need to be considered in poverty modelling, because interregional dependencies significantly affect poverty distribution patterns in South Sulawesi. Of the three models, the SARMA model shows the best combination of spatial effects with both spatial parameters (ρ and λ) being significant and consistent in direction, so this model can be considered the most representative in describing spatial interregional linkages. After estimating the parameters in the three spatial models, the next step is to form the regression model using the significant parameters.

Spatial Autoregressive Model (SAR)

$$\hat{y}_i = 0.0880 \sum_{j=1, i\neq j}^{n} w_{ij}\, y_j + 36.9500 - 0.3931 x_{1i}$$

Spatial Error Model (SEM)

$$\hat{y}_i = 42.4539 - 0.4458 x_{1i} + \mathrm{u}_i \text{ ; with } \mathrm{u}_i = 0.3653 \sum_{j=1; i\neq j}^{n} w_{ij}\, u_j + \varepsilon_i$$

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 7, No. 3, December 2025
2nd article, Pages: 292 – 307
DOI: https://doi.org/10.32493/sm.v7i3.54480

Spatial Autoregressive Moving Average (SARMA)

$$\hat{y}_i = -0.3029 \sum_{j=1, i\neq j}^{n} w_{ij}\, y_j + 38.3135 - 0.3554 x_{1i} + \mathrm{u}_i$$

$$\text{with } \mathrm{u}_i = 0.5572 \sum_{j=1; i\neq j}^{n} w_{ij}\, u_j + \varepsilon_i$$

After performing spatial regression modeling, the next step is to select the best model from the three models based on the AIC and BIC values.

3.7 Selecting the Best Model

The selection of the best model will be determined based on the AIC and BIC values. The model with the smallest AIC and BIC values will be selected as the best model for modelling the effect of the Human Development Index and Open Unemployment Rate on the Percentage of Poor Population in South Sulawesi in 2024. The results of the best model selection are presented in Table 10:

Table 10. Best Model Selection.

Regression Model	LogLik	AIC	BIC
OLS	-49.4582	106.9164	111.6287
SAR	-49.2636	108.5273	114.4176
SEM	-47.1364	104.2729	110.1632
SARMA	-45.3242	**102.6484**	**109.7167**

Based on Table 10, the smallest AIC and BIC values are possessed by the SARMA model (AIC = 102.6484; BIC = 109.7167). This indicates that the SARMA model is the best model compared to OLS, SAR, and SEM because it provides the most optimal balance between model accuracy and parameter complexity. Thus, the SARMA model was selected as the final spatial regression model in this study. This means that in South Sulawesi Province, the percentage of poor people is not only directly influenced by the Human Development Index (HDI) and Open Unemployment Rate (OUR) variables in the district/city, but also by the spatial influence of the surrounding area. In other words, the poverty rate of districts/cities in South Sulawesi can be influenced by the HDI and OUR conditions of neighbouring districts/cities, which indicates the existence of interregional linkages in the phenomenon of poverty in this province.

CONCLUSION

The results of this study conclude that there is spatial dependence on poverty levels in South Sulawesi Province in 2024. The best model to describe this relationship is Spatial Autoregressive Moving Average (SARMA). These results show that the poverty rate in a region is not only influenced by the Human Development Index (HDI) and Open Unemployment Rate (OUR) in that region, but also by the conditions of the surrounding areas. In general, the higher the HDI and the lower the OUR, the lower the percentage of poor people tends to be, both directly and through spatial influences between regions. This finding confirms that poverty in South Sulawesi is not only influenced by internal socioeconomic conditions, but also by the spatial effects of the surrounding regions. Thus, poverty alleviation efforts should be carried out in an integrated and cross-regional manner, taking into account the spatial interrelationships between regions in development planning.

StatMat
Jurnal Statistika dan Matematika

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 7, No. 3, December 2025
2nd article, Pages: 292 – 307
DOI: https://doi.org/10.32493/sm.v7i3.54480

REFERENCES

Anselin L. Spatial econometrics. Richardson (TX): Bruton Center, School of Social Sciences, University of Texas at Dallas; 1999.

Ardian R, Syahputra M, Dermawan D. Pengaruh pertumbuhan ekonomi terhadap tingkat pengangguran terbuka di Indonesia. EBISMEN: Jurnal Ekonomi, Bisnis dan Manajemen. 2022 Sep;1(3):190–198.

Asuero AG, Sayago A, González AG. The correlation coefficient: An overview. Critical Reviews in Analytical Chemistry. 2006;36(1):41–59.

Aswi A, Cramb S, Duncan E, Mengersen K. Detecting spatial autocorrelation for a small number of areas: a practical example. J Phys Conf Ser. 2021;1899(1):012098.

Badan Pusat Statistik (BPS). Provinsi Sulawesi Selatan dalam Angka 2025 / Sulawesi Selatan Province in Figures 2025. Makassar: BPS Provinsi Sulawesi Selatan; 2025.

Chatterjee S, Simonoff JS. Handbook of Regression Analysis. Hoboken (NJ): John Wiley & Sons; 2013.

Chen Y. Spatial autocorrelation equation based on Moran's index. Scientific Reports. 2023;13(1):19296.

Das M, Ghosh SK. A cost-efficient approach for measuring Moran's index of spatial autocorrelation in geostationary satellite data. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE; 2016. p. 5913–6.

Fortin M-J, James PMA, MacKenzie A, Melles SJ, Rayfield B. Spatial statistics, spatial regression, and graph theory in ecology. Spatial Statistics. 2012;1:100–109.

Halunga A, Orme CD, Yamagata T. A heteroskedasticity robust Breusch–Pagan test for contemporaneous correlation in dynamic panel data models. Journal of Econometrics. 2017; pp. 209–230.

LeSage JP. Spatial econometrics. Toledo (OH): Department of Economics, University of Toledo; 1998.

Mathur M. Spatial autocorrelation analysis in plant population: An overview. Journal of Applied and Natural Science. 2015;7(1):501–513.

Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis. 5th ed. Hoboken (NJ): John Wiley & Sons, Inc.; 2012.

Putra IKA, Arka S. Analisis pengaruh tingkat pengangguran terbuka, kesempatan kerja, dan tingkat pendidikan terhadap tingkat kemiskinan pada kabupaten/kota di Provinsi Bali. E-Jurnal EP Unud. 2018;7(3):416–444.

Razali NM, Wah YB. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. Journal of Statistical Modeling and Analytics. 2011;2(1):21–33.

Rinindah AE, Firdaus MA, Armayani S, Nadhilah W, Nurmayanti WP. Analisis faktor-faktor ketenagakerjaan di Indonesia dengan pendekatan regresi logistik biner. STATMAT (Jurnal Statistika dan Matematika). 2025;7(2):257–264.

Sari FM, Nasution HF, Sihombing PR. Pemodelan data kemiskinan Provinsi Sumatera Barat menggunakan regresi spasial. Infinity: Jurnal Matematika dan Aplikasinya (IJMA). 2021;2(1):51–61.