INFO ARTHA
Journal Homepage : www.jurnal.pknstan.ac.id/index.php/JIA

SPATIAL REGRESSION AND SPATIAL AUTOCORRELATION ANALYSIS OF THE DETERMINANTS OF POVERTY IN INDONESIA IN 2022

I Ketut Adi Chandra Wiguna
STIS Polytechnic of Statistics

Ni Komang Noviyanti
Bali Provincial Government

Kadek Shri Yogi Savitri
Bali Provincial Government
Correspondence Address: noviynti3@gmail.com

ABSTRACT ARTICLE INFORMATION

First Accepted
[03 08 2023]

Declared Accepted
[23 11 2023]

KEY WORDS:
Poverty, Spatial Regression, Spatial Autocorrelation, Province, Indonesia

JEL CLASSIFICATION:
I32

Poverty remains a solemn challenge in Indonesia, although the country has made significant progress in recent decades. Even though government efforts and various poverty alleviation programs have been carried out, most of the population in Indonesia lives below the poverty line. This research aims to examine variables that influence Indonesia's poverty rate by province in 2022 using spatial regression analysis and spatial autocorrelation. The secondary data used comes from 34 provinces in Indonesia and is obtained based on the results of publications by the Central Statistics Agency (BPS). The dependent variable is the percentage of poor people (P0), while the independent variables include average length of school (RLS), life expectancy (AHH), open unemployment rate (TPT), and Gini ratio. The analytical methods used include descriptive analysis, multiple linear regression analysis, and spatial regression analysis. It hopes this research can provide relevant policy recommendations for developing poverty alleviation policies in Indonesia.

Kemiskinan tetap menjadi tantangan serius di Indonesia meskipun negara ini telah mencapai kemajuan signifikan dalam beberapa dekade terakhir. Meskipun upaya pemerintah dan berbagai program pengentasan kemiskinan telah dilakukan, masih ditemukannya penduduk Indonesia yang hidup di bawah garis kemiskinan. Tujuan dari penelitian ini adalah untuk mengkaji variabel-variabel yang mempengaruhi angka kemiskinan Indonesia menurut provinsi pada tahun 2022 menggunakan analisis regresi spasial dan autokorelasi spasial. Data sekunder yang digunakan yaitu 34 provinsi di Indonesia dan didapat berdasarkan hasil publikasi Badan Pusat Statistik (BPS). Variabel dependen adalah persentase penduduk miskin (P0), sedangkan variabel independen meliputi RLS, AHH, TPT, dan GR. Penelitian ini diharapkan dapat memberikan rekomendasi kebijakan yang relevan dalam pengembangan kebijakan pengentasan kemiskinan di Indonesia.

1. INTRODUCTION

1.1. Background

Poverty remains a solemn challenge in Indonesia, even though the country has made significant progress in recent decades. Despite government efforts and various poverty alleviation programs, a large proportion of Indonesia's population is still considered poor. The Central Bureau of Statistics (BPS) reported that 26.36 million people, or 9.57% of the population, lived in poverty as of September 2022.

Another feature of poverty is limited access to public health services that receive little attention, resulting in weakened resilience to earn a living and stunted growth and development in children. Poverty affects the education system, which prevents the underprivileged from attending school effectively (Leleury and Tomasouw, 2019). It is due to the low quality of education and the scarcity of infrastructure and resources. In addition, there are few employment possibilities, capital, and skills. Poverty has a detrimental impact on quality of life, so an in-depth understanding of the contributing factors to poverty levels is essential to inform more effective public policies and poverty alleviation strategies.

In addition, Indonesia is a country that has significant geographic and socioeconomic diversity at the provincial level. Each province has unique characteristics that affect the poverty rate within it. Factors such as average years of schooling, life expectancy, labor force participation rate, and income inequality are substantial concerns in understanding and addressing poverty issues at the provincial level.

In this context, we will use 2022 provincial data and selected variables, such as average length of school (RLS), life expectancy (AHH), open unemployment rate (TPT), and Gini ratio. Spatial regression analysis emerged as an effective approach to understanding the factors that affect poverty rates in Indonesia by province. Spatial regression allows us to analyze the spatial relationship between relevant variables, considering geographical and spatial factors that may affect poverty rates. This approach provides a more comprehensive understanding of the patterns and determinants of poverty, which can help designing more targeted and sustainable policies to reduce poverty in Indonesia.

Thus, the aim of this research is to better understand Indonesia's poverty rate in 2022 and examine the variables that contribute to it using spatial regression analysis, identify spatial patterns and associated spatial autocorrelation, and provide relevant policy recommendations.

2. THEORETICAL FRAMEWORK AND HYPOTHESIS DEVELOPMENT

2.1. Poverty

According to the Central Bureau of Statistics, poverty is valued by using the concept of the ability to fulfill basic needs (basic needs approach), where poverty is an economic inability to satisfy basic food and non-food needs measured in terms of expenditure. A person can be said to be poor or living in poverty if their income or access to goods and services is relatively low compared to the average of other people in the regional economy. The definition of poor are people who have an average monthly expenditure below the poverty line (GK). In theory, the poverty line consists of two components: the food poverty line (FPL) and the non-food poverty line (NFPL). The GKM has the minimum food expenditure equivalent to 2,100 kilocalories per capita per day, while the GKNM has the minimum requirement for housing, clothing, education, and health.

2.2. Descriptive Analysis

Spatial Autocorrelation

Lembo (2006) stated that the purpose of the geographic analytic technique known as spatial autocorrelation is to identify the distribution of linear relationships (correlations) between places (observations). This approach is necessary to learn about the typical distribution patterns of an area and the relationships between these various places. This technique is also used to identify spatial models. One of the most widely used statistics to assess spatial autocorrelation is the Coefficient of Morans I.

The most popular approach to calculating global spatial autocorrelation is Moran's index (Moran's I). By using this technique, the beginning of spatial randomness is recognized. You can use Moran's index approach (Banerjee, 2004):

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})\, w_{ij}(x_j - \bar{x})}{\sum_{i \neq j}^{n} w_{ij} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

In the case of a conventional geographic weighting matrix, the range of Moran Index values is -1 to 1. A value of Moran index shows zero implies no clustering, while values of -1 I 0 and 0 I 1 indicate negative and positive spatial autocorrelation, respectively. If used-weight matrix is not standardized, the Moran index value does not guarantee measurement accuracy. The Moran Index's significance test is carried out to determine the presence or absence of spatial autocorrelation.

Hypothesis

$H_0: I = 0$ (No spatial dependency)

$H_1: I \neq 0$ (There is spatial dependency)

The test statistics can be formulated as:

$$Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}}$$

2.3. Inferential Analysis

Multiple Linear Regression

Multiple linear regression is a statistical technique used to model one dependent variable (the variable you want to predict) and two or more independent variables (the variables used to predict the dependent variable). The model can be represented by a linear equation that consists of the required coefficients and the independent variables in multiple linear regressions. The main goal of this technique is to find a linear equation that best captures the relationship between various variables. The equation for multiple linear regression often has the following form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$

In conducting multiple linear regression analysis, several factors should be considered, including:

1. Assumption testing
   Before conducting multiple linear regression analysis, it is necessary to check the basic assumptions of linear regression, such as the assumption of linear attachment between variables, the assumption of normality of residuals, and the assumption of homoscedasticity of residuals.
2. Model evaluation
   To assess the quality and fit of multiple linear regression models, it is necessary to evaluate them. Some evaluation methods include R-square (coefficient of determination), the coefficient significance test, and other assumption tests such as the Durbin-Watson test to check for autocorrelation.

Spatial Regression

a. Spatial Autoregressive Model (SAR)

This approach combines a direct regression model with geographic lags in the dependent variable using cross-sectional data. The following is a general SAR model, according to Baltagi and Liu (2012):

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

$H_0: \rho = 0$ (No spatial lag autocorrelation)
$H_1: \rho \neq 0$ (There is spatial lag autocorrelation)

b. Spatial Error Model (SEM)

This model is a geographic model with spatially correlated errors. The following equation describes the general SEM model:

$$y = X\beta + u$$
where: $u = \lambda W u + \varepsilon$
$$\varepsilon \sim N(0, \sigma^2 I)$$

$H_0: \lambda = 0$ (No autocorrelation of spatial residuals)
$H_1: \lambda \neq 0$ (There is autocorrelation of spatial residuals)

c. Spatial Autoregressive Moving Average (SARMA)

Anselin (1988) formulated a spatial regression model in the form of the Spatial Autoregressive Moving Average (SARMA) model with the following equation:

$$y = \rho W y + X\beta + u$$
$$u = \lambda W u + \varepsilon$$
$$\varepsilon \sim N(0, \sigma^2 I)$$

$H_0: \lambda = 0$ (No spatial lag dependence)
$H_1: \rho, \lambda \neq 0$ (There is spatial lag dependence)

d. Spatial Weighting Matrix

A spatial weighting matrix is created based on the information about the separation between two areas. According to LeSage (1999), there are many methods to calculate the value of Wij, including Rook Contiguity, Bishop Contiguity, Linear Contiguity, and Queen Contiguity. The location of a region within a region, along with the position of other regions around it, is used to calculate the spatial weight matrix. Rook contiguity is a term used to describe places that are close together. Bishop Contiguity is used for locations in which corner points are adjacent to other places. Linear contiguity is used for areas on the edge, either the left or right of another area. Meanwhile, Queen Contiguity is used for areas on the side or corner of another area.

3. RESEARCH METHODS

Descriptive statistical analysis and inferential statistical analysis are the types of analysis used in this study. All data analysis and conclusion-making begin with descriptive analysis (Bender, 2020). Descriptive data provides an overview of the distribution and behavior of research data. Another option is inferential analysis using multiple linear regression analysis with four independent variables and one dependent variable. Then, spatial regression analysis is used to identify derive policies from the most effective model.

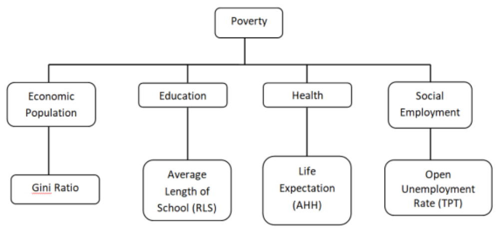


Diagram 1. Four independent variables

The secondary data was collected from 34 provinces in Indonesia in 2022 for this research. All statistics were obtained from the Central Bureau of Statistics (BPS) publications. The proportion of poor people (P0) is the dependent variable in this study. Average length of school (RLS), life expectation (AHH), open unemployment rate (TPT), and Gini ratio (GR) are some of the independent variables. Used applications for data processing are Microsoft Excel, R Studio, and a free-source software called GeoDa.

Table 1 Variable Name, Description, and Unit

No.	Variable Name	Description	Unit
1	P0	Percentage of poor population, as dependent variable	Percent

2	RLS	Average Years of Schooling	Year
3	AHH	Life Expectancy Rate	Year
4	TPT	Open Unemployment Rate	Percent
5	GR	Gini Ratio	Index

4. RESEARCH RESULTS

Descriptive Analysis

Figure 1 shows the distribution of poverty percentages in each province in Indonesia. The map shows five intervals to represent the poverty level in each province. The higher the poverty percentage in a province, the darker color of the map will be. Conversely, the lower the poverty percentage in a district or city, the darker color of the map will be.

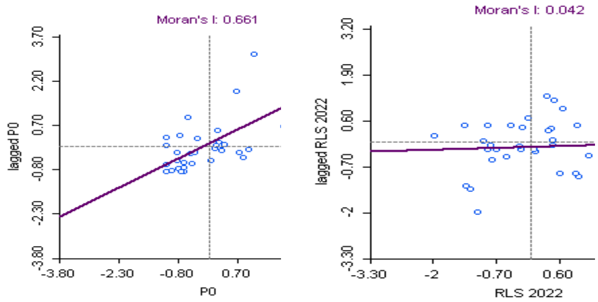
Figure 1: Poverty Percentage Distribution Map of Indonesia 2022

Some clusters of poverty percentages shown on the map indicate a spatial correlation in neighboring regions. Regions that have similar colors tend to be close to each other. It indicates that the poverty rate has an uneven distribution pattern across Indonesia.

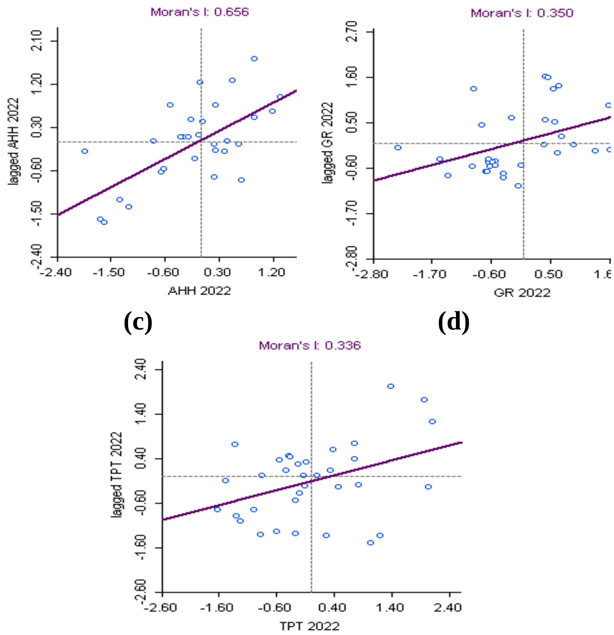
Data shows that high poverty rates tend to concentrate in eastern Indonesia. The provinces of Papua, West Papua, Maluku, NTB, and NTT are some of the provinces that show relatively high poverty rates. It shows darker-colors marking indicate a higher proportion of poverty in comparison to other provinces.

Global Autocorrelation

Moran's Index is used to determine whether there is global spatial autocorrelation between observations. The results of the global autocorrelation calculation assessed by the Moran Index are shown in Figure 2.



(a) (b)



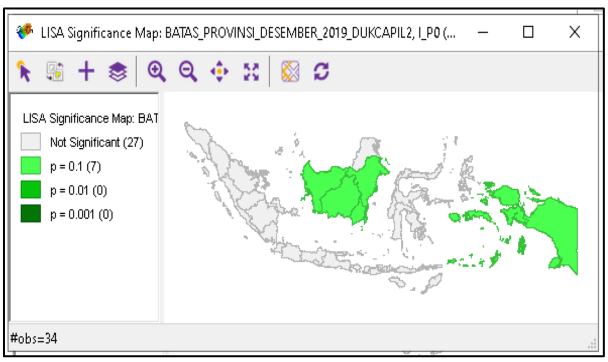
(e)

Figure 2. Moran's Index Scatter Plot

The Moran Index value of the poverty percentage variable (P0) of 0.681 indicates a strong positive spatial autocorrelation,ss indicating that an increase in the poverty rate in a province will also increase the poverty rate in its closest province. The Moran index of the RLS variable of 0.042 indicates a weak positive spatial autocorrelation, so an increase in the RLS variable in a province will increase the average years of schooling in its closest province. The Moran index of the life expectancy variable of 0.656 indicates strong-positive autocorrelation. This figure means that an increase in life expectancy in a province also increases life expectancy in the nearest province. The Moran index value of the Gini ratio variable of 0.350 indicates a positive-moderate autocorrelation. The open unemployment rate variable has a Moran index of 0.336, meaning that the variable has a positive but not too strong autocorrelation.

Local Autocorrelation

LISA is a tool for detecting autocorrelation. LISA also identifies early evidence of substantial spatial (regional) relationships. It indicates the existence of strong spatial connections for provinces that are close to each other, including high-high (H-H), low-low (L-L), high-low (H-L), and low-high (L-H). LISA cluster map is one effective tool to demonstrate the geographical relationship between Indonesian regions.



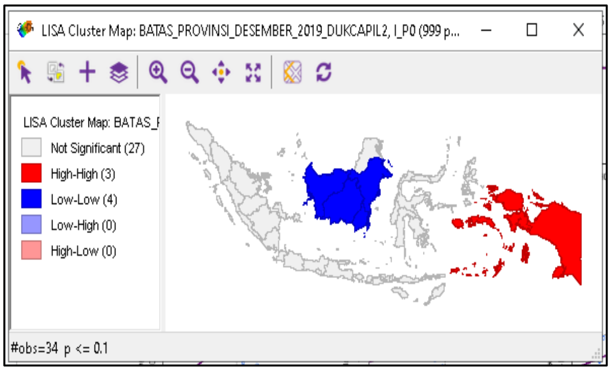


Figure 3. LISA Significance and Cluster Map

LISA identifies that Papua, West Papua, and Maluku are three Indonesian provinces with high spatial correlation (H-H). The H-H association indicates that provinces with high poverty rates are adjacent to another province with similar rates. In contrast, the four provinces of West Kalimantan, South Kalimantan, Central Kalimantan, and East Kalimantan have low-low (L-L) spatial interactions. According to the L-L geographical relationship, provinces with low poverty rates border other provinces with low poverty rates.

Inferential Analysis

Classical Regression Model (OLS)

The classical regression model is used to determine which independent factors significantly affect the dependent variable. Table 2 shows the Geoda output to see which variables significantly affect poverty.

Table 2 Geoda Output of Classical Regression Model

Variables	Coefficient	Std.Error	t-Statistic	Probability
Constant	85.083	19.008	4.468	0.000
RLS	-1.233	0.888	-1.389	0.018
AHH	-1.125	0.303	-3.710	0.001
TPT	-0.282	0.501	-0.563	0.578
GR	42.470	14.391	2.951	0.006

The results were obtained with a statistical significance level of 10 percent. Based on Table 2, Variables RLS, AHH, and GR have significant effects on poverty. With a probability value of 0.018, 0.001, and 0.006. While the TPT variable has no significant effect on the dependent variable.

Normality and Homoscedasticity Test

Several tests were performed to assess classical assumptions. They are normality, homoscedasticity, and multicollinearity tests. The test for error normality and homoscedasticity was conducted with the help of Geoda software.

Table 3 Geoda Output Test for Normality and Homoscedasticity

Test	DF	Value	Probability
Jarque-Bera	2	1.105	0.576
Breusch-Pagan	4	5.520	0.238

Table 3 presents the results. The Jarque-Bera normality test indicates that the error in the model is normally distributed. On the other hand, the Breusch-Pagan test, to ensure the homogeneity of the error variance of the regression model, shows that the variance of the error model is homogeneous.

Multicollinearity Test

SPSS software was used to perform non-multicollinearity tests. Table 4 presented the results, where VIF values that below 10 indicate no multicollinearity in all independent variables.

Table 4 R Studio Non-Multicollinearity Test

	Variables	VIF
RLS	Average Years of Schooling	1.564
AHH	Life Expectancy Rate	1.311
TPT	Open Unemployment Rate	1.496
	Gini Ratio	1.038

After confirming that classical assumption tests are met, we can proceed to the simple and spatial linear regression modeling.

Spatial Regression Model

The Lagrange Multiplier (LM) Spatial Regression Model is a method used, as an initial identification, to create an accurate spatial regression model. LM test is also used to detect spatial autocorrelation of lags, errors, or both (lags and errors) more precisely. Ignoring the latency and error of the LM at a certain test level leads to no dependency on lags or errors.

Table 5 Geoda Output Spatial Dependence Test

Spatial Dependence Test	MI/DF	Value	Probability
Moran`s I (Error)	0.254	2.016	0.044
LM (Lag)	1	6.790	0.010
Robust LM (Lag)	1	5.350	0.021
LM (Error)	1	2.094	0.148
Robust LM (Error)	1	0.654	0.419

The spatial dependency test in this study used the queen contiguity approach. This approach weights the provinces that cluster and are surrounded by each other in Indonesia. As to the LM test results in Table 5, there is spatial dependence as Moran's I probability value (p-value) is significant at the 10% level. The Lagrange Multiplier Lag p-value in table x is 0.010, or less than 10% alpha. The model will not proceed to the Spatial Error Model (SEM) because the Lagrange multiplier (error) is 0.148, which is more than 0.1. Therefore, the spatial autoregressive model (SAR) is the best regression model used to quantify poverty.

Estimation of the Best Regression Model

Spatial Autoregressive Model (SAR) is the best model, with RLS, TPT, and GR variables significant at the 10% significance level. If SAR modeling is performed using significant dependent variables, the results are shown in Table 6. So, the equation model formed is:

$$\hat{y}i = 52.682 - 0.337Wy - 1.240\,RLS - 0.661AHH + 35.249\,GR$$

Table 6 SAR Model Geoda Output

Variables	Coefficient	Std.Error	z-value	Probability
W_P0	0.337	112157	3.004	0.003
Constant	52.682	17.125	3.076	0.002
RLS	-1.240	0.711	-1.744	0.081
AHH	-0.661	0.268	-2.466	0.014
TPT	-0.387	0.401	-0.964	0.335
GR	35.249	11.854	2.974	0.003

The spatial regression analysis results offer a comprehensive understanding of the dynamics between different provinces, considering both local (independent variables in the province itself) and spatial effects (influence from neighboring provinces). The constant model value of 52.682 represents the percentage of people living in poverty with a scenario when all independent variables are zero. Moreover, the coefficients of independent variables provide valuable insights into the contribution of each variable to poverty levels. The negative coefficient (-1.240) for average years of schooling (RLS) depicts an inverse relationship between education levels and poverty rates. Provinces with lower education levels tend to exhibit higher poverty rates. Similarly, the negative coefficient (-0.661) for life expectancy (AHH) displays a negative correlation between life expectancy and poverty rates. Provinces with lower life expectancy levels witness higher poverty levels. In contrast, the positive coefficient (35.249) for the Gini Ratio (GR) indicates that higher levels of income inequality can lead to higher poverty rates in a province. The model also considers the spatial component, which highlights that the socioeconomic conditions of a province are influenced not only by independent variables within the province but also by the conditions of neighboring provinces.

Best Model Selection

Each regression equation resulting from two approaches is compared in regression modeling to obtain the optimal model. The optimal model was chosen by using the coefficient of determination and the AIC approach. The closer the predictor variable gets to one, the stronger its effect on the response variable, indicating a better fit between the model and the data. The AIC approach can help in selecting the best regression model by choosing the one with the lowest AIC value. Table 7 presents the comparison and AIC values.

Table 7 shows the comparison of R^2 and AIC values.

Table 7 Best Model Selection

MODEL	R^2	AIC
RLB	55.432%	191.215
SAR	66.478%	185.408

Based on table 7, the SAR model produces a higher overall value than the RLB model. In addition, the final AIC value of the SAR model is lower than the RLB model. The summary is that the SAR model is more accurate when used to predict Indonesia's poverty rate in 2022.

5. CONCLUSIONS AND SUGGESTIONS

The results of the analysis obtained in this study show that the poverty rate in Indonesia is highest in the eastern provinces of Indonesia, which include Papua, West Papua, Maluku, NTT, and NTB. And the highest is in Papua Province.

After conducting regression modeling to detect the most influential dependent variable results, a spatial regression model is more effective than the classical regression model by using the best model called the Spatial Autoregressive Model (SAR) model. It was proved by comparing which is higher between the smaller AIC value and the OLS model. R^2 is greater than the OLS model. It means that the independent variables can better explain the variance of the dependent variable in the SAR model compared to OLS. The variables that influence it significantly are average length of school, Gini ratio, and life expectation. It can be concluded that education, socio-demographics, and health issues have an important role in alleviating poverty in the regions.

In the spatial regression analysis, it shows that the location of the region has a significant effect on the poverty rate. Neighboring regions can influence the poverty rate in another region. A region is said to be neighboring if its boundaries are precisely attached to the boundaries of its neighboring regions. Spatial regression analysis also proves that neighborliness has a significant influence on average length of school, Gini ratio, and life expectation. The analysis has also produced a regression model with a relatively high value which is 0.6647.

Based on the regression analysis, it was found that three main variables have a significant influence on the poverty rate: average length of school, Gini ratio, and life expectation. Improvement in these variables can provide a significant reduction in the poverty rate. The study's findings recommend that the government should improve health and education facilities in the regions equally, such as the construction of health facilities that are easily accessible to the community, especially in areas with high poverty rates. The development of education facilities is needed to provide free education to people who did not complete the 9-year compulsory education program, particularly in provinces with the highest poverty rates. Finally, the government can strengthen social protection programs to reduce the Gini ratio. This can be achieved through social assistance, universal health insurance, or subsidies for the poor.

6. IMPLICATIONS AND LIMITATIONS

Our research has significant implications for understanding and addressing poverty in Indonesia. Spatial regression analysis and spatial autocorrelation allow researchers to identify significant factors in the poverty rate. It can help the government and related institutions design more targeted policies to reduce poverty in different regions.

However, some limitations need to be considered. The quality of the data used in the analysis and the methodology applied may affect the validity of the research results. In addition, the focus on 2022 may not cover long-term changes or events that may occur later. There may be other factors or variables that can affect poverty which have not been addressed in this study.

Before taking actions or policies based on the findings in this paper, it is necessary to consider these implications and limitations and understand the research results in a broader context. In addition, further research may be needed to delve deeper into the factors affecting poverty and long-term trends.

BIBLIOGRAPHY (*REFERENCES*)

Anselin L., (1988), Spatial Econometrics: Methods and Models, Dordrecht: Academic Publishers.

Central Bureau of Statistics. 2022. Indonesia in Figures 2022. DKI Jakarta: Central Bureau of Statistics.

Baltagi H. Badi & Long Liu. (2012). Testing for Spatial Lag and Spatial Error Dependence Using Double Length Artificial Regressions. Stat Papers (2014)55: 477-486

Banerjee, S. (2004). Hierarchical Modeling and Analysis for Spatial Data. Boca Raton: Chapman and Hall/CRC

Bender, F. E. (2020). Statistical Methods for Food and Agriculture. Taylor & Francis Group.

Leleury, Z. A., & Tomasouw, B. P. (2019). Clustering and Mapping Poverty Characteristics in Southwest Maluku Regency, Maluku Province using Self-Organizing Map and Biplot Analysis. BAREKENG: Journal of Mathematical and Applied Sciences, 13(2), 093-106.

Lembo, A.J. (2006). Spatial Autocorrelation.Cornell University

Yuliantari, P., Memi, N., & Sri, W. (2017). Spatial Autocorrelation Analysis of Hot Spots in East Kalimantan Using Moran Index and Local Indicator of Spatial Autocorrelation (LISA). Journal of EXPONENTIAL. 8(1), 63-70.